A Conserved Histidine in Insulin is Required for the Foldability of Human Proinsulin.

STRUCTURE AND FUNCTION OF AN AlaB5 ANALOG
Qing-xin Hua+, Ming Liu+, Shi-Quan Hu, Wenhua Jia, Peter Arvan,,* & Michael A. Weiss*
Department of Biochemistry, Case Western Reserve University, Cleveland, OH 44106
Division of Metabolism, Endocrinology & Diabetes, University of Michigan Medical Center,
Ann Arbor, Michigan 48109

Running title: Determinants of Proinsulin Foldability
+These authors contributed equally. *Address correspondence to: PA (parvan@umich.edu) or MAW (michael.weiss@case.edu).

The insulins of eutherian mammals contain histidines at positions B5 and B10. The role of HisB10 is well defined: although not required in the mature hormone for receptor binding, in the islet β cell this side chain functions in targeting of proinsulin to glucose-regulated secretory granules and provides axial zinc-binding sites in storage hexamers. The role of HisB5 is by contrast less well understood. Here, we demonstrate that its substitution by Ala markedly impairs insulin chain combination in vitro and blocks the folding and secretion of human proinsulin in a transfected mammalian cell line. The structure and stability of an AlaB5-insulin analog were investigated in an engineered monomer (DKP-insulin). Despite its impaired foldability, the structure of the AlaB5 analog retains a native-like T-state conformation. At the site of substitution inter-chain nuclear Overhauser enhancements (NOEs) are observed between the methyl resonance of AlaB5 and side chains in the A chain; these NOEs resemble those characteristic of HisB5 in native insulin. Substantial receptor-binding activity is retained (80 ± 10 % relative to the parent monomer). Although the thermodynamic stability of the AlaB5 analog is decreased (ΔΔGθ 1.7 ± 0.1 kcal/mole), consistent with loss of HisB5-related inter-chain packing and hydrogen bonds, control studies suggest that this decrement cannot account for its impaired foldability. We propose that nascent long-range interactions by HisB5 facilitate alignment of CysA7 and CysB7 in protein-folding intermediates: its conservation thus reflects mechanisms of oxidative folding rather than structure-function relationships in the native state.

INTRODUCTION
Insulin is a small globular protein containing two chains, A (21 residues) and B (30 residues). The mature hormone is the post-translational product of a single-chain precursor, proinsulin (1), in which a connecting domain extends from the C-terminal residue of the B domain (ThrB30) to the N-terminus of the A chain (GlyA1; Fig. 1A). In the pancreatic β cell proinsulin folds in the endoplasmic reticulum (ER) to form three specific disulfide bridges (A6-A11, A7-B7, and A20-B19; orange bars in Fig. 1A). Although the three-dimensional structure of proinsulin has not been determined, a variety of evidence indicates that it consists of a folded insulin moiety (red and blue ribbons in Fig. 1B) and disordered connecting region (dashed black line). Upon transit through the Golgi apparatus (GA) and entry into immature secretory granules (2), the C-peptide (black open circles in Fig. 1A) is excised by a specific set of prohormone convertases (3). The mature hormone is stored as Zn2+-stabilized hexamers within specialized secretory granules (4). Insulin hexamers dissociate on secretion into the portal circulation, enabling the circulating hormone to function as a Zn2+-free monomer. The stability and receptor-binding activity of insulin require maintenance of its three disulfide bridges (5-10).

Classical structure-function relationships in insulin have been inferred from the pattern of conservation and divergence among vertebrate sequences (11). The central portion of putative receptor-binding surface (spanning residues IleA2, ValA3, ValB12, PheB24, and PheB25; 11-13) is invariant (14). The importance of these residues has been extensively investigated by mutagenesis: substitutions markedly impair receptor binding (15-23). Likewise conserved are side chains integral to the structure of the hydrophobic core: LeuA16, TyrA19, LeuB11, and LeuB15 (11). Alanine scanning mutagenesis of such core side chains is associated with inefficient biosynthetic expression in Saccharomyces cerevisiae (24) and low yield of chemical synthesis by insulin chain combination.
Such perturbations are readily rationalized by structural models (25). The complex pathway of insulin biosynthesis, trafficking, assembly, and secretion may impose additional evolutionary constraints unrelated to the structure and function of the insulin monomer. Such hidden constraints may be revealed by analysis of conserved residues at sites tolerant of substitutions in vitro. An elegant example is provided by His$^{B10}$, which mediates axial zinc binding within the insulin hexamer (27). Whereas substitution of His$^{B10}$ by Asp enhances the stability and activity of the insulin monomer (28,29), the corresponding mutation in the human insulin gene leads to a syndrome of hyperproinsulinemia and diabetes mellitus (30). In the β cell Asp$^{B10}$-proinsulin exhibits impaired trafficking, leading to inefficient excision of the connecting peptide and failure to be efficiently stored within glucose-regulated secretory granules. His$^{B10}$ may thus contribute to subcellular targeting in the ER and GA. Recent studies suggest that the mutation may also impair the fidelity of disulfide pairing in the ER (31).

In this article we investigate the function of His$^{B5}$, the other histidine residue in the B chain. In the crystallographic T state – the predominant conformation of the insulin monomer in solution (5,32,33) – the B5 imidazole ring packs near the A7-B7 disulfide bridge at the edge of a solvated crevice between A- and B chains (11). Comparison of multiple independent T-state structures indicates consistent engagement of the B5 imidazole ring within a pocket lined by residues A6-A11 (Fig. 1C). Although the details of such packing differ among crystal structures, His$^{B5}$ appears to brace the A7-A11 disulfide bridge on the protein surface. The B5-related crevice is remote from the classical receptor-binding surface, and indeed, mutagenesis (34,35) and deletion analysis (11,36-38) suggest that His$^{B5}$ contributes only modestly to biological activity. His$^{B5}$ is nonetheless broadly conserved among eutherian mammals.2 To elucidate hidden biological functions of His$^{B5}$ that may constrain its divergence, we have undertaken an analysis of an Ala$^{B5}$-insulin variant in vitro in relation to effects of this and other B5 substitutions on the foldability of proinsulin in mammalian cell culture.

Our results are presented in two parts. We first describe the chemical synthesis and characterization of a monomeric Ala$^{B5}$ analog. Because the complex self-association properties of insulin ordinarily confound biophysical studies, a well-characterized monomeric template is provided by DKP-insulin (39-41). The three amino-acid substitutions in the B chain of DKP-insulin (His$^{B10}$→Asp (D), Pro$^{B28}$→Lys (K), and Lys$^{B29}$→Pro (P)) disallow formation of dimers and higher-order oligomers without perturbing its biological activity. Ala$^{B5}$-DKP-insulin retains substantial receptor-binding activity, and its structure closely resembles that of native insulin (5). Surprisingly, the pattern of long-range nuclear Overhauser enhancements (NOEs) indicates that the conformation of the B5-related crevice and neighboring disulfide bridges (cystines A7-B7 and A6-A11) are not significantly altered by the absence of the imidazole side chain. Because these initial studies were unrevealing in relation to evolutionary constraints, we next investigated foldability as probed by the efficiency of disulfide pairing in chain combination (25,26).

Substitution of His$^{B5}$ by Ala markedly impairs insulin chain combination in vitro and the folding of proinsulin in the ER of a mammalian cell. Because the protocol of chain combination (42) is generally robust (permitting the synthesis of hundreds of insulin analogs during the past 40 years), its exquisite sensitivity to a B5 substitution provides a striking contrast to the near-native structure and activity of the Ala$^{B5}$ analog. The biological importance of this synthetic block is demonstrated by transient-transfection analysis of proinsulin biosynthesis within the ER of living cells (43). In this assay a variety of amino-acid substitutions and deletions in proinsulin – in each case expected not to perturb bioactivity – have been found to perturb folding (44). The present studies strongly suggest that His$^{B5}$ functions in the ER to direct disulfide pairing in a kinetic pathway of protein folding. We propose that nascent packing of His$^{B5}$ against the A domain in a protein-folding intermediate orients Cys$^{A7}$ and Cys$^{B7}$ for disulfide bond formation and that such conserved interactions are dispensable once native folding has been achieved. We thus envisage that the His$^{B5}$→Ala substitution provides an example of a pathway mutation (45): its dramatic and unexpected effects on proinsulin biosynthesis illuminate the stringent structural requirements of
native oxidative protein folding in the endoplasmic reticulum.

MATERIALS AND METHODS

Synthesis of Insulin Analogs. Human insulin was kindly provided by Eli Lilly and Co (Indianapolis, IN). The tetra-S-sulfonate derivative of the human A chain was obtained by oxidative sulfitolysis (20). B-chain analogs were prepared by solid-phase chemical synthesis (20). DKP-insulin and AlaB5-DKP-insulin were prepared by chain combination (20,46). Predicted molecular masses were confirmed by matrix-assisted laser desorption ionization (MALDI-TOF) mass spectrometry (MS).

Proinsulin Expression and Mutagenesis. A human proinsulin cDNA was subcloned into the pcDNA3 vector (43). Variant proinsulin cDNAs encoding B5 mutations were created by three PCR reactions. Reaction 1 used forward and reverse primers to amplify a cDNA stretch encoding the signal peptide and the downstream site of the introduced mutation. Reaction 2 used another set of primers to amplify a cDNA encoding the mutation point as well as the remaining carboxyl-terminal region of proinsulin. The products from reactions 1 and 2 were designed to share 20-bp overlap in the region of the introduced mutation. Finally, a third PCR reaction used the primers at the 5' and 3' ends of the proinsulin coding sequence along with both gel-purified products from the first two PCR reactions as template, thereby generating the full-length mutant proinsulin cDNA. These were gel-purified and ligated into the pGEM T-vector (Promega). Mutations were confirmed by direct DNA sequencing, and the mutants subcloned into pcDNA3 using KpnI and EcoRI restriction sites (43).

Mammalian Cell Culture and Transfection. 293T cells were cultured in high-glucose Dulbecco's modified Eagle's medium plus 10% fetal bovine serum and 0.1% penicillin-streptomycin at 37 °C with 5% CO2. Cells were transfected with plasmid DNA 1 day after seeding using LipofectAMINE (InVitrogen) in OptiMEM for 5 h at 37 °C and then recovered overnight in medium containing 10% fetal bovine serum, with a change to normal growth medium at 24 h and assays of protein expression at 48 h.

Metabolic Labeling. Cells were pre-incubated in methionine/cysteine-deficient medium for 30 min and then metabolically labeled in the same medium containing 35S-labeled methionine and cysteine for 1 h, washed once in a large volume of complete chase medium, and then chased further in complete medium for 1 h. At the end of the chase, medium was collected, and the 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 (pH 7.4).

Immunoprecipitation and SDS-PAGE. Cell lysates and chase media that had been treated with a proteinase inhibitor cocktail (Roche Applied Science, Indianapolis, IN) were pre-incubated with Zysorbin (Zymed Laboratories Inc.) for 60 min at 4 °C and pre-cleared by centrifugation. Supernates were immunoprecipitated with guinea pig anti-insulin (Linco Diagnostics) plus fresh Zysorbin. Immunocomplexes were sedimented at 12,000 x g for 1 min, and pellets were washed twice with cell- lysis buffer and once in high-salt buffer (0.5 M NaCl, 1% Triton X-100, 10 mM EDTA, and 25 mM Tris (pH 7.4)). Samples were then analyzed by nonreducing Tris-Tricine-Urea-SDS-PAGE (43): samples were allowed to enter the gel at 50 V for 1 h and then run at 100 V overnight without cooling. Insulin gels were fixed initially in 20% trichloroacetic acid without alcohol, then in 12.5% trichloroacetic acid plus 50% methanol, then incubated briefly with water, and finally either phosphorimaged or incubated with 1 M sodium salicylate for 20 min and exposed to XAR film at -70 °C.

Receptor-Binding Assays. Relative activity is defined as the ratio of analog to wild-type human insulin required to displace 50 percent of specifically bound 125I-human insulin. A human placental membrane preparation containing the insulin receptor (IR) was employed as described (29). In all assays the percentage of tracer bound in the absence of competing ligand was less than 15% to avoid ligand-depletion artifacts.

Circular Dichroism. Far-ultraviolet (UV) CD spectra were obtained as described (8). Spectra, acquired with an Aviv spectropolarimeter (Aviv Biomedical, Inc., Lakewood, NJ), were normalized by mean residue ellipticity. Estimates of secondary structure were obtained by deconvolution (47). Samples were dissolved in 10 mM phosphate and 100 mM KCl (pH 7.4) at a
protein concentration of ca. 25 μM. For equilibrium denaturation studies samples were diluted to 5 μM; guanidine-HCl was employed as denaturant (26). Data were obtained at 4 °C and fitted by non-linear least squares to a two-state model (48).

NMR Spectroscopy. Spectra were obtained at 700 MHz at a protein concentration of 600 μM using a cryogenic 1H-NMR probe. The DKP-insulin template contains substitutions HisB10→Asp, ProB28→Lys, and LysB29→Pro (39-41). Three solution conditions were employed as described (5,22,26): in aqueous solution at pH 7.0 and 25 °C; at pH 8.0 and 32 °C; and in 20% v/v deuterioacetic acid at pH 1.9 and 25 °C. The latter co-solvent facilitates analysis of exchangeable amide resonances, otherwise incomplete at neutral pH due to base-catalyzed solvent exchange and conformational broadening (5; in this co-solvent insulin retains a native-like monomeric fold (49)). Resonance assignment was based on homonuclear 2D NOESY (mixing times 80 and 200 ms), total correlation spectroscopy (TOCSY; mixing time 55 ms), and double-quantum filtered correlated spectroscopy (DQF-COSY). Helix-related hydrogen bonds were inferred from the pattern of protected amide resonances as observed in D2O solution containing 20% deuterioacetic acid (50).

Histidine pKa Measurement. The pK_a of HisB5 in DKP-insulin was estimated based on the pH dependence of the chemical shift of its H_e imidazole resonance in D2O. The following pD values were employed: 2.5, 3.0, 3.5, 4.0, 6.3, 6.8 7.2, 7.5, 8.0, 8.8, 9.3 and 9.8 (direct meter reading).

Molecular Modeling. Structures were calculated by distance-geometry/simulated annealing (DG/SA) using program DG-II (51); restrained molecular dynamics (RMD) calculations were performed using X-PLOR (52).

RESULTS

We first describe the synthesis, activity, and solution structure of an AlaB5 analog of an engineered insulin monomer (DKP-insulin; 39,40). Such engineering (53) is required to avoid dimerization and higher-order protein assembly. We then investigate effects of B5 substitutions on the expression, folding, and secretion of proinsulin in a transfected mammalian cell line.

AlaB5 impairs insulin chain combination. Substitution HisB5→Asp was introduced into the DKP B chain; chain combination employed 40 mg of wild-type A chain and 20 mg of the variant B chain (each as S-sulfonate derivatives). Whereas this protocol ordinarily yields 4.0-4.4 mg of HPLC-purified DKP-insulin, the yield of AlaB5-DKP-insulin was 0.2-0.3 mg; similar reductions were observed in three combination reactions. Although the low yields may have been accentuated by less efficient recovery from the reverse-phase C8 column due to the smaller quantity of protein loaded, the AlaB5 substitution also reduced the prior insulin fraction (following CMC chromatography of the crude reaction mixture) by at least 10-fold. The limited efficiency of chain combination in this case stands in contrast to the general robustness of this reaction (26).

AlaB5 Insulin Analog is Active and Well Folded. Despite the poor yield of chain combination, repetition of the protocol enabled purification of sufficient quantities of AlaB5-DKP-insulin for analysis of activity and structure. The analog exhibits a receptor-binding affinity of 129 percent relative to native human insulin (the same value was obtained in two replicates; column 1 in Table 1). Because the relative affinity of DKP-insulin is 161 ± 19 under these conditions (enhanced binding is due to the AspB10 substitution; 28), specific effects of the B5 substitution on affinity relative to the parent monomer are 80 ± 10 percent. Retention of near-native receptor-binding affinity is in accord with the native in vivo potency of AlaB5-insulin in a rat hypoglycemic assay but higher than that originally reported by Sieber and coworkers in in vitro assays of insulin-stimulated glucose oxidation (21%) and lipolysis (31%) in isolated adipocytes (34). These lower values (obtained in 1979) may reflect incomplete purification of the analog by counter-current distribution prior to the advent of reverse-phase HPLC methods. The far-UV CD spectrum of AlaB5-DKP-insulin (dashed-dot line in Fig. 2A) resembles that of DKP-insulin (solid line in Fig. 2A) with similar but slightly attenuated α-helix-specific features at 222, 208, and 196 nm.

1H-NMR spectra of DKP-insulin and AlaB5-DKP-insulin likewise exhibit similar patterns of
resonance line widths and chemical-shift dispersion (Fig. 2C). The spectrum of the analog is tractable by homonuclear 2D-NMR methods (Fig. 3; ref. 54), permitting resonance assignment (Supplemental Material). Chemical shifts are essentially identical to those observed in DKP-insulin; changes of magnitude > 0.1 ppm are observed only at neighboring sites in the B5-related pocket. The novel AlaB5 spin system is well resolved in TOCSY and NOESY spectra (Fig. 3B, D). Analysis of secondary structure, based on diagnostic strings of d_{NN}, d_{AN}, d_{i,i+3} and d_{i,i+4} NOEs, is essentially identical to that of DKP-insulin (Supplemental Material). The pattern of protected amide protons in 20% deuterioacetic acid and 80% D2O is likewise the same as observed in DKP-insulin (5).

Inter-residue NOEs in DKP-insulin and AlaB5-DKP-insulin are similar as summarized by diagonal plots (Fig. 4A, B). Long-range NOEs are each case consistent with structures of T-state crystallographic protomers. Inter-chain long-NOEs from the β-CH3 resonance of AlaB5 (highlighted in red in Fig. 4B) are in accord with a rigid-body model derived from the wild-type structure. These include NOEs between AlaB5 and the CysA7 Hα (Fig. 3D) and IleA10 (Supplemental Material). The orientation of the N-terminal segment of the B chain (residues B1-B8) is further defined by native-like long-range NOEs involving PheB1, LeuB6, CysB7, and LeuA13.

The solution structure of AlaB5-DKP-insulin was calculated by distance-geometry and restrained molecular dynamics according to 674 NOE-, J-coupling-, and hydrogen-bond-related restraints (Supplemental Material). The ensemble is similar to the T-like conformation of DKP-insulin (Fig. 4C). The ensemble of A chains (red in Fig. 4C) and B chains (blue) closely matches the mean structure of DKP-insulin (ribbons in Fig. 4C). The root-mean-square deviation (RMSDs) between the two structures is similar to the pairwise RMSD within either ensemble alone. The respective orientations of the A- and B chains (Fig. 4C) and environment of the B5 side chain within an inter-chain crevice (Fig. 4D) are in each case similar. The imprecision of the DG/SA ensemble in this region (although in part due to informational uncertainty in the restraints) resembles the range of structural variation observed among wild-type crystal structures (Fig. 1C). Distances between the β-methyl carbon of AlaB5 to neighboring residues are similar to corresponding distances involving the β-carbon of HisB5 in the solution structure of DKP-insulin (Supplemental Material). Closer distances are observed between the β-carbon of AlaB5 and residue IleA10 than between the β-carbon of HisB5 and IleA10, presumably due to the smaller size of Ala than His. This proximity is enjoined in the DG/RMD calculation by an NOE between the methyl resonances of AlaB5 and IleA10. A similar trend toward smaller B5-related inter-residue distances in the analog is observed in the relationship between AlaB5/ThrA8 and AlaB5/cystine A7-B7; by contrast, the distance from AlaB5 to the α carbon of CysA6 is increased but consistent with the distribution of distances among crystal structures. Such local conformational adjustments are as expected in response to a loss-of-volume substitution. Maintenance of an overall native-like structure and receptor-binding surface is in accord with the substantial receptor-binding activity of the analog but contrast with the impaired efficiency of chain combination.

AlaB5 impairs protein stability. AlaB5-DKP-insulin exhibits decreased thermodynamic stability as inferred from guanidine denaturation studies (Fig. 3B). Fitting of these CD-detected denaturation curves by a two-state model (Table 1) yields ΔΔG values of 1.7 ± 0.1 kcal/mole. This decrement is consistent with loss of HisB5-specific inter-chain packing interactions and hydrogen bonds in an otherwise native-like structure. (In wild-type crystal structures the variable orientation of the B imidazole ring is associated with distinct hydrogen bonding schemes, involving either H-Nδ or H-Nε (but not both) and commonly to the carbonyl oxygens of either A7 or A9.) The existence of such hydrogen bonds in the solution structure of DKP-insulin and their contribution to its stability are in accord with a shift in the pKα of HisB5 by 1 pH unit (from ca. 6.4 to 7.4) as determined by pH titration of the 1H-NMR spectrum.

B5 substitutions impair proinsulin expression and secretion. Transient transfection of mammalian cells with a plasmid expressing human proinsulin provides a useful model for studying proinsulin folding within the endoplasmic
reticulum (Fig. 5A). Following transfection of 293T cells, we thus examined expression, disulfide isomer formation, and secretion of newly-synthesized proinsulin in cells radiolabeled with $^{35}$S-amino acids for 1 h and chased for a further hour. This model enables assessment of effects of 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. Although transfected 293T cells exhibit significant protein over-expression that may compromise the folding efficiency of proinsulin relative to that in β cells, the absence of endogenous proinsulin from these cells makes the detection of transfected proinsulin or variants straightforward.

Labeled proinsulin or variant proinsulins were immunoprecipitated with polyclonal anti-insulin and subjected to nonreducing Tris-Tricine-Urea-SDS-PAGE, which allows examination of distinct proinsulin disulfide isomers that form within the ER (Fig. 5B). In accord with prior studies (43) transfection of the wild-type proinsulin construct gives rise to robust expression, primarily of a fast-migrating species (arrow in Fig. 5B) that efficiently secreted from transfected cells (lane 3; "C") to medium (lane 4; "M"). The most-rapidly migrating species (arrow in Fig. 5B) is efficiently secreted from transfected cells (lane 3; "C") to medium (lane 4; "M"), which typically achieves >95% efficiency by 4 h chase. In addition, there are less-rapidly migrating disulfide isomers that generally represent a minor fraction of proinsulin: these exhibit a lower percentage secretion (highlighted by a bracket in Fig. 5B). Substitution of His$^{B5}$ by Ala results in decreased detection of the variant proinsulin with a marked decrease in the fraction of Ala$^{B5}$-proinsulin molecules achieving the native disulfide-bonded form (lanes 5 and 6 in Fig. 5B). Secretion of the variant proinsulin is undetectable (lane 6; asterisk). The impaired foldability of Ala$^{B5}$-proinsulin is in accord with the inefficiency of disulfide pairing encountered above in the chemical synthesis of Ala$^{B5}$-DKP-insulin.

To assess the generality of this result, we examined alternative B5 substitutions: Phe, Asp, Asn, and Gln. Rationales were as follows: Phe was tested as an alternative aromatic ring that (unlike His) is without hydrogen-bond acceptor or donor; Asn and Gln were chosen as polar nitrogenous side chains capable (like His) of donating and receiving hydrogen bonds; and Asp was included for comparison to Asn. Three replicates of these experiments are shown in Figure 6. In each case the B5 substitution markedly decreases recovery of newly-synthesized proinsulin and, unlike wild-type proinsulin, blocks release from cells ("C") into the medium ("M"). The stringency of the requirement for His at position B5 for folding and secretion is surprising given the marked structural variability observed in detail among wild-type crystal structures (Fig. 1C). It is possible that the B5 substitutions accelerate protein degradation rather than inhibit initial expression of the variant proinsulins, due to impaired formation of one or more disulfide bridges resulting in disulfide mispairing. Indeed, blocked formation of the B7-A7 disulfide bridge in proinsulin is the causative feature underlying the molecular pathogenesis of diabetes mellitus in the Akita mouse (55).

Stability and in vivo folding are uncorrelated. Biophysical studies of two-disulfide insulin- and proinsulin analogs lacking cystine A6-A11 indicate that such species form partial folds of low stability ($\Delta G_u = 1.9 \pm 0.3$ kcal/mole, implying that $\Delta G_u$ is ca. 3 kcal/mole; Table 1 (5,7)). In transfected CHO or 293T cells, removal of the A6-A11 disulfide bridge by mutation of these Cys residues to Ser does not block expression or secretion. Further, this pair-wise substitution produces little detectable change in band mobility by nonreducing Tris-Tricine-Urea-SDS-PAGE (31) despite the presumed loss of structural organization and low stability. This control experiment is illustrated in Figure 5B (lanes 7 and 8). Even more surprising, substitution of His$^{B10}$ by Asp – predicted to augment stability through optimization of segmental helical stability – leads to a paradoxical increase in the production of nonnative disulfide isomers in the ER (44). Together, these studies strongly suggest that the efficiency of folding and secretion in vivo is not correlated with extent of structure and stability in vitro. Impaired foldability of Ala$^{B5}$-proinsulin in 293T cells is thus unlikely to be a side consequence of the effect of the substitution on the thermodynamic stability of the protein once folded.
DISCUSSION

The present study has focused on the contribution of His\textsuperscript{B5} to the folding, structure, and stability of human insulin. Conserved within eutherian mammals, the imidizole ring packs within a solvated crevice at the edge of the protein surface (11). Direct contacts in this crevice, including inter-chain hydrogen bonds and a network of bound water molecules, constrain the orientation of the N-terminal segment of the B chain against the A chain. These interactions in part define the environment of the solvent-exposed A7-B7 disulfide bridge but do not contribute significantly to receptor binding (24,35). In an effort to understand what biological constraints might underlie the conservation of His\textsuperscript{B5}, we chose to investigate an Ala\textsuperscript{B5} analog. Because the analog exhibits near-native structure and activity, we undertook the comparative analysis of foldability in vitro and in cell culture.

Efficiency of Insulin Chain Combination

Chemical synthesis of Ala\textsuperscript{B5}-DKP-insulin was limited by inefficient disulfide pairing in the course of chain combination. Although a practical impediment to structural studies, such inefficiency is remarkable in light of the wide utility of this procedure in the synthesis of many unrelated analogs since its development by Katsoyannis in 1966 (42). The robustness of this reaction is highlighted by the efficient syntheses of two-disulfide analogs lacking either cystine A6-A11 or A7-B7 (5,7,8). These analogs are less well organized than native insulin, and their partial folds exhibit negligible thermodynamic stabilities. Similar efficiencies are observed in the synthesis of partially folded analogs in which the N-terminal A-chain α-helix is destabilized by multiple glycine substitutions (26). Conversely, synthetic yields are not significantly increased by substitutions that enhance stability through optimization of helical N- or C-capping residues (26,29). Lack of correlation between yield and stability provides evidence that insulin chain combination is under kinetic (rather than thermodynamic) control as originally proposed by Katsoyannis and coworkers (20).

Although synthetic failures are seldom reported, the following negative observations may relate to the reaction mechanism and hence to the associated folding pathway: (i) Substitution of invariant Gly\textsuperscript{B8} by L-amino-acids markedly impairs chain combination whereas D substitutions enhance its efficiency (41). The conformation of Gly\textsuperscript{B8} is integral to the direction of the B7-B10 β-turn. The positive B8 φ angle, located in a region of the Ramachandran plane ordinarily “forbidden” to L-amino acids, would be maintained by D substitutions but destabilized by L substitutions. Perturbation of this turn in a reaction intermediate might misalign Cys\textsuperscript{B7} and so impair formation of cystine A7-B7. L-amino-acid substitutions at residue B8 of a single-chain insulin precursor (mini-proinsulin) likewise impair its expression in yeast (56,57). (ii) Interchange of Leu\textsuperscript{B11} and Val\textsuperscript{B12} to create analog [Val\textsuperscript{B11}, Leu\textsuperscript{B11}]-insulin results in very low yields (20). Since substitution of Val\textsuperscript{B12} by Leu alone does not impair yield (21), this low yield is presumably due to the B11 substitution. In the wild-type structure Leu\textsuperscript{B11} packs in the core where it appears to stabilize the conformation of Leu\textsuperscript{B6} and cystine A6-A11. It is possible that Val\textsuperscript{B11} is associated with a packing defect that impairs alignment of Cys\textsuperscript{A6} and Cys\textsuperscript{A11} in a protein-folding intermediate. These examples suggest that insulin chain combination is guided by subtle nascent interactions in the folding chains.

Application to Folding of Proinsulin

Impaired chain combination of Ala\textsuperscript{B5}-DKP-insulin correlates with a block to the folding and secretion of Ala\textsuperscript{B5}-proinsulin and related analogs in a transfected rodent secretory cell line. Although such a correlation would seem natural (58-60), recent studies of proinsulin variants have shown that the biological requirements of foldability extend beyond chemical requirements. Substitution of His\textsuperscript{B10} by Asp, for example, impairs the fidelity of disulfide pairing in the ER despite the stabilizing effect of this substitution in vitro (44). Foreshortening of the connecting peptide in mini-proinsulin analogs, which enhances the efficiency of oxidative refolding in vitro (61), can likewise lead to scrambling of disulfide bridges in the ER (31). We imagine that such constructions perturb the enzymatic machinery of oxidative protein folding and/or the associated ER chaperones. Substitution of His\textsuperscript{B5} by
Ala (or the diverse set of other side chains tested in Fig. 6) is likely to introduce a kinetic block to native disulfide pairing, and hence degradation supervenes.

Because chain combination and in vivo foldability reflect kinetic processes, small decrements in thermodynamic stability cannot in themselves account for relative yields. The absence of thermodynamic control is demonstrated through control studies of an unstable two-disulfide analog in which cystine A6-A11 is pairwise substituted by Ser. Despite its partial unfolding and loss of a sigmoidal (cooperative) unfolding transition (7), chain combination and ER foldability are robust. We nevertheless suggest that the loss of HisB5-related inter-chain hydrogen bonds (apparent in the structure of AlaB5-DKP-insulin) and associated decrease in stability of the folded state (ΔΔGm 1.4 kcal/mole; inferred from guanidine denaturation studies) are relevant to impaired foldability. In particular, we imagine that these interactions are critical to the structure and metastability of a protein-folding intermediate and serve to guide the orientation of CysA7 and CysB7 for productive pairing. This intermediate is likely to contain cystine A20-B19, the first disulfide bridge to be identified in in vitro refolding studies of mini-proinsulin (62) and insulin-related growth factor I (63,64). Unlike pair-wise substitution of cystine A6-A11, mutations at A20 or B19 block productive folding and secretion in transfected cells (43).

In a reversible two-state folding reaction – a simpler process than multi-step oxidative folding – the distinction between effects of mutations on the native state versus the transition state is well understood (Fig. 7). Mutations that impair interactions involved in both states (Fig. 7A) increase the barrier height from the unfolded state (activation free energy ΔG*) and reduce native stability (ΔGm). By contrast, mutations at sites uninvolved with nascent interactions in the transition state (Fig. 7B) impair stability without effect on the barrier height from the unfolded state. These principles are likely to apply to the discrete steps in an oxidative folding pathway. We therefore imagine that HisB5 participates in both nascent inter-chain interactions preceding formation of cystine A7-B7 and stabilizes the subsequent disulfide-bridged species. Substitution of IleA2 by Gly, destabilizing but without effect on chain combination yield (26), might provide an example of the other class. It would be of future interest to investigate in the present 293T cell transfection assay whether the expression and secretion of proinsulin would be perturbed by this non-conservative substitution.

**Complementary Determinants of Foldability and Function**

Although the structure of an insulin monomer in solution resembles the crystallographic T state, in which HisB5 packs within an inter-chain crevice, zinc-insulin hexamers exhibit a ligand-dependent equilibrium among T6, T3R3 f, and R6 conformations (11). Whereas in the T state the B chain contains an extended N-terminal arm (residues B1-B6) and type II' β-turn (B7-B10) followed by the central α-helix (B9-B19), in the alternative R state the B chain contains an extended α-helix (residues B1-B19; 65,66). This structural transition thus alters the environment of HisB5. In a putative R-state monomer HisB5 and LeuB6 would extend from a novel helical surface; GlyB8 would exhibit a negative φ angle and thus escape from the “forbidden” right side of the Ramachandran plane. The relevance of an R-like structure to receptor binding has long been the subject of speculation (11,41,49).

Respective D- or L substitutions at B8 shift the equilibrium among T6, T3R3 f, and R6 hexamers in one direction or the other (41). D-amino acids at B8 favor the T state but block receptor binding; L substitutions destabilize the T state but can be highly active. Remarkably, among such chiral analogs efficiencies of chain combination are anti-correlated with receptor-binding activities. Although D substitutions cannot be tested in vivo, we imagine that the folding-competent conformation of proinsulin resembles the classical T state, including nascent interactions between HisB5 and the A chain. Whereas this conformation is retained in the free hormone, insulin may undergo a change in conformation on receptor binding: the bound state may exhibit R-like features (11,41,66). Because the TR transition is also associated with changes in the conformation and solvent accessibility of the A7-B7 disulfide bridge, it is plausible that these T-like or R-like conformations, if populated in protein-folding
intermediates, could favor or retard the kinetic mechanism of disulfide pairing. We thus envisage that His\textsuperscript{B5} and D-amino-acid substitutions at B8 contribute to the foldability of insulin by an analogous mechanism: stabilization of a local T-like structure near cystines A7 and B7.

**Concluding Remarks**

His\textsuperscript{B5} is conserved among eutherian mammals and birds but is not invariant among hystricomorph mammals. Whereas the guinea pig retains His\textsuperscript{B5}, for example, the divergent insulins of the coypu and chinchilla rat contain Arg. Similarly, whereas fish insulins frequently retain His\textsuperscript{B5} (including in the primitive hagfish), Arg is observed among ratfish, rabbit fish, and elephant fish. Arg\textsuperscript{B5} is also observed among reptiles. Accordingly, it would be of future interest to investigate the structure and foldability of Arg\textsuperscript{B5}-insulin analogs. In contrast to insulin itself, residue B5 is not conserved as His or Arg among insulin-related growth factors (67,68) or more distant members of the insulin-related superfamily (69), including among invertebrates (9).

The marked effects of B5 substitutions on foldability stand in contrast to the high biological activity of Ala\textsuperscript{B5}-DKP-insulin, once its folding has been achieved. We imagine that interactions between His\textsuperscript{B5} and the A chain are disrupted on binding of insulin to the insulin receptor. Such induced fit may be related to an R-like conformational change (41,66). The small changes in the free energy of binding observed on deletion of residues B1-B5 (\(\Delta\Delta G_\text{b} \approx 1\ \text{kcal/mole}\)) or mutation of B5 to Ala (\(\Delta\Delta G_\text{b} < 0.5\ \text{kcal/mole}\)) may reflect modulation of this conformational equilibrium or minor adjustments at the edge of the hormone-receptor interface. Conservation of His\textsuperscript{B5} among eutherian mammals is enjoined not by requirements of receptor binding but instead by the pathway of oxidative protein folding in the endoplasmic reticulum.

**Supplemental Material.** Four figures providing additional NMR spectra and summary of sequential assignment; and four tables giving selected inter-residue distances, restraint information and DG/SA statistics, resonance assignments and chemical-shift information.

**Protein Data Bank Accession Code.** The coordinates of Ala\textsuperscript{B5}-DKP-insulin will be deposited on acceptance in the RCSB Protein Data Bank (accession number pending).

**REFERENCES**

1. Steiner, D. F. (1967) Trans. N. Y. Acad. Sci. 30(1), 60-68
2. Huang, X. F., and Arvan, P. (1994) J. Biol. Chem. 269(33), 20838-20844
3. Steiner, D. F. (1998) Curr. Opin. Chem. Biol. 2(1), 31-39
4. Huang, G. S., and Oas, T. S. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 6878-6882
5. 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(2), 390-403
6. Dai, Y., and Tang, J. G. (1996) Biochim. Biophys. Acta 1296(1), 63-68
7. Weiss, M. A., Hua, Q.-X., Jia, W., Chu, Y.-C., Wang, R.-Y., and Katsoyannis, P. G. (2000) Biochemistry 39(50), 15429-15440
8. Hua, Q.-X., Nakagawa, S. H., Jia, W., Hu, S. Q., Chu, Y.-C., Katsoyannis, P. G., and Weiss, M. A. (2001) Biochemistry 40, 12299-12311
9. Hua, Q. X., Nakagawa, S. H., Wilken, J., Ramos, R. R., Jia, W., Bass, J., and Weiss, M. A. (2003) Genes Dev. 17, 826-831
10. Yan, H., Guo, Z. Y., Gong, X. W., Xi, D., and Feng, Y. M. (2003) Protein Sci. 12(4), 768-775
11. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., and Reynolds, C. D. (1988) Phil. Trans. Royal Soc. London Ser. 319(1195), 369-456
12. Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I. J., Blundell, T. L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J., and Gammeltoft, S. (1976) Nature 259(5542), 369-373
13. Liang, D. C., Chang, W. R., and Wan, Z. L. (1994) Biophysical chemistry 50(1-2), 63-71
14. De Meyts, P., and Whittaker, J. (2002) Nat. Rev. Drug Discovery 1, 769-783
15. Kitagawa, K., Ogawa, H., Burke, G. T., Chanley, J. D., and Katsoyannis, P. G. (1984) Biochemistry 23(7), 1405-1413
16. Nakagawa, S. H., and Tager, H. S. (1986) J. Biol. Chem. 261(16), 7332-7341
17. Mirmira, R. G., and Tager, H. S. (1989) J. Biol. Chem. 264(11), 6349-6354
18. Mirmira, R. G., Nakagawa, S. H., and Tager, H. S. (1991) J. Biol. Chem. 266(3), 1428-1436
19. Nakagawa, S. H., and Tager, H. S. (1992) Biochemistry 31(12), 3204-3214
20. Hu, S. Q., Burke, G. T., Schwartz, G. P., Fenderigos, N., Ross, J. B., and Katsoyannis, P. G. (1993) Biochemistry 32(10), 2631-2635
21. Nakagawa, S. H., Tager, H. S., and Steiner, D. F. (2000) Biochemistry 39(51), 15826-15835
22. Xu, B., Hua, Q. X., Nakagawa, S. H., Jia, W., Chu, Y. C., Katsoyannis, P. G., and Weiss, M. A. (2002) J. Mol. Biol. 316(3), 435-441
23. Huang, K., Xu, B., Hu, S. Q., Chu, Y. C., Hua, Q. X., Qu, Y., Li, B., Wang, S., Wang, R. Y., Nakagawa, S. H., Theede, A. M., Whittaker, J., De Meyts, P., Katsoyannis, P. G., and Weiss, M. A. (2004) J. Mol. Biol. 341(2), 529-550
24. Kristensen, C., Kjeldsen, T., Wiberg, F. C., Schaffer, L., Hach, M., Havelund, S., Bass, J., Steiner, D. F., and Andersen, A. S. (1997) J. Biol. Chem. 272(20), 12978-12983
25. Weiss, M. A., Nakagawa, S. H., Jia, W., Xu, B., Hua, Q. X., Chu, Y. C., Wang, R. Y., and Katsoyannis, P. G. (2002) Biochemistry 41(3), 809-819
26. Hua, Q. X., Chu, Y. C., Jia, W., Wang, R. Y., Katsoyannis, P. G., and Weiss, M. A. (2002) J. Biol. Chem. 277, 43443-43453
27. Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Mercola, D. A., and Vijayan, M. (1971) Nature 231(5304), 506-511
28. Schwartz, G. P., Burke, G. T., and Katsoyannis, P. G. (1987) Proc Natl Acad Sci U S A 84(18), 6408-6411
29. Weiss, M. A., Hua, Q.-X., Jia, W., Nakagawa, S. H., Chu, Y.-C., Hu, S.-Q., and Katsoyannis, P. G. (2001) J. Biol. Chem. 276, 40018-40024
30. Carroll, R. J., Hammer, R. E., Chan, S. J., Swift, H. H., Rubenstein, A. H., and Steiner, D. F. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 8943-8947
31. Liu, M., Ramos-Castaneda, J., and Arvan, P. (2003) J. Biol. Chem. 278(17), 14798-14805
32. Hua, Q. X., Shoelson, S. E., Kochoyan, M., and Weiss, M. A. (1991) Nature 354(6350), 238-241
33. Olsen, H. B., Ludvigsen, S., and Kaarsholm, N. C. (1996) Biochemistry 35(27), 8836-8845
34. Marki, F., de Gasparo, M., Eislер, K., Kamber, B., Riniker, B., Rittel, W., Sieber, P., and Hoppe Seylers, Z. (1979) Physiol. Chem. 360(11), 1619-1632
35. Burke, G. T., Hu, S. Q., Ohta, N., Schwartz, G. P., Zong, L., and Katsoyannis, P. G. (1990) Biochem. Biophys. Res. Commun. 173(3), 982-987
36. Schwartz, G., and Katsoyannis, P. G. (1978) Biochemistry 17(21), 4550-4556
37. Geiger, R., and Langner, D. (1981) Hoppe-Seyler's Zeit. Physiol. Chem. 362(6), 799-808
38. Nakagawa, S. H., and Tager, H. S. (1991) J. Biol. Chem. 266(18), 11502-11509
39. Weiss, M. A., Hua, Q. X., Lynch, C. S., Frank, B. H., and Shoelson, S. E. (1991) Biochemistry 30(30), 7373-7389
40. Shoelson, S. E., Lu, Z. X., Parlautan, L., Lynch, C. S., and Weiss, M. A. (1992) Biochemistry 31(6), 1757-1767
41. Nakagawa, S. H., Zhao, M., Hua, Q. X., Hu, S. Q., Wan, Z. L., Jia, W., and Weiss, M. A. (2005) Biochemistry 44(13), 4984-4999
42. Katsoyannis, P. G. (1966) Science 154(3756), 1509-1514
43. Liu, M., Li, Y., Cavener, D., and Arvan, P. (2005) J Biol Chem 280(14), 13209-13212
44. Zhang, B. Y., Liu, M., and Arvan, P. (2003) J. Biol. Chem. 278(6), 3687-3693
45. Betts, S., and King, J. (1999) Structure Fold Des. 7(6), R131-R139
46. Chance, R. E., Hoffman, J. A., Kroeff, E. P., Johnson, M. G., Schirmer, W. E., and Bormer, W. W. (1981) The production of human insulin using recombinant DNA technology and a new chain
recombination procedure. In: Rich, D. H., and Gross, E. (eds). Peptides: Synthesis, Structure and Function; Proceedings of the Seventh American Peptide Symposium, Pierce Chemical Co., Rockford, IL.

47. Sreerama, N., and Woody, R. W. (1993) Anal. Biochem. 209, 32-44
48. Sosnick, T. R., Fang, X., and Shelton, V. M. (2000) Meth. Enzymol. 317, 393-409
49. Hua, Q. X., and Weiss, M. A. (1991) Biochemistry 30(22), 5505-5515
50. Hua, Q. X., Jia, W., Frank, B. H., and Weiss, M. A. (1993) J. Mol. Biol. 230(2), 387-394
51. Xu, B., Hua, Q. X., Nakagawa, S. H., Jia, W., Chu, Y. C., Katsoyannis, P. G., and Weiss, M. A. (2002) Protein Science 11(1), 104-116
52. Brünger, A. T. (1993) XPLOR Manual, Version 3.1, Yale University Press, New Haven, CT
53. Brange, J., Ribel, U., Hansen, J. F., Dodson, G., Hansen, M. T., Havelund, S., Melberg, S. G., Norris, F., Norris, K., and Snel, L. (1988) Nature 333, 679-682
54. Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York, NY
55. Wang, J., Takeuchi, T., Tanaka, S., Kubo, S. K., Kayo, T., Lu, D., Takata, K., Koizumi, A., and Izumi, T. (1999) J. Clin. Invest. 103(1), 27-37
56. Zhang, Y. S., Feng, Y. M., and Dodson, G. G. (2000) Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai) 32(1), 1-3
57. Guo, Z. Y., Zhang, Z., Jia, X. Y., Tang, Y. H., and Feng, Y. M. (2005) Acta Biochim Biophys Sin (Shanghai) 37(10), 673-679
58. Katsoyannis, P. G., and Tometsko, A. (1966) Proc. Natl. Acad. Sci. U. S. A. 55(6), 1554-1561
59. Tang, J. G., and Tsou, C. L. (1990) Biochem. J. 268(2), 429-435
60. Heath, W. F., Belagaje, R. M., Brooke, G. S., Chance, R. E., Hoffmann, J. A., Long, H. B., Reams, S. G., Roundtree, C., Shaw, W. N., Slicker, L. J., and et al. (1992) J. Biol. Chem. 267(1), 419-425
61. Markussen, J. (1985) Int. J. Pept. Protein Res. 25(4), 431-434
62. Qiao, Z. S., Guo, Z. Y., and Feng, Y. M. (2001) Biochemistry 40(9), 2662-2668
63. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992) Biochemistry 31(6), 1749-1756
64. Miller, J. A., Narhi, L. O., Hua, Q. X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L., and Weiss, M. A. (1993) Biochemistry 32(19), 5203-5213
65. Bentley, G., Dodson, E., Dodson, G., Hodgkin, D., and Mercola, D. (1976) Nature 261(5556), 166-168
66. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, G. D., Sparks, C., and Swenson, D. (1989) Nature 338(6216), 594-596
67. Vajdos, F. F., Ultsch, M., Schaffer, M. L., Deshayes, K. D., Liu, J., Skelton, N. J., and de Vos, A. M. (2001) Biochemistry 40(37), 11022-11029
68. Brzozowski, A. M., Dodson, E. J., Dodson, G. G., Murshudov, G. N., Verma, C., Turkenburg, J. P., de Bree, F. M., and Dauter, Z. (2002) Biochemistry 41(30), 9389-9397
69. Eigenbrot, C., Randal, M., Quan, C., Burnier, J., O'Connell, L., Rinderknecht, E., and Kossiakoff, A. A. (1991) J. Mol. Biol. 221(1), 15-21
70. Bajaj, M., Blundell, T. L., Horuk, R., Pitts, J. E., Wood, S. P., Gowan, L. K., Schwabe, C., Wollmer, A., Gliemann, J., and Gammeltoft, S. (1986) Biochem J. 238(2), 345-351
71. Hellman, U., Wernstedt, C., Westermark, P., O'Brien, T. D., Rathbun, W. B., and Johnson, K. H. (1990) Biochem. Biophys. Res. Comm. 169(2), 571-577
72. Nishi, M., and Steiner, D. F. (1990) Mol. Endocrinol. 4(8), 1192-1198
73. Wan, Z., Xu, B., Chu, Y. C., Katsoyannis, P. G., and Weiss, M. A. (2003) Biochemistry 42(44), 12770-12783
74. Wan, Z. L., Huang, K., Xu, B., Hu, S. Q., Wang, S., Chu, Y. C., Katsoyannis, P. G., and Weiss, M. A. (2005) Biochemistry 44(13), 5000-5016
75. Kobayashi, M., Ohgaku, S., Iwasaki, M., Maegawa, H., Shigeta, Y., and Inouye, K. (1982) Biochem. Biophys. Res. Commun. 107(1), 329-336
FOOTNOTES

*Acknowledgements. This work was supported by NIH grant DK0697674 to P.A. and M.A.W. The authors are grateful for P. G. Katsoyannis for his advice and encouragement. We thank S. H. Nakagawa for receptor-binding assays; T. L. Blundell, G. G. Dodson, A. Ferscht, M. Karplus, and J. Whittingham for discussion, Shuhua Wang for assistance with peptide synthesis; and S. Price for assistance with the manuscript. This Article is a contribution from the Cleveland Center for Structural Biology.

1Abbreviations. CD, circular dichroism; DG, distance geometry; DKP-insulin, insulin analog containing three substitutions in B chain (Asp<sup>B10</sup>, Lys<sup>B28</sup> and Pro<sup>B29</sup>); DQF-COSY, double-quantum-filtered correlated spectroscopy; DTT, dithiothreitol; ER, endoplasmic reticulum; GA, Golgi apparatus; HPLC, high-performance liquid chromatography; IR, insulin receptor; kDa, kilo-dalton of mass; MALDI, matrix assisted laser desorption ionization; NMR, nuclear magnetic resonance; MS, mass spectrometry; NOEs, nuclear Overhauser enhancements; NOESY, NOE spectroscopy; RMD, restrained molecular dynamics; rp-HPLC, reverse-phase HPLC; SA, simulated annealing; RMSD, root-mean-square deviation; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TOCSY, total correlation spectroscopy; TOF, time of flight; and UV, ultra-violet. Δ<sub>Gb</sub> indicates change in free energy on ligand-receptor association (Δ<sub>Gb</sub> = RTln(K<sub>eq</sub>)) whereas Δ<sub>Gu</sub> indicates change in the free energy of a protein on unfolding; ΔG* indicates activation free energy. Amino acids are designated by standard one- and three-letter codes.

2Hystricomorph mammals exhibit divergent insulin sequences, including at B5. Such insulins exhibit low affinity and do not form Zn hexamers (70). Substitution of His<sup>B10</sup> by Asn prevents Zn-mediated assembly. Fibrillation of the monomeric insulin in the degu rat is associated with islet amyloidosis (71,72).

3Guanidine denaturation studies of Asp<sup>B10</sup>-insulin suggest that native-state stability is enhanced by 1.0 ± 0.1 kcal/mole (Table 1; ref. 29). Further, chain combination of Asp<sup>B10</sup>-insulin proceeds with an efficiency slightly higher than that of wild-type insulin (Table 1); by contrast with the formation of nonnative Asp<sup>B10</sup>-proinsulin isomers in transfected 293T cells, chemical synthesis of Asp<sup>B10</sup>-insulin does not yield isomers.

4The TR transition is also characterized by a change in the handedness of cystine A7-B7. The sulfur atoms of the latter are exposed in the T state but buried in a nonpolar crevice in the R state. We speculate that coupling between the B8 φ angle and handedness of cystine A7-B7 may account for the low yield of chain combination in synthesis of L-Ser<sup>B8</sup>-DKP-insulin.

5In classical structures the C-terminal B-chain β-strand covers Ile<sup>A2</sup> and Val<sup>A3</sup>. Analogs containing allo-Ile<sup>A2</sup> (in which the chirality of the β carbon is inverted; 19) or Leu<sup>A3</sup> exhibit native structure but low activity (22,73,74). Destabilization of the B-chain β-strand by substitution of Phe<sup>B24</sup> by D-Phe enhances activity (40,75).

FIGURE LEGENDS

Fig. 1. Structure of proinsulin and B5-related inter-chain crevice. (A) Sequence of human proinsulin: insulin moiety is shown in red (A chain) and blue (B chain). The connecting region is shown in black: flanking dibasic cleavage sites (filled circles) and C-peptide (open circles). (B) Structural model of insulin-like moiety and disordered connecting peptide (dashed black line). The A- and B domains are shown in red and blue, respectively. Cystines (orange) are labeled in yellow boxes. (C) Structure of B5-related crevice in multiple high-resolution crystal structures. His<sup>B5</sup> and surrounding residues A6-A11, B4,
B6 and B7 are shown in 15 independent T-state protomers. Structural variability reflects differences among crystal structures; two families of His$_{\text{B5}}$ side-chain orientations are observed. Residues in A chain are shown in red; residues in B chain are shown in blue; and disulfide bridges are shown in orange. Structural coordinates were obtained from Protein Databank entries: 1APH, 1BPH, 1CPH, 1DPH, 1G7A, 1LPH, 1MSO, 1TRZ, 1TYL, 1TYM, 1ZNI, 2INS and 4INS. Structures were aligned respect to the main-chain atoms of residues A6-A11 and B4-B7.

**Fig. 2.** Spectroscopic analysis of Ala$_{\text{B5}}$ analog of DKP-insulin. (A) Far-UV CD spectra: DKP-insulin (solid line) and Ala$_{\text{B5}}$ analog (open circles). Spectrum of analog exhibits native-like CD spectrum with slight attenuation of helix-related features, presumably due to dynamic instability rather than static loss of helix content. (B) Corresponding guanidine denaturation studies. Despite similarity in CD spectra, Ala$_{\text{B5}}$ analog exhibits a marked rightward shift in guanidine denaturation transition. (C) One-dimensional $^1$H-NMR spectra at pH 7.4 and 25 °C: (upper panel) Ala$_{\text{B5}}$ analog and (lower panel) DKP-insulin. Spectra exhibit similar patterns of chemical-shift dispersion and line widths.

**Fig. 3.** 2D-NMR identification of Ala$_{\text{B5}}$ spin system and B5-related NOEs. TOCSY spectra of (A) DKP-insulin and (B) Ala$_{\text{B5}}$-DKP-insulin in region containing A$_3$X spin systems of Ala side chains. (C and D) Portions of NOESY spectra indicating B5-related NOEs. (C) Contacts of DKP-insulin between His$_{\text{B5}}$ imidazole ring and A chain, and (D) analogous contacts involving B-CH$_3$ resonance of Ala$_{\text{B5}}$. Spectra were observed in D$_2$O at 700 MHz at 32 °C and pD 7.6 (direct meter reading); the mixing time was in each case 55 ms (TOCSY spectra) or 200 ms (NOESY spectra). Additional spectra are provided as Supplemental Material.

**Fig. 4.** Summary of inter-residue NOEs and structural models. (A and B) Diagonal plots showing inter-residue NOEs in DKP-insulin (A) and Ala$_{\text{B5}}$-DKP-insulin (B). Contacts between main-chain protons are shown at lower right; side-chain/side-chain and side-chain/main-chain contacts are shown at upper left. Contacts to B5 side chain are highlighted in each case in red. (C) DG/SA models of Ala$_{\text{B5}}$-DKP-insulin relative to mean structure of DKP-insulin (ribbon). The A chain is shown in red and B chain in blue. The mean secondary structures of the A- and B chains and their relative orientations are not perturbed by the Ala$_{\text{B5}}$ substitution. (D) Side-chain packing near Ala$_{\text{B5}}$ (black) relative to His$_{\text{B5}}$ (magenta); A- and B-chain side chains are otherwise shown in red and blue, respectively. Positions of corresponding side chains (excepting His$_{\text{B5}}$) in a representative model of DKP-insulin (closest to ensemble mean) are shown dark blue.

**Fig. 5.** Pathway of insulin biosynthesis and transient transfection assay for expression and secretion. (A) Nascent proinsulin folds as a monomer in ER wherein zinc-ion concentration is low; in Golgi apparatus zinc-stabilized proinsulin hexamer assembles, which is processed by cleavage of connecting peptide to yield mature insulin. Zinc-insulin crystals are observed in secretory granules. (B) Insulin hexamers dissociate in bloodstream to yield active monomers. (C) SDS-PAGE assay of proinsulin expression in ER (lanes labeled C; “cellular”) and secretion into medium (lanes labeled M; “media”). Transfection assay in 293T cells: wild-type (wt) human proinsulin and variants Ala$_{\text{B5}}$, Phe$_{\text{B5}}$, Asp$_{\text{B5}}$, Asn$_{\text{B5}}$, and Gln$_{\text{B5}}$. For each analog cellular proteins are shown in lane C and secreted proteins in lane M. The most-rapidly migrating species contains the native disulfide-pairing scheme; less rapidly-migrating species represent disulfide isomers. The empty-vector control is shown at far left (rows a and b). Cells were pulse-labeled with $^{35}$S-labeled amino acids for 1 h and chased for 1 h. The cells were lysed, and both lysates (C) and chase media (M) were immunoprecipitated with anti-insulin antiserum.

**Fig. 6.** Diverse B5 substitutions impair foldability and secretion of proinsulin in transfected cells. Three replicates (rows a-c) of transfection assay in 293 cells: wild-type (wt) human proinsulin and variants Ala$_{\text{B5}}$, Phe$_{\text{B5}}$, Asp$_{\text{B5}}$, Asn$_{\text{B5}}$, and Gln$_{\text{B5}}$. For each analog cellular proteins are shown in lane C and
secreted proteins in lane M. The most-rapidly migrating species contains the native disulfide-pairing scheme; less rapidly-migrating species represent disulfide isomers. Empty-vector controls are shown at far left (rows a and b). Cells were pulse-labeled with $^{35}$S-labeled amino acids for 1 h and chased for 1 h. The cells were lysed, and both lysates (C) and chase media (M) were immunoprecipitated with anti-insulin antiserum.

Fig. 7. Reaction coordinates for reversible two-state folding processes. (A) Mutations may destabilize both transition state (TS) and native state (N and N’), increasing the barrier height from the unfolded state (U). (B) Mutations at sites uninvolved in transition state may nonetheless destabilize the native state; no effect is observed on barrier height from the unfolded state. Guanidine denaturation experiments probe $\Delta G_u$ and $\Delta \Delta G_u$ (see Table 1). Although multi-step oxidative folding is more complex, analogous principles are proposed to underlie intermediate events.
Table 1. Synthetic Yields, Thermodynamic Stabilities and Activities of Insulin Analogs

| analog                  | yield a | ΔG_u c | ΔΔG_u d | C_mid e | m f | activity g |
|-------------------------|---------|--------|---------|---------|-----|------------|
| proinsulin              | NA b    | 3.9 ± 0.05 | -----   | 4.7 ± 0.1 | 0.77 ± 0.02 | ND          |
| insulin (In)            | 100     | 4.4 ± 0.05 |-------- | 5.3 ± 0.1 | 0.84 ± 0.01 | 100         |
| Asp^{B10}-In            | 110     | 5.4 ± 0.05 | 1.0 ± 0.1 | 6.0 ± 0.2 | 0.89 ± 0.03 | ND i         |
| DKP-Ing                 | 110     | 4.9 ± 0.05 | 0.5 ± 0.1 | 5.8 ± 0.1 | 0.84 ± 0.01 | 161 ± 19     |
| Ser^{A6, A11}-DKP-In    | 90      | 1.9 ± 0.3 h | 3.5 ± 0.4 | 2.9 ± 0.3 | 0.65 ± 0.07 | 0.1         |
| Ala^{B5}-DKP-In         | <10     | 3.2 ± 0.05 | 1.2 ± 0.1 | 4.6 ± 0.1 | 0.69 ± 0.01 | 129         |

aYield of chain combination is defined following purification of analogs by CM-cellulose chromatography.
bNA, not applicable. cΔG_u (kcal/mole) indicates apparent change in free energy on denaturation in guanidine-
HCl as extrapolated to zero denaturant concentration by a two-state model (48). dΔΔG_u (kcal/mole) indicates
difference in ΔG_u values relative to human insulin. Uncertainties in two-state fitting parameters do not include
possible systematic error due to non-two-state behavior. eC_mid is defined as that concentration of guanidine-HCl
at which 50% of the protein is unfolded. fThe m value (kcal/mol/M) is the slope obtained in plotting unfolding
free energy ΔG_u versus molar concentration of denaturant; this slope may be proportional to the protein surface
area exposed on unfolding. gActivity is defined by affinity for the human placental insulin receptor relative to
human insulin (100%); under these conditions the K_d for native insulin is 0.48 ± 0.06 nM. hDKP-insulin
contains three substitutions in the B chain to prevent self-assembly: the classical dimer interface is destabilized
by substitutions Pro^{B28} → Lys and Lys^{B29} → Pro whereas the hexamer interface is destabilized by His^{B10} → Asp
(39,40). iLack of sigmoidicity in transitions makes uncertain fitting of pre-transition base line (8). jND, not
determined. Past studies indicate that the activity of Asp^{B10}-insulin is similar to that of DKP-insulin (39,40).
Figure 2

A

B

C

ppm

$\lambda$ (nm)

$[\theta] \times 10^3$

$\Delta [\theta]_{222}$

$[\text{Gu HCl}]$ (M)

fractional

analog

parent

Ala$^B_5$

His$^B_5$
Figure 5

A diagram showing the distribution of Zn-insulin crystals in different cellular compartments. The diagram is divided into three sections:

- **A**: The rER (rough endoplasmic reticulum) and Golgi apparatus are shown, with arrows indicating the path of Zn-insulin crystals.
- **B**: The Golgi apparatus is labeled with a disassembly mark, suggesting a process related to crystal disassembly.
- **C**: A gel analysis showing bands representing isomers and native forms. The bands are labeled as follows:
  - Lane 1: Control (C)
  - Lane 2: Proinsulin (M)
  - Lane 3: Ala<sup>B5</sup> (C)
  - Lane 4: Ala<sup>B5</sup> (M)
  - Lane 5: [Ser<sup>A6</sup>, Ser<sup>A11</sup>] (C)
  - Lane 6: [Ser<sup>A6</sup>, Ser<sup>A11</sup>] (M)
  - Lane 7: Native (C)
  - Lane 8: Native (M)

The diagram illustrates the distribution and analysis of Zn-insulin crystals in different conditions, highlighting the role of the Golgi apparatus in crystal disassembly.
Figure 6

proinsulin

|            | empty vector | wt | Ala<sup>B5</sup> | Phe<sup>B5</sup> | Asp<sup>B5</sup> | Asn<sup>B5</sup> | Gln<sup>B5</sup> |
|------------|--------------|----|------------------|------------------|------------------|------------------|------------------|
|            | C            | M  | C                | M                | C                | M                | C                |

A [isomer native]

B [isomer native]

C [isomer native]
Figure 7

A

free energy

U

TS

ΔΔG^* ≈ ΔΔG_u

B

free energy

U

TS

ΔΔG^* = 0

reaction coordinate

ΔΔG_u
A conserved histidine in insulin is required for the foldability of human proinsulin.

Structure and function of an Alab5 analog

Qing-xin Hua, Ming Liu, Shi-Quan Hu, Wenhua Jia, Peter Arvan and Michael A. Weiss

J. Biol. Chem. published online May 25, 2006

Access the most updated version of this article at doi: 10.1074/jbc.M602617200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2006/05/26/M602617200.DC1