Evolution of insulin at the edge of foldability and its medical implications

Nischay K. Rege[a,1], Ming Liu[b,c,1], Yanwu Yang[d,1], Balamurugan Dhayalan[d], Nalinda P. Wickramasinghe[a], Yen-Shan Chen[d], Leili Rahimi[a,e], Huan Guo[c], Leena Haataja[c], Jinhong Sun[c], Faramarz Ismail-Beigi[a,e,f], Nelson B. Phillips[a], Peter Arvan[c], and Michael A. Weiss[a,d,e,2]

[a]Department of Biochemistry, Case Western Reserve University, Cleveland, OH 44106; [b]Department of Endocrinology and Metabolism, Tianjin Medical University General Hospital, 300052 Tianjin, China; [c]Division of Metabolism, Endocrinology & Diabetes, University of Michigan Medical School, Ann Arbor, MI 48105; [d]Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202; [e]Department of Medicine, Case Western Reserve University, Cleveland, OH 44106; and [f]Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, OH 44106

Edited by Barbara B. Kahn, Beth Israel Deaconess Medical Center, Boston, MA, and approved September 21, 2020 (received for review May 29, 2020)

Proteins have evolved to be foldable, and yet determinants of foldability may be inapparent once the native state is reached. Insight has emerged from studies of diseases of protein misfolding, exemplified by monogenic diabetes mellitus due to mutations in proinsulin leading to endoplasmic reticulum stress and β-cell death. Cellular foldability of human proinsulin requires an invariant Phe within a conserved crevice at the receptor-binding surface (position B24). Any substitution, even related aromatic residue Tyr^{B24}, impairs insulin biosynthesis and secretion. As a seeming paradox, a monomeric Tyr^{B24} insulin analog exhibits a native-like structure in solution with only a modest decrement in stability. Packing of Tyr^{B24} is similar to that of Phe^{B24}, adjoining core cystine B19–A20 to seal the core; the analog also exhibits native self-assembly. Although affinity for the insulin receptor is decreased ∼20-fold, biological activities in cells and rats were within the range of natural variation. Together, our findings suggest that the invariance of Phe^{B24} among vertebrate insulins and insulin-like growth factors reflects an essential role in enabling efficient protein folding, trafficking, and secretion, a function that is inapparent in native structures. In particular, we envision that the *para*-hydroxyl group of Tyr^{B24} hinders pairing of cystine B19–A20 in an obligatory on-pathway folding intermediate. The absence of genetic variation at B24 and other conserved sites near this disulfide bridge—excluded due to β-cell dysfunction—suggests that insulin has evolved to the edge of foldability. Nonrobustness of a protein's fitness landscape underlies both a rare monogenic syndrome and "diabesity" as a pandemic disease of civilization.

evolutionary medicine | unfolded protein response | protein structure | protein folding | folding efficiency

Insulin is a polypeptide hormone that regulates vertebrate metabolism (1). A therapeutic mainstay in diabetes mellitus (DM) (2) and long exploited as a model protein (3), insulin is the two-chain product of a single-chain precursor, proinsulin (4). Expressed in pancreatic β-cells, the precursor contains a connecting segment (C; 35 residues in human proinsulin) linking the C terminus of the B chain (30 residues) to the N terminus of the A chain (21 residues). Although refractory to crystallization, proinsulin's structure in solution contains a folded insulin moiety and flexible connecting segment (Fig. 1*A*) (5). This study has focused on an invariant Phe at position B24 (Phe^{B24}). Its side chain lies within a nonpolar crevice adjoining cystine B19–A20 (Fig. 1 *B–D*), the critical first disulfide bridge to pair in folding (6, 7).

Insulin and insulin-like sequences form a metazoan superfamily, including in invertebrates *Drosophila melanogaster* (8) and *Caenorhabditis elegans* (9) (*SI Appendix*, Fig. S1). Containing a distinctive pattern of disulfide bonds and framework nonpolar residues (10), structures of invertebrate insulin-like proteins (11) contain three α-helices whose orientation and packing resemble mammalian insulins (12, 13). Unlike invertebrate homologs, however, vertebrate domains are extended by a C-terminal B-chain β-strand (B24–B28). This element contains a conserved triplet of aromatic residues—Phe^{B24}, Phe^{B25}, and Tyr^{B26} in insulin—whose contribution to structure and function have been extensively investigated (14–16). Whereas Tyr may replace Phe at B25 and nonaromatic variation is occasionally observed at B26, Phe^{B24} is remarkable for its strict conservation (*SI Appendix*, Fig. S2) (15, 17).

Interest in Phe^{B24} was sharpened by the discovery of human mutations associated with DM (18, 19). Although early studies focused on relative affinities for the insulin receptor (IR), recent observations highlight potential contributions to folding efficiency (20). Proinsulin biosynthesis requires a complex series of chemical and cell-biological steps (21, 22) (Fig. 2), perturbable by diverse clinical mutations (23). Of particular interest are heterozygous mutations in proinsulin leading to toxic misfolding and impaired secretion (*in trans*) of WT insulin (24, 25). Aberrant aggregation triggers endoplasmic reticulum (ER) stress, leading in turn to β-cell dysfunction and eventual cell death (26). First observed in the Akita mouse (Cys^{A7}→Tyr) (27), corresponding human mutations cause permanent neonatal-onset DM (PNDM-*INS*) (28).

Significance

Protein sequences evolve under multiple constraints, including structure, stability, and bioactivity. Yet hidden constraints (inapparent in the native structure) may underlie informational content of protein sequences. An invariant aromatic residue in vertebrate insulins—a phenylalanine at its receptor-binding surface—is required for cellular folding efficiency. Any amino acid substitution impairs cellular biosynthesis, even tyrosine, whose related aromatic side chain preserves native structure and allows function within the range of natural variation. Our results suggest that sequences required for insulin's bioactivity (similar in all vertebrates) are frozen at the edge of nonfoldability. Whereas evolved regulatory networks are ordinarily robust, proinsulin's precarious foldability both underlies a rare monogenic form of diabetes and provides an evolutionary backdrop to the present obesity-related diabetes pandemic.

Author contributions: N.K.R., M.L., F.I.-B., N.B.P., P.A., and M.A.W. designed research; N.K.R., M.L., Y.Y., B.D., N.P.W., Y.-S.C., L.R., H.G., L.H., J.S., and F.I.-B. performed research; N.K.R., M.L., Y.Y., B.D., N.P.W., Y.-S.C., and M.A.W. analyzed data; and N.K.R., N.B.P., P.A., and M.A.W. wrote the paper.

Competing interest statement: M.A.W. has equity in Thermalin, Inc. (Cleveland, OH), where he serves as Chief Innovation Officer. N.B.P. is a consultant to Thermalin, Inc. F.I.-B. has equity in Thermalin, Inc. and is a consultant to Sanofi, Covance, and Novo Nordisk.

This article is a PNAS Direct Submission.

[1]N.K.R., M.L., and Y.Y. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: weissma@iu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2010908117/-/DCSupplemental.

First published November 5, 2020.

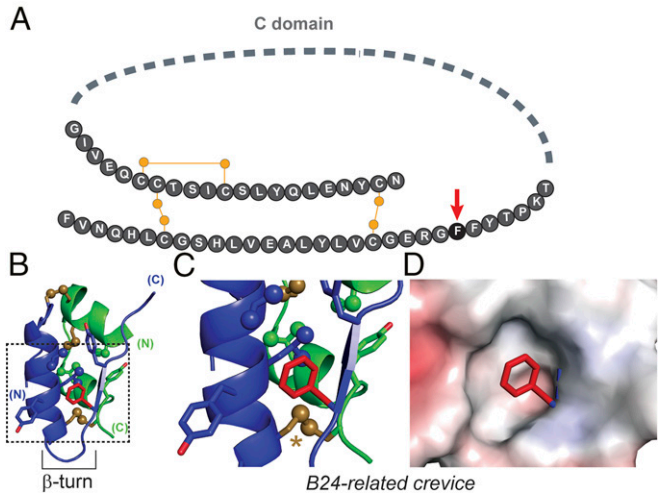

Fig. 1. Role of Phe^{B24} in proinsulin structure. (*A*) Protein sequence. The C domain is shown as a gray dashed line, and canonical disulfide linkages as yellow lines; yellow circles represent sulfur atoms. Phe^{B24} is shown as a black circle and marked by a red arrow. (*B*) Ribbon representation of the mature insulin hormone extracted from the crystal structure of T_6 insulin hexamer (PDB ID code 4INS). The A domain is green, the B domain blue. Disulfide bonds are shown as gold sticks; gold spheres represent sulfur atoms. Oxygen atoms are red, and methyl groups are represented as spheres (0.5 van der Waal radius) The B24 side chain is shown as red sticks. The B20–B23 β-turn is labeled. (*C*) An expanded view of the hydrophobic core of the insulin monomer demarcated by the dotted box in *B*, color coded as in *B*. The A20–B19 disulfide bond is highlighted by a gold asterisk. (*D*) Side chain of Phe^{B24} is shown in relation to the electrostatic potential surface of surrounding residues.

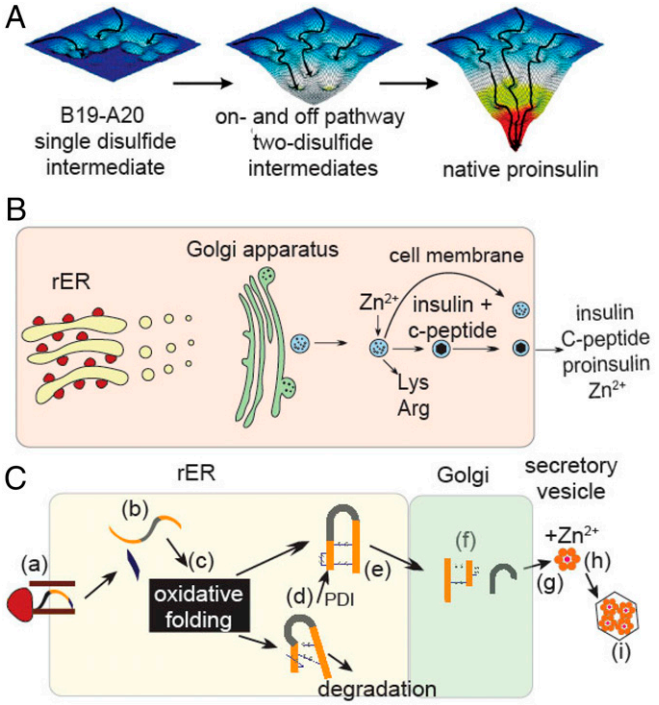

Fig. 2. Folding of proinsulin and its cellular pathway of biosynthesis. (*A*) Possible folding landscapes in nascent folding of proinsulin. Formation of cystine B19–A20 is proposed to represent a critical early step in folding. After its pairing, the landscape deepens, enabling successive formation of on- and off-pathway two-disulfide intermediates. Formation of off-pathway disulfide isomers (e.g., A7–A11 and A6–B7) represent "kinetic traps" within intermediate landscapes, depicted as "hanging valleys." The landscape "matures" when native disulfide pairing is achieved, enabling efficient refolding provided that native disulfide pairing is retained. Image is adapted from ref. 25. (*B*) Schematic depiction of proinsulin trafficking within a cell. (*C*) An expanded view of proinsulin biosynthesis and transport within organelles. Steps are labeled: (*a*) translation in the rough ER (rER); (*b*) translocation into ER with cleavage of signal sequence (29); (*c*) oxidative folding in ER with assistance of oxidoreductases to achieve native disulfide pairing (30); (*d*) quality control (associated with chaperone binding or ER-associated degradation); (*e*) export to the Golgi apparatus; (*f*) processing of the dibasic sites at the BC and CA junctions to yield the mature hormone and C peptide; (*g*) trafficking to glucose-regulated secretory vesicles; and (*h*) zinc-dependent assembly; and (*i*) microcrystallization (1).

Folding of proinsulin in the ER is coupled to native disulfide pairing, a process envisaged in relation to successive folding landscapes (25) (Fig. 2*A*) and requiring specialized ER machinery (Fig. 2 *B* and *C*). Two classes of clinical mutations impair disulfide pairing leading to ER stress. 1) The first involves gain or loss of a Cys, as respectively exemplified by Phe^{B24}→Cys and Cys^{A7}→Tyr (the latter recapitulating the Akita mouse), resulting in an unpaired thiol. Many such mutations have been identified in the A and B domains of proinsulin (20). 2) Non-Cys–related mutations are associated with variable ages of DM onset, ranging from neonatal (as exemplified by Gly^{B8}→Ser) (31, 32) to early childhood (Arg^{B22}→Gln) (19) and even in the third decade of life (Phe^{B24}→Ser) (18). The later the onset (as in maturity-onset diabetes of the young, MODY), the less severe the presumed folding defect, resulting in milder ER stress (33). Collectively, this set of syndromes is designated "mutant *INS*-gene diabetes of youth" (24).[§]

The present study focused on a seeming paradox: Evolutionary exclusion of sequence variation at B24 among vertebrate insulins and insulin-like growth factors (IGFs). In a cellular screen all 19 substitutions in human proinsulin introduced (despite robust initial translation) a general block to biosynthesis. This exclusion includes related aromatic substitution Phe^{B24}→Tyr. Like PNDM variant Cys^{B24}, Tyr^{B24} and representative nonpolar substitutions (Met^{B24}, Leu^{B24}, and Gly^{B24}) were found to induce ER stress and impair *in trans* the folding, trafficking, and secretion of WT human proinsulin. Despite such impaired biosynthesis, the solution structure of a monomeric Tyr^{B24} analog is essentially indistinguishable from that of native insulin (34), and the variant hormone can function in vivo to regulate metabolism. Receptor binding and cellular signaling are impaired but within the range of variation among vertebrate insulins.

Vertebrate insulins exemplify an ancestral protein under rigorous genetic selection: Variation is limited by the hormone's essential roles in development and metabolism. We envisage that Phe^{B24} functions in an oxidative intermediate to facilitate initial pairing of Cys^{B19} and Cys^{A20}, thereby defining an invariant determinant of foldability. Dominant interference with WT proinsulin biosynthesis and induction of ER stress by B24 variants, general features of a monogenic diabetes syndrome, enhance the stringency of foldability as an overarching evolutionary constraint. Because insulin's complex conformational lifecycle—operative in biosynthesis, self-assembly, and receptor binding—imposes independent structural requirements, we envision that its sequence is entrenched at the edge of toxic misfolding. This historical contingency provides an evolutionary backdrop to the emerging pandemic of type 2 diabetes mellitus (T2D) as a disease of civilization.

Results

Our study has two parts. We first focused on the cellular biosynthesis and secretion of variant proinsulins in culture. All

[§]Clinical mutation of the WT Phe (encoded by UUC) to Cys (codon UGC) leads to PNDM with negligible insulin secretion, whereas its mutation to Ser (UCC) is variably associated with MODY with the mutant and WT insulins both detectable in blood (18).

possible B24 substitutions were evaluated in relation to clinical mutations Cys^{B24} (PNDM) and Ser^{B24} (MODY). Cellular studies of nascent folding and secretability were then extended to Tyr^{B24}-proinsulin (chosen based on chemical similarity to Phe) by assays of ER stress and dominant interference with WT biosynthesis in a β-cell line. Next, we investigated the structure and in vitro properties of Tyr^{B24} insulin analogs, either monomeric (in the context of insulin *lispro*; Lys^{B28}, Pro^{B29}-insulin) or assembly competent (in the context of Orn^{B29}-insulin) (*SI Appendix*, Table S1). Preparation of these analogs exploited a native prefolded α-helical domain (*des*-octapeptide[B23–B30]-human insulin; DOI), which was extended by trypsin-catalyzed semisynthesis (15). This approach is robust to mutations in the B23–B30 segment that could otherwise impair nascent disulfide pairing.

B24 Substitutions Impair Biosynthesis. A complete set of B24 substitutions was constructed, and the variant proteins tested for expression and secretion in HEK293T cells. Impaired secretion was observed in each case (Fig. 3 and *SI Appendix*, Fig. S3), indicating a block to one or more prior steps in the biosynthetic pathway (Fig. 2 *B* and *C*). MODY-associated variant Ser^{B24} was least perturbed, which may rationalize its variable genetic penetrance (18). These findings highlight the unique contribution of Phe^{B24} to biosynthesis in accordance with its strict conservation. Furthermore, a chemically diverse subset of substitutions—aromatic (Tyr), aliphatic (Leu), sulfur-containing (Cys or Met), polar (Ser), or without side chain (Gly)—each impaired *in trans* WT cosecretion, a general feature of PNDM-*INS* (Fig. 4*A*).

Tyr^{B24} Impairs Folding and Secretion. We next focused on Tyr^{B24}-proinsulin given the similarity between Phe and Tyr. Pulse-chase assays enabled assessment of foldability in relation to WT proinsulin and clinical variants (Cys^{B24} and Ser^{B24}) (Fig. 4*B*). After biosynthesis for 1 h, ^{35}S-labeled proproteins were immunoprecipitated; expression and secretion were respectively determined by electrophoresis under reducing and nonreducing conditions. Proper oxidative folding was determined by comparing the intensity of the band corresponding to proinsulin-bearing native disulfide bonds (in the nonreducing gel, Fig. 4 *B*, *Upper*) to total proinsulin (reducing gel, Fig. 4 *B*, *Lower*). All variants displayed reduced foldability relative to WT (Fig. 4*C*). Cys^{B24}-proinsulin exhibited the most marked decrease in native fraction; nonnative disulfide isomers accumulated, visualized as more slowly migrating bands in the nonreducing gel. Ser^{B24} and Tyr^{B24} variants exhibited a larger native fraction than did Cys^{B24}. Quantitative analysis of secreted proteins demonstrated that each of the three B24 variants displayed diminished secretability (relative to WT) in rank order Cys^{B24} (most impaired), Tyr^{B24} (intermediate), and Ser^{B24} (least impaired) (Fig. 4*C*).

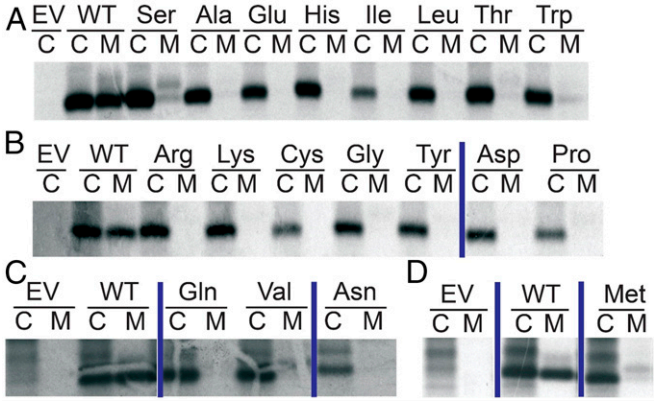

Fig. 3. Survey of substitutions at position B24 of human proinsulin. A secretion screen was performed in HEK293T cells following transient transfection of plasmids encoding WT human proinsulin, B24-variant proinsulins, or an empty vector (EV). The cells were metabolically labeled with ^{35}S-Met/Cys for 30 min followed by 90-min chase. At the end of the chase, media ("M") were collected and cells ("C") lysed. Cell lysates and media were immunoprecipitated with antiinsulin. Immunoprecipitates were analyzed with NuPage (Invitrogen) under reducing conditions. B24 substitutions are indicated above corresponding lanes at *Top*; the empty vector and WT human proinsulin are shown in each of the four gels. *A–D* Represent separate gels with vertical blue lines separating nonconsecutive lanes from the same gel; complete gels are shown in *SI Appendix*, Fig. S3.

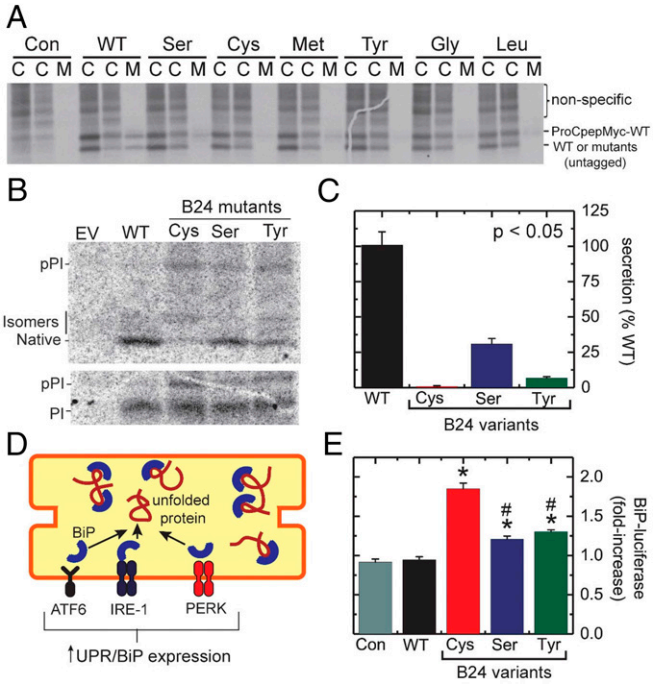

Fig. 4. Cellular folding and secretion properties of proinsulin variants in HEK293T cells. (*A*) Assessment of *trans*-dominance of B24 proinsulin variants in HEK293T cells. Gel depicts ^{35}S-labeled proteins in cell lysates at the end of 30-min ^{35}S-labeling (leftmost lanes marked "C") and after 90-min chase (middle lanes marked "C"). Samples collected from chase media were run in lanes marked "M." Bands correspond to Myc-tagged WT proinsulin and untagged proproteins are as labeled at right. Lanes marked "Con" (control) indicate nontransfected cells. Bands corresponding to myc-tagged proinsulin are visible in the media of cells cotransfected with WT-untagged proinsulin and, more faintly, on coexpression of Ser^{B24}, Gly^{B24}, Met^{B24}, and Leu^{B24} proinsulins. In contrast, the corresponding band is markedly attenuated (or not detectable) in media from cells cotransfected with Cys^{B24} or Tyr^{B24} variants. Portions of panel reprinted from ref. 35 with permission of the authors and included here for clarity and convenience. (*B*) Metabolic labeling of newly synthesized WT and variant preproinsulin (pPI) and proinsulin (PI) analyzed by Tris-tricine-urea SDS/PAGE under nonreducing (*Upper*) and reducing (*Lower*) conditions; bands in nonreducing SDS/PAGE represent natively folded proinsulin or nonnative isomers. (*C*) Secretability of WT or variant proinsulins in HEK293T cells. Differences from WT were in each statistically significant ($P < 0.05$). (*D*) Cartoon depicting BiP activation in unfolded-protein response (UPR). Accumulation of unfolded proteins (red coil) within the ER leads to the dissociation of BiP (blue semicircle) from ATF6 (brown), IRE-1 (purple), and PERK (orange). Liberated BiP then binds to an unfolded protein to prevent proteotoxic aggregation. Activated ATF6, IRE-1, and PERK initiate UPR, decreasing translation of most proteins within the cell and increasing ER-associated degradation (ERAD). Expression of ER-associated chaperones, including BiP, is up-regulated by UPR. (*E*) UPR activation by WT or variant proinsulins as monitored by a BiP-luciferase reporter in MIN6 cells; $^{*}P < 0.05$ relative to WT baseline, whereas $^{\#}P < 0.05$ relative to the elevated level of ER stress induced by Cys^{B24} proinsulin.

In separate experiments, ER stress associated with the misfolding of proinsulin variants was evaluated in mouse insulinoma (MIN6) cells using a BiP reporter assay (Fig. 4*D*). Cells were cotransfected with constructs expressing WT or variant proinsulins; luciferase was driven by a BiP promoter. Luciferase levels were determined at 48 h after transfection. Whereas expression of WT proinsulin stimulated expression of BiP-luciferase at a baseline level similar to that induced by an empty vector (control; "Con" in Fig. 4*E*), Cys^{B24}-proinsulin triggered a 2-fold increase in luciferase expression. Ser^{B24}- and Tyr^{B24}-proinsulin variants were each associated with a 1.5-fold increase in luciferase expression relative to WT (Fig. 4*E*). The difference in ER stress induced by Cys^{B24} versus Tyr^{B24} attained statistical significance ($P < 0.05$).

Tyr^{B24}-Proinsulin Is Trapped within the ER. Cellular trafficking of Tyr^{B24}-proinsulin was visualized in rat insulinoma (INS1) cells, which express and secrete endogenous insulins. Cells were transiently transfected to express WT, Tyr^{B24}, Ser^{B24}, or Cys^{B24} proinsulins. Although immunostaining indicated inefficient transfection (only a few cells per low-power field), cells that expressed WT proinsulin exhibited juxtanuclear staining consistent with Golgi localization. A small number of puncta were visualized beyond the juxtanuclear region, presumably from secretory granules. Cells expressing Tyr^{B24}, Ser^{B24}, and Cys^{B24} variants by contrast exhibited diffuse cytoplasmic localization; juxtanuclear localization was absent. These features suggest impaired Golgi trafficking (Fig. 5), an interpretation strengthened by colocalization of the variants with calnexin (red), an ER-associated chaperone. WT proinsulin displayed only modest calnexin colocalization, suggesting more rapid ER egress. Together, these differences suggest that the variants are retained in the ER, presumably due to a quality-control checkpoint (Fig. 2).

Tyr^{B24} Analog Exhibits Partial Activity. The affinity of Tyr^{B24}, Orn^{B29}-insulin for the detergent-solubilized, lectin-purified IR holoreceptor was determined by competition with ^{125}I-Tyr^{A14}-insulin (Fig. 6 *A* and *B* and Table 1). Consistent with previous studies (17), Tyr^{B24}, Orn^{B29}-insulin exhibited 20-fold–reduced affinities (relative to WT or Orn^{B29}-insulin) for IR isoforms A and B; binding to the type 1 IGF receptor (IGF-1R) was commensurately reduced (*SI Appendix*, Fig. S4). Cell-based assays demonstrated that Tyr^{B24}, Orn^{B29}-insulin retained partial signaling activities. Whereas in HepG2 cells such activity seemed substantial at a high hormone concentration [56 (±15)% and 59% (±12)% relative to Orn^{B29}-insulin at 50 nM as probed by IR autophosphorylation and downstream AKT phosphorylation] (western blots in Fig. 6 *C* and *D*), in a fluorescence-based pIR plate assay (*SI Appendix*, Fig. S5*A*) the analog's dose–response relationship in HepG2 cells was shifted to the right (Fig. 7 *A* and *B*). The degree of shift (~10-fold) was near-commensurate with the biochemical perturbation in IR affinities. A similar dose–response shift was observed in control studies of a natural species variant (*des*-B30 porcupine insulin) (Fig. 7 *B* and *C*). Proportionate attenuation of AKT phosphorylation at low hormone doses was also observed in MCF-7 breast cancer cells (5, 10, and 20 nM) (*SI Appendix*, Fig. S5) wherein the downstream signaling activity of Tyr^{B24}, Orn^{B29}-insulin was also similar to that of *des*-B30 porcupine insulin.

The in vivo potency of Tyr^{B24}, Orn^{B29}-insulin was next tested after intravenous injection in rats rendered diabetic by streptozotocin. Tyr^{B24}, Orn^{B29}-insulin exhibited reduced blood-glucose lowering capability after intravenous injection relative to Orn^{B29}-insulin. A nativelike pharmacodynamic profile was nonetheless recapitulated (in relation to the parent Orn^{B29}-insulin) at a twofold higher dose (Fig. 6*E*). Substantial preservation of in vivo potencies despite more marked in vitro perturbations has

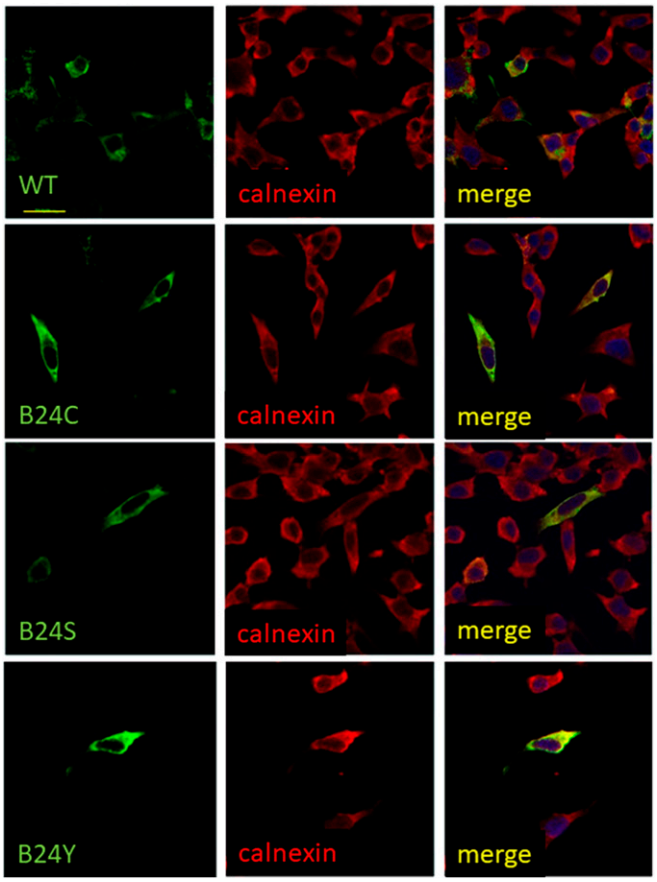

Fig. 5. Immunofluorescence studies of intracellular localization of WT or variant proinsulins in INS1 cells. The cells, which express endogenous rat proinsulins, were transiently transfected to express WT human proinsulin (*Top*) or B24 variants (Cys^{B24}, Ser^{B24} or Tyr^{B24} *Bottom*). Green fluorescence (*Left*) marks the human proteins, and red fluorescence (*Middle*) indicates ER-marker calnexin. In each row merger of these respective images is shown (*Right*). WT human proinsulin advances beyond the ER to accumulate in nascent secretory granules in the juxtanuclear Golgi region; in some cells, WT proinsulin traffics to mature secretory granules near the plasma membrane. By contrast, the variant proinsulins are retained in the ER. In these assays only one (or a few) cells per low-power field were transfected. (Scale bar, 20 μm.) Panels depicting WT- and Cys^{B24} proinsulin are reprinted from ref. 35 with permission of the authors and included here for clarity and convenience.

broadly been observed in studies of insulin analogs (*Discussion*) (36). A similar trend was seen in the context of insulin *lispro* (KP-insulin)[†], an syndromes is designated template (37) (*SI Appendix*, Fig. S6). Substantial in vivo activities have also been reported in studies of insulin analogs containing other B24 substitutions (Met^{B24}, Leu^{B24}, and Gly^{B24}) (*SI Appendix*, Fig. S7) (3, 15, 35, 38) shown above to impair proinsulin biosynthesis. In spite diminished relative affinity for IR in vitro (*SI Appendix*, Fig. S8 and Table S2), the in vivo activity of Tyr^{B24}, Orn^{B29}-insulin was similar to that of *des*-B30 porcupine insulin[‡] (*SI Appendix*,

[†]Insulin *lispro* (KP-insulin, containing substitutions Pro^{B28}→Lys and Lys^{B29}→Pro) is the active component of Humalog (Eli Lilly), a rapid-acting analog formulation in clinical use. The Lys-Pro switch destabilizes the dimer interface (37).

[‡]Porcupine insulin, like many of the divergent insulins in mammalian suborder *Hystricomorpha*, exhibits decreased stability, reduced IR affinity (25%), and diminished cellular activity in vitro (4%) (39).

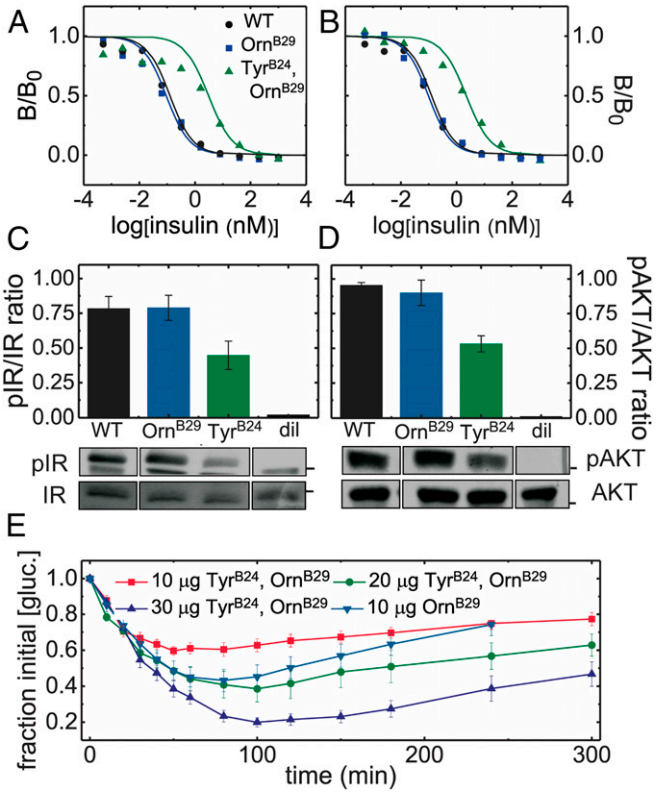

Fig. 6. Biological activity of Tyr^{B24}, Orn^{B29}-insulin in relation to control analogs. (*A* and *B*) Respective competitive receptor-binding assays IR-isoforms B and A (IR-B and IR-A). Data points are represented by colored symbols; corresponding lines indicate fitting; the color code is provided in *A*. Data in *B* is reprinted from ref. 35 with the permission of the authors and provided here for convenience. (*C* and *D*) Insulin signaling in a mammalian cell line: histogram depicting the (*C*) p-IR/IR ratio and (*D*) p-AKT/AKT ratio (determined by western blot) in MCF-7 cells treated with 50-nM insulin analogs or diluent vehicle ("dil"). Black bars indicate SEM. Bands corresponding to *C* p-IR and IR and *D* p-AKT and AKT are displayed underneath the corresponding columns of each histogram. Black boxes separating lanes indicate nonconsecutive lanes from the same gel. Black bars on IR/pIR and AKT/pAKT gels, respectively, indicate molecular mass of 100 kDa and 50 kDa. (*E*) A dose–response experiment assessing the in vivo activity of Tyr^{B24}, Orn^{B29}-insulin in relation to control Orn^{B29}-insulin in diabetic rats. Blood-glucose values averaged 400 mg/dL (n = 5).

Fig. S9), and so was within the range allowed by natural variation (39, 40).

Tyr^{B24} Analog Retains Native Secondary Structure with Reduced Stability. Circular dichroism (CD) spectroscopy was utilized to assess the structure and stability of two Tyr^{B24}-insulin analogs. The spectrum of Tyr^{B24}, Orn^{B29}-insulin (at both 25 and 37 °C) exhibited attenuated α-helical features relative to WT or native Orn^{B29}-insulin (*SI Appendix*, Figs. S10*A* and S11*A*). Mitigation of such relative attenuation on 10-fold dilution suggested that this difference reflects impaired dimerization (*SI Appendix*, Fig. S12). This conclusion was corroborated by CD studies of corresponding monomeric KP-insulin analogs (*SI Appendix*, Fig. S10*C*).

Guanidine denaturation studies indicated decreased thermodynamic stabilities of Tyr^{B24} analogs. Whereas WT and Orn^{B29}-insulin had similar free energies of unfolding (ΔG_u as inferred from a two-state model), the stability of Tyr^{B24}, Orn^{B29}-insulin was decreased by 0.9 (±0.2) kcal/mol at both 25 and 37 °C (Table 2 and *SI Appendix*, Figs. S10*B* and S11*B*, and Table S3*A*). In the context of KP-insulin, the destabilizing effect of Tyr^{B24} was slightly smaller [$\Delta\Delta G_u$ 0.6 (±0.2) kcal/mole] (Table 2 and *SI Appendix*, Fig. S10*D*) and similar to the respective decrements in stability between *des*-B30 human insulin and *des*-B30 porcupine insulin [$\Delta\Delta G_u$ 0.7 (±0.2) kcal/mole] (*SI Appendix*, Fig. S13 and Table S3*B*) and between human insulin and IGF1 (*SI Appendix*, Table S3*C*). These changes are not expected in themselves to impair cellular trafficking and secretion (*SI Appendix*, Fig. S11 *C* and *D* and Table S3*D*).

Table 1. Receptor affinities of insulin analogs

Analog	IR-A (nM)	IR-B (nM)	IGF-1R* (nM)
WT insulin	0.08 ± 0.03	0.08 ± 0.03	2.30 ± 0.39
Orn^{B29} -insulin†	0.07 ± 0.04	0.08 ± 0.02	2.49 ± 0.41
Tyr^{B24}, Orn^{B29}-insulin	1.73 ± 0.57	2.63 ± 0.44	22.5 ± 4.0

Affinities were determined using a detergent-solubilized and immobilized receptor by competitive displacement of ^{125}I-Tyr^{A14}-human insulin.
*IGF-1R denotes Type 1 IGF receptor.
†Substitution of Lys^{B29} by related basic residue ornithine (Orn) was introduced to simplify trypsin-mediated semisynthesis.

Tyr^{B24} Preserves Native-like 3D Structure. The solution structure of a monomeric Tyr^{B24}–DKP-insulin analog (DKP-insulin; Asp^{B10}, Lys^{B28}, Pro^{B29}) was determined using heteronuclear 3D/4D-NMR methods. ^{1}H, ^{13}C, and ^{15}N chemical shifts exhibited similar patterns in the parent and Tyr^{B24} analog (*SI Appendix*, Figs. S14 and S15). The side chain of Leu^{B15}, which packs near B24, selectively exhibited attenuated secondary shifts (*SI Appendix*, Fig. S15*A* and Table S4*A*). Main-chain ^{1}H-^{15}N HSQC fingerprints of the two analogs are essentially identical; slightly attenuated H_N secondary shifts were observed at positions B17–B19 and B21 (*SI Appendix*, Fig. S15*B* and Table S4*B*). The Tyr^{B24} analog's solution structure (Fig. 8*A*) was essentially identical to its Phe^{B24} parent (Fig. 8*B* and *SI Appendix*, Tables S5 and S6), including near B24 (Fig. 8 *C* and *D*). Corresponding side-chain dihedral angles were similar in detail; for Ile^{A2}, Tyr^{A19}, Leu^{B11}, Val^{B12}, Leu^{B15}, Tyr^{B16}, Tyr^{B24} (or Phe^{B24}), and Tyr^{B26}, any such differences were generally <10° (*SI Appendix*, Table S7). Structural comparisons are further described in *SI Appendix*, *Supplemental Methods* (*SI Appendix*, Figs. S14–S18 and Tables S4–S8).

Thermodynamic differences between Tyr^{B24} and Phe^{B24} analogs were corroborated in KP-insulin through measurement of ^{1}H-^{2}H exchange rates in 10 mM deuterio-acetic acid at pH 3.0 and 25 °C (41) (*Methods* and *SI Appendix*, Figs. S17 and S18). In each case protons within insulin's core, comprising the internal surfaces of the C-terminal A-chain α-helix and central B-chain α-helix, were most protected; exchange of amide protons closest to internal cystine B19–A20 requires global unfolding (42) (Fig. 9). Protection factors (PF) at these sites were generally decreased in the Tyr^{B24} analog, leading to an estimated $\Delta\Delta G_u$ of 0.35 (±0.04) kcal/mol (*SI Appendix*, Table S8). Although Tyr^{B24}-KP insulin retained the same residue-specific pattern of subglobal exchange sites as in its parent, these sites generally exhibited more rapid exchange (reduced PFs) in the variant. This was most pronounced at subglobal sites within the central B-chain α-helix (42), indicating that Tyr^{B24} reduces its segmental stability. These subtle findings suggest that the *para*-OH group of Tyr^{B24} frustrates packing within the U-shaped supersecondary interface (engagement of Leu^{B11}, Val^{B12}, Leu^{B15}, and Cys^{B19} by Tyr^{B24} and Tyr^{B26}) (3), leading to enhanced conformational fluctuations.

Tyr^{B24} Weakens Dimerization without Perturbing the Kinetic Stability of the R_6 Hexamer. Dimerization was probed by size-exclusion chromatography (SEC): Analogs were made 0.6 mM in zinc-free solution and applied to a gel-filtration column (*SI Appendix*, Fig. S19*A*). Whereas the elution profile of KP-insulin

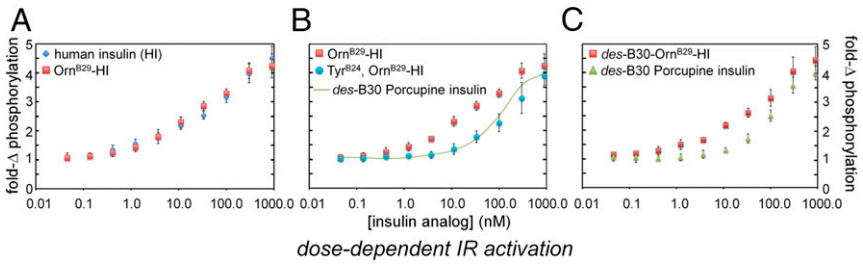

Fig. 7. Insulin signaling and dose–response relationships. (*A*) Similar activities of WT insulin (HI; blue diamond) and Orn^{B29}-insulin (Orn^{B29}-HI; red square). Ligand concentrations are shown in horizontal axis. (*B*) The dose–response relationship of Tyr^{B24}, Orn^{B29}-insulin (blue circle) is shifted to the right relative to Orn^{B29} parent (red square). Superimposed as a green line is the activity of *des*[B30]-porcupine insulin. (*C*) Dose–response relationships of *des*[B30]-porcupine insulin (green triangle) relative to *des*[B30]-Orn^{B29} human insulin (red square). Ligand doses: 0.05, 0.15, 0.5, 5, 10, 30, 100, 300, and 900 nM; assays were performed in triplicate.

corresponded to monomeric insulin (calculated molecular mass: 4.2 kDa) (*SI Appendix*, Fig. S19*B*), the elution peak of Tyr^{B24}, Orn^{B29}-insulin contained a dimeric component but was largely monomeric (4.7 kDa). WT (6.8 kDa) and Orn ^{B29}-insulin (6.6 kDa) eluted as dimers (Table 3, right column). The lifetime of zinc-coordinated, phenol stabilized (R_6) hexamers formed by the variant was measured using Co^{2+}-absorption spectroscopy (*SI Appendix*, Fig. S19 *C* and *D*). The tetrahedral Co^{2+} absorbance band at 574 nm (characteristic of an R_6 hexamer) of Tyr^{B24}, Orn^{B29}-insulin was essentially identical to those of native Orn^{B29}-insulin and WT insulin (*SI Appendix*, Fig. S19*E*). Kinetic sequestration studies (16) of native Orn^{B29}-insulin yielded a dissociation half-life ($t_{1/2}$) of 8.6 (±1.3) min. WT insulin and Tyr^{B24}, Orn^{B29}-insulin exhibited similar dissociation rates [$t_{1/2}$ = 7.7 (±1.3) and 7.6 (±0.3) min, respectively]; these values were within one SD of Orn^{B29}-insulin. Any difference in dissociation rates between Tyr^{B24}, Orn^{B29}-insulin and Orn^{B29}-insulin was small relative to KP-insulin [$t_{1/2}$ = 4.6(±0.3) min] (Table 3, middle column, and *SI Appendix*, Fig. S19*F*).

Susceptibility of Tyr^{B24} Analog to Fibrillation Is within Species Variation. The lag time of Tyr^{B24}, Orn^{B29}-insulin was measured to probe relative susceptibility to proteotoxic aggregation. Tyr^{B24}, Orn^{B29}-insulin exhibited a lag time of 2.1 (±0.3) h. Although shorter than that of WT insulin or native Orn^{B29}-insulin [lag times 4.8 (±2.5) and 4.3 (±0.5) h, respectively], the lag time of the Tyr^{B24} analog was longer than that of bovine insulin [1.2 (±0.2) h] (Table 4 and *SI Appendix*, Fig. S20) and indistinguishable from that of *des*-B30 porcupine insulin (*SI Appendix*, Fig. S21 and Table S9), species variants chosen to represent allowed natural variation.

Table 2. Thermodynamic stabilities of insulin analogs

Analog	ΔG_u* (kcal mol^{-1})	C_{mid} (M)	$m^{\dagger}$ (kcal mol^{-1} M^{-1})
Human insulin	3.2 ± 0.1[‡]	5.2 ± 0.1	0.62 ± 0.02
Orn^{B29} -insulin	3.2 ± 0.1	5.0 ± 0.1	0.64 ± 0.01
Tyr^{B24}, Orn^{B29}-insulin	2.3 ± 0.1	4.7 ± 0.2	0.49 ± 0.03
Bovine insulin[§]	3.3 ± 0.1	4.6 ± 0.1	0.72 ± 0.02
Insulin *lispro*[¶]	2.9 ± 0.1	4.9 ± 0.2	0.59 ± 0.01
Tyr^{B24}-insulin *lispro*	2.3 ± 0.1	4.9 ± 0.2	0.47 ± 0.02

*Parameters were inferred from CD-detected guanidine denaturation data by application of a two-state model; uncertainties represent fitting errors for a given dataset.

[†]The *m*-value [slope Δ(G)/Δ(M)] correlates with extent of hydrophobic surfaces exposed on denaturation.

[‡]Estimates of ΔG_u pertain to insulin in 50 mM KCl and 10 mM KPi (pH 7.4) at 25 °C.

[§]Bovine insulin differs from human insulin at three sites (Thr^{B30}→Ala, Thr^{A8}→Ala and Ile^{A10}→Val), representative of natural variation among species; further variation is provided by porcupine insulin (*SI Appendix*, Fig. S14 and Table S9).

[¶]Insulin *lispo* (the active component of Humalog; Lilly) contains paired substitutions Pro^{B28}→Lys (K) and Lys^{B29}→Pro (P); the analog is thus designated KP-insulin.

Discussion

Evolution of the vertebrate insulin-IGF family provides a model for study of genetic variation in a protein under strict selection. The present study has defined a cryptic yet essential contribution of an invariant aromatic residue (Phe^{B24}) to nascent foldability and demonstrated that this contribution is as important as the residue's direct engagement at the hormone–receptor interface (43).

Oxidative refolding of proinsulin in vitro occurs in steps (44). The first is preoxidative: Nascent structure in the A and B domains coalesce to favor formation of the A20–B19 disulfide bridge (the preferred second step) (45). This and successive disulfide bridges steepen the folding landscape ("landscape maturation"). Although productive folding in vivo (46) is facilitated by ER oxidoreductases, including protein disulfide isomerase (47), specific disulfide pairing nonetheless poses a baseline biological challenge: Indeed, in the majority of human cell lines even WT biosynthesis is inefficient (24). A specialized β-cell ER proteome presumably evolved to mitigate this challenge. PNDM mutations in proinsulin severely impair foldability, as exemplified by mutations at B8, an invariant Gly. "Chiral mutagenesis" of Gly^{B8} implicated the sign of the B8 ϕ dihedral angle as a critical parameter affecting foldability (48, 49). D-amino acids at B8 enhance folding yields in vitro but markedly impair activity once the native state has been reached. Foldability and function may thus impose distinct evolutionary constraints and even be at odds.

Survey of B24 Substitutions. In accordance with the strict conservation of Phe^{B24}, we first observed in HEK293T cells that each of the 19 possible amino acid substitutions at B24 impairs the biosynthesis and secretion of human proinsulin. This finding highlights the importance of Phe^{B24} in a folding intermediate. Equilibrium models [i.e., single-chain insulin analogs and related peptides containing only cystine B19–A20 (46)] provided evidence that the B24 side chain participates in a native-like cluster of conserved nonpolar residues near this bridge, including Leu^{B15}, Leu^{A16}, and Tyr^{A19}; their interactions are a defining feature of the native core (44, 46). We imagine that diffusion and collision of nascent α-helices in the B and A domains of reduced proinsulin favor such nativelike clustering and, in turn, B20–A19 disulfide pairing. A subset of non-Cys–related PNDM mutations occur at these sites and at Gly^{B23}, whose conformational properties may enable Phe^{B24} to join the nascent cluster (50). This mechanism, foreshadowing native B-chain supersecondary structure, raises the possibility that substitutions at B24 either 1) decrease the likelihood that B24 joins the cluster or 2) perturb the immediate redox chemistry underlying B19–A20 disulfide bonding.

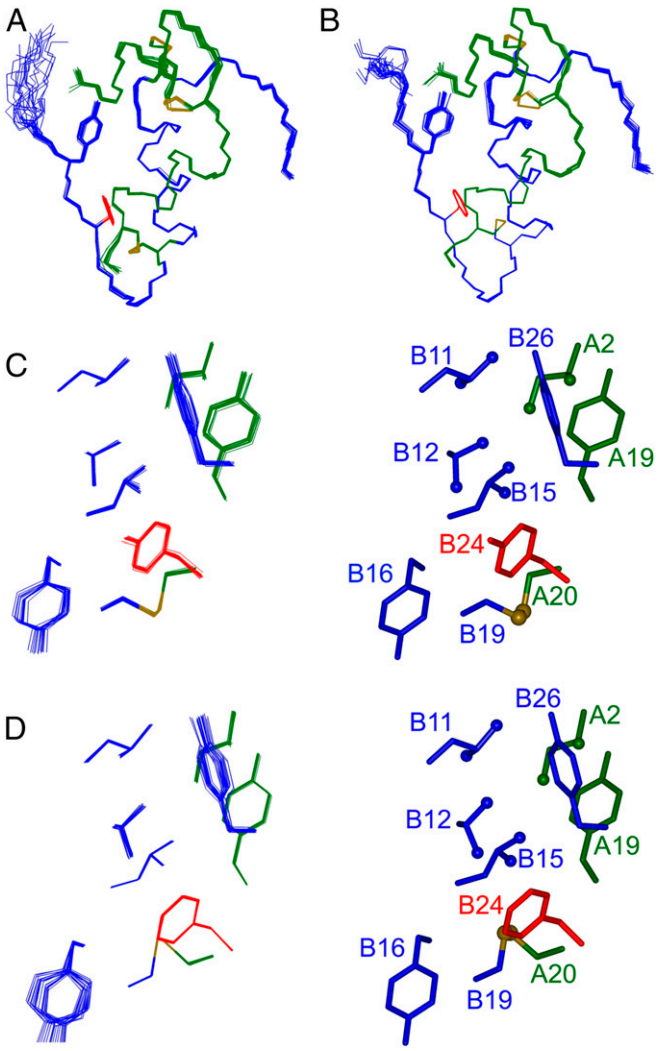

Fig. 8. Solution structure of Tyr^{B24} insulin monomer. (*A*) Tyr^{B24} DKP-insulin and (*B*) parent DKP-insulin. In each case the A chain is green, and B chain blue; B24 and B26 side chains (sticks) are red and blue; disulfide bridges are gold. (*C* and *D*) Expanded view of B24 environment in the ensemble (*Left*) and in a representative stick model (*Right*): (*C*) Tyr^{B24} and (*D*) Phe^{B24}. The color scheme is as in *A*. Methyl groups represented as green or blue spheres; and sulfur atoms, gold spheres. The solution structures were derived from heteronuclear NMR analysis.

Because of the chemical similarity of Phe and Tyr (often interchangeable in protein families), we next focused on Tyr^{B24}-proinsulin in an effort to understand its strict evolutionary exclusion. Immunostaining was undertaken of INS-1 cells expressing this variant in relation to mild or severe clinical mutations: Ser^{B24} (MODY) or Cys^{B24} (PNDM). The variants each exhibited an altered immunofluorescence pattern relative to WT. Whereas WT proinsulin accumulated in juxtanuclear puncta and mature secretory granules, the variants generally exhibited a diffuse localization colocalizing with the ER-chaperone calnexin. Similar features were described in studies of the Akita mutation (Cys^{A7}→Tyr) (33, 51) and an unrelated MODY mutation (Arg^{B22}→Glu) (50). These findings indicate ER dysfunction but do not provide a quantitative probe of disease severity. In HEK293T cells Cys^{B24} and Tyr^{B24} proinsulin (but not the milder Ser^{B24} proinsulin) interfered *in trans* with WT folding and trafficking, a general feature of PNDM-*INS* mutations (52). Such dominance appears due to nascent dimerization and higher-order association between the variant and WT prohormones. In cells lacking critical ER chaperones (53) or components of the ER exit machinery, WT proinsulin exhibits analogous entanglement (54). In the natural history of nonsyndromic T2D, overexpression of WT proinsulin (in the face of peripheral insulin resistance) may likewise lead to misfolding, entanglement, and chronic ER stress (55).

Phe^{B24} Provides an "Aromatic Anchor" in Diverse Processes. In native insulin Phe^{B24} exhibits three distinct structural environments: 1) In the monomer the aromatic ring rests within a crevice to stabilize the B-chain U-turn and seal the core; 2) in the dimer it packs within an antiparallel β-sheet to stabilize self-assembly; and 3) in the hormone-receptor complex it packs within a conserved interfacial pocket. Substitution of Phe^{B24} by Tyr perturbs each of these structures but within the range of natural variation among vertebrates. The enigma of Tyr's evolutionary exclusion is deepened by the invariance of Phe^{B24} also among vertebrate IGFs, whose lower stability and ambiguous disulfide pairing properties in vitro (56) are compensated in vivo by specific

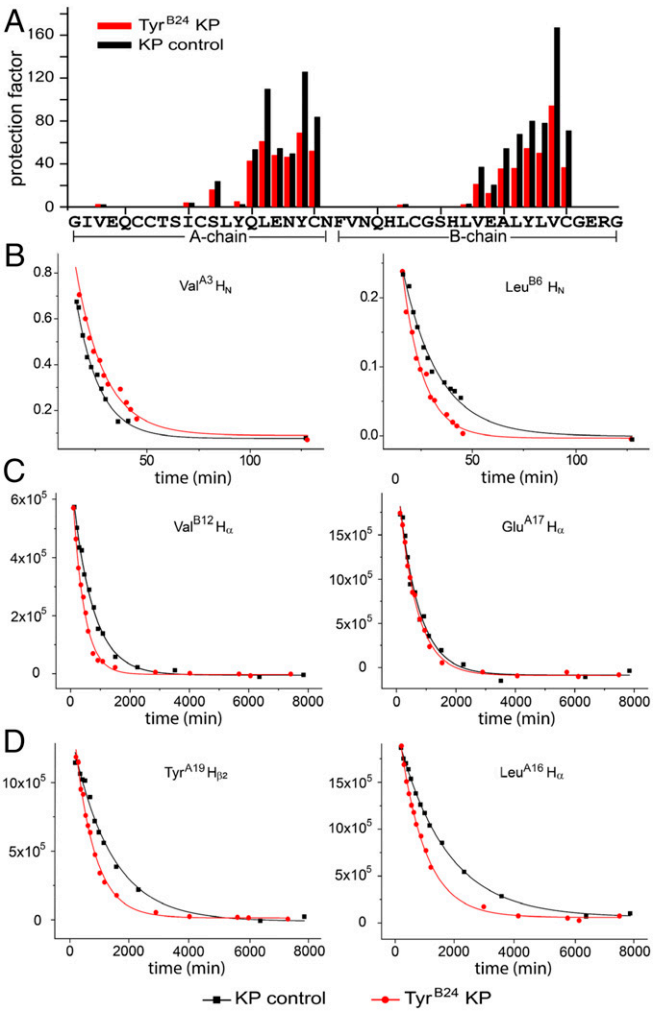

Fig. 9. Amide-proton ^1H-^2H exchange kinetics. (*A*) Histogram of residue-specific PFs in Tyr^{B24}-insulin *lispro* (red bars) and parent insulin *lispro* (black bars). Assays were performed in 10 mM deuterio-acetic acid (pH 3.0). (*B–D*) Representative examples of exponential ^1H-^2H exchange at specific residues associated with (*B*) local, (*C*) subglobal, or (*D*) global exchange kinetics as defined in the parent monomer (42). Analysis of global exchange corroborates decrease in ΔG_u due to B24 substitution (*SI Appendix*, Table S7). Additional ^1H-NMR spectra are given in *SI Appendix*, Fig. S5.

Table 3. Self-association properties of insulin analogs

Analog	$t_{1/2}$ Hexamer dissociation* (min ± SD)	Calculated molecular mass by SEC† (kDa ± fit error)
Human insulin (ins)	7.7 ± 1.3	6.8 ± 3.1
Ins *lispro*‡	4.6 ± 0.3	4.2 ± 2.0
Orn^{B29} -ins§	8.6 ± 1.3	6.6 ± 3.0
Tyr^{B24}, Orn^{B29}-ins	7.6 ± 0.3	4.7 ± 2.2

*$t_{1/2}$ and SD were calculated from three replicates; two-sided *P* values for *lispro*, Orn^{B29}-ins, and Tyr^{B24}, Orn^{B29}-ins against $t_{1/2}$ of human insulin were 0.06, 0.44, and 0.91, respectively.
†Proteins were made 0.6 mM in a buffer containing $ZnCl_2$ at a ratio of two zinc ions per insulin hexamer and applied to SEC column (*Materials and Methods*); masses were calculated based on protein standards (*SI Appendix*, Fig. S19*B*).
‡Insulin lispro (KP-insulin) contains substitutions Pro^{B28}→Lys (K) and Lys^{B29}→Pro (P).
§Substitution Lys^{B29}→Orn was introduced for synthetic convenience.

IGF-binding proteins (57). We discuss these levels of structure in turn.

Insulin monomer. The structure of Tyr^{B24}-insulin *lispro* (as an engineered monomer in solution) is essentially identical to WT (58). In particular, the conformation and structural environment of Tyr^{B24} are similar to those of Phe^{B24}; the *para*-OH group is largely exposed to solvent. Thus, the structure of the native state, once achieved, does not inform how or why Tyr^{B24} perturbs the folding of proinsulin. Subtle ^{1}H-NMR dynamic differences were observed between Phe^{B24} and Tyr^{B24} insulin analogs, including more rapid ^{1}H-^{2}H amide proton exchange at subglobal sites (*SI Appendix*, *Supplemental Biophysical Discussion*). Although of biophysical interest, these may be of limited functional significance in the native state. In contrast, we imagine that such dynamic perturbations provide a critical echo of frustrated side-chain interactions in an on-pathway protein-folding intermediate. Enhanced conformational fluctuations in the Tyr^{B24} analog may also rationalize its shorter fibrillation lag time.

Although Tyr^{B24} attenuates the stability of an engineered insulin monomer [$\Delta\Delta G_u$ 0.6 (±0.2) kcal/mole relative to insulin *lispro*], quality-control mechanisms in the ER do not "measure" such stabilities. Indeed, a "molten globule" variant of proinsulin (containing pairwise substitution of cystine A6–A11 by Ala) (46, 51) exhibits efficient biosynthesis and secretion, suggesting that even a partial fold of marginal stability can pass quality-control criteria. How Tyr^{B24} impairs folding efficiency may be considered in relation to model folding landscapes: Successive funnels to the native state (59) coupled to sequential disulfide pairing (60) (Fig. 2*A*). Because paths from rim to base in each landscape encounter local maxima and minima—respectively associated with kinetic barriers or kinetic traps (61)—it is possible that the same small thermodynamic decrement in a metastable folding intermediate [i.e., $\Delta\Delta G_u$ 0.6 (±0.2) kcal/mole relative to a lower baseline stability] can markedly impede folding efficiency (62).

Insulin dimer and hexamer assembly. In the context of a dimerization-competent analog (Orn^{B29}-insulin), Tyr^{B24} weakens self-association but remains compatible with native R_6 hexamer assembly. It may seem surprising that Tyr^{B24} weakens dimerization, and yet nascent Tyr^{B24}-proinsulin can still interfere with WT proinsulin biosynthesis. We imagine that conformational features of a "molten" interface in the ER would impose less-stringent constraints on side-chain structures than is imposed by the native insulin dimer (3). Imprecision in an ER folding intermediate, for example, could make Phe^{B24} and Tyr^{B24} equally able to participate in transient encounters among partial folds. The critical difference would derive from a specific Tyr^{B24}-associated kinetic block to disulfide pairing, leading to greater exposure of nonpolar surfaces and unpaired thiol groups to form aberrant intermolecular disulfide bridges (52).

Hormone–receptor complex. Receptor-binding assays demonstrated a 20-fold reduction in IR affinity of Tyr^{B24}, Orn^{B29}-insulin. The analog nonetheless displayed substantial glucose-lowering activity in relation to Orn^{B29}-insulin (or WT) in diabetic rats. That in vivo activities are maintained despite more marked in vitro perturbations has been broadly observed in studies of insulin analogs (36) and ascribed to delayed receptor-mediated clearance of the hormone from the bloodstream (36). Such coupling between receptor binding and clearance provides a physiologic compensation, which has presumably enhanced the evolutionary robustness of metabolic regulation.

The twofold difference in the in vivo activity of Tyr^{B24}, Orn^{B29}-insulin is unlikely to be of physiologic significance as reference ranges for blood-insulin concentrations vary three- to sixfold in nondiabetic individuals, and in principle β-cells could readily augment insulin secretion twofold to compensate (63). Moreover, the in vitro affinity of Tyr^{B24}, Orn^{B29}-insulin for IR and its activities in cellular assays in in vivo fall within the range of species variants (Fig. 7 and *SI Appendix*, Fig. S5 and Table S10) and so would not in themselves pose insuperable evolutionary barriers. The substantial activity of Tyr^{B24} insulin analogs is in accordance with the recent demonstration that Gly^{B24}-insulin can be used to treat DM as effectively as WT (35) and by the substantial in vivo activities of diverse nonfoldable yet stable variants (Met^{B24}, Leu^{B24}, and Ser^{B24} analogs) (Fig. 4*A* and *SI Appendix*, Table S11).

Evolution and Human Genetics. The baseline rate of de novo mutations in vertebrate reproduction implies that variation in the *INS* gene is ongoing. Although rare mutations like Cys^{B24} are increasingly being ascertained in genetic clinics, these mutations represent evolutionary dead ends (excepting modern medical care). The invariance of Phe^{B24} and other key residues in the insulin/IGF family (e.g., Gly^{B8}, Val^{B12}, Leu^{B11}, Leu^{B15}, Val^{A3}, Leu^{A16}, and Tyr^{A19}) (3) presumably reflects the stringency of purifying selection for proper hormonal control of developmental processes and metabolic homeostasis.

The milder Ser^{B24} variant (lacking *trans* interference in accordance with blood levels of WT insulin and Ser^{B24} insulin in the proband) (18) is also excluded from the vertebrate insulin/IGF family. Although this mutation is inheritable, its evolutionary exclusion presumably reflects decreased reproductive fitness over several generations (excepting modern medical care) in accordance with principles of population genetics (64). We posit that the age of onset or severity of DM due to mutations in the insulin gene (23) may correlate with extent of ER stress (65)

Table 4. Fibrillation lag times of insulin analogs

Analog	Fibrillation lag time (h ± SD)
Human insulin (ins)	3.9 ± 1.1*
Bovine ins	1.2 ± 0.2
Orn^{B29}-ins	4.3 ± 0.5
Tyr^{B24}, Orn^{B29}-ins	2.1 ± 0.3

*For human insulin, mean and SD were calculated without inclusion of an extreme outlier; with this point included, estimate is 4.8 ± 2.7.

BIOCHEMISTRY

or Golgi stress (66). In the natural history of PNDM and MODY due to *INS* mutations, impaired glucose tolerance and onset of DM precedes β-cell death or apoptosis. The present comparison of ER-stress levels induced by Tyr^{B24}-proinsulin relative to WT, Ser^{B24}, and Cys^{B24} proinsulins suggests that the phenotype of a human patient (or other mammal) expressing Tyr^{B24}-proinsulin would be more severe than associated with Ser^{B24} (MODY) but less severe than associated with Cys^{B24} (PNDM) (i.e., onset of DM in the first decade of life) (67). This phenotype would impose significant selection pressure against the evolutionary appearance and maintenance of this variant. Such a patient may nonetheless be identified in pending clinical screens because a Tyr codon can be obtained from the Phe codon with a single base change.

Concluding Remarks. The informational content of protein sequences can have many layers of meaning. Patterns of conservation and divergence provide a starting point to decipher evolutionary constraints. Our findings suggest that Tyr^{B24} perturbs an on-pathway folding intermediate but is well tolerated once the native state is reached. The modest decrements observed in thermodynamic stability, conformational stability, self-assembly, and amyloidogenicity are likely to be within the range of species variants. The existence of such "pathway mutations" was foreshadowed in studies of a phage tail-spike (68). Whereas the latter is a multisubunit β-helix whose folding is coupled to assembly, proinsulin folds as a monomeric domain (21). We envisage that the aromatic ring of Phe^{B24} stabilizes a nascent folding nucleus containing a conserved cluster of nonpolar side chains. This nucleus foreshadows structural features of the native state. Remarkable for its nonrobustness, the foldability of proinsulin thus depends on subtle features of invariant residues: An unadorned aromatic ring (Phe^{B24}), neighboring packing efficiency (e.g., Leu^{A16}) (69), and the thin thread of a single dihedral angle (Gly^{B8}) (48, 70).

In the radiation of vertebrates the sequence and structure of insulin have been constrained not only by hormonal function, but also by the efficiency of proinsulin's conformational search leading to the native state (71). We envisage that toxic misfolding has limited genetic variation. Given that continuous, high-level biosynthesis of insulin is required in vertebrates for metabolic homeostasis, it is remarkable that the hormone's biosynthetic pathway operates at the seeming edge of toxic misfolding (72). Such nonrobustness has medical implications. Although clinical mutations in the insulin gene are uncommon, in nonsyndromic T2D chronic ER stress—induced by compensatory overexpression of the hormone in the face of peripheral resistance—may contribute to progressive β-cell dysfunction (73, 74). Considered in the cultural context of "diabesity," biophysical constraints to proinsulin's foldability highlight themes of evolutionary medicine in a pandemic disease of civilization (75).

Materials and Methods

Further information regarding experimental procedures are detailed in *SI Appendix*.

Materials. Human insulin was purchased from BioDel. Bovine insulin was purchased from Sigma Aldrich. Reagents for peptide synthesis were as described (76).

Preparation of Insulin Analogs. Variant insulins were prepared by trypsin-catalyzed semisynthesis or solid-phase peptide synthesis and purified by rp-HPLC (*SI Appendix*, Fig. S22); predicted molecular masses were confirmed by mass spectrometry (76).

Cell-Based Assays of Folding and Secretion. HEK-293T cells were transfected with constructs expressing WT proinsulin or variants. At 48-h post-transfection, the cells were labeled with ^{35}S-Met/Cys followed by a 0- or 90-min chase as indicated. Samples were immunoprecipitated with antiinsulin antibodies (35) and analyzed using Tris-tricine-urea SDS/PAGE under nonreducing conditions versus reducing conditions (35, 52). The secretion of proinsulin was analyzed using 4 to 12% NuPAGE (Invitrogen) using reducing condition.

Cell-Based Assessment of *trans*-Dominance. HEK293 cells were plated as above and cotransfected with plasmid encoding untagged human proinsulin-WT and plasmids encoding Myc-tagged human proinsulin-WT or variants as previously described (52). Cells were pulse-labeled as above followed by 0- or 90-min chase. Cell lysates and chase media were immunoprecipitated with anti-insulin and analyzed using 4 to 12% NuPAGE (Invitrogen) under reducing conditions (30).

BiP-Driven Luciferase Assay. Min6 cells were plated into 24-well plates 1 d before transfection. Cells were cotransfected with BiP-firefly-luciferase reporter plasmid, CMV-*Renilla*-luciferase plasmid (Promega), and human WT or mutant proinsulin at a DNA ratio of 1:2:5, respectively. At 48-h post-transfection, cell extracts were prepared for the dual-luciferase reporter assay (Promega) with BiP-luciferase normalized to *Renilla* luciferase activity (35).

Immunofluorescence Studies of Proinsulin Biosynthesis. Rat pancreatic β-cell line INS1 were transiently transfected with human proinsulin WT or variants. Cells were fixed and then stained with primary anti-human proinsulin monoclonal antibody and rabbit anticalnexin. Thereafter, sections were rinsed and incubated with secondary antibodies conjugated to Alexa Fluor 488 or 568 (Invitrogen). Cells were imaged by epifluorescence in an Olympus FV500 confocal microscope with a 60× objective.

Receptor-Binding Assays. Affinities for detergent-solubilized IR-B, IR-A, or IGF-1R holoreceptor were measured by a competitive-displacement assay (16). To obtain dissociation constants, data were analyzed by nonlinear regression (77).

Cell Signaling Assays. For Western blot assays, MCF-7 cells (ATCC) were cultured to 70 to 75% confluence, serum-starved for 24 h, and then exposed to insulin analogs at doses 5, 10, 20, and 50 nM. Blotting protocols were as described previously (41). The plate-based in-cell fluorescence-based immunoblotting pIR assay employed HepG2 cells (ATCC; ~8,000 cells per well). Fluorescence signals were monitored by a LI-COR Infrared Imaging system (Odyssey). Insulin doses were in the range 0.5 to 900 nM. Western blot and plate-based assays were performed in triplicate.

Rat Studies. Effects of insulin analogs formulated in Lilly buffer (16) on blood-glucose concentration following intravenous injection in streptozotocin-rendered diabetic rats were assessed in relation to Orn^{B29}-insulin (16). Rats were anesthetized with isoflurane, and insulin or analog was administered via tail-vein injection.

CD Spectroscopy. Spectra were acquired in 10 mM potassium phosphate (pH 7.4) and 50 mM KCl (15). Free energies of unfolding (ΔG_u) were inferred at 25 °C from two-state modeling (15).

NMR Spectroscopy. NMR spectra were acquired at 700 MHz at pH or pD 7.4 (direct meter reading) at 25 °C (41); these included through-bond correlation experiments (78), 3D ^{13}C-NOESY (79, 80), and 4D time-shared NOESY (80). Main-chain dihedral angle restraints were generated using *TALOS*$^+$. Structures were calculated using *X-PLOR-NIH* (81) as described previously (41). Ensembles were visualized using *insightII* and *molmol* software (82). ^{1}H-^{2}H exchange was performed in 100% D_2O with 10 mM deuteio-acetic acid (pD 3.0). Protection factors were calculated from the ratio of the measured-to-intrinsic exchange rates (41).

Size-Exclusion Chromatography. Analogs were made 0.6 mM in 10 mM Tris·HCl (pH 7.4) (76). Samples (20 μL) were loaded on an Enrich SEC70 column (10 mm × 300 mm with fractionation range 3 to 70 kDa). Elution times were monitored by absorbance at 280 nm. Molecular masses and void volume (V_0) were inferred in reference to standard proteins (16).

Hexamer Disassembly Assays. Disassembly of phenol-stabilized Co^{2+}-substituted R_6 insulin hexamers was monitored as described previously (16) using absorbance at 574 nm (83); Co^{2+} sequestration was initiated by addition of EDTA. Data were fit to a monoexponential decay equation (84).

Fibrillation Assays. Proteins were made 60 μM in PBS with 0.02% sodium azide and 16 μM thioflavin T (ThT). Samples were incubated in a 96-well plate at 37 °C with continuous shaking at 400 RPM in a Biotech plate reader. ThT fluorescence at 480 nm (excitation 450 nm) was assessed at 5-min intervals. Time of initial ThT fluorescence defined lag time.

Data Availability. The NMR chemical shifts and molecular coordinates of Tyr^{B24}-DKP-insulin have been deposited in the Protein Data Bank (PDB ID code 6X4X) (58).

ACKNOWLEDGMENTS. We thank Y. Liu and L. Broadwater (Thermalin, Inc.) for generously providing ^{13}C, ^{15}N-labeled Asp^{B10}-DOI; Q.-X. Hua for advice concerning NMR methods; L. Whittaker and J. Whittaker for advice regarding receptor-binding studies; K. Carr, R. Grabowski, P. Macklis, and M. Swain for assistance with rat studies; M. C. Lawrence and V. Pandyarajan for helpful discussion; P. DeMeyts (Novo Nordisk) for a gift of radiolabeled insulin and IGF-1; 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 C. M. Beall (Case Western Reserve University), D. P. Goldenberg (University of Utah), M. Karplus (Harvard University), J. A. King (Massachusetts Institute of Technology), P. Princehouse (Case Western Reserve University), D. Ron (Cambridge, United Kingdom), T. Sosnick (University of Chicago), and the late D. F. Steiner (University of Chicago) for helpful discussion in the early stages of this work. M.A.W. is a Fellow of the Origins Institute at Case Western Reserve University. This work was supported in part by the NIH Grant R01 DK040949 (to M.A.W.). Early stages of this work were also supported by NIH Grants R01 DK069764 (to M.A.W. and P.A.) and R01 DK48280 (to P.A.). N.K.R. was a predoctoral Fellow of the NIH Medical Scientist Training Program, supported by 5T32GM007250-38 (institutional) and Fellowship 1F30DK112644. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. N.B.P. was supported in part by American Diabetes Association Grants 7-13-IN-31 and 1-08-RA-149.

1. G. Dodson, D. Steiner, The role of assembly in insulin's biosynthesis. *Curr. Opin. Struct. Biol.* **8**, 189–194 (1998).
2. A. N. Zaykov, J. P. Mayer, R. D. DiMarchi, Pursuit of a perfect insulin. *Nat. Rev. Drug Discov.* **15**, 425–439 (2016).
3. E. N. Baker *et al*., The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **319**, 369–456 (1988).
4. D. F. Steiner *et al*., Proinsulin and the biosynthesis of insulin. *Recent Prog. Horm. Res.* **25**, 207–282 (1969).
5. Y. Yang *et al*., Solution structure of proinsulin: Connecting domain flexibility and prohormone processing. *J. Biol. Chem.* **285**, 7847–7851 (2010).
6. M. A. Weiss *et al*., Hierarchiacal protein "un-design": Insulin's intrachain disulfide bridge tethers a recognition α-helix. *Biochemistry* **39**, 15429–15440 (2000).
7. Q. X. Hua *et al*., Mechanism of insulin chain combination. Asymmetric roles of A-chain α-helices in disulfide pairing. *J. Biol. Chem.* **277**, 43443–43453 (2002).
8. J. Varghese, S. F. Lim, S. M. Cohen, Drosophila miR-14 regulates insulin production and metabolism through its target, sugarbabe. *Genes Dev.* **24**, 2748–2753 (2010).
9. Q. X. Hua *et al*., A divergent INS protein in *Caenorhabditis elegans* structurally resembles human insulin and activates the human insulin receptor. *Genes Dev.* **17**, 826–831 (2003).
10. Z. Y. Guo, Z. S. Qiao, Y. M. Feng, The in vitro oxidative folding of the insulin superfamily. *Antioxid. Redox Signal.* **10**, 127–139 (2008).
11. S. Blumenthal, From insulin and insulin-like activity to the insulin superfamily of growth-promoting peptides: A 20th-century odyssey. *Perspect. Biol. Med.* **53**, 491–508 (2010).
12. J. M. Conlon, Evolution of the insulin molecule: Insights into structure-activity and phylogenetic relationships. *Peptides* **22**, 1183–1193 (2001).
13. J. M. Conlon, Molecular evolution of insulin in non-mammalian vertebrates. *Am. Zool.* **40**, 200–212 (2000).
14. R. G. Mirmira, S. H. Nakagawa, H. S. Tager, Importance of the character and configuration of residues B24, B25, and B26 in insulin-receptor interactions. *J. Biol. Chem.* **266**, 1428–1436 (1991).
15. V. Pandyarajan *et al*., Aromatic anchor at an invariant hormone-receptor interface function of insulin residue B24 with application to protein design. *J. Biol. Chem.* **289**, 34709–34727 (2014).
16. V. Pandyarajan *et al*., Contribution of Tyr^{B26} to the function and stability of insulin. Structure-activity relationships at a conserved hormone-receptor interface. *J. Biol. Chem.* **291**, 12978–12990 (2016).
17. R. G. Mirmira, H. S. Tager, Role of the phenylalanine B24 side chain in directing insulin interaction with its receptor: Importance of main chain conformation. *J. Biol. Chem.* **264**, 6349–6354 (1989).
18. S. E. Shoelson, K. S. Polonsky, A. Zeidler, A. H. Rubenstein, H. S. Tager, 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 (1984).
19. J. Stoy *et al*., Clinical and molecular genetics of neonatal diabetes due to mutations in the insulin gene. *Rev. Endocr. Metab. Disord.* **11**, 205–215 (2010).
20. M. Liu *et al*., Mutant *INS*-gene induced diabetes of youth: Proinsulin cysteine residues impose dominant-negative inhibition on non-mutant proinsulin transport. *PLoS One* **5**, e13333 (2010).
21. Z. S. Qiao, C. Y. Min, Q. X. Hua, M. A. Weiss, Y. M. Feng, In vitro refolding of human proinsulin. Kinetic intermediates, putative disulfide-forming pathway, folding initiation site, and protential role of C-peptide in folding process. *J. Biol. Chem.* **278**, 17800–17809 (2003).
22. L. Haataja *et al*., Disulfide mispairing during proinsulin folding in the endoplasmic reticulum. *Diabetes* **65**, 1050–1060 (2016).
23. M. A. Weiss, Proinsulin and the genetics of diabetes mellitus. *J. Biol. Chem.* **284**, 19159–19163 (2009).
24. M. Liu *et al*., Proinsulin misfolding and diabetes: Mutant *INS* gene-induced diabetes of youth. *Trends Endocrinol. Metab.* **21**, 652–659 (2010).
25. M. A. Weiss, Diabetes mellitus due to the toxic misfolding of proinsulin variants. *FEBS Lett.* **587**, 1942–1950 (2013).
26. M. Liu *et al*., INS-gene mutations: From genetics and beta-cell biology to clinical disease. *Mol. Aspects Med.* **42**, 3–18 (2015).
27. T. Izumi *et al*., Dominant negative pathogenesis by mutant proinsulin in the Akita diabetic mouse. *Diabetes* **52**, 409–416 (2003).
28. N. Herbach *et al*., Dominant-negative effects of a novel mutated *Ins2* allele causes early-onset diabetes and severe β-cell loss in Munich $Ins2^{C95S}$ mutant mice. *Diabetes* **56**, 1268–1276 (2007).
29. H. W. Davidson, (Pro)Insulin processing: A historical perspective. *Cell Biochem. Biophys.* **40**, 143–158 (2004).
30. M. Liu, Y. Li, D. Cavener, P. Arvan, Proinsulin disulfide maturation and misfolding in the endoplasmic reticulum. *J. Biol. Chem.* **280**, 13209–13212 (2005).
31. J. Stoy *et al*., Insulin gene mutations as a cause of permanent neonatal diabetes. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15040–15044 (2007).
32. C. Colombo *et al*., Seven mutations in the human insulin gene linked to permanent neonatal/infancy-onset diabetes mellitus. *J. Clin. Invest.* **118**, 2148–2156 (2008).
33. M. Liu, I. Hodish, C. J. Rhodes, P. Arvan, Proinsulin maturation, misfolding, and proteotoxicity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15841–15846 (2007).
34. Q. X. Hua *et al*., Mapping the functional surface of insulin by design: Structure and function of a novel A-chain analogue. *J. Mol. Biol.* **264**, 390–403 (1996).
35. N. K. Rege *et al*., "Register-shift" insulin analogs uncover constraints of proteotoxicity in protein evolution. *J. Biol. Chem.* **295**, 3080–3096 (2020).
36. U. Ribel, P. Hougaard, K. Drejer, A. R. Sorensen, Equivalent in vivo biological activity of insulin analogues and human insulin despite different in vitro potencies. *Diabetes* **39**, 1033–1039 (1990).
37. D. N. Brems *et al*., Altering the association properties of insulin by amino acid replacement. *Protein Eng.* **5**, 527–533 (1992).
38. R. K. Assoian, N. E. Thomas, E. T. Kaiser, H. Tager, [Leu^{B24}]insulin and [Ala^{B24}]insulin: Altered structures and cellular processing of B24-substituted insulin analogs. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5147–5151 (1982).
39. R. Horuk *et al*., A monomeric insulin from the porcupine (*Hystrix cristata*), an Old World hystricomorph. *Nature* **286**, 822–824 (1980).
40. S. P. Wood, T. L. Blundell, A. Wollmer, N. R. Lazarus, R. W. Neville, The relation of conformation and association of insulin to receptor binding; X-ray and circular-dichroism studies on bovine and hystricomorph insulins. *Eur. J. Biochem.* **55**, 531–542 (1975).
41. M. D. Glidden *et al*., Solution structure of an ultra-stable single-chain insulin analog connects protein dynamics to a novel mechanism of receptor binding. *J. Biol. Chem.* **293**, 69–88 (2018).
42. Q. X. Hua, W. Jia, M. A. Weiss, Conformational dynamics of insulin. *Front. Endocrinol. (Lausanne)* **2**, 48 (2011).
43. J. G. Menting *et al*., Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3395–E3404 (2014).
44. Q. X. Hua *et al*., 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 (1996).
45. M. A. Weiss *et al*., 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 (2002).
46. Q. X. Hua *et al*., Hierarchical protein folding: Asymmetric unfolding of an insulin analogue lacking the A7-B7 interchain disulfide bridge. *Biochemistry* **40**, 12299–12311 (2001).
47. G. Rajpal, I. Schuiki, M. Liu, A. Volchuk, P. Arvan, Action of protein disulfide isomerase on proinsulin exit from endoplasmic reticulum of pancreatic β-cells. *J. Biol. Chem.* **287**, 43–47 (2012).
48. S. H. Nakagawa *et al*., Chiral mutagenesis of insulin. Foldability and function are inversely regulated by a stereospecific switch in the B chain. *Biochemistry* **44**, 4984–4999 (2005).
49. Y. Sohma *et al*., Comparative properties of insulin-like growth factor 1 (IGF-1) and [Gly7D-Ala]IGF-1 prepared by total chemical synthesis. *Angew. Chem. Int. Ed. Engl.* **47**, 1102–1106 (2008).
50. J. Stoy *et al*., In vivo measurement and biological characterisation of the diabetes-associated mutant insulin p.R46Q (GlnB22-insulin). *Diabetologia* **60**, 1423–1431 (2017).
51. S. U. Gorr, X. F. Huang, D. J. Cowley, R. Kuliawat, P. Arvan, Disruption of disulfide bonds exhibits differential effects on trafficking of regulated secretory proteins. *Am. J. Physiol.* **277**, C121–C131 (1999).
52. J. Sun *et al*., Role of proinsulin self-association in mutant *INS* gene-induced diabetes of youth. *Diabetes* **69**, 954–964 (2020).

53. J. Wright *et al*., Endoplasmic reticulum oxidoreductin-1α (Ero1α) improves folding and secretion of mutant proinsulin and limits mutant proinsulin-induced endoplasmic reticulum stress. *J. Biol. Chem.* **288**, 31010–31018 (2013).
54. J. Fang *et al*., COPII-dependent ER export: A critical component of insulin biogenesis and β-cell ER homeostasis. *Mol. Endocrinol.* **29**, 1156–1169 (2015).
55. K. L. Lipson *et al*., Regulation of insulin biosynthesis in pancreatic beta-cells by an endoplasmic reticulum-resident protein kinase IRE1. *Cell Metab.* **4**, 245–254 (2006).
56. L. O. Narhi *et al*., Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: Genetic models of protein-folding intermediates. *Biochemistry* **32**, 5214–5221 (1993).
57. B. A. Magee, G. K. Shooter, J. C. Wallace, G. L. Francis, Insulin-like growth factor I and its binding proteins: A study of the binding interface using B-domain analogues. *Biochemistry* **38**, 15863–15870 (1999).
58. Y. Yang, M. A. Weiss, B24Y DKP insulin. RCSB PDB (*Protein Data Bank*). https://www.rcsb.org/structure/6X4X. Deposited 24 May 2020.
59. J. N. Onuchic, P. G. Wolynes, Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
60. Q. X. Hua, W. Jia, B. H. Frank, N. B. Phillips, M. A. Weiss, A protein caught in a kinetic trap: Structures and stabilities of insulin disulfide isomers. *Biochemistry* **41**, 14700–14715 (2002).
61. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
62. K. A. Dill, H. S. Chan, From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
63. Y. Ikeda, T. Suehiro, T. Nakamura, Y. Kumon, K. Hashimoto, Clinical significance of the insulin resistance index as assessed by homeostasis model assessment. *Endocr. J.* **48**, 81–86 (2001).
64. M. Nishi, K. Nanjo, Insulin gene mutations and diabetes. *J. Diabetes Investig.* **2**, 92–100 (2011).
65. J. Sun *et al*., Proinsulin misfolding and endoplasmic reticulum stress during the development and progression of diabetes. *Mol. Aspects Med.* **42**, 105–118 (2015).
66. R. N. Bone *et al*., A computational approach for defining a signature of β-cell Golgi stress in diabetes mellitus. *Diabetes*, 10.2337/db20-0636 (2020).
67. I. Hodish *et al*., Misfolded proinsulin affects bystander proinsulin in neonatal diabetes. *J. Biol. Chem.* **285**, 685–694 (2010).
68. S. Betts, J. King, There's a right way and a wrong way: In vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. *Structure* **7**, R131–R139 (1999).
69. M. Liu *et al*., Crystal structure of a "nonfoldable" insulin impaired folding efficiency despite native activity. *J. Biol. Chem.* **284**, 35259–35272 (2009).
70. M. Avital-Shmilovici, J. Whittaker, M. A. Weiss, S. B. Kent, 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 (2014).
71. O. Ashenberg, L. I. Gong, J. D. Bloom, Mutational effects on stability are largely conserved during protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 21071–21076 (2013).
72. M. Liu *et al*., Deciphering the hidden informational content of protein sequences: Foldability of proinsulin hinges on a flexible arm that is dispensable in the mature hormone. *J. Biol. Chem.* **285**, 30989–31001 (2010).
73. J. A. Rodríguez-Pérez, U. M. Marigorta, A. Navarro, Integrating genomics into evolutionary medicine. *Curr. Opin. Genet. Dev.* **29**, 97–102 (2014).
74. L. Leroux *et al*., Compensatory responses in mice carrying a null mutation for Ins1 or Ins2. *Diabetes* **50**, S150 (2001).
75. M. Watve, M. Diwekar-Joshi, What to expect from an evolutionary hypothesis for a human disease: The case of type 2 diabetes. *Homo* **67**, 349–368 (2016).
76. V. Pandyarajan *et al*., Biophysical optimization of a therapeutic protein by non-standard mutagenesis. Studies of an iodo-insulin derivative. *J. Biol. Chem.* **289**, 23367–23381 (2014).
77. Z. X. Wang, An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.* **360**, 111–114 (1995).
78. A. Bax, Multidimensional nuclear magnetic resonance methods for protein studies. *Curr. Opin. Struct. Biol.* **4**, 738–744 (1994).
79. L. E. Kay, NMR methods for the study of protein structure and dynamics. *Biochem. Cell Biol.* **75**, 1–15 (1997).
80. Y. Xu, D. Long, D. Yang, Rapid data collection for protein structure determination by NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 7722–7723 (2007).
81. C. D. Schwieters, J. J. Kuszewski, G. M. Clore, Using Xplor-NIH for NMR molecular structure determination. *Prog. Nuc. Mag. Res. Spec.* **48**, 47–62 (2006).
82. R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55, 29–32 (1996).
83. M. Roy *et al*., Spectroscopic signatures of the T to R conformational transition in the insulin hexamer. *J. Biol. Chem.* **264**, 19081–19085 (1989).
84. D. T. Birnbaum, M. A. Kilcomons, M. R. DeFelippis, J. M. Beals, Assembly and dissociation of human insulin and $Lys^{B28}Pro^{B29}$-insulin hexamers: A comparison study. *Pharm. Res.* **14**, 25–36 (1997).