

## Age of diabetes onset in the mutant proinsulin syndrome correlates with mutational impairment of protein foldability and stability

Balamurugan Dhayalan<sup>‡,^\*</sup>, Yen-Shan Chen<sup>‡,^</sup>, Yanwu Yang<sup>‡</sup>, Mark Jarosinski<sup>‡</sup>, Deepak Chatterjee<sup>‡</sup>, Rachel Grabowski<sup>§</sup>, Shayan Poordian<sup>§</sup>, Nelson F.B. Phillips<sup>§</sup>, Peter Arvan<sup>¶</sup>, Faramarz Ismail-Beigi<sup>§,†</sup>, and Michael A. Weiss<sup>‡,\*</sup>

From the <sup>‡</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202 USA

Departments of <sup>§</sup>Biochemistry and <sup>†</sup>Medicine, Case Western Reserve University School of Medicine, Cleveland, OH 44106 USA

<sup>¶</sup>Division of Metabolism, Endocrinology and Diabetes, University of Michigan Medical School, Ann Arbor 48105 MI USA

\*To whom correspondence should be addressed: Tel.: +1 317 274 1717 and E-mail, [bdhayal@iu.edu](mailto:bdhayal@iu.edu); Tel.:+1 317 274 7151 and E-mail, [weissma@iu.edu](mailto:weissma@iu.edu). <sup>^</sup>These authors contributed equally to this work.

**Keywords:** Hormone; Protein folding; Folding efficiency; diabetes; insulin

---

Diverse heterozygous mutations in the human insulin gene cause a monogenic diabetes mellitus (DM) syndrome due to toxic misfolding of the variant proinsulin. Whereas mutations that add or remove cysteines (thereby leading to an odd number of thiol groups) generally lead to neonatal-onset DM, non-Cys-related mutations can be associated with a broad range of ages of onset. Here, we compare two mutations at a conserved position in the central B-chain  $\alpha$ -helix: one neonatal in DM onset (Val<sup>B18</sup>→Gly) and the other with onset delayed until adolescence (Ala<sup>B18</sup>). The substitutions were introduced within a 49-residue single-chain insulin precursor optimized for folding efficiency (Zaykov, A., et al. *ACS Chem. Biol.* 9, 683-91 (2014)). Although mutations are each unfavorable, Gly<sup>B18</sup> (a) more markedly perturbs DesDi folding efficiency *in vitro* than does Ala<sup>B18</sup> and (b) more severely induces endoplasmic reticulum (ER) stress in cell-based studies of the respective proinsulin variants. In corresponding two-chain hormone analogs, Gly<sup>B18</sup> more markedly perturbs structure, function and thermodynamic stability than does Ala<sup>B18</sup>. Indeed, the Gly<sup>B18</sup>-insulin analog forms a molten globule with attenuated  $\alpha$ -helix content whereas the Ala<sup>A18</sup> analog retains a natively cooperative structure

with reduced free energy of unfolding ( $\Delta\Delta G_u$ , 1.2( $\pm$ 0.2) kcal/mole relative to Val<sup>B18</sup> parent). We propose that mutations at B18 variably impede nascent pairing of Cys<sup>B19</sup> and Cys<sup>A20</sup> to an extent correlated with perturbed core packing once native disulfide pairing is achieved. Differences in age of disease onset (neonatal or adolescent) reflect relative biophysical perturbations (severe or mild) of an obligatory on-pathway protein folding intermediate.

---

Foldability is an evolved biophysical property of globular protein sequences (1). Whereas amyloid is envisaged as a polypeptide's universal thermodynamic ground state (as a class of heteropolymers) (2,3), biological selection yields sequences capable of efficient folding to form a diversity of globular structures (4). Protection from amyloidogenesis is provided by the local thermodynamic stability of the "native state" and by kinetic barriers limiting its access to a coupled landscape of toxic aggregates (5,6). The informational content of protein sequences thus not only encodes native architectures—as visualized by the mature technologies of structural biology—but also efficient folding trajectories that can avoid kinetic traps and non-native aggregation (5,7). Residue-specific patterns of sequence conservation in a protein reflect

a combination of such evolutionary constraints: some evident in the structure and function of the native state, others unseen as cryptic safeguards against off-pathway trajectories (8).

Interest in foldability and its associated safeguards has been stimulated by the recognition of diseases of protein misfolding (2,9). Classical examples are provided by extracellular amyloidogenesis (10,11), pathological deposits associated with over-secretion of immunoglobulin light chains or  $\beta_2$ -microglobulin (12); similar deposits comprise mutant forms of small globular domains (transthyretin and lysozyme);(13,14). Recent attention has focused on toxic misfolding within the cell (15,16). Mutations that impair folding efficiency can cause monogenic diseases even when the native state, if reached, would be capable of normal function (17). This paradigm pertains to the most common allele of cystic fibrosis (CF)<sup>1</sup>, which contains a deletion of a conserved Phe (Phe508) in the CFTR channel (18). A set of allele-specific small molecules (elixacaftor, tezacaftor, and ivacaftor; Trikafta [Vertex]) has been identified that restores foldability and function in such patients (a therapeutic approach given *Breakthrough* designation by the US Food and Drug Administration) (19). The key role of Phe508 in CFTR's folding mechanism (as unmasked by medical genetics) is not apparent from structural analysis of the native state (20). Defining the cryptic roles of such conserved "safeguards" define biophysical problems of broad translational importance.

Insulin has long provided a model for the biophysical study of protein structure, stability and evolution (21-23). In the past decade a large and diverse collection of heterozygous mutations in the insulin gene (*INS*) has been found in association with non-classical diabetes mellitus (DM) (24-27). Although specific mutations can perturb essentially any step in the complex biosynthetic pathway of insulin or its hormonal function (from Golgi trafficking to prohormone processing to receptor binding)(28), the majority impair the foldability of proinsulin (the single-chain precursor of insulin (29,30)) in the rough endoplasmic reticulum (ER)(31-33). Impaired foldability of proinsulin causes ER stress (32), leading in turn to  $\beta$ -cell dysfunction and eventual  $\beta$ -cell death (34,35). Ages of disease onset of such "proinsulinopathies" from neonatal to early adulthood (31).<sup>2</sup> Mutations that add or remove a Cys (thereby giving rise to an odd number of thiol groups

in nascent proinsulin) ordinarily give rise to neonatal-onset DM (24,25). Dominance of the mutant allele in this syndrome reflects aberrant protein-protein interactions between the wild-type (WT) and variant proinsulins in the ER (36,37), including intermolecular disulfide pairing (38).

The present study focuses on a conserved Val in the central B-chain  $\alpha$ -helix of insulin as a site of DM-associated clinical mutations (39,40). Val<sup>B18</sup> adjoins a critical internal disulfide bridge (cysteine B19-A20), proposed to pair as a key early step in the folding of proinsulin (41-45). The isopropyl side chain of Val<sup>B18</sup> packs within an inter-chain crevice to abut the conserved aliphatic side chains of Leu<sup>A13</sup> and Leu<sup>A16</sup> (Fig. 1 and Supplemental Fig. S3) (46,47). None of these five core side chains (B18, B19, A13, A16 or A20) pack at the primary hormone-receptor interface (Fig. 2) (48-50). Our interest was stimulated by the marked difference in ages of onset between variants Ala<sup>B18</sup> (adolescence) (40,51) and Gly<sup>B18</sup> (infancy) (39). To probe the molecular origins of this difference, we undertook a synthetic study of single-chain insulin precursors and comparative biophysical study of Val<sup>B18</sup>-, Ala<sup>B18</sup>-, and Gly<sup>B18</sup>-insulin analogs. Our approach was extended through cell-biological studies of ER stress on expression of the corresponding variant proinsulins.

To our knowledge, this study represents the first interrogation of a cryptic inter-chain crevice in the insulin molecule. Our findings suggest that native packing of branched aliphatic side chains in this potential space is critical to the foldability of proinsulin, presumably by promoting nascent pairing of cysteine B19-A20. These conserved side chains also buttress and stabilizes the mature hormone's receptor-binding surface. Age of DM onset and ER stress thus correlate with both extent of impaired foldability *in vitro* and extent of structural perturbation in the native state, once reached. We envisage that this correlation will apply generally in the mutant proinsulin syndrome (52).

## Results

### *Chemical synthesis provides an assay of foldability*

To investigate the folding efficiencies of variant insulin analogs, we have chosen DesDi as a synthetic model (53). DesDi (a mini-proinsulin) is a single-chain 49-residue polypeptide t with substitution Pro<sup>B28</sup>→Lys<sup>3</sup> and a Lys<sup>B28</sup>-Gly<sup>A1</sup> peptide bond; it

hence lacks residues B29-B30. Lys<sup>B28</sup> provides a cleavage site (for a Lys-specific endoprotease) to yield an active two-chain insulin analog (Lys<sup>B28</sup>-*des*-dipeptide[B29,B30]-human insulin). DesDi exhibits higher folding efficiency than proinsulin or other mini-proinsulins, presumably due to optimization of the B28-A1 linkage to bias the unfolded-state ensemble toward on-pathway conformations. This template thus enables preparation of mutant insulin analogs not otherwise amenable to insulin chain combination (22,44).

Peptides were obtained by solid-phase peptide synthesis (SPPS) with a heat-assisted coupling/deprotection protocol (see Experimental Procedures and Supplemental Methods). The crude polypeptides were subjected to oxidative folding reactions to assess relative yields. The parent DesDi (with Val<sup>B18</sup>) thus provided a baseline for comparative studies of the clinical variants (Ala<sup>B18</sup> and Gly<sup>B18</sup>). Folding was monitored by analytical reverse-phase high-performance liquid chromatography (rp-HPLC). Ala<sup>B18</sup> DesDi exhibited a ~twofold reduction in folding efficiency whereas Gly<sup>B18</sup> exhibited ~eightfold reduction (Fig. 3). The ratio of yields following semi-preparative rp-HPLC purification followed the same trend (See Supplemental Table S1).

Once folded, the three single-chain precursors were treated with Endo Lys-C endoprotease to obtain corresponding two-chain *des*-dipeptide[B29, B30] insulin analogs (Supplemental Fig. S4-S9). Although Lys<sup>B28</sup>-*des*-dipeptide[B29,B30]-insulin is less stable than WT insulin (and impaired in dimerization), this template enabled comparison of B18 side chains (Val, Ala or Gly) with respect to relative changes in structure, stability and function.

To gain additional insight into *in-vitro* foldabilities (44,54), we next exploited chain combination. The pairing reaction of isolated A- and B chains is ordinarily tolerates a wide variety of substitutions: over the past five decades hundreds of insulin analogs have been so prepared, demonstrating that the information required for proinsulin folding is contained within the A- and B-chain sequences (55). Under the present reaction conditions (see Experimental Procedures and Supplemental Methods), the WT chains (as S-sulfonate derivatives) combined to yield WT insulin in expected yield. In striking contrast, combination of the WT A chain

with either of the two variant B chains (Ala<sup>B18</sup> and Gly<sup>B18</sup>) each gave only trace product (Supplemental Fig. S13-15). Although this failure of chain combination highlights the present mutational perturbations, relative yields did not correlate with respective ages of onset, presumably due to the lower baseline robustness of chain combination relative to the folding of single-chain precursors.

### CD studies uncover mutational perturbations

Far-ultraviolet (UV) circular dichroism (CD) spectra of single-chain analogs essentially had a similar pattern in 200-255 nm range except for Gly<sup>B18</sup> which had attenuated 195 nm peak (Fig. 4A). Deconvolution by SELCON algorithm (56) indicated these analogs with similar helical propensities and Ala being slightly better compared to Val or Gly. (See Supplemental Table S2) Free energies of unfolding at 37 °C as inferred from two-state modeling of chemical denaturation indicated the decremental values in the order Val (more stable) > Ala > Gly (less stable) (Fig. 4B). Noticeable changes were seen in spectra of respective two chain analogs (Fig. 4C). While Gly<sup>B18</sup> found to be completely disordered, Ala<sup>B18</sup> had better helical structure even compared to Val<sup>B18</sup> presumably due to its superior helical propensity when compared to native valine (57,58) (See Table 2 and Supplemental Table S2). SELCON results clearly showed the Gly<sup>B18</sup> being unstructured with 20.6% of  $\alpha$ -helical content compared to Val<sup>B18</sup> (45.2%) or Ala<sup>B18</sup> (47.5%). (Table S2). Guanidine denaturation studies of two chain analogs at 25 °C indicated decreased thermodynamic stabilities for Ala<sup>B18</sup> analog compared to Val<sup>B18</sup> ( $\Delta\Delta G_u$  1.2 ( $\pm 0.2$ ) kcal/mole] (Table 1). Gly<sup>B18</sup> had a parabolic transition indicative of molten globule like structure (Fig. 4D).

Temperature-dependent CD spectra indicated no significant changes in the 200-255 nm range for single-chain analogs (See supplemental Fig. S16, left panel). The peak at 195 nm however found to attenuate going from 4 °C to 25 °C to 37 °C. In the case of two chain analogs (supplemental Fig. S16, right panel), noticeable changes were seen in Gly<sup>B18</sup> where disordered structure at 25 °C or 37 °C gained significant structural features at 4 °C. Whereas Val<sup>B18</sup> and Ala<sup>B18</sup> showed marginal effect with temperature.

### NMR spectra probe hydrophobic-core residues.

Two chain versions of DesDi-Val<sup>B18</sup>, Ala<sup>B18</sup> and Gly<sup>B18</sup> were analyzed by 1D <sup>1</sup>H-NMR spectra, 2D TOCSY, NOESY and <sup>1</sup>H-<sup>13</sup>C HSQC correlation spectra to gain structural insights. The assignments of <sup>1</sup>H and <sup>13</sup>C chemical shift were established through unique through-bond connections in TOCSY and specific NOE pattern in NOESY spectra, and confirmed in <sup>13</sup>C-HSQC correlation spectra (Fig. 5 and 6; also see Supplemental Fig. S17). The Val<sup>B18</sup> analog exhibits <sup>1</sup>H-NMR spectral property of native-like insulin as observed in the aromatic region and in the upfield-shifted methyl region (Fig. 5, far right). The well-resolved methyl resonances indicated as a: Ile<sup>A2</sup>  $\gamma_2$ -CH<sub>3</sub>; b: Ile<sup>A10</sup>  $\gamma_2$ -CH<sub>3</sub>; c: Leu<sup>B15</sup>  $\delta_2$ -CH<sub>3</sub> d: Ile<sup>A2</sup>  $\delta_1$ -CH<sub>3</sub>; e: Ile<sup>A10</sup>  $\delta_1$ -CH<sub>3</sub>; f: Leu<sup>B15</sup>  $\delta_1$ -CH<sub>3</sub>. The signature long-range NOEs from the aromatic protons of Phe<sup>B24</sup> and Tyr<sup>B26</sup> residues to methyl protons of Leu<sup>B15</sup> residue, and that from the aromatic protons of Tyr<sup>A19</sup> to methyl protons of Ile<sup>A2</sup> and Leu<sup>A16</sup>, were also observed, indicating native-like conformation of DesDi-Val<sup>B18</sup> (Fig. 6A,D). In comparison to DesDi-Val<sup>B18</sup>, 1D and 2D homonuclear <sup>1</sup>H-NMR spectra of DesDi-Ala<sup>B18</sup> indicated similar, but not identical, chemical-shift dispersion and typical native-like packing of hydrophobic residues in the hydrophobic core of insulin (Fig. 5, middle panel). The secondary <sup>1</sup>H-NMR chemical shifts of the upfield-shifted side chain of Leu<sup>B15</sup> (a characteristic signature of the Phe<sup>B24</sup> ring current in the native B-chain super-secondary structure) and that of Phe<sup>B24</sup> aromatic protons were attenuated (shifted toward random-coil direction) in the Ala<sup>B18</sup> variant. Similar signature long-range NOEs between methyl protons and those of the aromatic side chain were also observed, except signal broadening related to Phe<sup>B24</sup> probably due to presumptive conformational exchange or self-association (Fig. 6B,E). 2D heteronuclear <sup>1</sup>H-<sup>13</sup>C HSQC correlation spectrum can provide structural footprint at the atomic level. Patterns of nature abundance <sup>1</sup>H-<sup>13</sup>C HSQC NMR spectra of DesDi-Val<sup>B18</sup> and DesDi-Ala<sup>B18</sup> also exhibit typical spectral characteristics of native-like insulin both in aromatic and methyl region (See Supplemental Fig. S17), as observed in 1D and 2D <sup>1</sup>H-NMR spectra. These qualitative <sup>1</sup>H and <sup>13</sup>C spectra features provide evidence of a native-like structure of DesDi-Ala<sup>B18</sup> with a local change in structure and/or dynamics near the site of substitution transmitted to the central B-chain  $\alpha$ -helix and to  $\beta$ -strand portion of the B-chain. The DesDi-Gly<sup>B18</sup> lost its secondary chemical shift dispersion in both aromatic and methyl region (Fig. 5,

top panel) and the signature long-range NOEs between methyl and aromatic protons were not observed (Fig. 6C,F), suggesting a unfolded structure.

The temperature-dependent NMR spectra were also acquired at 12, 22 and 32 °C, respectively (See Supplemental Fig. S18). The chemical shift dispersion of DesDi-Ala<sup>B18</sup> and DesDi-Gly<sup>B18</sup> analogs slightly increased when lowering the temperature. Interestingly, one most upfield shifted peak in the DesDi-Ala<sup>B18</sup> <sup>1</sup>H-NMR spectrum was observed at 12 °C around 0.27 ppm, possibly from Leu<sup>B15</sup>  $\delta_1$ -CH<sub>3</sub> and/or Ile<sup>A10</sup>  $\gamma_2$ -CH<sub>2</sub> protons. Due to the spectral overlap and line broadening of Leu<sup>B15</sup>-related resonances at 12 °C, it was difficult to assign. Therefore, the <sup>1</sup>H-NMR spectrum of DesDi-Ala<sup>B18</sup> was acquired at ~9-times lower concentration at 32 °C (See Supplemental Fig. S19). Well-resolved and upfield-shifted signals were observed in the concentration of ~0.03 mM and the spectral pattern both in aromatic and methyl region was very similar to that of DesDi-Val<sup>B18</sup> analog (Supplemental Fig. S19). Those peaks shifted toward upfield and the chemical shift dispersion increased upon temperature decrease.

#### **Age of disease onset correlates with ER stress**

The endoplasmic reticulum (ER) is the central site for protein folding, post-translational modifications, and transportation (59). The accumulation of misfolded proteins in the ER would induce the expression of glucose-regulated proteins (GRPs) including GRP78/heavy chain binding protein (BiP) and enhanced the expression of the C/EBP Homologous Protein (CHOP) (60). This response is regulated by the signaling pathway initiated by the phosphorylation of PERK (Fig. 7A).

To examine the ER stress effect induced by misfolded proinsulin-driven signaling pathway in mammalian cells, the HEK 293T cell lines were transiently transfected with WT (Val<sup>B18</sup>) and two of mutations (Ala<sup>B18</sup> and Gly<sup>B18</sup>), we investigated the accumulation of ER stress markers in protein level. The phosphorylated PERK kinase exhibited significantly increasing in cells expressing Ala<sup>B18</sup> and Gly<sup>B18</sup> proinsulin. The accumulations of the other two representative marker proteins, CHOP and BiP (61), also increased comparing to the WT proinsulin control and a function-defected mutation without structural impairment (62,63) (Leu<sup>A3</sup>; Fig. 7B). The ER stress also activates *BiP* promoter leading to its

increased mRNA transcription to resolve the stress by properly folding these proinsulins (61). Since these mutated insulin proteins still fail to be folded properly, the Integrated Stress Response (ISR) is induced to attenuate global protein synthesis and the *CHOP* gene is activated for triggering apoptosis. Our result shows that the gene expressions of *CHOP* and *BiP* exhibit significantly increasing (Fig. 7C), which is consistent with the phenomenon of ER stress responses.

### **Mutations at B18 perturb hormonal activity**

Insulin's hinge-opening at a dimer-related interface is coupled to closure of the insulin receptor (IR) ectodomain legs. This engagement leads to TK trans-phosphorylation and receptor activation. Activation of signal pathway was probed by a kinase cascade in HepG2 cells. Quantitative dose-dependent IR autophosphorylation was evaluated in a 96-well plate assay (Fig. 8A,B). Insulin *lispro* (KP)<sup>[2]</sup> and Val<sup>B18</sup> DesDi (WT) insulin triggered a robust autophosphorylation readouts whereas the mutations of Ala<sup>B18</sup> and Gly<sup>B18</sup> DesDi exhibited significantly reduced activities. These data indicate that the requirement of high-dose exposure of Ala<sup>B18</sup> and Gly<sup>B18</sup> DesDi analogs to reach native-like activities (>3000 nM, comparing to KP and WT DesDi for only 50 nM of treatment; Fig. 8B). An overview of phosphorylation of IR and downstream phosphorylation of Ser-Thr protein kinase AKT (protein kinase B; ratio p-AKT/AKT) at different hormone doses (50 nM and 3000 nM) were provided by western blot (Fig. 8C). These data support the findings that these two mutations retain little activities and at least 60-fold higher doses are required for comparable IR-activation function.

The *in vivo* potencies of these analogs, evaluated by subcutaneous (SQ) injection in diabetic rats (Fig. 9) indicated Ala<sup>B18</sup> as having reduced blood glucose lowering capability when compared to Val<sup>B18</sup> (p-value <0.05) (Fig. 9A,B). On the other hand, Gly<sup>B18</sup> had very low activity even at twofold or fourfold higher doses per kilogram (Fig. 9C,D). Although the individual p-values were 0.07 and 0.22, respectively, combining the data supported that hypothesis that Gly<sup>B18</sup> impairs activity with  $p < 0.05$ . It is not known whether the patients with these monogenic diabetes syndromes secrete the mutant insulin (39,40) and thus whether their reduced potencies contribute to disease onset of progression.

## **DISCUSSION**

Insulin sequences have been highly conserved in the evolutionary radiation of vertebrates: critical roles in fetal development and metabolic homeostasis have evidently imposed rigorous functional selection for more than 500 million years (64,65). Such conservation is particularly marked in the hydrophobic core wherein two internal disulfide bridges (cystines B19-A20 and A6-A11) nestle among the invariant side chains of Leu<sup>B6</sup>, Leu<sup>B15</sup>, Phe<sup>B25</sup>, Ile<sup>A2</sup>, Leu<sup>A16</sup> and Tyr<sup>A19</sup> (47). This invariant framework is shared by insulin-like growth factors (IGF-1 and IGF-2), homologs of proinsulin arising from gene duplication in the transition from protochordates at an early stage of vertebrate evolution (65,66). Purifying selection in this family is envisioned at general steps in protein biosynthesis and function: from oxidative folding in the ER (enabling specific disulfide pairing) to intracellular trafficking from the ER to secretory granules and from exocytosis to target-cell signaling (33). A large literature pertains to sequence determinants of receptor binding and activation (67,68). Additional selective pressures presumably accompany distinctive properties of insulin not shared by IGFs (and vice versa): *e.g.*, proteolytic processing of proinsulin to liberate the two-chain hormone and its self-assembly in the secretory granules of pancreatic  $\beta$ -cells (30) (and conversely IGF-specific binding to partner proteins; IGFBPs (69)). The invariant structural framework of insulin and IGFs provides a scaffold for specialization of cognate receptor-binding surfaces in the co-evolution of homologous receptor tyrosine kinases (65).

The mutant proinsulin syndrome has highlighted sequence determinants of nascent folding efficiency (52,70). Whereas classical comparison of species variants enabled study of divergent positions (71), advances in medical genetics has provided (and continues to provide) an ever-expanding database of *INS* mutations as unsuccessful "experiments of nature" (For recent review, see (27)). Such mutations generally affect conserved sites (47). Guided by clinical correlations, the present study thus focused on position B18, broadly conserved among vertebrate insulins as Val (but not invariant).<sup>[4]</sup> Patients have been identified with mutations Ala<sup>B18</sup> (DM onset in adolescence) (40) and Gly<sup>B18</sup> (DM onset in infancy) (39). Because of this salient difference in ages of onset, we sought to investigate the relative impact of

these substitutions on folding, structure, stability and function. Our approach combined chemical protein synthesis with cell-based interrogation of ER stress. Foreshortened single-chain insulin precursor polypeptides ("mini-proinsulins") were prepared by SPPS using the "DesDi" framework established by DiMarchi and colleagues; in this 49-residue precursor Pro<sup>B28</sup> is substituted by Lys and directly followed by Gly<sup>A1</sup> (residues B29 and B30 are absent) (53). Folding of DesDi is more robust than longer precursors (including proinsulin) to unfavorable substitutions, presumably because the B28-A1 peptide bond facilitates formation of nascent structure that guides on-pathway disulfide pairing. In this framework Ala<sup>B18</sup> and Gly<sup>B18</sup> perturbed folding yields (relative to Val<sup>B18</sup> DesDi) by ~twofold and ~8-fold, respectively. A similar pattern of relative perturbations to the thermodynamic stabilities of the respective single-chain analogs was also observed in CD-detected chemical denaturation experiments (Table 1). Although native or near-native  $\alpha$ -helix contacts were preserved in the single-chain analogs, the guanidine-induced unfolding transition of the Gly<sup>B18</sup> analog exhibited marginal cooperativity and a marked reduction in *m*-value (column 3 in table 1). Because the latter probes extent of exposure of nonpolar surfaces on denaturation, the low *m*-value of Gly<sup>B18</sup> DesDi indicates aberrant pre-exposure of such surfaces in the folded state. These *in vitro* studies were extended through measurements of ER stress induced on transient transfection of WT or variant proinsulins in a human cell line: in these assays Gly<sup>B18</sup> led to more robust induction of ER-stress signaling than Ala (Table S3). Ages of DM onset therefore correlate with the relative folding yields, native-state stabilities and degree of cellular ER stress.

To gain additional insight into structure and function, the single-chain synthetic precursors were cleaved by endo-Lys C to yield two-chain insulin analogs. The parent analog (*des*[B29, B30]-Lys<sup>B28</sup>-insulin) exhibits natively-like structure and function. Unlike CD spectra of the single-chain precursors, spectra of the two-chain analogs exhibited significant differences. Although Ala<sup>B18</sup> was associated with a slight accentuation of  $\alpha$ -helical features, this two-chain analog nonetheless exhibited impaired stability ( $\Delta\Delta G_u$  1.2( $\pm$ 0.2) kcal/mole). Strikingly, Gly<sup>B18</sup> induced a twofold attenuation of  $\alpha$ -helix content with complete loss of cooperativity; the partial fold is thus

a molten globule (72). Whereas <sup>1</sup>H- and <sup>13</sup>C-NMR studies of the Ala<sup>B18</sup> two-chain analog corroborated preservation of natively-like chemical shifts and inter-residue NOEs, spectra of the Gly<sup>B18</sup> analog exhibited limited chemical-shift dispersion and attenuated inter-residue NOEs. A subset of long-range NOEs was nonetheless retained (as inferred based on presumptive assignments), suggesting that the Gly<sup>B18</sup> molten globule is not a random coil but instead maintains on average some native-like features (73). We speculate that maintenance of the three canonical disulfide bridges enforces a subset of natively-like tertiary contacts, at least on average in a flexible ensemble of structures. The low but not negligible folding yield of the Gly<sup>B18</sup> single-chain precursor is likely enabled by the B28-A1 peptide bond, which is in effect a fourth bridge that enforces native-like structural relationships (74). Cleavage of the B28-A1 peptide bond further relaxes this constraint leading to a molten partial fold. Because the 35-residue connecting domain of proinsulin (together with residues B29-B30) provides only a long flexible tether between the main-chain atoms of residues B28 and A1, we anticipate that (a) Gly<sup>B18</sup>-proinsulin would likewise form a molten globule and (b) the present DesDi-based synthetic studies of a Gly<sup>B18</sup> single-chain precursor (however markedly perturbed) over-estimate the degree to which this mutation would block the oxidative folding of full-length proinsulin (75,76). It would be of future interest to evaluate the folding yields of Gly<sup>B18</sup> single-chain precursor polypeptides (relative to corresponding Val<sup>B18</sup> and Ala<sup>B18</sup> analogs) as a function of connecting-domain length in the range 0-35 residues.

The mechanistic origins of the perturbations observed in the Ala<sup>B18</sup> and Gly<sup>B18</sup> analogs (relative to WT Val) may be considered in relation to general biophysical principles. Three properties may be pertinent: (a) relative hydrophobic transfer free energies; (77-79) (b) relative helical propensities (57,58,80) and (c) creation of a destabilizing cavity or crevice (81,82) created by the mutations as "large-to-small" substitutions (82). The first two properties reflect intrinsic features of these amino-acid types; the third depends also on structural details of the insulin molecule. This three-part analysis is given in Table 2. The mild perturbation introduced by Ala<sup>B18</sup> is the net result of favorable and unfavorable terms. Whereas Ala exhibits a higher helical propensity than Val (as a  $\beta$ -branched side chain) (57,58,80), its

hydrophobic transfer free energy is lower than that of Val (79). Because Val<sup>B18</sup> is largely buried, net destabilization is predicted on account of Ala's smaller volume, which leads to a cavity penalty (81,82). This cavity penalty is even larger for Gly<sup>B18</sup> as is the hydrophobic transfer free-energy penalty; unlike in the Ala<sup>B18</sup> analog, these unfavorable terms are not compensated or offset by a more favorable helix-propensity term. Our findings are thus in accordance with general models of protein stability (83). Global loss of helix content and organized structure in the Gly<sup>B18</sup> two-chain molten globule is presumably driven by entropic terms, i.e., minimization of the entropic cost of folding to a native state whose putative stability (in the context of the WT structure) would suffer three independent perturbations (a-c).

### **Concluding Remarks**

The diverse mutations associated with the mutant insulin syndrome exhibit a broad range of ages of onset (84). The mildest, with onset in early adulthood, may even exhibit variable genetic penetrance (38,85); the most severe present in the neonatal period (24,25). The mutations typically exhibit genetic dominance such that expression of the variant proinsulin interferes with biosynthesis and secretion of WT proinsulin (33). Although such patients currently require insulin replacement therapy (86), it is possible that in the future such monogenic forms of DM may be treated through pharmacologic or gene-based therapies designed either to attenuate  $\beta$ -cell ER stress or to restore folding to the variant proinsulin (87).

We imagine that a panel of variant proinsulins (beyond Gly and Ala at position 18) might provide spectrum of relative foldabilities--in essence a calibrated "genetic rheostat" of ER stress levels--as a model for development of novel therapeutics: a case study in precision medicine (88). Given the general role of chronic ER stress in non-syndromic Type 2 DM (i.e., as induced by compensatory over-expression of WT proinsulin in the face of obesity-related insulin resistance), such  $\beta$ -cell-directed therapies might address the growing pandemic of prediabetes leading to "diabesity" as a pending crisis in global health (89,90).

### **Experimental procedures**

Additional details pertaining to materials and methods are provided in the Supplement.

#### ***Chemical peptide synthesis***

Peptides were synthesized on Tribute (Gyros Protein technologies) 2-Channel peptide synthesizer using 6-Cl-HOBt and DIC (1:1, equimolar with respect to Fmoc-protected amino acids) as the coupling agents and 20% piperidine in DMF as deprotection agent. After cleavage, peptide was precipitated in ether and dried.

#### ***Folding of DesDi mini-proinsulin precursor and its two-chain conversion***

The crude 49-mer peptides, after the Fmoc-SPPS, was subjected to folding conditions (53): 0.1 mM peptide, 20 mM glycine, 2 mM cysteine at pH 10.5 for 16 hr.

Single-chain DesDi analogs were treated with Endo Lys-C enzyme (91) in 25 mM Tris base, 100 mM urea buffer (pH 8.5) at 12 °C water bath for 24 h. After HPLC indicated two chain conversion (typically 60-80%), the reaction mixture was acidified and purified on semi-preparative HPLC. Fractions containing clean protein were pooled and lyophilized.

#### **Chain Combination**

Wild-type A- and B chains were obtained by sulfitolysis of commercially available human insulin. Variant B chains were made by solid-phase peptide synthesis and converted to their respective S-sulfonate derivatives. For chain-combination reactions, a stock solution of WT A-chain S-sulfonates was first prepared and split three ways before mixing with respective B-chain S-sulfonate analogs (WT, Ala<sup>B18</sup> or Gly<sup>B18</sup>); the reaction proceeded in 0.1 M glycine buffer at pH 10.5 using 1 equiv. of dithiothreitol and stirred at 4 °C for 2 days; details are provided in the Supplement.

#### ***Circular Dichroism (CD)***

Samples were quantitated using Genesis 150 UV-Vis spectrophotometer (Thermo Scientific) using one-cm cuvette by measuring peptide absorbances at 280 nm. The extinction coefficient used was: 6335 M<sup>-1</sup> cm<sup>-1</sup>.

Far-ultraviolet (UV) spectra were obtained on a Jasco J-1500 spectropolarimeter equipped with an automated syringe-driven titration unit. Spectra were obtained from 195-255 nm as described (92). Insulin analogues were made 50  $\mu$ M in a buffer containing 10

mM potassium phosphate (pH 7.4) and 50 mM KCl. Helix-sensitive wavelength 222 nm was used as a probe of denaturation of protein (at 5  $\mu$ M) by guanidine hydrochloride. Thermodynamic parameters were obtained by application of a two-state model as described (93).

### ***NMR spectroscopy***

$^1\text{H}$  NMR spectra of DesDi two chain analogs were acquired at a proton frequency of 700 MHz at pD 7.4 (direct meter reading) at 32 °C.  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear single-quantum coherence (HSQC) spectra were acquired at natural abundance as described. The spectra were obtained at  $^{13}\text{C}$  frequency of 176 MHz at constant temperature of 305 K using the “hsqcetgp” Bruker pulse sequence as described by the vendor. Acquisition with FID size 2048 x 128, 800 scans, 1.0 sec relaxation delay, sweep widths 11 ppm ( $^1\text{H}$ ) and 70 ppm ( $^{13}\text{C}$ ) with offset 4.7 and 40 ppm for the  $^1\text{H}$  and  $^{13}\text{C}$  dimension, respectively. Data were processed with Topspin 4.0.6 (Bruker Biospin) and analyzed with Sparky software (94) using a 90° shifted-sine window function to a total of 2048  $\times$  1024 data points (F2  $\times$  F1), followed by automated baseline- and phase correction. All NMR data were acquired using a BRUKER 700 MHz spectrometer equipped with  $^1\text{H}$ ,  $^{19}\text{F}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  quadruple resonance cryoprobe.

### ***Cell culture***

Human embryonic kidney 293T cell was cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin as recommended by the American Type Culture Collection. Human hepatocellular carcinoma cell line HepG2 was also cultured in DMEM with 10% FBS and 1% penicillin/streptomycin. The protocol of signaling assays, which employed 24-h serum starvation, was adapted from previously published protocol (23). After starvation, cells were treated in parallel with a set of insulin analogs in serum-free medium.

### ***Transient transfections and ER stress assays***

Transfections were performed using Lipofectamine 3000 as described by the vendor (Invitrogen). Transfected HEK 293T cells were subjected to the Bio-Rad one-step real-time qPCR protocol. Readouts were provided by the up-regulation of ER stress markers *CHOP* and *BiP*. mRNA (messenger ribonucleic acid) abundances were measured in triplicate. In Western blot assay probing ER stress

markers (60,95), after 24 hour post transient transfection, cells were lysed by RIPA buffer (Cell Signaling Technology; CST). Protein concentrations in lysates were measured by BCA assay (Thermo)- and subjected to 4-20% SDS-PAGE and WB using anti-pPERK. Anti-PERK, anti-BiP and anti-CHOP antibodies (CST) at a dilution ratio of 1:1000; GAPDH provided a loading control.

### ***In-cell pIR immunoblotting***

The insulin-dependent IR activation was probed via fluorescent readouts. HepG2 cells were seeded (~8000 cells/well) into a 96-well black plate. After serum starvation, serial analog dilutions (100  $\mu$ l) the following treatments were applied to each well as described (23). Fixed cells were then exposed to the primary antibody (10  $\mu$ L anti-pTyr 4G10) and the secondary antibody (anti-mouse-IgG-800-CW antibody (Sigma) in 25 ml Blocking Buffer) was added to enable measurement of the fluorescence signals (details in supplemental information and ref (23)).

### ***Cell-based insulin signaling assays***

HepG2 cell line was subjected to 24-h serum-free starving, followed by addition of medium containing testing analogs (50 nM for ordinary treatment, additional 3000 nM for Ala<sup>B18</sup> and Gly<sup>B18</sup> DesDi two chain analogs). After lysed by RIPA buffer, total protein concentrations of lysate were determined by BCA assay (Thermo). Blotting protocols were modified from previous publication (23) (for detail, see the supplemental information).

### ***Rat Experiments***

Animals were maintained in accredited facility of Case Western Reserve University School of Medicine. All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) office of the University. Animal care and use was monitored by the University's Veterinary Services.

### ***Measurement of the Glucose-Lowering Effect of Insulins in Diabetic Rats***

Male Lewis rats (average body mass of ~300 g) were rendered diabetic by streptozotocin (STZ), as previously described (96). DesDi two chain insulins were dissolved in Lilly® Diluent buffer with the specified dose and injected in 100  $\mu$ L/300 g rat. Control rats received the appropriate volume of the



Lilly buffer. For these subcutaneous (SQ) experiments, rats were injected under the skin into the soft tissue in the posterior aspect of the neck. Following injection, blood glucose was measured using a small drop of blood obtained from the clipped tip of the rat's tail using a clinical glucometer (EasyMax® V Glucose Meter, Oak Tree Health, Las Vegas, NV). Blood-glucose concentrations were measured at  $t=0$ , and every 10 min for the first hour (h), every 20 min for the second h, every 30 min for the third h, and then each h for the rest of the experiment.

### Statistical Analysis

Data analysis was performed using MATHLAB Online (MathWorks). Areas under the curve (AUC) of glucose values in the diluent- and insulin-injected rats were calculated (Fig. S20). Statistical significance was determined using ANOVA; a  $p$  value of  $<0.05$  was considered significant whereas  $p$  values just above this threshold may suggest a trend whose validation would require larger studies.

*Acknowledgments.* This work was supported in part by the National Institutes of Health Grant (R01 DK040949 to M.A.W.). Early stages of this work were also supported by a grant from the NIH to M.A.W. and P.A. (R01 DK069764) and to P.A. (R01 DK48280). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Y.-S.C. was supported in part by a foundation grant from the Diabetes Research Connection. N.F.B.P. was supported in part by the American Diabetes Association Grants 7-13-IN-31 and 1-08-RA-149. We thank M.C. Lawrence and N. Rege for helpful discussion; and G.I. Bell and L.H. Philipson (University of Chicago) for summary of clinical mutations in the human insulin gene. M.A.W. thanks M. Karplus (Harvard), T. Sosnick (University of Chicago) and the late P.G. Katsoyannis (Mt. Sinai Medical Center, NY,

NY) for helpful discussion in the early stages of this work. We dedicate this article to the memory of the late Prof. G.G. Dodson (University of York, York, UK).

*Disclosures.* M.A.W. has equity in Thermalin, Inc. (Cleveland, OH) where he serves as Chief Innovation Officer; he has also been consultant to Merck Research Laboratories and DEKA Research and Development Corp. N.B.P. and F.I.-B. are consultants to Thermalin, Inc. and also have options, warrants or equity. F.I.-B. is a consultant to Sanofi and has received grants from Novo-Nordisk. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### Author Contributions

B.D. and M.J. undertook chemical peptide syntheses and folding assays; B.D. performed CD studies; Y.S.C. performed cell-based assays; Y.S.C., and M.A.W. designed and interpreted functional assays; Y.Y. and M.A.W. designed, performed and interpreted NMR studies; B.D., D.C. and M.A.W. interpreted cryo-EM structures of insulin-receptor model complexes; R.G., N.F.B.P. and F.I.-B. performed rat studies; B.D., Y.S.C., Y.Y., D.C., and M.A.W. prepared figures; B.D. and M.A.W. conceptualized the program of research. All authors contributed commented on the manuscript.

## FOOTNOTES

<sup>1</sup>*Abbreviations:* AKT, protein kinase B; AUC, areas under the curve in insulin tolerance tests; BiP, binding immunoglobulin protein; CHOP, C/EBP homologous protein; CD, circular dichroism; CFTR, cystic fibrosis transmembrane conductance regulator; ER, endoplasmic reticulum; DM, diabetes mellitus; GRP, glucose-regulated proteins; HEK, human embryonic kidney cell line; HepG2, human liver cancer cell line; HPLC, high-performance liquid chromatography; HSQC, heteronuclear single quantum coherence; IGF, insulin-like growth factor; IGFBP, insulin-like growth factor binding protein; SPPS, solid-phase peptide synthesis; ISR, integrated stress response; IR, insulin receptor; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; PDB, Protein Data Bank; PERK, protein kinase R-like ER kinase; qPCR, quantitative polymerase chain reaction; STZ, streptozotocin; SQ, subcutaneous; TK, tyrosine kinase; TOCSY, total correlation spectroscopy; WT, wild-type. WT represents human insulin whereas the “parent” DesDi precursor is Lys<sup>B28</sup>-des-dipeptide[B29,B30]-mini-proinsulin. Amino acids are designated by standard one- or three-letter codes.

<sup>2</sup>The mutant proinsulin syndrome can thus present as permanent neonatal DM or as a subtype of the clinical syndrome *maturity-onset diabetes of the young* (MODY) (97); this spectrum of monogenic DM is collectively designated “mutant *ins* gene-induced diabetes of youth” (MIDY) (26,51).

<sup>3</sup>Lys<sup>B28</sup> is a substitution in insulin *lispro*, the active component of Humalog<sup>®</sup>, a prandial analog formulation (Eli Lilly and Co.) (98). In broad clinical use, this analog contains paired substitutions in the B chain (Pro<sup>B28</sup>→Lys and Lys<sup>B29</sup>→Pro) that destabilize the zinc insulin hexamer (99) leading to accelerated SQ absorption (100).

<sup>4</sup>Rare species variants include related branched-chain residues Ile (in some snakes) and Thr in the amyloidogenic insulin of an hystricomorph rodent (the degu rat; [Supplemental Fig. S1](#)); Ile is also observed at position B18 of relaxin-related members of the vertebrate insulin superfamily ([Supplemental Fig. S2](#)).

**Table 1.** CD denaturation studies

<i>analog</i>	$\Delta G_u$ (kcal/mol) <sup>a</sup>	<i>m</i> -value <sup>b</sup>	$C_{mid}$ (M) <sup>c</sup>
<i>single-chain precursors (37 °C)</i>			
DesDi-Val <sup>B18</sup> (parent)	7.5 ± 0.4 <sup>d</sup>	1.10 ± 0.06	6.8 ± 0.4
DesDi-Ala <sup>B18</sup>	4.0 ± 0.1	0.60 ± 0.03	6.7 ± 0.2
DesDi-Gly <sup>B18</sup>	0.6 ± 0.1	0.49 ± 0.02	1.2 ± 0.1
<i>two-chain insulin analogs (25 °C)</i>			
human insulin	3.3 ± 0.1	0.69 ± 0.01	4.8 ± 0.1
insulin <i>lispro</i>	3.0 ± 0.1	0.63 ± 0.01	4.7 ± 0.1
DesDi-Val <sup>B18</sup> (WT)	2.9 ± 0.1	0.64 ± 0.02	4.5 ± 0.1
DesDi-Ala <sup>B18</sup>	1.7 ± 0.1	0.46 ± 0.01	3.7 ± 0.1
DesDi-Gly <sup>B18</sup>	N.D. <sup>e</sup>	N.D.	N.D.

<sup>a</sup>Estimates of free-energy changes were inferred from two-state modeling of chemical-denaturation titrations as described (23); assays were performed in 10 mM potassium phosphate (pH 7.4) and 50 mM KCl.

<sup>b</sup>The *m*-value is the slope of unfolding free energy  $\Delta G_u$  versus molar concentration of denaturant

<sup>c</sup> $C_{mid}$  is the guanidine denaturant concentration at which 50% of the protein is unfolded.

<sup>d</sup>The parent single-chain analog (Lys<sup>B28</sup>-*des*-dipeptide[B29,B30]-miniproinsulin (53)) is too stable to unfold completely in 8 M guanidine-HCl and 37 °C; this limited modeling of the post-transition baseline and introduced a larger uncertainty in model parameters.

<sup>e</sup>This insulin analog exhibited a non-cooperative unfolding transition.

**Table 2.** Comparison of predicted and observed thermodynamic parameters.

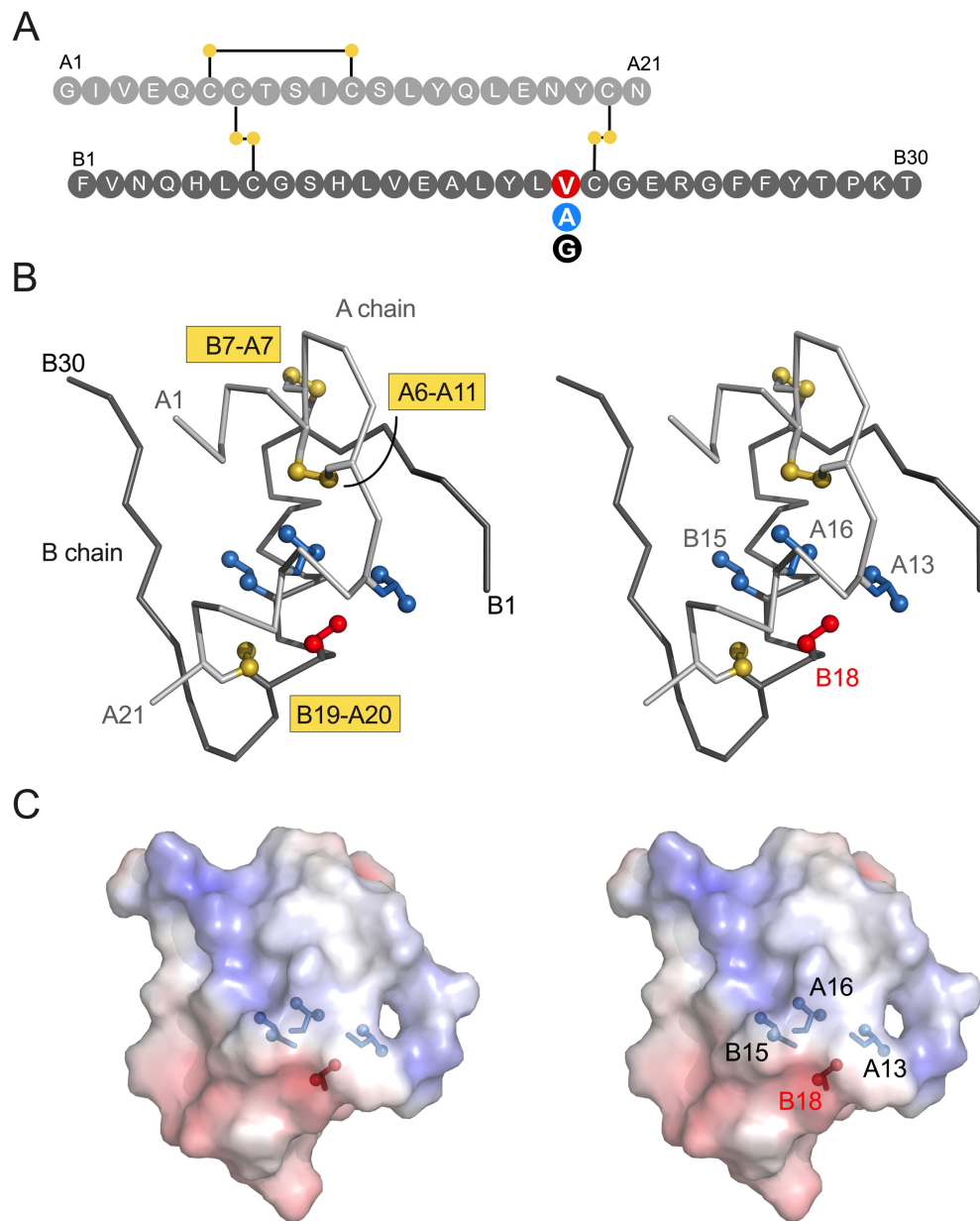
	$\Delta\Delta G$ (Cavity penalties) (kcal/mol) <sup>a</sup>	$\Delta\Delta G$ (hydrophobic- transfer free energy penalty) <sup>b</sup> (kcal/mol)	$\Delta\Delta G$ (Helical Propensity) <sup>c</sup> (kcal/mol)	Sum of predicted $\Delta\Delta G$	$\Delta\Delta G_u$ (kcal/mol) <sup>d</sup>
Ala <sup>B18</sup>	0.89	1.19	-1.06	1.02 ± 0.2	1.2 ± 0.2
Gly <sup>B18</sup>	1.63	2.04	0.83	4.50 ± 0.2	>2.9

<sup>a</sup> Calculated cavity penalty using Brian Matthews' formula by multiplying cavity volume by a factor of 22 cal/mol/Å<sup>3</sup> (81). The Cavity volumes for Ala and Gly were 40.3 and 73.9 Å<sup>3</sup> respectively. These were calculated by CASTp 3.0 as reported by Jie Liang and co-workers to measure geometric and topological properties of protein structures (101).

<sup>b</sup> Hydrophobic transfer-free energy penalties were calculated from the estimated hydrophobic effect for residue burial (Val, 3.38; Ala, 2.15; Gly, 1.18 kcal/mol) (79) after taking solvent exposure into account followed by subtracting from WT residue. The solvent exposure at B18 for Val, Ala and Gly respectively were 19%, 28% and 41%.

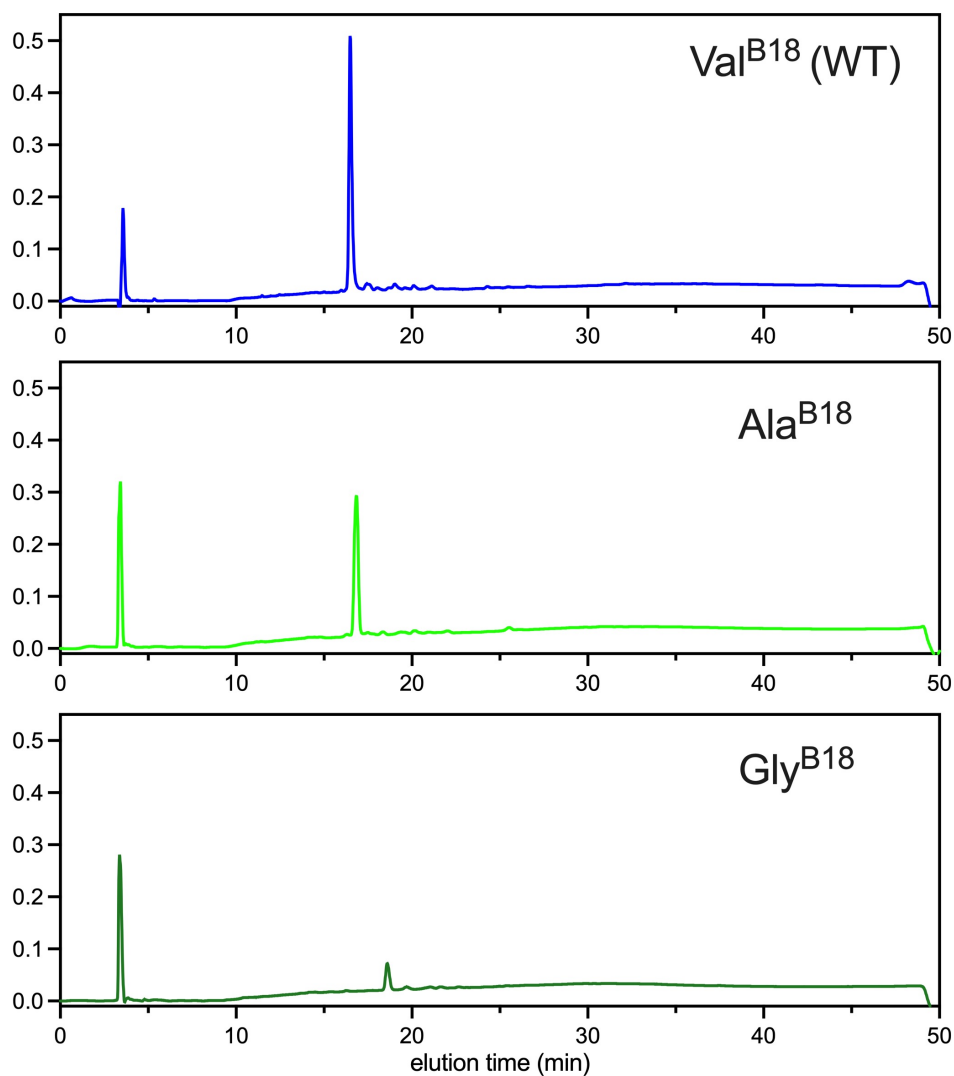
<sup>c</sup> The absolute helical propensities for Ala and Gly were 1.61 and 0.05 relative to 0.23 for Val (58).

<sup>d</sup> From two-state transition fits of CD-guanidine titrations (93) performed at 25 °C in pH 7.4 phosphate buffer containing 50 mM KCl.

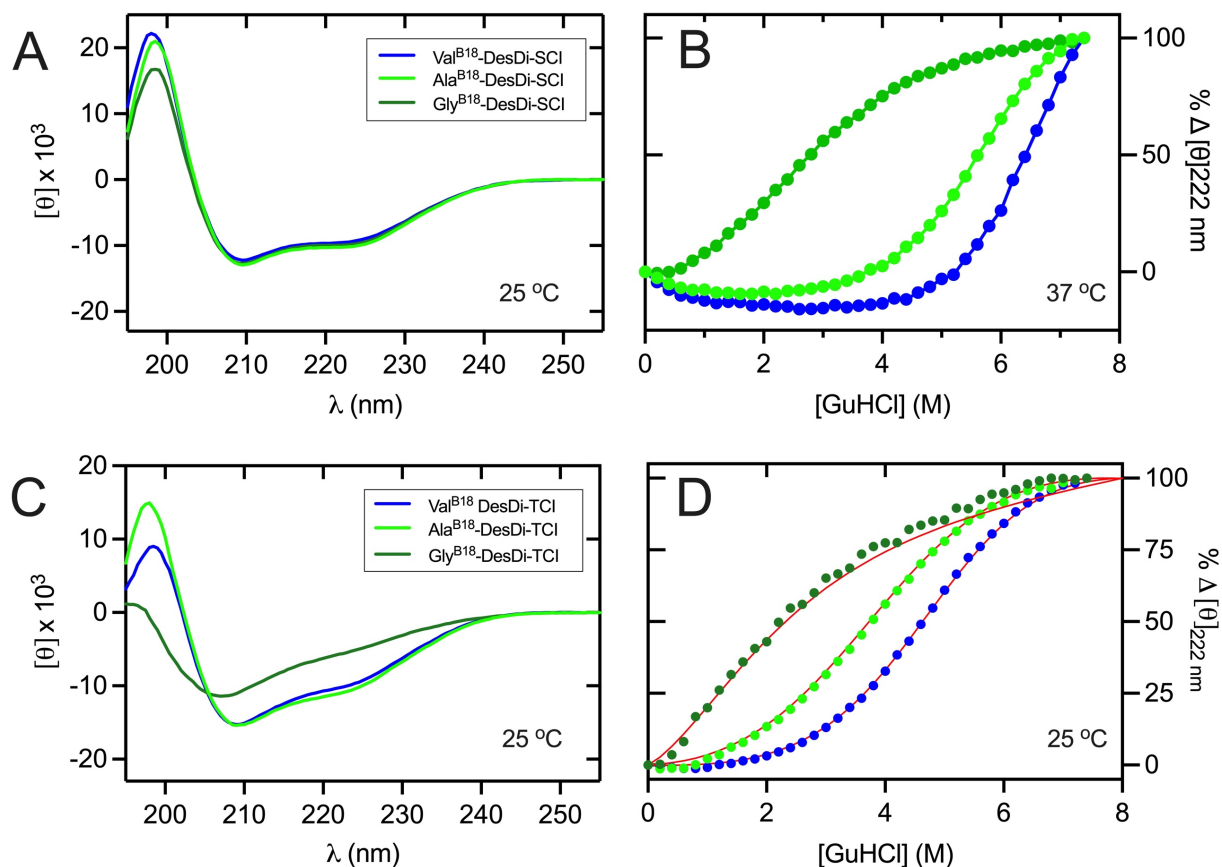


**Figure 1.** Sequence and structure of insulin. (A) Sequence of insulin showing clinical mutations at position B18 (highlighted by solid red circle). Residues are labeled by standard single letter code (white). The A chain is shown as light gray circles (upper sequence), and the B chain as dark gray circles (lower sequence). Neonatal onset or delayed onset is indicated by filled red or blue circles, respectively. (B) Stereo view of insulin monomer ( $C_{\alpha}$ -trace ribbon model; PDB entry 4INS (47)). Side chains of Leu<sup>B15</sup> (blue), Val<sup>B18</sup> (red), Leu<sup>A13</sup> (blue), Leu<sup>A16</sup> (blue) and disulfide bridges (golden) are shown as sticks. Terminal methyl groups of these side chains and sulfur atoms are shown as spheres (one-third Van der Waals radii). The A- and B chain ribbons are shown in light and dark gray, respectively. (C) Spatial environments of residue B18 (color coding and drawing same as in panel B) shown as electrostatic surfaces that are calculated in the absence of B18 side chain. blue and red surfaces are coded by positive or negative electrostatic potential.



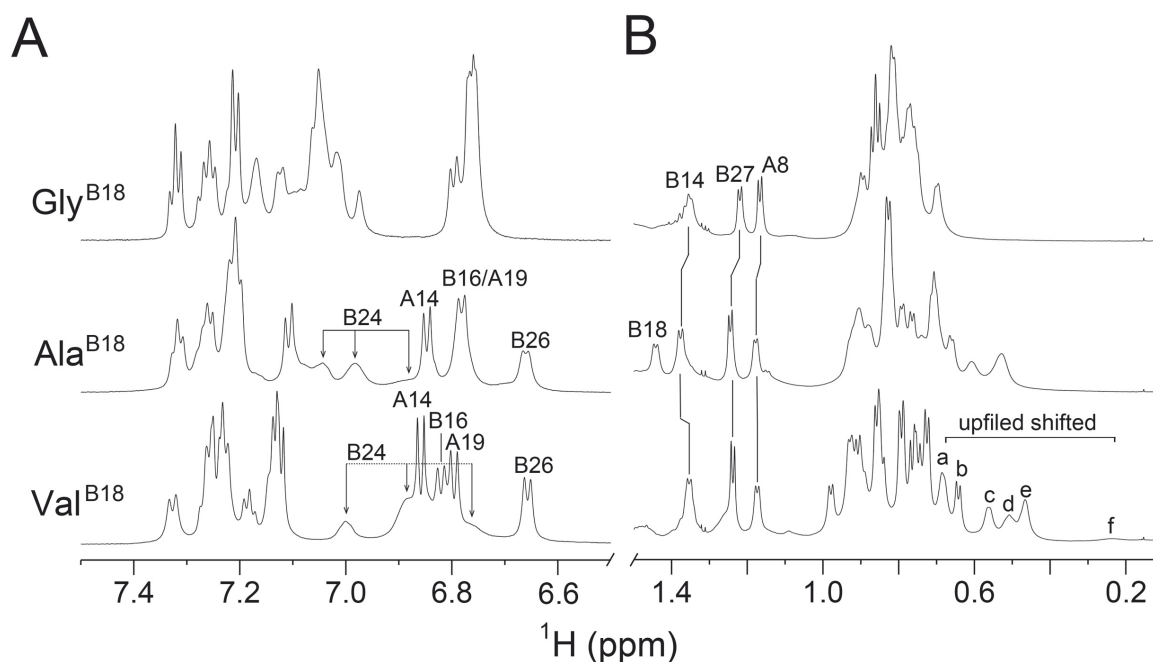


**Figure 3.** Comparison of folding profiles of wild-type (Val) and clinical mutations (Ala and Gly) at position B18 using DesDi synthetic precursor. Folding reactions were carried out using crude polypeptides at 100 mg scale (~18  $\mu$ mol) at 0.1 mM peptide concentration in 20 mM glycine, 2 mM Cysteine at pH 10.5 aqueous buffer under atmosphere of air (4 °C).

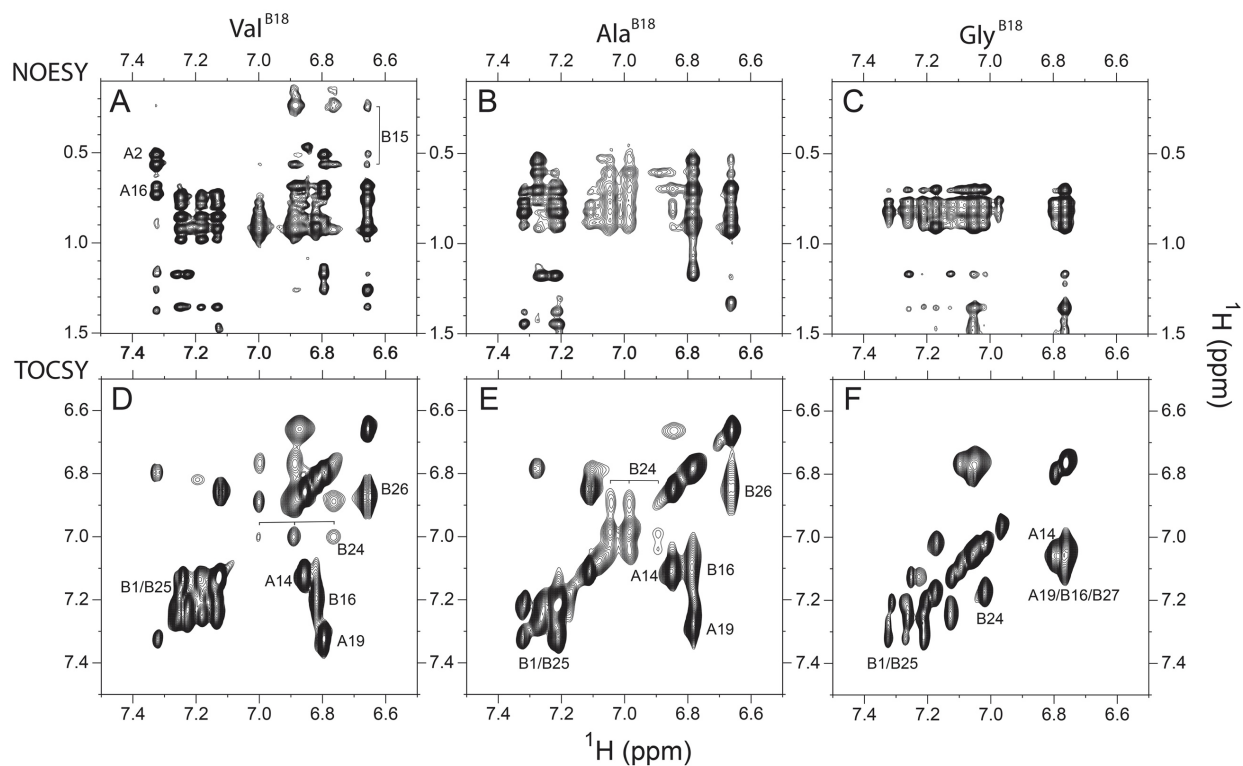


**Figure 4.** CD and guanidine denaturation studies of DesDi insulin analogs. (*A, B*) Far-UV spectra of Val<sup>B18</sup> (WT), Ala<sup>B18</sup>, and Gly<sup>B18</sup>-DesDi single-chain insulin analogs (*A*) and corresponding CD-detected unfolding transitions monitored by ellipticity at 222 nm (*B*). (*C, D*) Spectra of respective two chain analogs (*C*) and corresponding unfolding transitions (*D*). Color codes are defined in panels A and C. All the studies were performed at 25 °C except for the guanidine denaturation studies of single-chain precursors which was done at 37 °C owing to their relative higher stability at 25 °C. The resulting stability values are given in [Table 1](#).

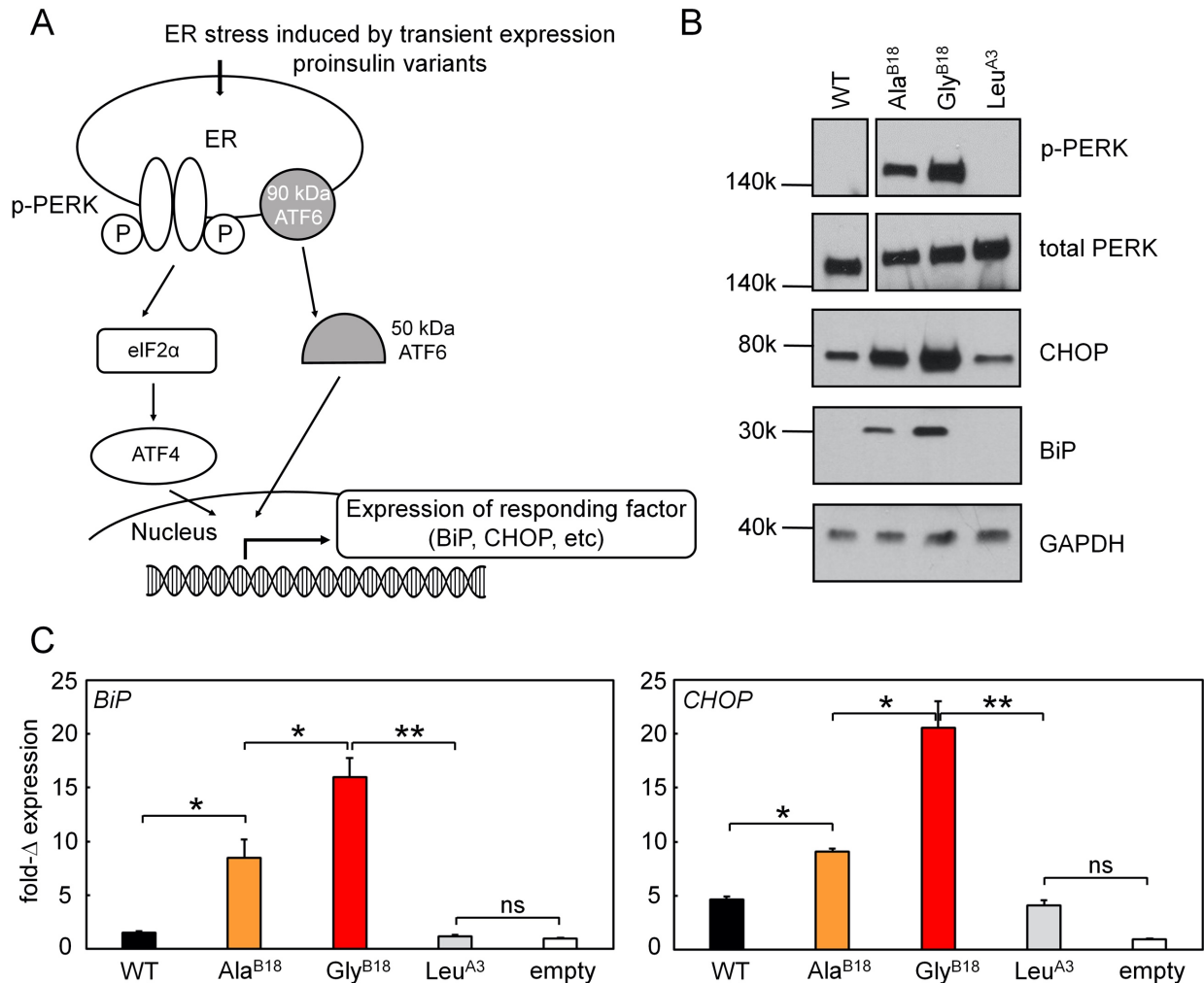




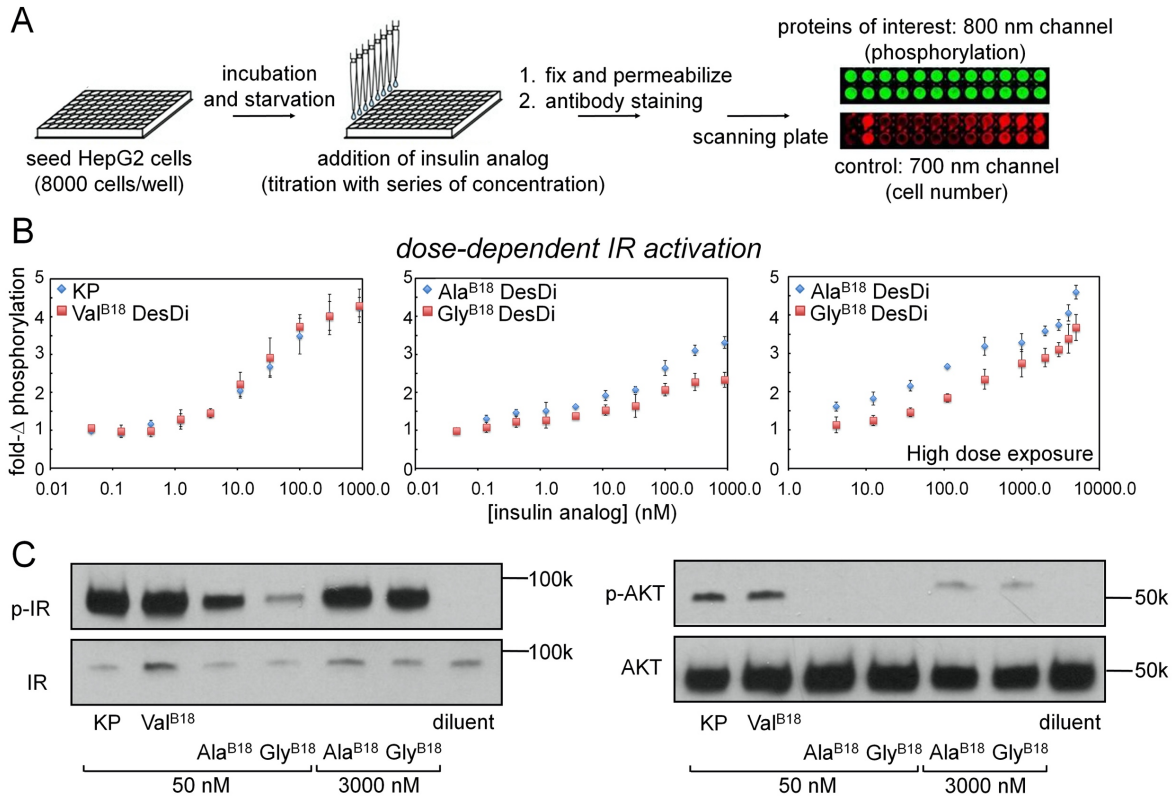
**Figure 5.** Stack plot of 1D <sup>1</sup>H-NMR spectra of DesDi insulin analogs. (A) aromatic and (B) methyl region. Bottom panel is parent DesDi Val<sup>B18</sup> analog, middle panel is Ala<sup>B18</sup> analog and top panel is Gly<sup>B18</sup> analog. The Val<sup>B18</sup> analog exhibits spectral property of native-like insulin as observed in the aromatic region and in the upfield-shifted methyl region (*far right*). The well-resolved methyl resonances indicated as a: Ile<sup>A2</sup>  $\gamma_2$ -CH<sub>3</sub>; b: Ile<sup>A10</sup>  $\gamma_2$ -CH<sub>3</sub>; c: Leu<sup>B15</sup>  $\delta_2$ -CH<sub>3</sub>; d: Ile<sup>A2</sup>  $\delta_1$ -CH<sub>3</sub>; e: Ile<sup>A10</sup>  $\delta_1$ -CH<sub>3</sub>; f: Leu<sup>B15</sup>  $\delta_1$ -CH<sub>3</sub>. For Ala<sup>B18</sup> analog, aromatic and methyl resonances of signature residues Phe<sup>B24</sup> and Leu<sup>B15</sup> shifted to downfield, as well as exhibited a reduction in chemical-shift dispersion. Spectra were acquired at pD 7.4 (direct meter reading) at 32 °C in D<sub>2</sub>O.



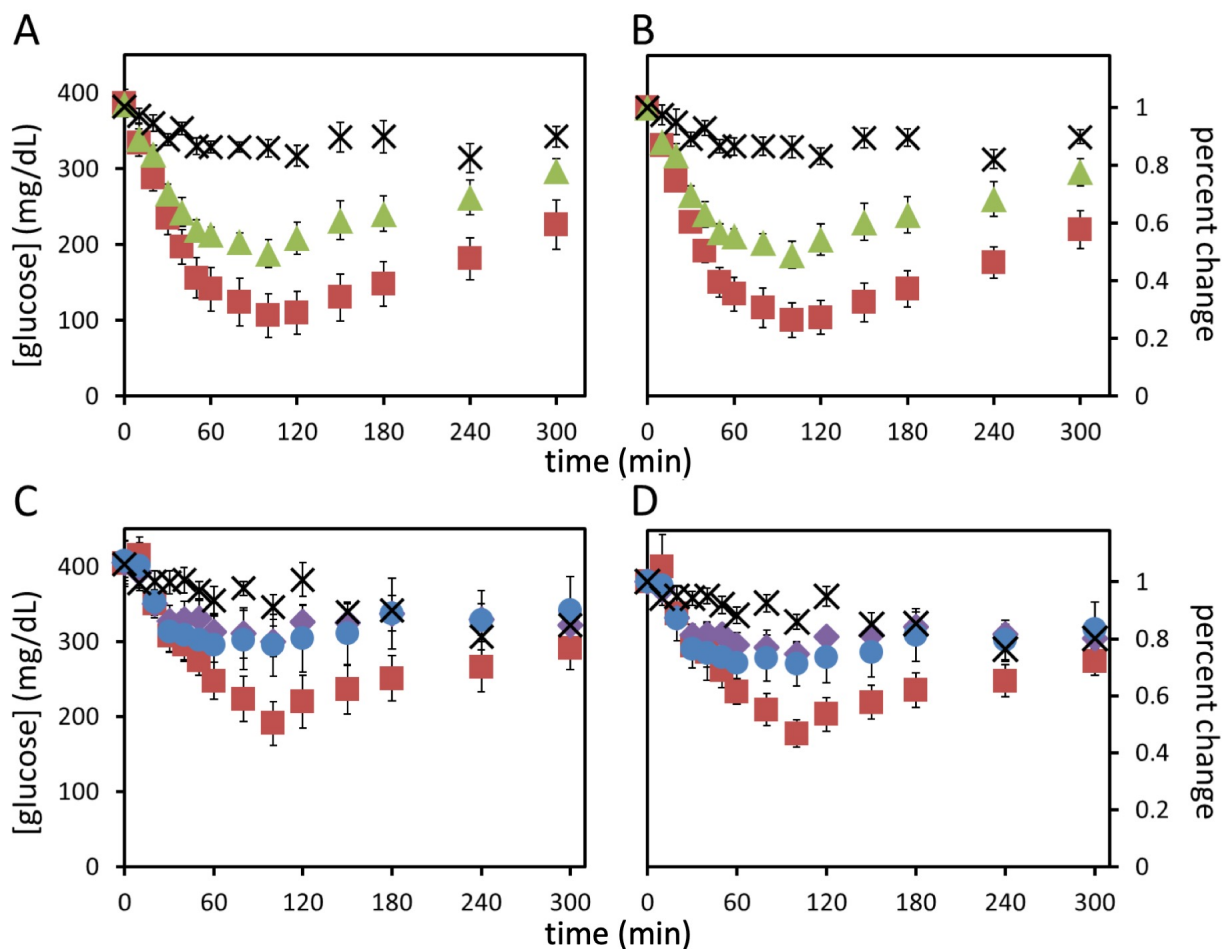
**Figure 6.** 2D NMR studies. Top panel shows NOEs from aromatic to aliphatic protons in 2D NOESY spectra of DesDi Val<sup>B18</sup> analog (A), Ala<sup>B18</sup> analog (B) and Gly<sup>B18</sup> analog (C). Bottom panel shows aromatic resonances of 2D TOCSY spectra of DesDi Val<sup>B18</sup> analog (D), Ala<sup>B18</sup> analog (E) and Gly<sup>B18</sup> analog (F). Spectra were acquired at pD 7.4 (direct meter reading) at 32 °C in D<sub>2</sub>O.



**Figure 7.** Proinsulin mutants induce the ER stress. (A) Schematic figure presenting the ER stress induced by expression of proinsulin analogs in cells (102). Mutant proinsulin variants give rise to accumulation of unfolded protein in the ER resulting in stress, which causes phosphorylation of PERK and induction of ISR. CHOP induction is a downstream response of the ISR and BiP is a chaperon, which would be activated and lead to its increased expression. The protein levels of p-PERK and the accumulations of ER stress markers BiP and CHOP were tested by Western blot approach (B). The transcription responses of these two ER stress markers were also monitored by qPCR assay (C). (B) Western-blot assays probing the ER stress markers: p-PERK/PERK, CHOP, and BiP, induced by the expression of proinsulin variants. Expression of WT and Leu<sup>A3</sup> proinsulin provides the controls showing inactive ER stress. (C) Real-time qPCR assay probing the transcription responses of *BiP* and *CHOP* genes induced by WT and variants of proinsulin. Gene markers for ER stress were significantly activated by the expressions of Gly<sup>B18</sup> (red) and Ala<sup>B18</sup> (orange) proinsulin. WT and Leu<sup>A3</sup> proinsulin served as controls. Asterisks (\*) and (\*\*) indicate p-value < 0.05 and < 0.01. The “ns” indicates p-value > 0.05.



**Figure 8.** Dose-response studies of insulin activity. (A) Schematic outline of in-cell assays probing IR autophosphorylation in exposure of various insulin analogs. Flow chart shows the procedure to monitor hormone-induced IR signaling via in-cell illumination assay using the ratio of readouts between 800 nm (phosphorylation) and 700 nm (cell number control). (B) The IR autophosphorylation on hormone binding. Insulin lispro (labeled as KP) and Val<sup>B18</sup> (WT) DesDi insulin share similar activity in all the titration conditions (left). Mutations (Ala and Gly) in B18 position significantly reduced the activities (middle) and only reach the WT DesDi-like activities when exposed in high dose condition (>3000 nM; right). (C) Western-blot assays probing the p-IR/IR and p-AKT/AKT in the ordinary (50 nM) and high dose (3000 nM) of analog Ala<sup>B18</sup> and Gly<sup>B18</sup> in medium.



**Figure 9.** Rat studies of DesDi two chain insulin analogues. (A) time course of [blood glucose] following SQ injection. Val<sup>B18</sup> (red squares, n=6); Ala<sup>B18</sup> (green triangle, n=6); diluent (X, 100  $\mu$ L, n=6). The dose of WT or mutant insulin was 6 nmol/kg rat. (C) Dose-response studies of Gly<sup>B18</sup> relative to parent Val<sup>B18</sup> analog at a 6 nmol/kg rat dose (red squares; N=5). Doses of the Gly<sup>B18</sup> analog: 12 nmol/kg rat (blue circles; N=5) and 24 nmol/kg rat (purple diamond; N=5). (B) and (D), normalized blood glucose profiles for the data in panels A and C, respectively. Standard errors of mean are given by black vertical error bars. Quantitative and statistical analyses are provided in [Supplemental Fig. S14](#).

## References

1. Dobson, C. M., Šali, A., and Karplus, M. (1998) Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed.* **37**, 868-893
2. Dobson, C. M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329-332
3. Jimenez, J. L., Nettleton, E. J., Bouchard, M., Robinson, C. V., Dobson, C. M., and Saibil, H. R. (2002) The protofilament structure of insulin amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9196-9201
4. Zeldovich, K. B., and Shakhnovich, E. I. (2008) Understanding protein evolution: from protein physics to Darwinian selection. *Annu. Rev. Phys. Chem.* **59**, 105-127
5. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426**, 884-890
6. Hartl, F. U., and Hayer-Hartl, M. (2009) Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574
7. Hirsh, A. E., and Fraser, H. B. (2001) Protein dispensability and rate of evolution. *Nature* **411**, 1046-1049
8. Thomas Jr, G. J., Becka, R., Sargent, D., Yu, M. H., and King, J. (1990) Conformational stability of P22 tailspike proteins carrying temperature-sensitive folding mutations. *Biochemistry* **29**, 4181-4187
9. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507-511
10. Stefani, M., and Dobson, C. M. (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678-699
11. Chiti, F., and Dobson, C. M. (2017) Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27-68
12. Ohashi, K. (2001) Pathogenesis of  $\beta$ 2-microglobulin amyloidosis. *Pathol. Int.* **51**, 1-10
13. Granel, B., Valleix, S., Serratrice, J., Cherin, P., Texeira, A., Disdier, P., Weiller, P. J., and Grateau, G. (2006) Lysozyme amyloidosis: report of 4 cases and a review of the literature. *Medicine (Baltimore)* **85**, 66-73
14. Johnson, S. M., Connelly, S., Fearn, C., Powers, E. T., and Kelly, J. W. (2012) The transthyretin amyloidoses: from delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug. *J. Mol. Biol.* **421**, 185-203
15. Arunagiri, A., Haataja, L., Cunningham, C. N., Shrestha, N., Tsai, B., Qi, L., Liu, M., and Arvan, P. (2018) Misfolded proinsulin in the endoplasmic reticulum during development of beta cell failure in diabetes. *Ann. N. Y. Acad. Sci.* **1418**, 5
16. Johnston, H. E., and Samant, R. S. (2020) Alternative systems for misfolded protein clearance: life beyond the proteasome. *FEBS J.* **288**, 4464-4487
17. Du, K., Sharma, M., and Lukacs, G. L. (2005) The  $\Delta$ F508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat. Struct. Mol. Biol.* **12**, 17-25
18. Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., O'Riordan, C. R., and Smith, A. E. (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **63**, 827-834
19. Middleton, P. G., Mall, M. A., Dřevínek, P., Lands, L. C., McKone, E. F., Polineni, D., Ramsey, B. W., Taylor-Cousar, J. L., Tullis, E., and Vermeulen, F. (2019) Elexacaftor-tezacaftor-ivacaftor for cystic fibrosis with a single Phe508del allele. *N. Engl. J. Med.* **381**, 1809-1819
20. Lukacs, G. L., and Verkman, A. (2012) CFTR: folding, misfolding and correcting the  $\Delta$ F508 conformational defect. *Trends Mol. Med.* **18**, 81-91
21. Weiss, M. A., Nakagawa, S. H., Jia, W., Xu, B., Hua, Q. X., Chu, Y. C., Wang, R. Y., and Katsoyannis, P. G. (2002) Protein structure and the spandrels of San Marco: insulin's receptor-

- binding surface is buttressed by an invariant leucine essential for protein stability. *Biochemistry* **41**, 809-819
22. Liu, M., Wan, Z.-l., Chu, Y.-C., Aladdin, H., Klaproth, B., Choquette, M., Hua, Q.-x., Mackin, R. B., Rao, J. S., and De Meyts, P. (2009) Crystal structure of a “nonfoldable” insulin impaired folding efficiency despite native activity. *J. Biol. Chem.* **284**, 35259-35272
  23. Rege, N. K., Liu, M., Yang, Y., Dhayalan, B., Wickramasinghe, N. P., Chen, Y.-S., Rahimi, L., Guo, H., Haataja, L., and Sun, J. (2020) Evolution of insulin at the edge of foldability and its medical implications. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 29618-29628
  24. Støy, J., Edghill, E. L., Flanagan, S. E., Ye, H., Paz, V. P., Pluzhnikov, A., Below, J. E., Hayes, M. G., Cox, N. J., and Lipkind, G. M. (2007) Insulin gene mutations as a cause of permanent neonatal diabetes. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15040-15044
  25. Colombo, C., Porzio, O., Liu, M., Massa, O., Vasta, M., Salardi, S., Beccaria, L., Monciotti, C., Toni, S., and Pedersen, O. (2008) Seven mutations in the human insulin gene linked to permanent neonatal/infancy-onset diabetes mellitus. *J. Clin. Investig.* **118**, 2148-2156
  26. Liu, M., Sun, J., Cui, J., Chen, W., Guo, H., Barbetti, F., and Arvan, P. (2015) INS-gene mutations: from genetics and beta cell biology to clinical disease. *Mol. Aspects Med.* **42**, 3-18
  27. Støy, J., De Franco, E., Ye, H., Park, S.-Y., Bell, G. I., and Hattersley, A. T. (2021) In celebration of a century with insulin—Update of insulin gene mutations in diabetes. *Mol. Metab.*, 101280
  28. Liu, M., Weiss, M. A., Arunagiri, A., Yong, J., Rege, N., Sun, J., Haataja, L., Kaufman, R. J., and Arvan, P. (2018) Biosynthesis, structure, and folding of the insulin precursor protein. *Diabetes Obes. Metab.* **20**, 28-50
  29. Steiner, D. F. (1967) Evidence for a precursor in the biosynthesis of insulin. *Trans. N. Y. Acad. Sci.* **30**, 60-68
  30. Steiner, D., Clark, J., Nolan, C., Rubenstein, A., Margoliash, E., Aten, B., and Oyer, P. (1969) Proinsulin and the biosynthesis of insulin. *Recent Prog. Horm. Res.* **25**, 207
  31. Liu, M., Hodish, I., Haataja, L., Lara-Lemus, R., Rajpal, G., Wright, J., and Arvan, P. (2010) Proinsulin misfolding and diabetes: mutant INS gene-induced diabetes of youth. *Trends Endocrinol. Metab.* **21**, 652-659
  32. Sun, J., Cui, J., He, Q., Chen, Z., Arvan, P., and Liu, M. (2015) Proinsulin misfolding and endoplasmic reticulum stress during the development and progression of diabetes. *Mol. Aspects Med.* **42**, 105-118
  33. Liu, M., Weiss, M. A., Arunagiri, A., Yong, J., Rege, N., Sun, J., Haataja, L., Kaufman, R. J., and Arvan, P. (2018) Biosynthesis, structure, and folding of the insulin precursor protein. *Diabetes Obes. Metab.* **20**, 28-50
  34. Zuber, C., Fan, J.-Y., Guhl, B., and Roth, J. (2004) Misfolded proinsulin accumulates in expanded pre-Golgi intermediates and endoplasmic reticulum subdomains in pancreatic beta cells of Akita mice. *FASEB J.* **18**, 917-919
  35. Liu, M., Hodish, I., Rhodes, C. J., and Arvan, P. (2007) Proinsulin maturation, misfolding, and proteotoxicity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15841-15846
  36. Hodish, I., Liu, M., Rajpal, G., Larkin, D., Holz, R. W., Adams, A., Liu, L., and Arvan, P. (2010) Misfolded proinsulin affects bystander proinsulin in neonatal diabetes. *J. Biol. Chem.* **285**, 685-694
  37. Sun, J., Xiong, Y., Li, X., Haataja, L., Chen, W., Mir, S. A., Lv, L., Madley, R., Larkin, D., and Anjum, A. (2020) Role of proinsulin self-association in mutant INS gene-induced diabetes of youth. *Diabetes* **69**, 954-964
  38. Liu, M., Haataja, L., Wright, J., Wickramasinghe, N. P., Hua, Q.-X., Phillips, N. F., Barbetti, F., Weiss, M. A., and Arvan, P. (2010) Mutant INS-gene induced diabetes of youth: proinsulin cysteine residues impose dominant-negative inhibition on wild-type proinsulin transport. *PLoS One* **5**, e13333

39. Sun, F., Du, W., Ma, J., Gu, M., Wang, J., Zhu, H., Song, H., and Gao, G. (2020) A Novel c. 125 T> G (p. Val42Gly) Mutation in The Human INS Gene Leads to Neonatal Diabetes Mellitus via a Decrease in Insulin Synthesis. *Exp. Clin. Endocrinol. Diabetes* **128**, 182-189
40. Piccini, B., Artuso, R., Lenzi, L., Guasti, M., Braccesi, G., Barni, F., Casalini, E., Giglio, S., and Toni, S. (2016) Clinical and molecular characterization of a novel INS mutation identified in patients with MODY phenotype. *Eur. J. Med. Genet.* **59**, 590-595
41. 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) Oxidative refolding of insulin-like growth factor 1 yields two products of similar thermodynamic stability: a bifurcating protein-folding pathway. *Biochemistry* **32**, 5203-5213
42. Hua, Q. X., Gozani, S. N., Chance, R. E., Hoffmann, J. A., Frank, B. H., and Weiss, M. A. (1995) Structure of a protein in a kinetic trap. *Nat. Struct. Biol.* **2**, 129-138
43. Qiao, Z. S., Guo, Z. Y., and Feng, Y. M. (2001) Putative disulfide-forming pathway of porcine insulin precursor during its refolding in vitro. *Biochemistry* **40**, 2662-2668
44. Hua, Q. X., Chu, Y. C., Jia, W., Phillips, N. B., Wang, R. Y., Katsoyannis, P. G., and Weiss, M. A. (2002) Mechanism of insulin chain combination. Asymmetric roles of A-chain  $\alpha$ -helices in disulfide pairing. *J. Biol. Chem.* **277**, 43443-43453
45. Yan, H., Guo, Z. Y., Gong, X. W., Xi, D., and Feng, Y. M. (2003) A peptide model of insulin folding intermediate with one disulfide. *Protein Sci.* **12**, 768-775
46. Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D., Mercola, D., and Vijayan, M. (1971) Atomic positions in rhombohedral 2-zinc insulin crystals. *Nature* **231**, 506-511
47. Baker, E. N., Blundell, T. L., Cutfield, J. F., Dodson, E. J., Dodson, G. G., Hodgkin, D. M. C., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., and Sakabe, K. (1988) The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **319**, 369-456
48. Menting, J. G., Whittaker, J., Margetts, M. B., Whittaker, L. J., Kong, G. K.-W., Smith, B. J., Watson, C. J., Žáková, L., Kletvíková, E., and Jiráček, J. (2013) How insulin engages its primary binding site on the insulin receptor. *Nature* **493**, 241-245
49. Menting, J. G., Yang, Y., Chan, S. J., Phillips, N. B., Smith, B. J., Whittaker, J., Wickramasinghe, N. P., Whittaker, L. J., Pandeyarajan, V., and Wan, Z.-l. (2014) Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3395-E3404
50. Weis, F., Menting, J. G., Margetts, M. B., Chan, S. J., Xu, Y., Tennagels, N., Wohlfart, P., Langer, T., Müller, C. W., and Dreyer, M. K. (2018) The signalling conformation of the insulin receptor ectodomain. *Nat. Commun.* **9**, 1-10
51. Haataja, L., Arunagiri, A., Hassan, A., Regan, K., Tsai, B., Dhayalan, B., Weiss, M. A., Liu, M., and Arvan, P. (2021) Distinct states of proinsulin misfolding in MIDY. *Cell. Mol. Life Sci.* **78**, 6017-6031
52. Dhayalan, B., Chatterjee, D., Chen, Y.-S., and Weiss, M. A. (2021) Structural Lessons from the Mutant Proinsulin Syndrome. *Front. Endocrinol. (Lausanne)* **in press**
53. Zaykov, A. N., Mayer, J. P., Gelfanov, V. M., and DiMarchi, R. D. (2014) Chemical synthesis of insulin analogs through a novel precursor. *ACS Chem. Biol.* **9**, 683-691
54. Katsoyannis, P. (1966) Synthesis of insulin. *Science*, 1509-1514
55. Tang, J.-G., and Tsou, C.-L. (1990) The insulin A and B chains contain structural information for the formation of the native molecule. Studies with protein disulphide-isomerase. *Biochem. J.* **268**, 429-435
56. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from CD spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252-260
57. O'Neil, K. T., and DeGrado, W. F. (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250**, 646-651



58. Chakrabartty, A., Kortemme, T., and Baldwin, R. L. (1994) Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**, 843-852
59. Trombetta, E. S., and Parodi, A. J. (2003) Quality control and protein folding in the secretory pathway. *Annu. Rev. Cell. Dev. Biol.* **19**, 649-676
60. Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., and Ron, D. (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* **6**, 1099-1108
61. Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P., and Ron, D. (2000) Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat. Cell Biol.* **2**, 326-332
62. Nanjo, K., Sanke, T., Miyano, M., Okai, K., Sowa, R., Kondo, M., Nishimura, S., Iwo, K., Miyamura, K., and Given, B. (1986) Diabetes due to secretion of a structurally abnormal insulin (insulin Wakayama). Clinical and functional characteristics of [LeuA3] insulin. *J. Clin. Invest.* **77**, 514-519
63. Kobayashi, M., Takata, Y., Ishibashi, O., Sasaoka, T., Iwasaki, M., Shigeta, Y., and Inouye, K. (1986) Receptor binding and negative cooperativity of a mutant insulin,[LeuA3]-insulin. *Biochem. Biophys. Res. Commun.* **137**, 250-257
64. Steiner, D. F., and Chan, S. J. (1988) An overview of insulin evolution. *Horm. Metab. Res.* **20**, 443-444
65. Chan, S. J., Cao, Q.-P., and Steiner, D. F. (1990) Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9319-9323
66. Jin Chan, S., and Steiner, D. F. (2000) Insulin through the ages: phylogeny of a growth promoting and metabolic regulatory hormone. *Am. Zool.* **40**, 213-222
67. De Meyts, P. (1994) The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. *Diabetologia* **37**, S135-S148
68. Murray-Rust, J., McLeod, A., Blundell, T., and Wood, S. (1992) Structure and evolution of insulins: implications for receptor binding. *Bioessays* **14**, 325-331
69. Allard, J. B., and Duan, C. (2018) IGF-binding proteins: why do they exist and why are there so many? *Front. Endocrinol.* **9**, 117
70. Weiss, M. A. (2013) Diabetes mellitus due to the toxic misfolding of proinsulin variants. *FEBS Lett.* **587**, 1942-1950
71. Blundell, T., and Humbel, R. (1980) Hormone families: pancreatic hormones and homologous growth factors. *Nature* **287**, 781-787
72. Ptitsyn, O. (1995) Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83-229
73. Hua, Q. X., Narhi, L., Jia, W., Arakawa, T., Rosenfeld, R., Hawkins, N., Miller, J. A., and Weiss, M. A. (1996) Native and non-native structure in a protein-folding intermediate: spectroscopic studies of partially reduced IGF-I and an engineered alanine model. *J. Mol. Biol.* **259**, 297-313
74. Derewenda, U., Derewenda, Z., Dodson, E., Dodson, G., Bing, X., and Markussen, J. (1991) X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule: A completely inactive analogue. *J. Mol. Biol.* **220**, 425-433
75. Luisier, S., Avital-Shmilovici, M., Weiss, M. A., and Kent, S. B. (2010) Total chemical synthesis of human proinsulin. *Chem. Commun.* **46**, 8177-8179
76. Avital-Shmilovici, M., Whittaker, J., Weiss, M. A., and Kent, S. B. (2014) Deciphering a molecular mechanism of neonatal diabetes mellitus by the chemical synthesis of a protein diastereomer,[D-AlaB8] human proinsulin. *J. Biol. Chem.* **289**, 23683-23692
77. Tanford, C. (1962) Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* **84**, 4240-4247

78. Chothia, C. (1974) Hydrophobic bonding and accessible surface area in proteins. *Nature* **248**, 338-339
79. Karplus, P. A. (1997) Hydrophobicity regained. *Protein Sci.* **6**, 1302-1307
80. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., and Baldwin, R. L. (1990) Relative helix-forming tendencies of nonpolar amino acids. *Nature* **344**, 268-270
81. Eriksson, A. E., Baase, W. A., Wozniak, J. A., and Matthews, B. W. (1992) A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **355**, 371-373
82. Xu, J., Baase, W. A., Baldwin, E., and Matthews, B. W. (1998) The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.* **7**, 158-177
83. Matthews, B. W. (1993) Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **62**, 139-160
84. Edghill, E. L., Flanagan, S. E., Patch, A.-M., Boustred, C., Parrish, A., Shields, B., Shepherd, M. H., Hussain, K., Kapoor, R. R., and Malecki, M. (2008) Insulin mutation screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a rare cause of diabetes diagnosed in childhood or adulthood. *Diabetes* **57**, 1034-1042
85. Shoelson, S., Polonsky, K., Zeidler, A., Rubenstein, A., and Tager, H. (1984) Human insulin B24 (Phe---Ser). Secretion and metabolic clearance of the abnormal insulin in man and in a dog model. *J. Clin. Invest.* **73**, 1351-1358
86. Vetere, A., Choudhary, A., Burns, S. M., and Wagner, B. K. (2014) Targeting the pancreatic  $\beta$ -cell to treat diabetes. *Nat. Rev. Drug Discov.* **13**, 278-289
87. Yong, J., Johnson, J. D., Arvan, P., Han, J., and Kaufman, R. J. (2021) Therapeutic opportunities for pancreatic  $\beta$ -cell ER stress in diabetes mellitus. *Nat. Rev. Endocrinol.*, 1-13
88. Greeley, S. A., Tucker, S. E., Naylor, R. N., Bell, G. I., and Philipson, L. H. (2010) Neonatal diabetes mellitus: a model for personalized medicine. *Trends Endocrinol Metab* **21**, 464-472
89. Farag, Y. M., and Gaballa, M. R. (2011) Diabetes: an overview of a rising epidemic. *Nephrol. Dial. Transplant.* **26**, 28-35
90. Khan, M. A. B., Hashim, M. J., King, J. K., Govender, R. D., Mustafa, H., and Al Kaabi, J. (2020) Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. *J. Epidemiol. Glob. Health* **10**, 107
91. Glidden, M. D., Aldabbagh, K., Phillips, N. B., Carr, K., Chen, Y.-S., Whittaker, J., Phillips, M., Wickramasinghe, N. P., Rege, N., and Swain, M. (2018) An ultra-stable single-chain insulin analog resists thermal inactivation and exhibits biological signaling duration equivalent to the native protein. *J. Biol. Chem.* **293**, 47-68
92. Sreerama, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32-44
93. Sosnick, T. R., Fang, X., and Shelton, V. M. (2000) Application of circular dichroism to study RNA folding transitions. *Methods Enzymol.* **317**, 393-409
94. Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327
95. Yan, W., Frank, C. L., Korth, M. J., Sopher, B. L., Novoa, I., Ron, D., and Katze, M. G. (2002) Control of PERK eIF2 $\alpha$  kinase activity by the endoplasmic reticulum stress-induced molecular chaperone P58IPK. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15920-15925
96. Pandeyarajan, V., Phillips, N. B., Rege, N., Lawrence, M. C., Whittaker, J., and Weiss, M. A. (2016) Contribution of TyrB26 to the function and stability of insulin: structure-activity relationships at a conserved hormone-receptor interface. *J. Biol. Chem.* **291**, 12978-12990
97. Riddle, M. C., Philipson, L. H., Rich, S. S., Carlsson, A., Franks, P. W., Greeley, S. A. W., Nolan, J. J., Pearson, E. R., Zeitler, P. S., and Hattersley, A. T. (2020) Monogenic Diabetes: From Genetic Insights to Population-Based Precision in Care. Reflections From a Diabetes Care Editors' Expert Forum. *Diabetes Care* **43**, 3117-3128

98. DiMarchi, R. D., Mayer, J. P., Fan, L., Brems, D. N., Frank, B. H., Green, J. K., Hoffman, J. A., Howey, D. C., Long, H. B., Shaw, W. N., Shields, J. E., Sliker, L. J., Su, K. S. E., Sundell, K. L., and Chance, R. E. (1992) Synthesis of a fast-acting insulin based on structural homology with insulin-like growth factor I. in *Peptides: Proceedings of the Twelfth American Peptide Symposium* (Smith, J. A., and Rivier, J. E. eds.), ESCOM Science Publishers B. V., Leiden, The Netherlands. pp 26-28
99. Ciszak, E., Beals, J. M., Frank, B. H., Baker, J. C., Carter, N. D., and Smith, G. D. (1995) Role of C-terminal B-chain residues in insulin assembly: the structure of hexameric Lys<sup>B28</sup>Pro<sup>B29</sup>-human insulin. *Structure* **3**, 615-622
100. Birnbaum, D. T., Kilcomons, M. A., DeFelippis, M. R., and Beals, J. M. (1997) Assembly and dissociation of human insulin and Lys<sup>B28</sup>Pro<sup>B29</sup>-insulin hexamers: a comparison study. *Pharm. Res.* **14**, 25-36
101. Tian, W., Chen, C., Lei, X., Zhao, J., and Liang, J. (2018) CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363-W367
102. Todd, D. J., Lee, A.-H., and Glimcher, L. H. (2008) The endoplasmic reticulum stress response in immunity and autoimmunity. *Nat. Rev. Immunol.* **8**, 663-674