Deciphering the Hidden Informational Content of Protein Sequences

FOLDABILITY OF PROINSULIN HINGS ON A FLEXIBLE ARM THAT IS DISPENSABLE IN THE MATURE HORMONE

Received for publication, June 8, 2010, and in revised form, July 21, 2010 Published, JBC Papers in Press, July 27, 2010, DOI 10.1074/jbc.M110.152645

Ming Liu1, Qing-xin Hua1, Shi-Quan Hu1, Wenhua Jia1, Yanwu Yang1, Sunil Evan Saith1, Jonathan Whittaker1, Peter Arvan2, and Michael A. Weiss3

From the 1Division of Metabolism, Endocrinology & Diabetes, University of Michigan Medical Center, Ann Arbor, Michigan 48109 and the 2Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106

Protein sequences encode both structure and foldability. Whereas the interrelationship of sequence and structure has been extensively investigated, the origins of folding efficiency are enigmatic. We demonstrate that the folding of proinsulin requires a flexible N-terminal hydrophobic residue that is dispensable for the structure, activity, and stability of the mature hormone. This residue (PheB1 in placental mammals) is variably dispensable in the native state. Such arm-deletion impaired insulin chain combination and led to cell culture to formation of non-native disulfide isomers with impaired secretion of the variant proinsulin. Cellular folding and secretion were maintained by hydrophobic substitutions at B1 but markedly perturbed by polar or charged side chains. We propose that, during folding, a hydrophobic side chain at B1 anchors transient long-range interactions by a flexible N-terminal arm (residues B1–B8) to mediate kinetic or thermodynamic partitioning among disulfide intermediates. Evidence for the overall contribution of the arm to folding was obtained by alanine scanning mutagenesis. Together, our findings demonstrate that efficient folding of proinsulin requires N-terminal sequences that are dispensable in the native state. Such arm-dependent folding can be abrogated by mutations associated with B-cell dysfunction and neonatal diabetes mellitus.

The efficiency of protein folding poses a fundamental problem at the intersection of biophysics, cell biology, and medicine (1, 2). Because the existence of a unique and accessible ground state is unrepresentative of polypeptides as a class of heteropolymers, foldability is an evolved property of biological sequences (3). Current kinetic models envisage a funnel-shaped free-energy landscape, enabling multiple trajectories to the native state (4–6). What distinguishes foldable from non-foldable sequences (7), and how are bottlenecks avoided (8–10)? The salience of these questions has been reinforced by recognition of proteotoxicity as a general pathological mechanism underlying diverse diseases (11, 12). Here, we describe a cryptic folding element in a protein that is dispensable once the native state has been reached.

A model is provided by insulin, a globular protein central to the regulation of vertebrate metabolism (13). Its impaired biosynthesis causes B-cell dysfunction and permanent neonatal-onset diabetes mellitus (DM)4 (14–17). The insulin gene encodes a single-chain precursor, preproinsulin (Fig. 1, A, top) (18). A signal peptide (gray bar) is cleaved on translocation into the endoplasmic reticulum (ER) to yield proinsulin. The precursor contains successive sequence motifs, defining B, C, and A domains (blue, black, and red, respectively, in Fig. 1A) (19). Whereas the translocated polypeptide is reduced and unfolded, oxidative folding in the ER yields a well organized A-B (insulin-like) core and disordered C-domain (dashed black segment in Fig. 1B) (20–26). Folding is coupled to disulfide pairing (A6–A11, A7–B7, and A20–B19; gold in Fig. 1, A and B). Proinsulin isomers are formed at low concentrations in B-cells (27), and their accumulation may be linked to B-cell dysfunction (28, 29).

Insulin is obtained from proinsulin by proteolytic processing. After transit through the Golgi apparatus and entry into immature secretory granules (30), specific prohormone convertases excise the C-peptide at conserved dibasic sites (BC and CA junctions; green in Fig. 1, A and B), liberating the bioactive hormone (31–33). Insulin thus contains two chains, designated A (21 residues) and B (30 residues), and is stored as Zn2+-stabilized hexamers within specialized secretory granules (34). The hexamers dissociate on secretion; the circulating hormone functions as a Zn2+-free monomer. Because the structure of...
A Cryptic Folding Element in Proinsulin

FIGURE 1. Proinsulin and its biosynthetic pathway. A, pathway of insulin biosynthesis beginning with preproinsulin (top): signal peptide (gray), B-domain (blue), dibasic BC junction (green), C-domain (black), dibasic CA junction (red), and A-domain (red). In the ER the unfolded prohormone undergoes specific disulfide pairing to yield native proinsulin (middle panels). Cleavage of BC and CA junctions (by prohormone convertases PC1 and PC2 and by carboxypeptidase E) leads to mature insulin and the C-peptide (bottom). B, structural model of insulin-like moeity and disordered connecting peptide (dashed line). The A- and B-domains are shown in red and blue, respectively; the disordered connecting domain is shown in yellow. Cystines are labeled in yellow boxes. Structural model of proinsulin has recently been verified by heteronuclear NMR studies (26).

The N-terminal arm of proinsulin provides an example of a cryptic folding element, highlighting foldability as an implicit constraint underlying biological selection of polypeptide sequences. The multistep pathway of insulin biosynthesis from nascent folding to assembly and secretion evidently imposes evolutionary constraints unrelated to the structure and function of the mature hormone. Arm-dependent folding may be abridged by mutations in proinsulin associated with β-cell dysfunction. Strikingly, deletion of PheB1 blocks cellular folding of proinsulin, whereas des-PheB1-insulin retains native-like properties. Three lines of investigation: studies of B1 substitutions, Ala scanning of the arm, and construction of proinsulin/IGF-1 chimeras, together demonstrate that foldability requires a flexible N-terminal hydrophobic anchor.

MATERIALS AND METHODS

Chemical Synthesis—Insulin, KP-insulin (containing substitutions ProB28 → Lys and LysB29 → Pro, which prevent dimerization (41, 61, 62)), proinsulin, and KP-proinsulin were provided by Eli Lilly and Co. (Indianapolis, IN); S-sulfonate B-chain derivatives were obtained by oxidative sulfitolysis (51). A- and B-chain analogs were otherwise prepared by solid-phase synthesis (63). Insulin analogs were prepared by chain combination (63) and purified as described (51). Predicted molecular masses were confirmed by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MS). Insulin analogs were monocomponent by reverse phase high-performance liquid chromatography (RP-HPLC). Synthetic yields were calculated...
A Cryptic Folding Element in Proinsulin

Receptor Binding Assays—Dissociation constants for binding of insulin analogs to IR to be determined in competitive radioligand binding assays with \([^{125}\text{I}-\text{Tyr}^{A14}]\) human insulin (64). The assay employed the B isoform of IR (IR-B). Assays were performed with isolated IR-B with a C-terminal FLAG tag using a microtiter plate antibody capture technique as described (65). The plates (Nunc Maxisorb) were incubated over night at 4 °C with FLAG M2 IgG (100 μg/ml) in phosphate-buffered saline. In all assays the percentage of tracer bound in the absence of competing ligand was <15% to avoid ligand-depletion artifacts. Dissociation constants of analogs were obtained by non-linear regression analysis (66) employing a model describing competitive binding of two different ligands to a receptor. Control studies of cellular extracts in the absence of competing ligand was obtained by analogy to published assignments (69).

Mammalian Cell Culture—HEK293T cells were cultured as described (38) at 37 °C in high-glucose Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum and 0.1% penicillin/streptomycin with 5% CO₂. For transfections, cells were plated into 6-well plates 1 day before transfection. Plasmid DNA (2 μg) was transfected into each well using Lipofectamine (Invitrogen) as described (38, 67).

Metabolic Labeling and PAGE—At 40 h post-transfection the cells were preincubated in methionine/cysteine-deficient medium for 30 min, metabolically labeled in the same medium containing \(^{35}\text{S}\)-labeled Met and Cys for 1 h, washed once with complete medium, and chased for indicated times. After chase, medium 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 medium were immunoprecipitated with guinea pig anti-insulin antisemur (LINCO Diagnostics) and analyzed by Tris-Tricine-urea-SDS-PAGE under reducing and non-reducing conditions (27, 38, 67).

Spectroscopy—Circular dichroism (CD) spectra, obtained using an Aviv spectropolarimeter equipped with an automated titration unit, were measured at protein concentrations of 25 (far-ultraviolet spectra) or 5 μM (denaturation studies) in 50 mM potassium phosphate (pH 7.4) at 4 °C (56). \(^1\)H NMR spectra were obtained at 700 MHz in H₂O or D₂O solutions containing 10 mM deuterioacetic acid (pH 3.0, direct meter reading) at 25 °C. Spatial relationships within insulin- and proinsulin analogs were probed by nuclear Overhauser effect (NOE). The following two-dimensional NMR spectra were acquired: double quantum-filtered correlation spectroscopy (DQF-COSY), NOE spectroscopy (NOESY), and total correlation spectroscopy (TOCSY). Resonance assignment of KP-insulin (complete) and KP-proinsulin (partial) was obtained by standard methods (supplemental materials) (68). Presumptive resonance assignment of des-Phe\(^{61}\)-insulin (supplemental Table S1) was obtained by analogy to published assignments (69).

Thermodynamic Modeling—CD-detected guanidine denaturation data (as monitored at 222 nm) were fitted by non-linear least squares to a two-state model as described (70). In brief, CD data \(\theta(x)\) were fitted by a nonlinear least-squares program according to Equation 1,

\[
\theta(x) = \frac{\theta_A + \theta_B e^{-\Delta G_{\text{HCl}} - mx/RT}}{1 + e^{-\Delta G_{\text{HCl}} - mx/RT}}
\]  
(Eq. 1)

where \(x\) is the concentration of guanidine hydrochloride and \(\theta_A\)
and $\theta_p$ are baseline values in the native and unfolded states. These baselines were approximated by pre- and post-transition lines $\theta_p(x) = \theta_B + m_p x$ and $\theta_p(x) = \theta_B + m_p (x - \theta_B)$. Fitting the original CD data and baselines simultaneously circumvents artifacts associated with linear plots of $\Delta G$ as a function of denaturant (70, 71).

**RESULTS**

The Arm of Insulin and Proinsulin Exhibit T-like Conformations—$^1$H NMR studies of engineered insulin monomers have defined a spectroscopic signature of a T-like conformation based on inter-residue NOEs (41, 42). As illustrated in the spectrum of KP-insulin (supplemental Fig. S1), these include long-range contacts between the side chains of Phe$^{B1}$–Leu$^{A13}$ and His$^{B5}$–Ile$^{A10}$ within a shallow inter-chain crevice; additional NOEs between Leu$^{B6}$ and Leu$^{B11}$ (not shown) reflect formation of an intervening $\beta$-turn at the base of the arm (residues B7–B10). Inconsistent with an R-like $\alpha$-helical conformation, these and related diagnostic NOEs are retained in the spectrum of proinsulin analogs (supplemental Figs. S1 and S2). In spectra of engineered insulin analogs and proinsulin analogs, NOEs are not observed between the aromatic rings of Phe$^{B1}$ and Tyr$^{A14}$, which among crystallographic T-state protomers of insulin exhibit a broad range of distances (supplemental Fig. S3) (72). Similarly, no NMR evidence has been observed for stable maintenance of a hydrogen bond between the B1 carboxyl oxygen and the side chain NH$_2$ of Gln$^{B4}$ as observed in a minority of crystal structures (39).

Despite observation of long-range NOEs, the side chains of Phe$^{B1}$ and Val$^{B2}$ in monomeric analogs of insulin and proinsulin exhibit near random-coil $^1$H NMR chemical shifts and motional narrowing. These trends hold in spectra acquired under a variety of conditions (in aqueous solution at neutral pH, in 10 mM deuterioacetic acid at pH 3.0, and in 20% deuterioacetate, pH 1.9). An example is provided by the ortho and meta $^1$H NMR secondary chemical shifts of Phe$^{B1}$ in KP-insulin at neutral pH (0.11 and 0.12 ppm, respectively; defined as the difference between observed chemical shifts and tabulated random-coil values): these values are markedly smaller than the corresponding secondary shifts of an analogous Phe ring elsewhere in insulin that stably packs against a nonpolar surface (Phe$^{B24}$, 0.61 and 0.43 ppm).

Additional evidence for the conformational variability of the N-terminal arm has been obtained by heteronuclear NMR spectroscopy. Whereas the main chain $^{13}$C NMR chemical shifts of residues B9–B19 in DKP-proinsulin and DKP-insulin (26, 73) exhibit canonical helical values (74), for example, the $^{13}$C chemical shifts of residues B1–B3 conform to a pattern associated with segmental flexibility (see “Discussion” and supplemental Table S2) (75). Furthermore, although the amide resonances of Phe$^{B1}$ and Val$^{B2}$ are not observable due to rapid solvent exchange, those of Asn$^{B3}$, Gln$^{B4}$, and His$^{B5}$ exhibit a gradient of attenuated $^1$H,$^1$H,$^1$H neutron coupling constants of 0.068 $\pm$ 0.009 ppm (analog) and 0.073 $\pm$ 0.010 ppm (wild-type). Far UV CD spectra of des-$B1$- and wild-type insulin are likewise similar (Fig. 4B). The CD-detected guanidine denaturation studies at 4 °C suggest that the two proteins exhibit similar thermodynamic stabilities (Fig. 4C). Application of a two-state model (supplemental Table S3) yields similar estimates of free energies of unfolding ($\Delta G_u$) of 4.1 $\pm$ 0.1 kJ/mol (des-$B1$-insulin) versus 4.0 $\pm$ 0.1 kJ/mol (wild-type). The analog nonetheless exhibits a small left shift ($\Delta T_{m}$, 4 °C) in its thermal unfolding transition (i.e., toward lower temperatures in the

**Deletion of Phe$^{B1}$ Impairs Chain Combination but Is Well Tolerated in the Mature Hormone—**Classical studies of the total chemical synthesis of insulin have demonstrated that the isolated A- and B-chains contain sufficient information to specify native disulfide pairing (76). Robust to diverse amino acid substitution (56), chain combinations has enabled the preparation of >100 insulin analogs by academic and pharmaceutical laboratories (77). Although yield is limited by side reactions (disulfide-bridged cyclic chains, B-chain dimers and polymers), formation of insulin disulfide isomers is ordinarily negligible (78). Surprisingly, combination of the wild-type A-chain with des-$B1$ B-chain is perturbed; its yield is reduced 3-fold relative to wild-type chain combination. Although the predominant low molecular weight product is des-$B1$-insulin (elution time 45.6 min in the HPLC chromatogram shown in Fig. 3A), a major contaminant is a disulfide isomer with elution time delayed by 6 min (arrow in Fig. 3A); the additional elution peaks represent the expected side products. Despite such reduced yield, des-$B1$-insulin may readily be isolated (Fig. 3B), enabling its characterization.

**Binding of des-$B1$-insulin to the isolated insulin receptor (isofrom B; ▲ and dotted line in Fig. 4A) is essentially indistinguishable from that of wild-type insulin (■ and solid line in Fig. 4A). Curve-fitting yields respective estimates of dissociation constants of 0.068 $\pm$ 0.009 and 0.073 $\pm$ 0.010 nm (wild-type). Far UV CD spectra of des-$B1$- and wild-type insulin are likewise similar (Fig. 4B).**
broad temperature range 30–70 °C) as monitored by mean residue ellipticity at 222 nm (Fig. 4D).

1H NMR studies of human insulin and des-PheB1-insulin were undertaken as dimers in 10 mM deuterioacetic acid (pD 3.0) at 25 °C (79). Resonance assignments of des-B1-insulin are provided under supplemental Table S1. The aliphatic spectrum of wild-type insulin (Fig. 5B) provides a fingerprint of the T-state-specific packing of PheB1 against the A-chain (Fig. 5A). The spectrum of des-PheB1-insulin (Fig. 5C) exhibits a similar overall envelope of resonances but with selected changes in chemical shift due to the absence of the PheB1 magnetic ring current (ValB2, IleA10, and LeuA13; black labels). These trends in 1H NMR chemical shifts were well resolved in corresponding two-dimensional TOCSY spectra (Fig. 6); whereas the majority of spin systems in insulin are unperturbed by the removal of PheB1 (black labels), selective changes in chemical shift are prominent within the N-terminal arm of the B-chain and within or adjoining its T-state-specific docking site (red labels). Spatial relationships within this docking site (visualized in Fig. 7, A and B) were probed by comparison of two-dimensional NOESY spectra (Fig. 7C, D). The region shown contains contacts between aliphatic protons (ω1; horizontal axis) and aromatic protons (ω2; vertical axis). Whereas B1-related cross-peaks in the wild-type spectrum (Fig. 7C) are absent as expected in the variant spectrum (Fig. 7D), T-state-specific NOEs from the imidazole ring of HisB15 are retained within an inter-chain crevice (wild-type cross-peaks l–q versus variant cross-peaks r–u). Constraints between the aromatic ring of TyrA19 and neighboring methyl groups (IleA2, LeuB11, and LeuB15; not labeled in the figure), diagnostic of helix-helix packing in the core of insulin (41, 42), are essentially identical in the two proteins.

Deletion of PheB1 Perturbs Cellular Folding—Transient transfection of human cells with a plasmid expressing proinsulin provides a model for studying its folding within the endoplasmic reticulum, subcellular trafficking, and secretion (Fig. 8A). Following transfection of 293T cells, we thus examined the relative expression of wild-type or variant proinsulins. Following pulse labeling of newly synthesized proteins with 35S-amino acids, labeled wild-type or variant proinsulins were immunoprecipitated with polyclonal anti-insulin antiserum and subjected to nonreducing Tris–Tricine–urea–SDS-PAGE, which allows examination of distinct proinsulin disulfide isoforms that form within the ER. The absence of endogenous proinsulin in these cells makes detection of the transfected proteins straightforward. As previously demonstrated (27, 38, 67), resolution of discrete proinsulin bands in this gel system provided an assay for extent of native folding, competing disulfide-
A Cryptic Folding Element in Proinsulin

**FIGURE 6. Two-dimensional NMR analysis of des-Phe**

### A

**des-B1**

- **ω₁ (ppm)**
  - **ω₂ (ppm)**

- **ω₁ (ppm)**
  - **ω₂ (ppm)**

**insulin**

- **ω₁ (ppm)**
  - **ω₂ (ppm)**

isomer formation, and efficiency of secretion.

Perturbed cellular folding of des-B1-proinsulin is in accordance with perturbed chain combination of des-B1-insulin (above). Because band intensities were attenuated after the 1 h chase, the studies were repeated with chase times of 0 and 2 h (Fig. 9, A and B, respectively). Furthermore, because impaired folding could be associated with formation of high molecular weight complexes (due to aberrant formation of mispaired intermolecular disulfide bridges), PAGE analysis was undertaken with and without reduction by dithiothreitol. On reduction a single band was observed, presumably reflecting total expression of the variant proinsulin. Without chase, initial overall expression of des-B1-proinsulin was substantial (Fig. 9A, lane 6) albeit less than that of wild-type proinsulin (lane 5). The distribution of disulfide isomers favored non-native species (Fig. 9A, lane 3). Following an extended chase period of 2 h, a substantial intracellular accumulation of des-B1-proinsulin was observed (reduced band in lane 14 in Fig. 9B) without secretion (lane 15). In the absence of reduction any bands corresponding to low molecular weight species were faint (Fig. 9B, lane 10), implying that the des-B1-proinsulin polypeptides are sequestered in aberrant high molecular weight complexes. Under these conditions wild-type proinsulin predominantly undergoes native folding and secretion (Fig. 9B, lanes 8–9 and 12–13).

Control studies suggest that the impaired foldability of des-B1-proinsulin in 293T cells is unlikely to reflect thermodynamic instability of the protein once folded. Prior biophysical studies have established that two-disulfide insulin- and proinsulin analogs lacking cystine A6–A11 form in vitro partial folds of low stability (ΔΔG° > 2 kcal/mol) (41, 80). In transfected 293T cells removal of the A6–A11 disulfide bridge by pairwise mutation of Cys to Ser did not block expression or secretion (Fig. 8C, lanes 13 and 14). The substitutions likewise caused little detectable change in band mobility (81) despite the presumed loss of structural organization. Folding and secretion are likewise robust to the destabilizing mutation Ileα2 → Gly (Fig. 8C, lanes 15 and 16); this substitution in the hydrophobic core was found to impair stability (ΔΔG° of an insulin analog by at least 1.6 kcal/mol (82). The impaired foldability of des-B1-proinsulin thus stands in marked contrast to the native-like in vitro stability of des-B1-insulin.

**Cellular Folding of Proinsulin Requires a Hydrophobic Residue at B1**—To further probe the nature of the B1 folding determinant, we next examined multiple substitutions (Fig. 8, B–E). Substitution of PheB1 by Asp, for example, resulted in robust expression with a marked decrease in the fraction of AspB1-proinsulin molecules achieving the native disulfide-bonded form (Fig. 8B, lane 5); secretion of the variant proinsulin was

---

This cellular model enabled assessment of effects of arm substitutions on disulfide pairing in the ER and subsequent trafficking, but does not recapitulate β-cell-specific prohormone processing or formation of microcrystalline insulin storage depots in glucose-regulated secretory vesicles.

---

7 This cellular model enabled assessment of effects of arm substitutions on disulfide pairing in the ER and subsequent trafficking, but does not recapitulate β-cell-specific prohormone processing or formation of microcrystalline insulin storage depots in glucose-regulated secretory vesicles.
The corresponding intracellular samples (Fig. 9C, lanes 18 and 20) indicate preferential formation of a disulfide isomer. Previous studies demonstrated that AlaB5 likewise impairs folding and secretion (67).

Proinsulin and IGF-I are homologous proteins with divergent arms and different refolding properties (83–87). IGF-I lacks a B1 residue and exhibits successive arm residues, GPGQTLGG; residues B6–B8 (boldface) are conserved in proinsulin. To test whether an IGF-I arm could support the folding of proinsulin, two chimeric analogs were constructed: [des-B1, GlyB2, ProB3, GlnB4]proinsulin (Fig. 8F, lanes 41 and 42 hybrid 1) and [des-B1, GlyB2, ProB3, GlnB4, ThrB5]proinsulin (hybrid 2, lanes 43 and 44). Each of these “arm-swapped” analogs exhibited preferential formation of non-native isomers within the ER with impaired secretion. Native folding of IGF-I in vivo is presumably stabilized by specific IGF-binding proteins that contact its divergent and foreshortened arm (64, 88, 89).

These biological results motivated chemical synthesis [des-B1, GlyB2, ProB3, GlnB4]insulin. In synthesis of des-B1-insulin, the variant arm impaired the yield of chain combination (also by 3-fold). Although a predominant product was formed, analysis of reverse phase HPLC-resolved side products by MS demonstrated formation of two competing disulfide isomers (presumably containing aberrant disulfide pairings (A6–B7, A7–A11, A20–B19) and (A6–A7, A11–B7, and A20–B19); designated swap and swap2 in Ref. 90. Evidence that the predominant product contained native disulfide bridges was provided by its non-negligible biological activity (IR-B dissociation constant 0.136 ± 0.019 nM; only 3-fold lower than wild-type insulin in the same assay).

**DISCUSSION**

The recent (and continuing) exponential increase in the size of the data base of protein structures poses the challenge of functional annotation. Whereas residues involved in substrate binding or catalysis may be readily recognizable, determinants of folding efficiency may not be apparent in the native state (3, 7). Such residues may nonetheless impose key evolutionary constraints (38) and emerge as sites of mutation associated with human genetic diseases (60).

The present study has focused on the N-terminal arm of proinsulin (FVNQH; residues B1–B5). In classical studies of insulin (39) a seeming paradox was posed by the conservation of these
residues despite their dispensability for receptor binding (58) and marked structural variability (59). Could the arm have a hidden biological function? An interdisciplinary set of studies was thus undertaken to investigate the contribution of the arm to the efficiency of protein folding. Although such a contribution seemed unlikely given that the arm is only partially ordered in an engineered proinsulin monomer (26), we posited that a transient role in folding might be hidden once the native state is reached. Such “hidden” contributions of specific residues to folding efficiency have been extensively explored in a trimeric β-helix (91).

Insulin has been extensively studied since its isolation in 1922 and landmark clinical application (92). It may seem surprising, therefore, that a new function of a sequence motif has only now been recognized. Yet the role of the N-terminal arm of the B-chain has long been enigmatic. Indeed, its structure undergoes a fundamental change in secondary structure, from extended (T) to α-helical (R), as part of a long-range allosteric reorganization of insulin hexamers, designated the TR transition (93). Whereas the conformation of an insulin monomer in solution resembles the T-state (40), adoption of R-like features on receptor binding has been proposed (39, 59, 94). An intriguing hypothesis envisions that induced fit of the arm represents a switch between folding-competent and active conformations (36). This model thus represents the potential biological utility of a chameleonic sequence in a globular protein (43).

Proinsulin contains an insulin-like moiety (the A and B domains) and flexible connecting segment (the C domain). A recent heteronuclear NMR study of an engineered proinsulin monomer has provided evidence that its N-terminal arm (residues B1–B8) exhibits partial disorder as indicated by a trend toward N-terminal attenuation of amide-related [¹H]–[¹³C] heteronuclear NOEs (26). Such studies employed substitutions in the classical dimer interface (Pro⁸⁹ → Lys and Lys⁹⁸ → Pro) and trimer interface (His¹¹⁰ → Asp) to enable NMR characterization at neutral pH (26, 41). Although such signals at B1 and B2 were not observable, complete [¹³C] NMR resonance assignment was obtained. Trends in main chain [¹³C] secondary chemical shifts (Fig. 10A) provides an estimate of disorder as inferred from a random-coil index (Fig. 10B) and predicted residue-specific order parameters on the picosecond-nanosecond time scale (Fig. 10C) (75). In Fig. 10 chemical-shift index values and predicted dynamic parameters of the B, C, and A domains are shown in blue, black, and red, respectively, in relation to helical segments (spirals at bottom of panel A). Whereas the A domain exhibits consistent low values of the random-coil index (Fig. 10B) and high values of the order parameter (Fig. 10C), the B domain is remarkable for “ramps” from B1–B8 and B25–B30. These trends predict that the B domain should exhibit a progressive increase in disorder on the picosecond-nanosecond time scale toward the N terminus and BC junction, respectively. In the future it would be of interest to obtain quantitative estimates of such disorder by measurement of [¹⁵N] and [¹³C] NMR relaxation times and their interpretation in relation to spectral density functions.

Determinants of Foldability—The N-terminal arm of the B-chain consists of residues proximal to the α-helical domain of the hormone (residues B5–B8) and residues distal to this domain (B1–B4). Evidence for the biological importance of the proximal portion of the arm has been provided by clinical observations that mutations at B5, B6, or B8 can cause permanent neonatal-onset diabetes mellitus, presumably due to toxic misfolding of the mutant proinsulin in the ER of pancreatic β-cells (14–17). Studies of insulin chain combination had earlier shown that these residues were critical for disulfide pairing.
A Cryptic Folding Element in Proinsulin

even when receptor binding was not significantly impaired (36, 37, 67, 95).

The pertinence of the T-state to folding and the efficiency of disulfide pairing is supported by stereospecific effects of substitutions of GlyB8. Stabilization (or destabilization) of the T-state to folding and the efficiency of disulfide pairing may be rendered inefficient due to either conformational search of the polypeptide leading to native properties. Ala scanning of the disulfide isomers (although to varying extents). Because PheB1 (and indeed, residues B1-B4 (58)) may be deleted without loss of activity, the distal arm provides an example of a folding element that is dispensable once the native state has been reached.

We imagine that the pattern of non-polar and polar side chains in the distal arm contributes to the efficiency and fidelity of disulfide pairing.9 General clustering of N-terminal hydrophobic side chains PheB1 and ValB2 with A-chain side chains could, for example, enhance the probability of collisions between the thiol (or thiolate) groups of CysB7 and CysA7. Productive alignment may then be enhanced by the proximal arm as a positive φ angle at B8 enables specific long-range interactions by the side chains of HisB5 and LeuB6. Because the coarseness of our experimental probes nonetheless prevents unambiguous atomic-scale interpretation, multiple molecular models may account for our findings. In the absence of the wild-type arm, the nascent conformational search of the polypeptide leading to native disulfide pairing may be rendered inefficient due to either destabilization of on-pathway intermediates or stabilization of off-pathway interstates. Furthermore, apparent bottlenecks may reflect thermodynamic or kinetic traps. Arm mutations may even create barriers not pertinent to the folding mechanism of the wild-type protein.

A Cryptic Folding Element in Proinsulin

T-state residues B1–B4 pack loosely within an inter-chain crevice adjoining CysB7 and CysA7. The present study has shown that deletion of PheB1 blocks cellular folding of proinsulin whereas the two-chain model and mature product, des-PheB1-insulin, retains native-like properties. Ala scanning of the distal arm further demonstrated that each side chain influences the efficiency of folding and secretion (although to varying extents). Because PheB1 (and indeed, residues B1-B4 (58)) may be deleted without loss of activity, the distal arm provides an example of a folding element that is dispensable once the native state has been reached.

We envisage that, within the partial folds of protein-folding intermediates, nascent T-like local structure in the proximal arm enhances the efficiency and fidelity of disulfide pairing. In particular, the conformation of residues B5–B8 (including inter-domain hydrogen bonding by the imidazole ring of HisB5 (37)) would profoundly alter the position of CysB7, its orientation with respect to CysA6 and CysA7, and in turn the trajectory of the distal arm once a disulfide bond was established. In the

Figure 9. Pulse-chase studies of proinsulin analogs in HEK293T cells. A, the newly synthesized wild-type or des-PheB1-human proinsulin from transfected HEK293T cells (C) were immunoprecipitated as described in the legend to Fig. 8 and analyzed by Tris-Tricine-urea-SDS-PAGE under non-reducing (lanes 1–3) and reducing (lanes 4–6) conditions. con (lanes 1 and 4) denotes empty vector control. des-PheB1-human proinsulin presented mostly as misfolded disulfide isomers. Reduction leads to coalescence of monomers, des-PheB1-human proinsulin exhibit similar electrophoretic mobilities; extent of expression of variant (lanes 1–3) was reduced relative to wild-type (lane 5). B, aliquots of cells from panel A were chased for 2 h and analyzed under non-reducing and reducing conditions. Whereas the majority of wild-type proinsulin was secreted from cells (C) to medium (M) after a 2-h chase, the des-PheB1-human proinsulin was barely detectable in the cells (lane 10) under non-reducing conditions and was not secreted. Under reducing conditions, however, the variant protein was recoverable (lane 14), suggesting that it formed aberrant disulfide-linked complexes and was retained in the ER. C, corresponding studies of AlaB4, AlaB2, and AlaB3 variant proinsulins analyzed under non-reducing conditions demonstrate a range of perturbations: AlaB2, severe; AlaB3, moderate, and AlaB4, partial impairment of biosynthesis and secretion. D, control Western blot for specificity of anti-serum. Aliquots (100 ng) of human insulin (lane 26), des-PheB1-insulin (lane 27), and human proinsulin (lane 28) were resolved under non-reducing conditions, transferred onto a nitrocellular membrane, and blotted with guinea pig anti-porcine insulin antibody. This Western blot indicated that the antibody recognized the three proteins with similar efficiencies.

Although the present set of substitutions tested at B1 highlights the importance of its non-polar character, ArgB1 occurs as a rare variant in non-mammalian insulin sequences. An example is provided by the divergent insulin of the hagfish (103), a member of a primitive lineage of marine invertebrates that lack a vertebral column. We speculate that the aliphatic portion of ArgB1 may pack against an analogous arm-related groove, whereas its charged guanidinium moiety projects into solvent. Analogous proximal side chain packing may underlie the incomplete block to folding and secretion imposed by LysB1 relative to AspB1 in the present studies.
IGF-I disulfide isomer on redox-coupled refolding (64, 98). Such divergent folding properties reflect co-evolution of specific IGF-binding proteins (64, 88, 89). The incomplete folding information of IGF-I and IGF-II highlights the breach in the autonomous foldability of proinsulin. We thus speculate that co-evolution of IGF-binding proteins has enabled IGF-I and IGF-II to explore regions of sequence space forbidden to insulin due to constraints of autonomous foldability.

Concluding Remarks—Insulin is one of the most studied globular proteins and yet among the least well understood. Understanding its conformational lifecycle will require extensive future studies, from deciphering the molecular mechanisms by which proinsulin folds to determining structures of hormone-receptor complexes. Recent advances in human genetics have highlighted the direct connection between the biophysical chemistry of proinsulin and the pathogenesis of β-cell dysfunction in monogenic forms of DM (14, 60).

The present study was designed to test the hypothesis that the N-terminal arm of proinsulin functions, despite its conformational variability in the mature hormone, as a cryptic folding element. Experimental design has integrated assays of protein
folding and trafficking in mammalian cell culture with in vitro studies of protein structure, stability, and activity. Evidence has been provided that mutations in the arm can cause a broad range of effects on folding efficiency with possible clinical implications. Although no mutations associated with neonatal DM have been found to date in the distal arm (14, 60), it is possible that such mutations would induce less marked ER stress and β-cell dysfunction than those associated with neonatal-onset DM (which cluster in the proximal arm or within the α-helical domain) and so would present later in life (17). Evidence for mutation-specific ages of DM onset associated with the extent of biosynthetic impairment has recently been provided by comparison of mutations at arm position B6 (99).

Whereas a neonatal mutation (LeuB6 → Pro) is predicted to distort both packing of the arm and its main chain conformation, a less perturbing substitution (LeuB6 → Met) is associated with onset of DM in adolescence and adulthood (17–36 years of age) (99). Distal arm mutations might thus present as a form of maturity-onset diabetes of the young or as polymorphisms conferring susceptibility to adult-onset β-cell dysfunction in the context of obesity (28, 100).

The genetics of neonatal DM and other diseases of misfolding suggest that protein evolution has been constrained not only by structure and function, but also by folding efficiency and, in the breach, the associated risk of toxic misfolding. The present studies of des-B1-insulin and arm analogs of proinsulin have demonstrated that efficient folding may require the transient function of a conserved folding element and that this essential role may be structurally unapparent once the ground state is reached. The contributions of PheB1 and ValB15 to cellular foldability are particularly striking in light of their high thermal B-factors in crystals (39) and 1H NMR motional narrowing in solution (40–42). Whereas these and other nonpolar side chains (including LeuB6, IleA10, and LeuA13) may contribute to hydrophobic collapse near cysteines (at B7, A6, A7, and A11), the efficiency of disulfide pairing is likely to require specific structural features of HisB5 and GlnB8 as visualized in the classical T-state of insulin (39). It is also possible that the aromatic ring of TyrA14 contributes to folding efficiency despite its variable positioning among crystal structures and flexibility in solution (72).

The crystallographic TR transition in zinc insulin hexamers has long provided a model for the transmission of conformational change in proteins (39). Although chameleon sequences analogous to the N-terminal arm of proinsulin are uncommon in the overall crystallographic data base, structures of native states, once reached, may mask the extent of conformational plasticity among protein-folding intermediates. We anticipate that functional annotation of protein structures will in general require multidisciplinary efforts to decipher the folding information hidden in their sequences.

10 A mutation has also been described in the B20–B24 β-turn of proinsulin (ArgB22 → Gln) that is presented in the second decade as maturity-onset diabetes of the young (MODY) (17). Chronic elevation of ER stress in β-cells of this patient presumably led to a slow but progressive loss of β-cell mass during childhood. ER stress may likewise contribute to the pathogenesis of insulin Los Angeles (PheB24 → Ser), a classical insulinopathy of variable penetrance presenting in adulthood (104).

A Cryptic Folding Element in Proinsulin

Acknowledgments—We thank Eli Lilly and Co. for generously providing materials, S. Yadav and the staff of the peptide core facility at the Cleveland Clinic Foundation, S. Wang for assistance with insulin chain combination, K. Huang and the staff of the Center for Proteomics and Structural Biology for assistance with MS assays, L. Whitaker for assistance with receptor-binding studies, M. Lin for plasmid construction, N. B. Phillips and P. G. Katsoyannis for experimental advice, and G. G. Dodson, B. H. Frank, F. Ismail-Beigi, S. B. Kent, and D. F. Steiner for discussion.

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. Lazardis, 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. Danner, 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. Oliveira, M., and Wolynes, P. G. (2005) Q. Rev. Biophys. 38, 245–288
11. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zardo, J., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) Nature 416, 507–511
12. Dobson, C. M. (1999) Trends Biochem. Sci. 24, 329–332
13. Dobson, G., and Steiner, D. (1998) Curr. Opin. Struct. Biol. 8, 189–194
14. Stay, J., Edghill, E. L., Flanagan, S. E., Ye, H., Paz, Y. P., Pluzhnikov, A., Below, J. E., Hayes, M. G., Cox, N. J., Lipkind, G. M., Lipton, R. B., Greeley, S. A., Patch, A. M., Ellard, S., Steiner, D. F., Hattersley, A. T., Philipson, L. H., and Bell, G. I. (2007) Proc. Natl. Acad. Sci. U.S.A. 104, 15040–15044
15. Colombo, C., Porzio, O., Liu, M., Massa, O., Vasta, M., Salardi, S., Becarria, L., Monciotti, C., Toni, S., Pedersen, O., Hansen, T., Federici, L., Pesavento, R., Cadario, F., Federici, G., Ghirri, P., Arvan, P., Iafusco, D., Barbetti, F., and Early Onset Diabetes Study Group of the Italian Society of Pediatric Endocrinology and Diabetes (2008) J. Clin. Invest. 118, 2148–2156
16. Edghill, E. L., Flanagan, S. E., Patch, A. M., Bousted, C., Parrish, A., Shields, B., Shepherd, M. H., Hussain, K., Kapoor, R. R., Malecki, M., MacDonald, M. J., Stay, J., Steiner, D. F., Philipson, L. H., Bell, G. I., Hattersley, A. T., and Ellard, S. (2008) Diabetes 57, 1034–1042
17. Molven, A., Ringdal, M., Nordbo, A. M., Raeder, H., Stay, J., Lipkind, G. M., Steiner, D. F., Philipson, L. H., Bergmann, I., Aarskog, D., Undlien, D. E., Joner, G., Sevik, O., Bell, G. I., and Njolstad, P. R. (2008) Diabetes 57, 1131–1135
18. Steiner, D. F., and Chan, S. J. (1988) Horm. Metab. Res. 20, 443–444
19. Steiner, D. F. (1967) Trans. N. Y. Acad. Sci. 30, 60–68
20. Frank, B. H., and Veros, A. J. (1968) Biochem. Biophys. Res. Commun. 32, 155–160
21. Pekar, A. H., and Frank, B. H. (1972) Biochemistry 11, 4013–4016
22. Frank, B. H., Pekar, A. H., and Veros, A. J. (1972) Biochemistry 21, 486–491
23. Snell, C. R., and Smyth, D. G. (1975) J. Biol. Chem. 250, 6291–6295
24. Brems, D. N., Brown, P. L., Heckenlively, L. A., and Frank, B. H. (1990) Biochemistry 29, 9289–9293
25. Weiss, M. A., Frank, B. H., Khait, I., Pekar, A., Heiney, R., Shoelson, S. E., and Neuringer, L. I. (1990) Biochemistry 29, 8389–8401
26. Yang, Y., Hua, Q. X., Liu, J., Shimizu, E. H., Choquette, M. H., Mackin, R. B., and Weiss, M. A. (2010) J. Biol. Chem. 285, 7847–7851
27. Liu, M., Li, Y., Cavenner, D., and Arvan, P. (2005) J. Biol. Chem. 280,
A Cryptic Folding Element in Proinsulin

31000

A. C. Beall, G. A. P. Berthoud, J. R. Blandon, and M. A. Weiss (2008) J. Biol. Chem. 283, 21918–21920

B. H. B. S. K. C. N. (1991) J. Biol. Chem. 266, 11502–11509

C. H. D. A. L. K. J. E. (1992) Trends Biochem. Sci. 16, 341–345

D. R. D. C. S. T. (1994) Biochem. J. 284, 19159–19163

E. R. D. L. R. H. E. (1993) Protein Sci. 2, 4344–43453

F. L. D. H. A. L. K. J. E. (1997) J. Biol. Chem. 272, 12978–12983

G. A. P. Berthoud, J. R. Blandon, and M. A. Weiss (2008) J. Biol. Chem. 283, 21918–21920

H. B. S. K. C. N. (1991) J. Biol. Chem. 266, 11502–11509

J. C. J. F. (1976) Nature 261, 1749–1756

K. Y. E. T. H. H. A. L. K. J. E. (1997) J. Biol. Chem. 272, 12978–12983

L. D. H. A. L. K. J. E. (2000) J. Biol. Chem. 281, 2459–24909

M. A. N. V. T. S. (1992) Trends Biochem. Sci. 16, 341–345

N. A. L. D. S. T. K. (1999) Structure 7, R131–R139

O. H. S. B. L. D. (1999) FEBS Lett. 443, 271–276

P. A. N. V. T. S. (1999) FEBS Lett. 443, 271–276

Q. H. D. L. R. T. (2000) Proc. Natl. Acad. Sci. U.S.A. 107, 6578–6581

R. M. D. L. R. H. E. (1992) NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York

S. M. C. J. D. (1998) J. Biol. Chem. 273, 14479–14485

T. A. B. D. R. D. F. (1992) J. Biol. Chem. 267, 1156–1157

U. L. D. E. (1996) Science 273, 4344–43453

V. M. D. S. T. K. (1999) Structure 7, R131–R139

W. M. D. S. T. K. (1999) Structure 7, R131–R139

X. F. S. K. A. H. E. (1993) Biochem. J. 301, 263–270

Y. L. D. S. T. (1993) J. Mol. Biol. 231, 431–432

Z. M. D. S. T. K. (1999) Nature 393, 40–45

A. C. Beall, G. A. P. Berthoud, J. R. Blandon, and M. A. Weiss (2008) J. Biol. Chem. 283, 21918–21920

B. H. B. S. K. C. N. (1991) J. Biol. Chem. 266, 11502–11509

C. H. D. A. L. K. J. E. (1997) J. Biol. Chem. 272, 12978–12983

D. R. D. C. S. T. (1994) Biochem. J. 284, 19159–19163

E. R. D. L. R. H. E. (1993) Protein Sci. 2, 4344–43453

F. L. D. H. A. L. K. J. E. (2000) J. Biol. Chem. 272, 12978–12983

G. A. P. Berthoud, J. R. Blandon, and M. A. Weiss (2008) J. Biol. Chem. 283, 21918–21920

H. B. S. K. C. N. (1991) J. Biol. Chem. 266, 11502–11509

J. C. J. F. (1976) Nature 261, 1749–1756

K. Y. E. T. H. H. A. L. K. J. E. (1997) J. Biol. Chem. 272, 12978–12983

L. D. H. A. L. K. J. E. (2000) J. Biol. Chem. 281, 2459–24909

M. A. N. V. T. S. (1992) Trends Biochem. Sci. 16, 341–345

N. A. L. D. S. T. K. (1999) Structure 7, R131–R139

O. H. S. B. L. D. (1999) FEBS Lett. 443, 271–276

P. A. N. V. T. S. (1999) FEBS Lett. 443, 271–276

Q. H. D. L. R. T. (2000) Proc. Natl. Acad. Sci. U.S.A. 107, 6578–6581

R. M. D. L. R. H. E. (1992) NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York

S. M. C. J. D. (1998) J. Biol. Chem. 273, 14479–14485

T. A. B. D. R. D. F. (1992) J. Biol. Chem. 267, 1156–1157

U. L. D. E. (1996) Science 273, 4344–43453

V. M. D. S. T. K. (1999) Structure 7, R131–R139

W. M. D. S. T. K. (1999) Structure 7, R131–R139

X. F. S. K. A. H. E. (1993) Biochem. J. 301, 263–270

Y. L. D. S. T. (1993) J. Mol. Biol. 231, 431–432
A., Fetita, S., Tarasov, A. L., Guillausseau, P. J., Boesgaard, T. W., Pedersen, O., Hansen, T., Polak, M., Gautier, J. F., Froguel, P., Rutter, G. A., and Vaxillaire, M. (2010) *Diabetes* **59**, 653–661

100. Florez, J. C. (2008) *Diabetologia* **51**, 1100–1110

101. Sieber, P. S., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Märki, F., and de Gasparo, M. (1978) *Hoppe Seyler’s Z. Physiol. Chem.* **359**, 113–123

102. 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

103. Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Emdin, S. F., and Reynolds, C. D. (1979) *J. Mol. Biol.* **132**, 85–100

104. Shoelson, S. E., Polonsky, K. S., Zeidler, A., Rubenstein, A. H., and Tager, H. S. (1984) *J. Clin. Invest.* **73**, 1351–1358