Toxic Misfolding Constrains Protein Evolution

Nischay K. Rege, Ming Liu, Balamurugan Dhayalan, Yen-Shan Chen, Nicholas A. Smith, Leili Rahimi, Jinhong Sun, Huan Guo, Yanwu Yang, Leena Haataja, Nelson F. B. Phillips, Jonathan Whittaker, Brian J. Smith, Peter Arvan, Faramarz Ismail-Beigi, & Michael A. Weiss

Purpose of Supplement

Seventeen Supplemental Figures and two Supplemental Tables are provided. These provide additional data and details concerning SARs, stabilities and biosynthesis of insulin analogs.

Table of Contents

Figure S1. Alignment of vertebrate insulin sequences……..…….....3

Figure S2. Ribbon model of an insulin dimer……...4

Figure S3. B24-B26 binding surfaces in insulin-μIR complex ……..……...…….....5

Figure S4. Biological activities of Gly^{B24}- and Ser^{B24}-insulin ……....……...……..6

Figure S5. Experimental design and register-shift paradigm……...…….…….……..7

Figure S6. Schematic of assays for the assessment of hormone-induced IR signaling…….…….....8

Figure S7. Assessment of insulin-driven transcription activation in mammalian cell lines…….....9

Figure S8. CD spectra and thermodynamic studies ……..…..…….…….....10

Figure S9. Calculation of molecular weight using size-exclusion chromatography……..…….....11

Figure S10. Zinc coordination within the R_6 insulin hexamer……..……...12

Figure S11. Dot plot of fibrillation lag times…….……...…..13

Figure S12. Pathway of insulin biosynthesis……...……..…….14

Figure S13. Folding and secretion of Ser^{B24}-proinsulin.....…...……..……..15

Figure S14. Proinsulin- and ER-localizing constructs.….....…….…..……..16

Figure S15. Intracellular distribution of a neonatal-onset MIDY variant ……....……..17

Figure S16. Cartoon representation of BiP activation in response to unfolded protein…….....18

Figure S17. MD simulation of B24 cavity size in Gly^{B24}-μIR complex……..……19

Table S1. Thermodynamic fitting parameters..……...……..…..……...20

Table S2. Fibrillation lag times..……..…..……...……..…….21
Supplemental References....……...22

mammalian

	15	.	.	.	.	20	.	.	.	.	25	.	.	.	.	30
Human	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	T
Dog	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	A
Cat	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	A
Pig	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	A
Rat	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	M	S
Cattle	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	A
Sheep	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	A
Beluga Whale	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	A
Porcupine	L	Y	L	V	C	G	N	D	G	F	F	Y	R	P	K	A
Guinea Pig	L	Y	S	V	C	Q	D	D	G	F	F	Y	I	P	K	D
Octadon degus	L	Y	M	T	C	G	R	S	G	F	Y	R	-	P	H	D
Microcavia niata	L	Y	E	V	C	R	D	K	G	F	F	S	R	P	K	-

non-mammalian

	15	.	.	.	.	20	.	.	.	.	25	.	.	.	.	30
Human	L	Y	L	V	C	G	E	R	G	F	F	Y	T	P	K	T
Bullfrog	L	Y	L	V	C	G	D	R	G	F	F	Y	S	P	R	S
Salamander	L	R	V	V	C	G	E	R	G	F	H	S	-	P	T	S
Rattlesnake	L	F	L	I	C	G	E	R	G	F	Y	Y	S	P	R	S
Sea Turtle	L	Y	L	V	C	G	E	R	G	F	F	Y	S	P	K	A
Alligator	L	Y	L	V	C	G	E	R	G	F	F	Y	S	P	K	G
Gecko	L	I	L	V	C	G	D	R	G	F	Y	W	G	P	K	E
Skink	L	Y	L	V	C	G	E	R	G	F	Y	Y	S	P	K	A
Cobra	L	F	L	I	C	G	E	R	G	F	Y	Y	S	P	R	A
Turkey Vulture	L	Y	L	V	C	G	E	R	G	F	F	Y	S	P	K	A
Chicken	L	Y	L	V	C	G	E	R	G	F	F	Y	S	P	K	A
Finch	L	Y	L	V	C	G	E	R	G	F	F	Y	Q	P	K	A
Hummingbird	L	Y	L	V	C	G	E	R	G	F	F	Y	S	P	K	A
Wild Turkey	L	Y	L	V	C	G	E	R	G	F	F	Y	S	P	K	A
Zebrafish	L	Y	L	V	C	G	P	T	G	F	F	Y	N	P	K	R
Shark	L	Y	F	V	C	G	E	R	G	F	F	Y	S	P	K	Q
Lamprey	L	Y	V	V	C	G	D	R	G	F	F	Y	T	P	S	K
Hagfish	L	Y	I	A	C	G	V	R	G	F	F	Y	D	P	T	K

Figure S1. Alignment of vertebrate insulin sequences. Human insulin and mammalian insulin sequences are shown in the *upper* panel. *Octadon degus* and *Microcavia niata* species are known to have aberrant insulin sequences. Non-mammalian sequences are shown in the *lower* panel. *Black boxes* indicate residues that are conserved among all known insulin sequences; *dashed gray boxes* indicate residues that are conserved in >90% of known insulin sequences. Residue Phe^{B24} (conserved among all validated vertebrate insulin sequences) is highlighted within the *red box*.

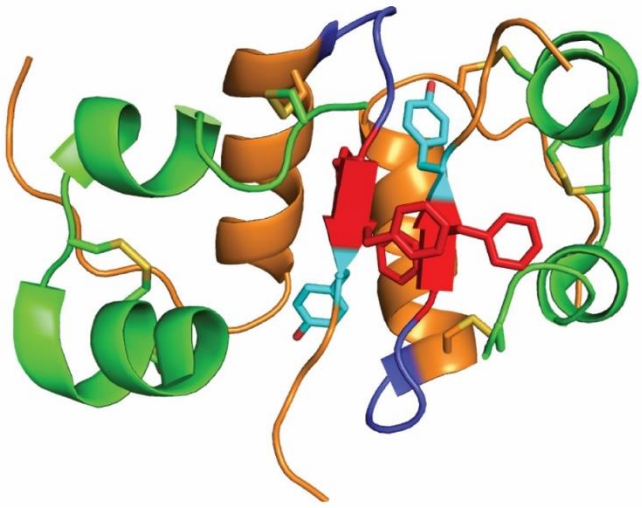

Figure S2. Ribbon model of an insulin dimer. A constituent dimer of the T_6 insulin hexamer (PDB: 4INS). The A chain is shown as *green ribbon*; residues B1 to B19 are shown as *orange ribbon*. The B20-B23 β-turn is shown as *blue coil*. Phe^{B24} and Phe^{B25} are shown as *red sheet* (main chain) and *red sticks* (side chains); Tyr^{B26} side chains are shown as *cyan sticks*. Residues B27-B30 are shown as *orange line* (main chain).

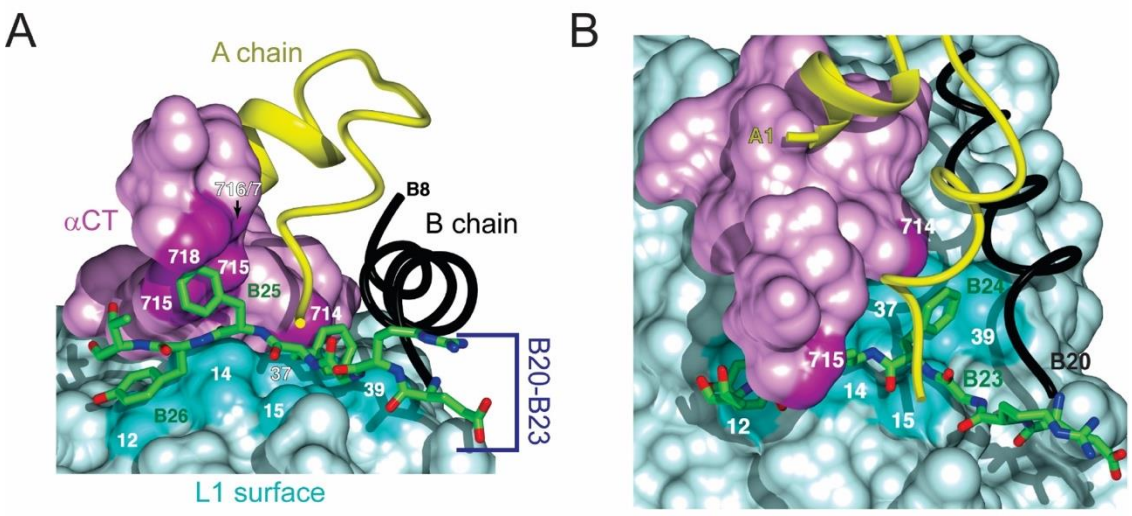

Figure S3. B24-B26 binding surfaces in insulin-μIR complex. The L1 domain is represented as a *cyan surface*, and the αCT peptide as a *purple* surface. Two orientations are shown in panels *A* and *B*. IR residue numbers are labeled in *white*, insulin residues B24-B26 in *green*, and B20-B23 bracketed in *blue*. Coordinates were obtained from PDB entry 4OGA.

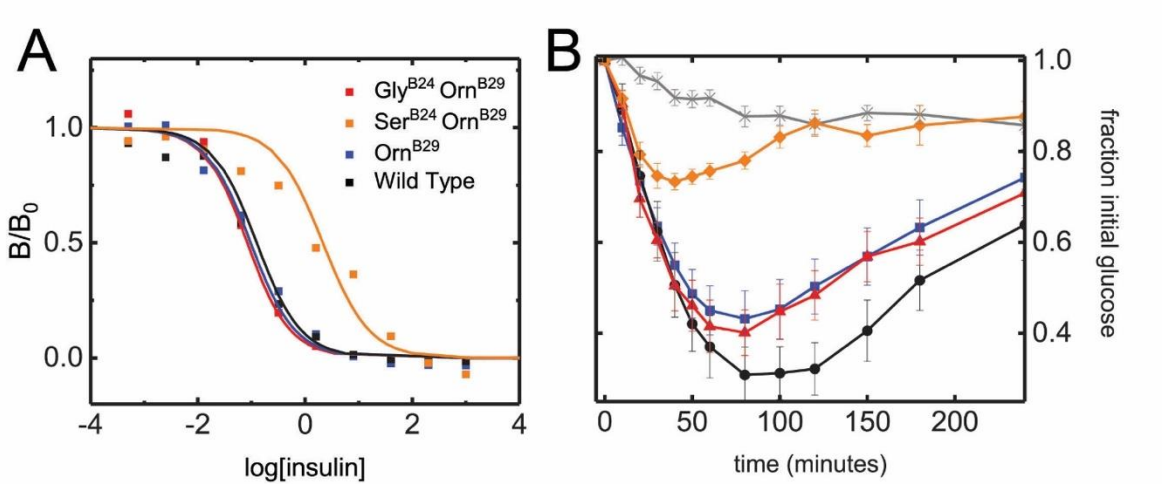

Figure S4. Biological activities of Gly^{B24}- and Ser^{B24}-insulin. (*A*) *In vitro* affinities of [Gly^{B24}, Orn^{B29}]-insulin, [Ser^{B24}, Orn^{B29}]-insulin, and control analogs for detergent-solubilized IR-A were determined by competitive displacement of ^{125}I-Tyr^{A14}-insulin. Colored squares represent % tracer bound at increasing concentrations of analog; corresponding colored curves are fitted (see Table 1 in main text). Colors are defined within inset. (*B*) Time-course of fraction of initial blood-glucose concentration (mean glycemia *ca*. 400 mg/dL) following intravenous (IV) injection of 10-μg doses of insulin analogs in STZ rats. Color code is as in panel *A*; control data obtained following IV injection of insulin-free diluent (vehicle) is shown as a *gray line* with *gray 'x.'*

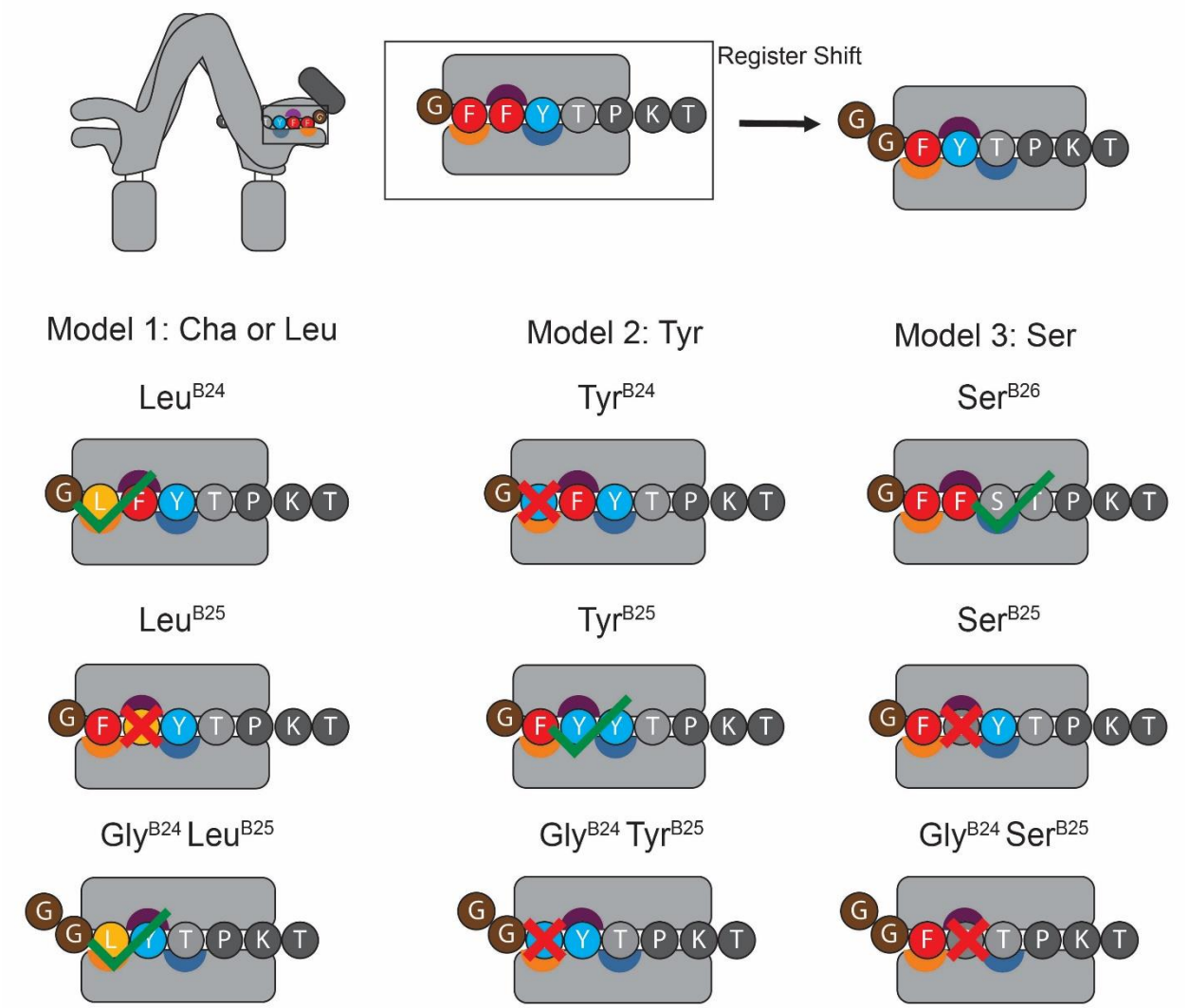

Figure S5. Experimental design and register-shift paradigm. Residues Phe^{B24}, Phe^{B25} and Tyr^{B26} occupy distinctive binding surfaces or pockets at the hormone-receptor interface (denoted by *orange*, *purple*, and *blue-gray semicircles*); this interface comprises ectodomain elements αCT and L1 (inset). In the "register-shift" model extrusion of Gly^{B24} would enable Phe^{B25} to occupy the B24-binding pocket, Tyr^{B26} to pack against the B25-binding surface, and Thr^{B27} to bind the B26-binding surface (*black arrow* next to inset).

Critical predictions of the register-shift model were tested herein through a series of position-specific mutations. *Model 1*: an analog containing Gly^{B24} and Cha- or Leu^{B25} substitutions would be predicted to be active due to packing of Cha or Leu (denoted by the *yellow circle*) into the B24-binding pocket (*orange semicircle*) rather than onto the B25-binding surface (*purple semicircle*). Conversely, in *Model 2* an analog with Gly^{B24} and Tyr^{B25} would be predicted to exhibit reduced activity due to insertion of the Tyr into the B24-binding pocket. In *Model 3* a Gly^{B24} mutation in the context of Ser^{B26}-insulin would be predicted to attenuate activity.

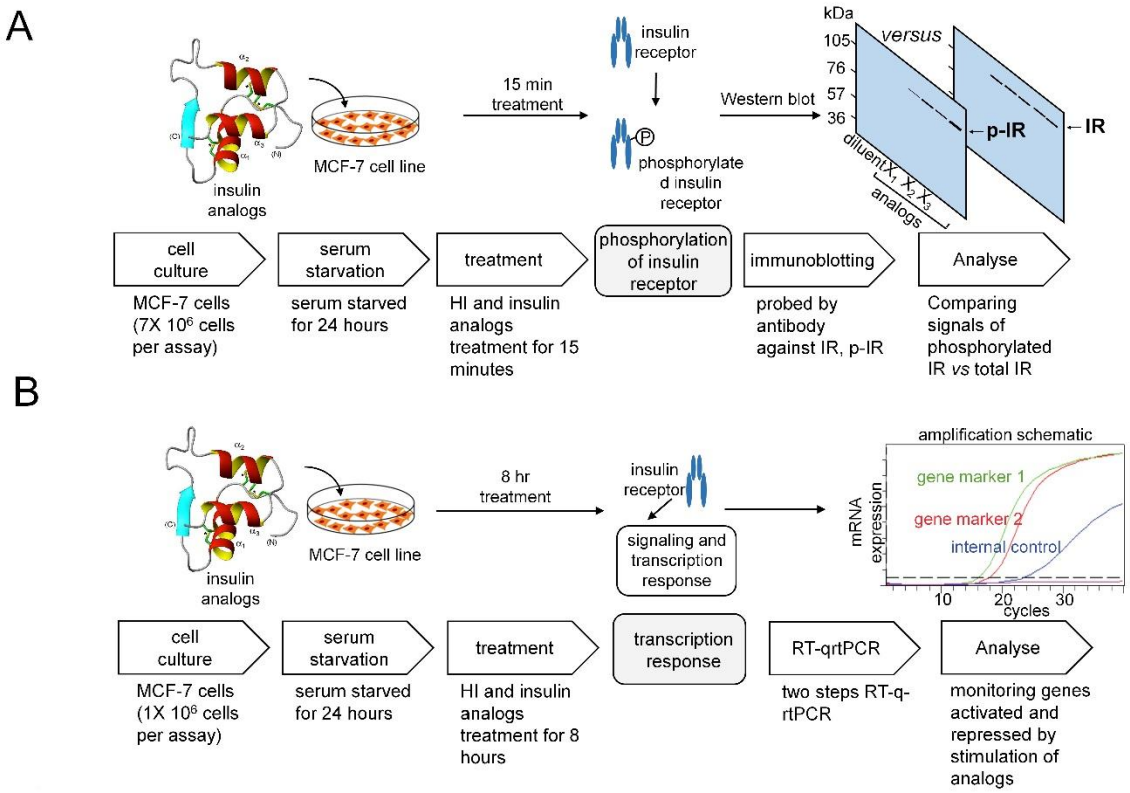

Figure S6. Systematic assessment of hormone-induced IR signaling. The present collection of insulin analogs were characterized with respect to (*A*) insulin-dependent IR autophosphorylation and (*B*) post-receptor transcriptional regulation (activation and repression). MCF-7 human breast-cancer cells expressing both IR isoforms (IR-A and IR-B) and high levels of IGF-1R (1,2) were treated with each of the present set of insulin analogs. After 15 min treatment Tyr-phosphorylation of IR was assessed by Western blotting (3) as illustrated in schematic form in panel (*A*). To analyze insulin-regulated transcriptional activities, cells were treated for 8 h, and the lysates collected. The transcriptional responses (activation or repression) were monitored as probed by the genes encoding cell-cycle factors *cyclin D1* and *cyclin G2* (panel B). Readouts were assessed by q-rtPCR (schematic adapted from: Bio-Rad Labs. [2017; http://www.bio-rad.com/en-us/applications-technologies/qpcr-assay-designoptimization]). Reciprocal responses—increased accumulation of cyclin D1 mRNA and decreased cyclin G2 mRNA—provided readouts for insulin-induced activation of cellular proliferation pathways.

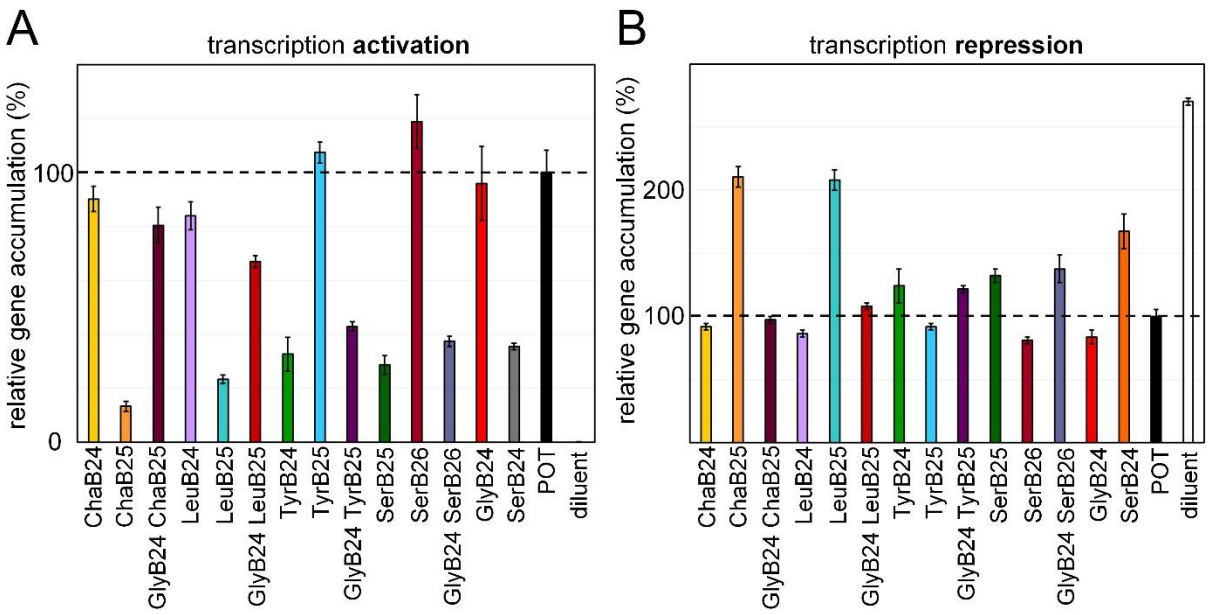

Figure S7. Assessment of insulin-driven transcription activation in mammalian cell lines. As illustrated in schematic form in Figure S6, q-rtPCR was used to assess the transcriptional activation of *cyclin D1* and transcriptional repression of *cyclin G2* on treatment with the present collection of insulin analogs: (*A*) relative changes in mRNA abundance of *cyclin D1* and (*B*) relative changes in mRNA abundance of *cyclin G2* are shown in relation to controls (Orn^{B29}-insulin and diluent).

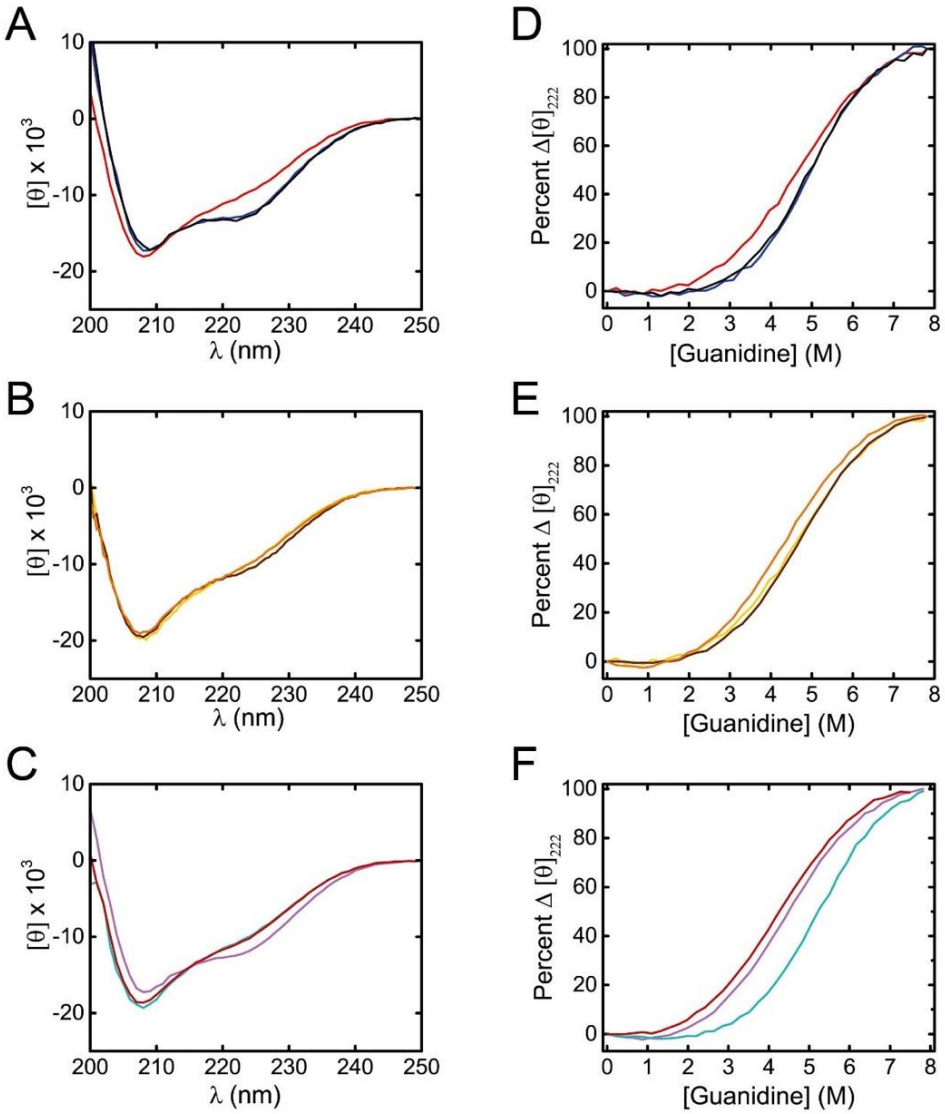

Figure S8. Circular dichroism (CD) spectra and thermodynamic studies of Gly^{B24}- and parent analogs. (*A*-*C*) Far-UV spectra of insulin analogs provided probes of protein secondary structure. (*D*-*F*) Fractional unfolding of analogs as a function of increasing concentration of guanidine-HCl (denaturant) was monitored by CD ellipticity at 222 nm. Data were fit to a two-state unfolding model (for inferred parameters ΔG_u, C_{mid} and *m*, see Table S1). Color code is as follows: (*A*, *D*) WT insulin, *black*; Orn^{B29}-insulin, *blue*; and [Gly^{B24}, Orn^{B29}]-insulin, *red*; data are also shown in main text Fig. 5*A* and 5*B*, respectively, and provided here for comparison with data shown in figure B, C, E, and F. (*B*, *E*) [Cha^{B24}, Orn^{B29}]-insulin, *yellow*; [Cha^{B25}, Orn^{B29}]-insulin, *orange*; and [Gly^{B24}, Cha^{B25}, Orn^{B29}]-insulin, *brown*. (*C*, *F*) [Leu^{B24}, Orn^{B29}]-insulin, *lavender*; [Leu^{B25}, Orn^{B29}]-insulin, *teal*; and [Gly^{B24}, Leu^{B25}, Orn^{B29}]-insulin, *red*.

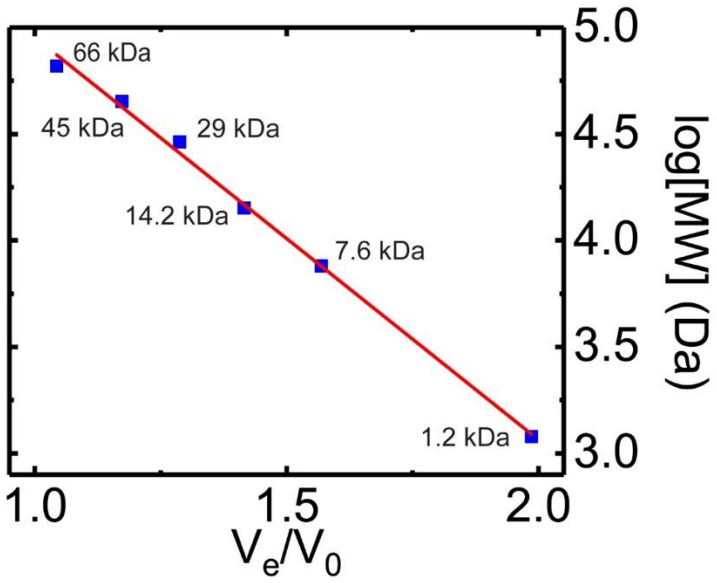

Figure S9. Calculation of molecular mass using size-exclusion chromatography (SEC). Molecular masses of analogs were inferred using a linearized plot of log molecular mass *versus* V_e/V_0 of standard proteins. V_e is the elution volume, and V_0 the column void volume (*black arrow* in Fig. 5C, main text). Estimated masses are given in Table 4. The following standards were used: thyroglobulin (669Da, V_0), BSA (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (29 kDa), lactalbumin (14.2 kDa), IGF-1 (7.6 kDa), and a synthetic peptide (1.2 kDa). Elution times of standards plotted against log[molecular weight] are represented as *blue squares*. Linear regression was used to determine the molar mass of insulin species is shown as a *red line* with the equation $-1.89[V_e/V_0]+6.84 = \log[MW]$; the coefficient of determination (R^2) was 0.998.

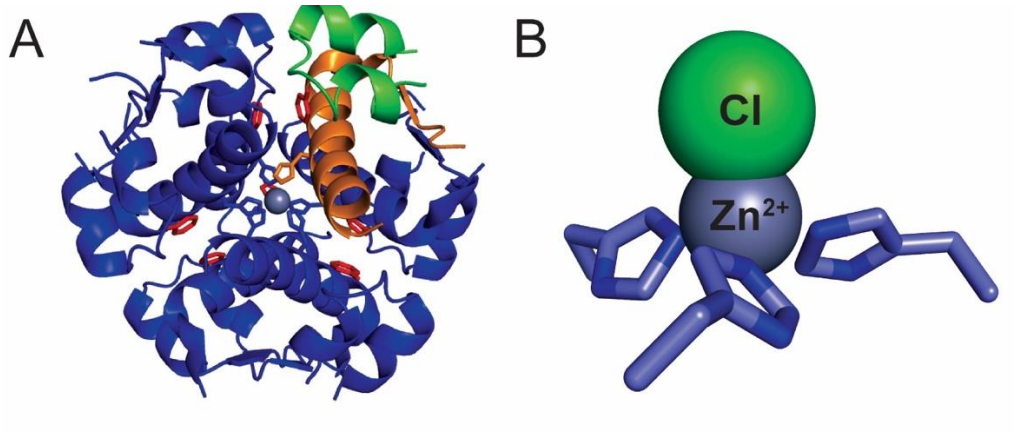

Figure S10. Zinc coordination within R_6 insulin hexamer. The zinc-coordinated, phenol-stabilized (R_6) insulin hexamer. (*A*) A single constituent insulin monomer is highlighted with the A- and B chains represented as *green* and *orange ribbons*, respectively. The remaining monomers are depicted as *blue ribbons*. Residue His^{B10} in each monomer is shown in a *stick* representation. Zinc ions are represented as *gray sphere* (overlying in this orientation). Phenol molecules, which bind to the insulin hexamer as allosteric ligands, are shown as *red sticks*. (*B*) Ball-and-stick model of one tetrahedral Zn^{2+}-coordination site within R_6 insulin hexamer. In each component trimer, three His^{B10} side chains (*blue sticks)* and one chloride ion (*green sphere*) coordinate a divalent zinc ion (*gray sphere*).

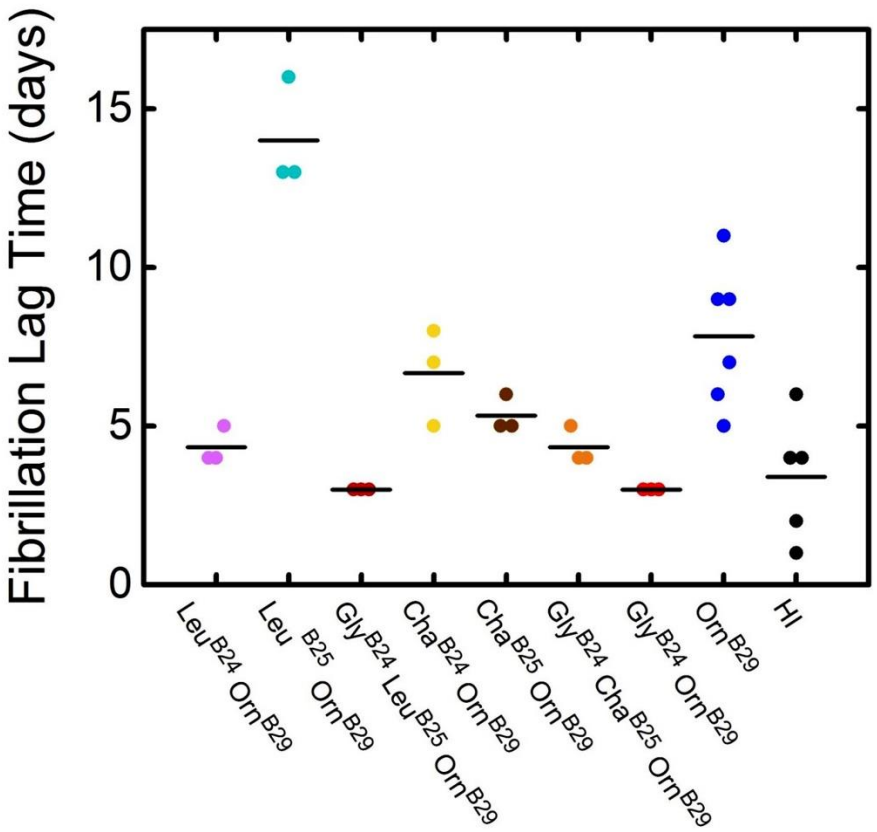

Figure S11. Fibrillation lag times of insulin analogs. A dot plot of lag times (in days) is shown. The lag time (defined prior to onset of fibrillation at room temperature) was characterized using a "slow sloshing" technique to mimic pharmaceutical formulations of insulin in a vial. Onset of fibrillation was defined as two-fold enhancement of ThT fluorescence. Amino-acid substitutions in analogs are given at bottom. Mean lag times are shown as black horizontal lines (see also Table S2). This assay was designed to better model real-world conditions for patients than the rapid-shaking accelerated assay (as illustrated in the main text) but requires more material and is slower. WT insulin is abbreviated "HI" at *bottom right*.

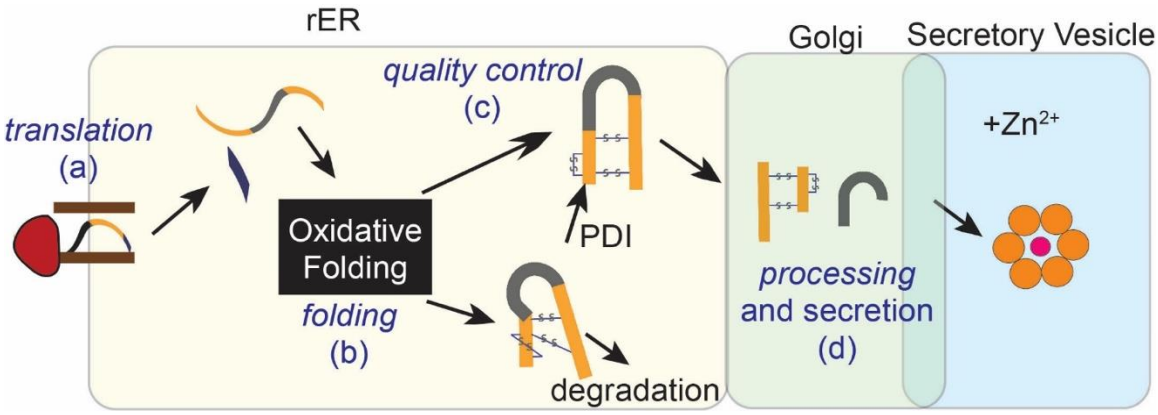

Figure S12. Schematic illustration of insulin biosynthesis and secretion. *Inset* symbols: (*a*) Preproinsulin, the primary ribosomal product, is co-translationally translocated into rough endoplasmic reticulum; the signal peptide (*blue strip*) is cleaved during this process. (*b*) The resulting proinsulin molecule undergoes oxidative folding to form three canonical disulfide bridges. This process is facilitated by ER chaperones and oxidoreductases. (*c*) Unfolded or misfolded insulin, in which cysteine residues remain unpaired or have formed non-canonical pairings (forming disulfide isomers), is isomerized into native proinsulin by protein disulfide isomerase (PDI) or degraded via ERAD. (*d*) Natively folded proinsulin is trafficked into the Golgi apparatus, proteolytically processed into two-chain insulin and packaged into secretory vesicles where it forms zinc-coordinated hexamers.

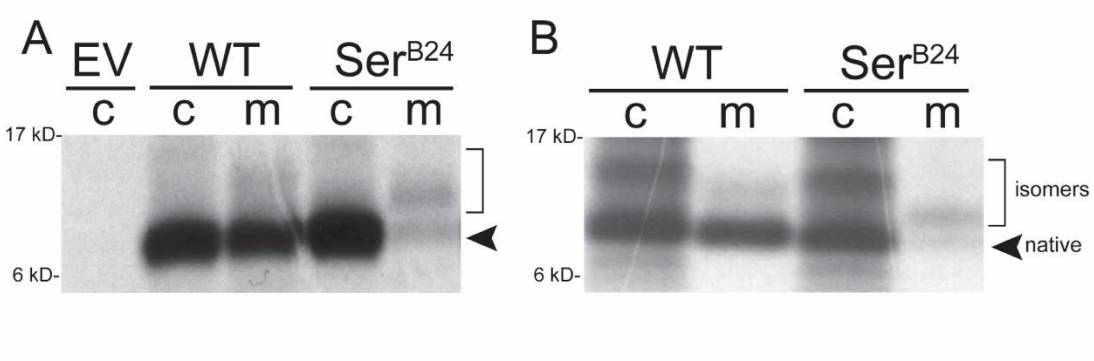

Figure S13. Aberrant folding and secretion of Ser^{B24}-proinsulin. (*A*) Non-reducing SDS-PAGE gel comparing folding and secretion of Ser^{B24}-proinsulin to that of WT proinsulin in HEK293T cells. A lane corresponding to a non-transfected control (EV) is also shown. Bands corresponding to native proinsulin and disulfide isomers are as labeled. Lanes representing samples collected from cell media are labeled "m", whereas those representing cell lysate samples are labeled "c." (*B*) A separate experiment comparing the folding and secretion of Ser^{B24}-proinsulin to a WT-insulin control yielded a pattern similar to that in panel *A*.

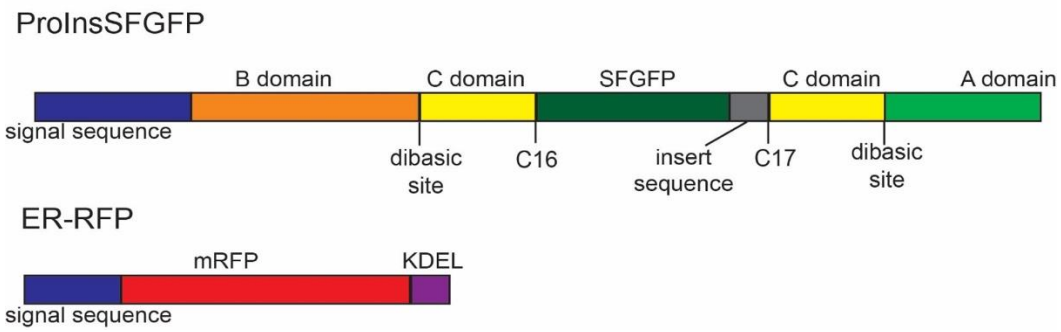

Figure S14. Domain structure of chimeric proinsulin- and ER-localizing constructs. Proinsulin was localized within cells using ProIns-GFP, a construct containing a fluorescent SFGFP domain, a super-folding green-fluorescent protein that is able to achieve fluorescence within the oxidizing environment of the ER. SFGFP was introduced within the proinsulin C domain (*yellow rectangle*) immediately downstream of residue C16 (*dark green*). A six amino-acid insert domain (sequence Ala-Ala-Ala-Ala-Gly-Pro; *gray*) separated the C-terminal residue of SFGFP from proinsulin residue C17. The remaining motifs of native human proinsulin (including the N-terminal signal sequence (*blue*), B domain (*orange*), C-terminal A domain (*light green*), and the dibasic proteolytic site separating the C domain from the A and B domains) remained intact. Mutations were introduced at the B24 position using site-directed mutagenesis. The ER-RFP construct was created by introducing the ER-localizing signal sequence from bovine prolactin (*blue*) to the N terminus of a monomeric red fluorescent protein (mRFP) domain (*red*). A KDEL ER-retention sequence (*violet*) was introduced to sequester the fluorescent chimeric protein within the ER.

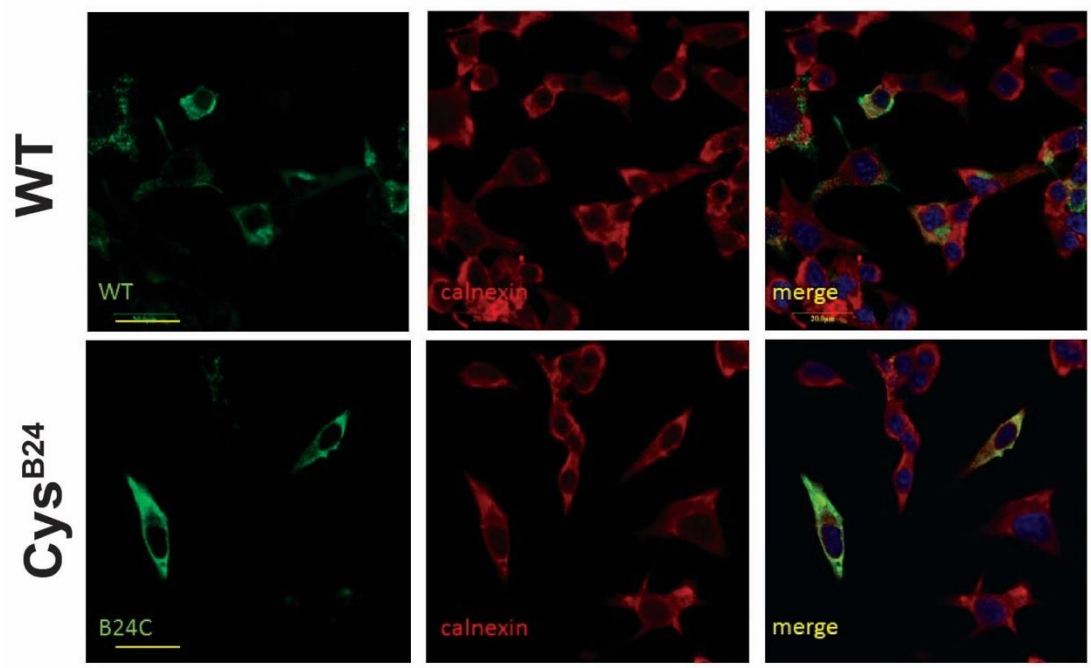

Figure S15. Intracellular distribution of a neonatal-onset MIDY-associated variant. INS1 cells expressing only rat proinsulin were transiently transfected to express human proinsulin or permanent neonatal-onset diabetes-associated proinsulin variant Phe^{B24}→Cys. Cells were stained for the native ER chaperone calnexin (*red* fluorescence) and human proinsulin (*green*). Non-transfected cells (absent *green* signal) provide a negative control for proinsulin co-localization. Whereas WT human proinsulin advances beyond the ER, the Cys^{B24} variant demonstrates a high degree of co-localization with calnexin, indicating a defect in intracellular proinsulin transport. Scale bars indicate a length of 20.0 μM.

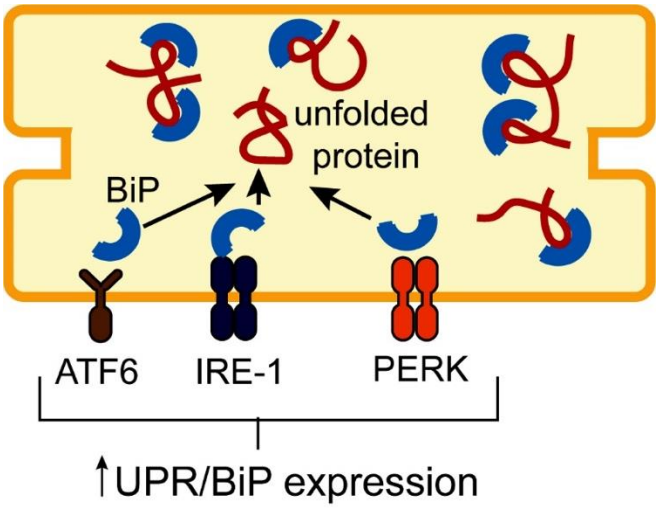

Figure S16. Cartoon representation of BiP activation in response to unfolded protein. Unfolded protein (*red coil*) within the ER activates the unfolded protein response (UPR), BiP (*blue semicircle*) dissociates from activating transcription factor 6 (ATF6; *brown*), inositol-requiring kinase 1 (IRE-1; *purple*), and protein kinase R-like endoplasmic reticulum kinase (PERK; *red-orange*). Liberated BiP binds to unfolded protein to prevent proteotoxic aggregation.

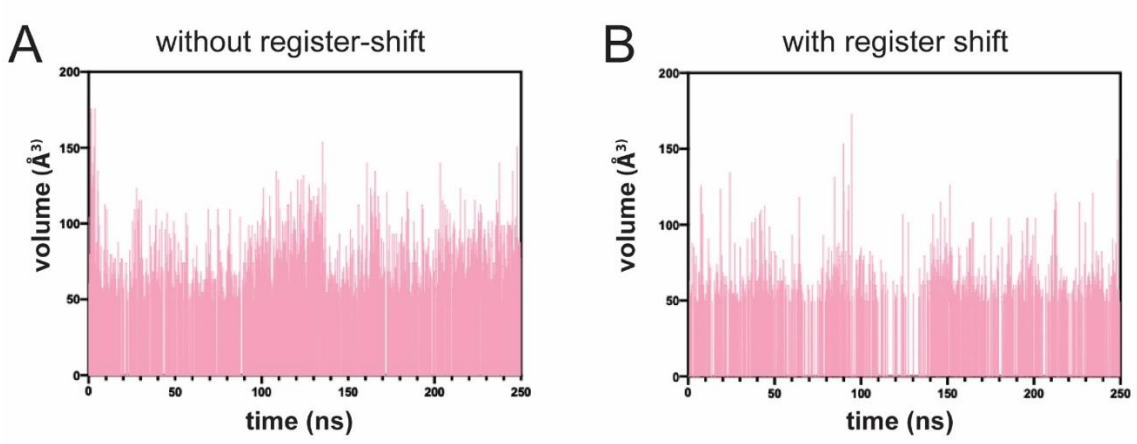

Figure S17. MD simulation of B24 cavity size in Gly^{B24}-µIR complex. Molecular dynamics (MD) simulations were undertaken based on a model of Gly^{B24}-insulin bound to micro-receptor (µIR); the model was created using the crystal structure of the insulin-µIR complex (PDB entry 4OGA). (*A*) "Naïve" model. The size of the B24 cavity within the B24 binding pocket in the L1 domain of µIR was monitored during the 250-ns simulation in a model in which Gly^{B24}-insulin is bound to µIR in the standard binding mode. (*B*) Register-shift model. Corresponding cavity sizes in the MD simulation of a model in which Gly^{B24}-insulin is bound to µIR in the register-shifted mode.

Table S1. Thermodynamic fitting parameters of insulin analogs

Analog	ΔG_u[a] (kcal mol^{-1})	C_{mid}[b] (M)	m[c] (kcal mol^{-1} M^{-1})
Leu^{B24},Orn^{B29}	2.3 ± 0.1	4.6 ± 0.09	0.5 ± 0.01
Leu^{B25},Orn^{B29}	2.9 ± 0.1	4.8 ± 0.16	0.6 ± 0.02
Gly^{B24},Leu^{B25},Orn^{B29}	1.7 ± 0.1	4.3 ± 0.32	0.4 ± 0.03
Cha^{B24},Orn^{B29}	2.5 ± 0.1	5.0 ± 0.30	0.5 ± 0.03
Cha^{B25},Orn^{B29}	2.7 ± 0.1	4.5 ± 0.08	0.6 ± 0.01
Gly^{B24},Cha^{B25},Orn^{B29}	2.1 ± 0.1	4.2 ± 0.25	0.5 ± 0.03
Gly^{B24},Orn^{B29}	2.3 ± 0.1	4.6 ± 0.28	0.5 ± 0.03
Orn^{B29}	3.6 ± 0.2	4.9 ± 0.21	0.7 ± 0.03
HI[d]	3.4 ± 0.1	5.0 ± 0.20	0.7 ± 0.03

[a]From two-state modeling of CD-detected guanidine titrations (25 °C and pH 7.4).

[b] C_{mid} is the guanidine denaturant concentration at which 50% of the protein is in the unfolded state.

[c] The *m*-value is the slope of unfolding free energy (ΔG_u) *versus* molar concentration of denaturant.

[d] WT human insulin is abbreviated as "HI."

Table S2. Fibrillation lag times

Analog	lag time[a] (days ± SD (N))
Leu^{B24},Orn^{B29}	4.7 ± 0.5 (3)
Leu^{B25},Orn^{B29}	14.0 ± 1.7 (3)
Gly^{B24},Leu^{B25},Orn^{B29}	3.0 ± 0.1[b] (3)
Cha^{B24},Orn^{B29}	7.0 ± 1.3 (3)
Cha^{B25},Orn^{B29}	5.0 ± 0.1[b] (3)
Gly^{B24},Cha^{B25},Orn^{B29}	4.3 ± 0.6 (3)
Gly^{B24},Orn^{B29}	3.0 ± 0.1[b] (3)
Orn^{B29}	7.8 ± 2.2[b] (6)
HI	3.4 ± 1.7 (5)

[a] Fibrillation lag times pertain to zinc-free WT insulin and analogs (in a monomer-dimer-higher order equilibrium); each protein was made 60 μM in PBS (pH 7.4). Samples were gently rocked at room temperature to protract fibrillation lag times relative to vigorous shaking at an elevated temperature. A twofold increase over baseline in ThT fluorescence provided a criterion for onset of fibrillation.

[b] All individual samples in this set exhibited the same lag time. As the method employed in this study could not distinguish lag times at a resolution < 1 day, some variance was added to the data by adding ± 0.1 to individual data points in order to obtain standard deviations.

Supplemental References

1. Knudsen, L., Hansen, B. F., Jensen, P., Pedersen, T. A., Vestergaard, K., Schaffer, L., Blagoev, B., Oleksiewicz, M. B., Kiselyov, V. V., and De Meyts, P. (2012) Agonism and antagonism at the insulin receptor. *PLoS One* **7**, e51972

2. Listov-Saabye, N., Jensen, M. B., Kiehr, B., Svendesen, J. E., Lundby, A., Holm, G. M., and Oleksiewicz, M. B. (2009) MCF-7 human mammary adenocarcinoma cells exhibit augmented responses to human insulin on a collagen IV surface. *J. Appl. Toxicol.* **29**, 470-477

3. Glidden, M.D., Aldabbagh, K., Phillips, N.B., Carr, K., Chen, Y.S., Whittaker, J., Phillips, M., Wickramasinghe, N.P., Rege, N., Swain, M., Peng, Y., Yang, Y., Lawrence, M.C., Yee, V.C., Ismail-Beigi, F., & Weiss, M.A. (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