SI Appendix (Menting *et al.*)

Supplementary Text

1. Materials and Methods

A. Preparation of Materials.

Synthesis of [^{13}C, ^{15}N]-labeled peptides and proteins for TROSY studies. [^{13}C, ^{15}N]-labeled peptides were synthesized by solid-phase peptide chemistry using (U-^{13}C; ^{15}N)-N-Fmoc-protected amino acids at selected sites (Cambridge Isotope Laboratories, Andover, MA). (a) *Labeled αCT peptide*. A 17-residue fragment derived from the αCT segment of the insulin receptor was prepared with sequence R-KTFEDYLHNVVFVPR-K (*i.e.*, insulin receptor residues 703-717; extended by an N-terminal Arg and C-terminal Lys to enhance solubility); labeled residues are *underlined*. The peptide was purified by reverse-phase high-performance liquid chromatography (rp-HPLC) using a semi-preparative Proto 300 C8 column (250x10 mm; Higgins Analytical, Inc., Mountain View, CA) with a 0.1% trifluoroacetic acid (TFA)/acetonitrile solvent system. Identity and purity were assessed by analytical C18 rp-HPLC (Waters) and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF-TOF MS; Applied Biosystems MS model 4800). MS of the final product (2181.2 Da) was in accordance with the calculated mass (2181.5 Da). A corresponding unlabeled peptide was synthesized and purified similarly. (b) *Labeled insulin peptides*. A [^{13}C, ^{15}N]-labeled octapeptide was synthesized and purified as above with sequence GFFYTKPA (*i.e.,* corresponding to B-chain residues B23-B30 with substitutions Pro^{B28}→Lys and Lys^{B29}→Pro to prevent dimerization of the intended insulin analogs (1, 2)); labeled residues are *underlined*. At the C-terminal position of the octapeptide, threonine was substituted by [^{13}C, ^{15}N]-alanine (as at position B30 of porcine insulin) to reduce the cost of labeled reagents. The molecular mass of the final product (955.5 Da) was in accordance with calculated mass (957 Da). The octapeptide was employed in trypsin-mediated semi-synthesis (see below) (3) to prepare a monomeric porcine insulin analog (KP-porcine insulin) that contained isotopic labels at positions B23, B24, B26 and B30.

Alanine scanning mutagenesis of residues B24-B26. Insulin analogs containing successive alanine substitutions at positions B24-B26 were prepared by trypsin-mediated semi-synthesis essentially as described (3). This method, exploiting the tryptic site between Arg^{B22} and Gly^{B23}, employs a synthetic octapeptide to introduce modifications in the C-terminal segment of the B chain. To simplify this protocol, a competing tryptic site within the octapeptide was eliminated by substitution of Lys^{B29} by ornithine (Orn). Such analogs are designated POT (Pro^{B28}-Orn^{B29}-Thr^{B30}); the wild-type human insulin sequence is Pro^{B28}-Lys^{B29}-Thr^{B30}. The Orn^{B29} substitution does not affect activity or stability. The analogs display a single peak on analytical reverse-phase HPLC and were found by MALDI-TOF mass spectrometry to exhibit the predicted molecular mass.

Synthesis of analog 1. The protocol for solid-phase synthesis is as described (4). The native

A-chain was obtained by oxidative sulfitolysis of human insulin as described (5, 6). D-Ala^{B20}, D-Ala^{B23}]-insulin and [D-Ala^{B20}, D-Ala^{B23}]-KP B chains were prepared by automated 9-fluorenylmethoxycarbonyl (Fmoc)-based synthesis and derivatized by S-sulfonation. Insulin chain combination was achieved as described (5). Purification of the [D-Ala^{B20}, D-Ala^{B23}] analogs was accomplished in each case by cation-exchange chromatography (CM-52, 24x1.2 cm, Whatman, Clifton, NJ) using acetate buffer pH 3.3 and a NaCl gradient. Final purification was by rp-HPLC using a C18 column with a 0.1% TFA/acetonitrile solvent system. Combination of 40 mg of A-chain and 20 mg of the variant B-chain resulted in a final yield of 1.1 mg. The corresponding yield of native chain combination under these conditions is *ca* 4.0 mg. The purified analog was observed to be a single component on analytical rp-HPLC (C18 column, 25x0.46 cm) using two independent elution gradients, acetonitrile in 0.1% TFA and methanol in 0.1% TFA. Since insulin isomers exhibit very low affinity for the insulin receptor, native disulfide pairing was implied by the higher residual receptor-binding affinity of the analog (**Fig. S4*B***). The yield of chain combination was reduced four-fold relative to native chain combination; no non-native isomers were encountered. The matrix-assisted laser desorption-ionization (MALDI) mass spectrum of the product was in accordance with its predicted value (5770.4 Da).

Chemical synthesis of peptides containing ΔPhe. Octapeptides with ΔPhe in the *Z* isomeric configuration were synthesized in the Molecular Biotechnology Core at the Cleveland Clinic (OH), using a Liberty peptide synthesizer (CEM Corp, NC). The sequences of B-chain insulin B23-B30 octapeptides containing ΔPhe at B24 or B25 were G-ΔPhe-FYTKPT and GF-ΔPhe-YTKPT. Preloaded Fmoc-L-Thr(tBu)-Wang resin and Fmoc-2,3-dehydrophenylalanine were purchased from Chem-Impex International, Inc. (Wood Dale, IL). The solvents and reagents including TFA, dimethylformamide, 1-methyl-2-pyrrolidone, *t*-butyl methyl ether, triisopropylsilane, 3,6-dioxa-1,8-octanedithiol and piperidine were purchased from Sigma-Aldrich (St Louis, MO). Peptides were synthesized by standard solid phase chemistry as described (4, 7). Peptides were purified by preparative rp-HPLC using a C18 Vydac column (250x10 mm) (Hesperia, CA). Identity and purity of the peptides were assessed by analytical C18 rp-HPLC (Beckman), and MALDI-TOF-TOF MS using a mass spectrometer model 4800 from Applied Biosystems). MS of the final products gave observed molecular masses of 958.3 Da (calculated mass of 958 Da) and 958.5 Da (calculated mass of 958 Da) for the ΔPhe^{B24} octapeptide and ΔPhe^{B25} octapeptide, respectively. The ΔPhe-containing octapeptides were also synthesized by Prof. Chauhan's group in New Delhi as described (8, 9) and generously provided for initial studies.

Semi-synthesis of analogs 2 and 3. The ΔPhe analogs of KP-insulin (or DKP-insulin) were prepared by trypsin-catalyzed semi-synthesis and purified by rp-HPLC (3). This protocol employs (i) synthetic octapeptides (N)-G-<u>ΔPhe</u>-FYT<u>KP</u>T or (N)-GF-<u>ΔPhe</u>-YT<u>KP</u>T with modified residue ΔPhe and "KP" substitutions *underlined*, and (ii) truncated analog *des*-octapeptide[B23-B30]-insulin (or *des*-octapeptide[B23-B30]-Asp^{B10}-insulin) generated by trypsin cleavage of human insulin (or an Asp^{B10} derivative of a mini-proinsulin precursor expressed in yeast *Pichia pastoris*). Because the octapeptide differs from the native B23-B30 sequence GFFYTPKT by interchange of Pro^{B28} and Lys^{B29} (rendering the KP bond resistant to trypsin) protection of the lysine ε-amino group is not required during trypsin treatment. In brief, *des*-octapeptide (15 mg) and octapeptide (15 mg) were dissolved in a mixture of dimethylacetamide/1,4-butandiol/0.2 M Tris-acetate (pH 8) containing 10 mM calcium acetate and 1 mM ethylenediaminetetraacetic acid (EDTA) (35:35:30 v/v, 0.4 mL). The final pH was adjusted to 7.0 with 10 µL of *N*-methylmorpholine. The solution was cooled to 12 °C, and 1.5 mg of L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated (TPCK) trypsin was added and incubated for 2 days at 12 °C. An additional 1.5 mg of TPCK-treated trypsin was added after 24 hr. The reaction was acidified with 0.1% TFA and purified by preparative C4 rp-HPLC. MS using MALDI-TOF-TOF (Applied Biosystems; Foster City, CA) in each case gave expected values of 5805.6 Da (calculated mass of 5805 Da) and 5805.7 Da (calculated mass of 5805 Da) for the ΔPhe^{B24}-KP-insulin and ΔPhe^{B25}-KP-insulin, respectively; similar MS verification was obtained in the course of preparing analogs of DKP-insulin. Efficiency of semi-synthesis of ΔPhe^{B24} analogs was *ca* 20%, whereas the efficiency of semi-synthesis of ΔPhe^{B25} analogs was *ca* 60%. Crystallization trials employed the KP version of analog 2 such that His^{B10} would be available for zinc-ion coordination as in native insulin; crystals were readily obtained under conditions associated with classical crystallization of zinc insulin hexamers (*33*).

Isotopic labeling of analogs 2 and 3. Biosynthetic labeling was accomplished in yeast *Pichia pastoris* through over-expression and secretion of a single-chain insulin precursor (10-12). Yield was enhanced by use of the stabilizing substitution $His^{B10} \rightarrow Asp$. In minimal medium the predominant carbon source was provided by ^{13}C-methanol and the predominant nitrogen source was provided by ^{15}N ammonium sulfate. In brief, the over-expressing strain of yeast cells were inoculated on day 1 in 10 ml of unlabeled YPS medium (containing 1% Yeast Extract, 2% Peptone, and 0.1 M Sorbitol) plus 100 μg/ml Zeocin in 50-ml conical tubes and cultured at 30 °C while shaking at 225 rpm. On day 4, 10 OD of the cell culture was inoculated into 2 L flasks containing 300 ml of YPS and cultured at 30 °C while shaking at 225 rpm. On day 4, ^{15}N ammonium sulfate was added to each flask to a final concentration of 0.05 g/100 ml (0.05%). After 6 hrs, cells were centrifuged and re-suspended in 300 ml of a basal medium containing 0.17% Yeast Nitrogen Base without amino acids, 0.5% ^{15}N-labeled ammonium sulfate (*i.e.*, 0.5 g/100 ml) and 1% (v/v) ^{13}C-methanol. Cells were cultured at 22 °C shaking at 225 rpm. On day 5, ^{13}C-methanol was added to a total concentration of 0.5% (v/v). Sample was harvested on the morning of day 7.

Semi-synthesis of analog 4. The Gly^{B24}-KP-insulin and Gly^{B24}-DKP-insulin were prepared by trypsin-catalyzed semi-synthesis and purified as above by rp-HPLC (3). This protocol employed the same synthetic octapeptide (N)-G-$\underline{G}$-FYT$\underline{KP}$T with modified residue Gly^{B24} and "KP" substitutions *underlined*, and (ii) truncated analog *des*-octapeptide-[B23-B30]-insulin (or *des*-octapeptide[B23-B30]-Asp^{B10}-insulin) generated by trypsin cleavage of human insulin (or an Asp^{B10} derivative of a mini-proinsulin precursor expressed in yeast *Pichia pastoris*). The analogs were purified and characterized as described above.

B. Functional Assays.

Insulin analogs 1-3 (see **Table 1** in ***main text***) were characterized at six successive levels of molecular, cellular and physiological function: (i) stability of a ternary complex consisting of the insulin analog, αCT peptide and the IR310T fragment; (ii) equilibrium binding affinities for the isolated insulin receptor (13, 14); (iii) ability to stimulate autophosphorylation of the transmembrane β-subunit of the isolated insulin receptor by its cytoplasmic tyrosine-kinase domain (designated TK activity) (14); (iv) ability to stimulate autophosphorylation of the cellular insulin receptor in mammalian cell culture (15, 16); (v) ability to activate a post-receptor signaling pathway leading to tyrosine phosphorylation of the kinase AKT (15, 17); and (vi) ability to reduce the blood-glucose concentration of rats (18) rendered diabetic by streptozotocin (19). Analog 4 was assessed with respect to receptor-binding affinity and biological activity in STZ rats.

μIR stability assays. To assess the relative stabilities of native and variant μIR complexes, we adapted the polyethylene glycol (PEG) precipitation method with tracer ^{125}I-[Tyr^{A14}]-insulin (Perkin-Elmer). The IR310.T fragment was incubated overnight at 4 °C in a 100-μl volume containing 10^{-5} M αCT peptide, $3x10^{-11}$ M ^{125}I-[Tyr^{A14}]-insulin and varying concentrations of unlabeled insulin analog in binding buffer [100 mM HEPES (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$, 50 μg/ml BSA, and 0.025% Triton-X100]. 0.5 ml of 0.2% γ-globulin and 0.5 ml 33% PEG-8000 were added, kept on ice for 10 min and microcentrifuged for 10 min; the supernatant was removed by aspiration, and radioactivity in pellet was determined in a gamma-counter. Entrapped radioactivity (cpm in the absence of IR310.T) was 5-7% of input, and sufficient IR310.T was added to obtain total binding of approximately 20-25% of added cpm.

Holoreceptor binding assays. Binding affinities to IR-B, IR-A and IGF-1R were obtained with a modified competitive displacement assay done in a microtiter plate antibody capture assay. Briefly, microtiter strip plates (Nunc Maxisorb) were incubated at 4 °C overnight with 100 μl/well with a stock solution of 40 μg/ml anti-FLAG IgG. Detergent-solubilized lysates of 293 PEAK cells transfected with cDNAs encoding the B or A isoform of the insulin receptor (IR-B/IR-A) or the Type 1 IGF receptor (IGF-1R) with C-terminal FLAG-tags were purified using wheat-germ agglutinin chromatography (20). Partially purified receptors were then immobilized in the anti-FLAG IgG coated plates. Plates were extensively washed and competitive binding assays using labeled ^{125}I-[Tyr^{A14}]-insulin or labeled IGF-I tracer ^{125}I-[Tyr^{31}]-IGF-I and unlabeled insulin analogs were carried out as described (13, 14). Data for homologous- and heterologous receptor binding were analyzed as described (21).

TK activation assays. Relative activities of insulin analogs were evaluated in an assay of hormone-stimulated receptor autophosphorylation as described (14). In brief, lectin-purified and detergent-solubilized insulin receptors (isoform B) were incubated with varying concentrations of insulin in 0.1 ml of 0.15 M NaCl, 10 mM $MgCl_2$, 25 mM HEPES (pH 7.8), 0.05% Triton X-100 (v/v) for 2 hrs at 25 °C. Autophosphorylation was initiated by addition of ATP, Na_2VO_4 and $MgCl_2$ to final concentrations of 1 mM, 2 mM and 10 mM, respectively.

The reaction was quenched after 20 min by the addition of EDTA and ethyleneglycoltetraacetic acid to final concentrations of 25 mM. The reaction solution was then transferred to Delfia microtiter assay plates (PerkinElmer Life Sciences) coated with 83-7 anti-insulin receptor IgG (22) and incubated overnight at 4 °C to immobilize the insulin receptor. After washing with phosphate-buffered saline containing 0.02% Tween 20, receptors were incubated with europium-labeled PY20 (PerkinElmer Life Sciences) according to the manufacturer's directions. After further washing, the phosphotyrosine-containing receptor was quantitated by time-resolved fluorescence using a Victor 3 (PerkinElmer Life Sciences) plate reader. Dose-response curves were obtained by nonlinear least squares analysis as implemented in Prism (GraphPad software).

Cellular studies of insulin signaling. Insulin- and insulin analog-stimulated insulin receptor and AKT phosphorylation were assessed using Pathscan Sandwich ELISA kits (Cell Signaling Technology; Danvers, MA) in an IGF-1R-deficient mouse embryo fibroblast cell line (23) expressing the human insulin receptor (isoform B; IR-B), kindly provided by Prof. Leah Cosgrove, CSIRO, Adelaide, South Australia. Cells at 80% of confluence in 6-well plates were serum starved overnight, treated with varying concentrations of insulin, analog or diluent for 5 min and lysed on ice with 1× Cell Lysis Buffer provided in the PathScan Phospho-Insulin Receptor (panTyr) or phosphor-AKT(Ser473) Sandwich ELISA kits (Cell Signaling Technology; Danvers, MA). Lysates were then sonicated, centrifuged, and the supernatant collected. Protein concentration was determined using DC Protein Assay (Bio-Rad; Hercules, CA); 50 µg of lysate were used to perform the ELISA.

Glucose uptake assays. Male Lewis rats (mean body mass ~300 g) were rendered diabetic by treatment with streptozotocin (STZ) as described (19). To test the *in vivo* potency of insulin analogs in relation to native human insulin, protein solutions containing native human insulin, an analog, or buffer alone (protein-free sterile diluent obtained from Eli Lilly and Co. Indianapolis, IN; composed of 16 mg glycerin, 1.6 mg *meta*-cresol, 0.65 mg phenol, and 3.8 mg sodium phosphate pH 7.4) were injected subcutaneously, and resulting changes in blood glucose concentration were monitored by serial measurements using a clinical glucometer (Hypoguard Advance Micro-Draw meter). In certain experiments, the insulin sample (or diluent) was injected through the lateral tail vein. The intravenous (IV) protocol was used to eliminate any differences in observed potency (or duration) that could be attributed to absorption from the subcutaneous depot. To ensure uniformity of formulation, insulins were each re-purified by rp-HPLC, dried to powder, dissolved in diluent at the same maximum protein concentration (300 µg/mL) and re-quantified by analytical C4 rp-HPLC; dilutions were made using the above buffer. Rats were injected subcutaneously at time *t*=0 with 6.7, 20, or 36 μg insulin or insulin analog in 100 µl of buffer per 300 g rat. This middle dose corresponds to *ca* 67 μg/kg body weight, which in international units (IU) implies 2 IU/kg body weight. Dose-response studies of native insulin indicated that at this dose a near-maximal rate of glucose disposal was achieved during the first hour following injection. Blood was obtained from the clipped tip of the tail at time *t*=0 and then every 10 min up to 90 min. The efficacy of insulin action to reduce blood glucose concentration was calculated using

(a) the change in concentration over time (using least-mean squares and initial region of linear fall) divided by the concentration of insulin injected and (b) the integrated area between the glucose time-dependence curve and a horizontal line at the starting blood glucose concentration. Assessment of statistical significance was performed using a Student's *t*-test.

C. X-ray Crystallography.

Crystallographic analysis of Complex A. Crystals of the quaternary complex of IR310.T, human insulin, the insulin receptor αCT peptide 704-719 and Fab 83-7 ("Complex A") were grown using protocols effectively identical to those described previously (24). Diffraction data of varying quality were collected from more than forty crystals at beamline MX2 (25) of the Australian Synchrotron (Melbourne). All data were collected at 0.9537 Å wavelength and at 100 K; crystals were cryo-protected as described previously (24). Where multiple data sets were collected from a single crystal, the X-ray beam was repositioned to an unexposed region of the crystal between data set acquisition. All data sets were processed with XDS (v30-Mar-2013) (26) and six data sets (from four crystals) were judged to diffract to at least 4.0 Å resolution. These were then assessed for their ability to merge with the originally-reported Complex A Data Set 1 (24) (the latter herein termed D_1; re-processed to the highest resolution at which $CC_{1/2}$ remained significant at the p=0.001 level of probability (27)). Variation in cell dimension was assessed as insignificant at a resolution limit of 3.5 Å if the difference in the body diagonal length of the unit cell of any one crystal from the mean body diagonal length of the others was less than half the maximum resolution limit (*i.e.*, 1.75 Å). For each data set, the high-resolution cut-off was set equal to that of the highest resolution shell in which the statistic $CC_{1/2}$ remained significant at the p=0.001 level of confidence. Each of the six new data sets were then merged individually with D_1 and the merged two-crystal data set which showed the greatest improvement in $<I/\sigma(I)>$ (and concomitantly in $CC_{1/2}$) in the outer resolution shell was retained as data set D_2. Each of the remaining five data sets was then merged individually with D_2 and the merged three-crystal data set that gave the greatest improvement in $<I/\sigma(I)>$ (and concomitantly in $CC_{1/2}$) was retained as data set D_3. Statistics for the above processes are in **Table S3**. Attempts at merging the four remaining data sets individually with D_3 led to no improvement in the outer shell statistics; these remaining data sets were thus discarded.

Crystallographic refinement of Complex A began with PDB entry 3W11 (24) as the starting model, which was then refined against data set D_3 using autoBUSTER v2.10 (28). Inspection of the resultant $2F_{\mathrm{obs}}$-F_{calc} difference electron density map obtained after the first round of crystallographic refinement revealed two polypeptide-like segments, the first running approximately parallel to the first strand of the central β-sheet (L1-β_2) of the receptor L1 domain and the second extending onwards from insulin receptor residue 715 (the last modeled residue of the αCT segment in PDB entry 3W11). More detailed inspection then suggested that insulin residue the A-chain C-terminal residue $\mathrm{Asn}^{\mathrm{A21}}$ was incorrectly positioned and complicating interpretation of these segments. Further refinement was then conducted with this residue omitted from the model. Segmentation of a $2F_{\mathrm{obs}}$-F_{calc} difference map was undertaken using CHIMERA v1.8.1 (29), having first masked out all density within 1.7 Å of any atom within the model or its symmetry-related counterparts within the unit cell (**Fig. 1*C***). The resultant maps allowed interpretation of the above two density segments as corresponding to insulin residues B22-B27 and αCT residues 716-719, respectively, and also revealed the correct location for Asn^{A21}. These additional residues were then included in the model and the resultant model iteratively refined with manual rebuilding in COOT (30). Negative *B*-factor sharpening (31) assisted map interpretation during the above process. Refinement protocols included (a) local structure similarity restraints (32) of the IR310.T fragment to PDB entry 2HR7_A (33), the 83-7 light chain to PDB entry 1IL1_B (34) and the 83-7 heavy chain to PDB entry 1FNS_H (35), (b) TLS refinement (six domains: IR310.T, insulin A chain, insulin B chain, Fab 83-7 light chain, Fab 83-7 heavy chain and αCT), (c) allowance for unmodeled segments of polypeptide and carbohydrate through use of a map-based non-uniform prior, (d) stereochemically-restrained individual atomic coordinate refinement, and (e) restrained individual isotropic *B*-factor refinement (28, 36, 37). Final refinement statistics are presented in **Table S3**; Ramachandran plot statistics are: 91.1% of residues lie the preferred region, 4.8 % in the allowed region and 4.1% in the outliers region; region definitions are those implemented in COOT v0.7.1 (30).

Crystallographic analysis of Analog 1. Zinc-free crystals of the β-turn-stabilized analog were initially obtained unintentionally during attempts to co-crystallize the analog in complex with IR310.T in the presence of the insulin receptor αCT peptide (residues 704-719) and Fab 83-7. The refined conditions under which [D-Ala^{B20}, D-Ala^{B23}]-human insulin crystallized employed 0.7 M tri-sodium citrate plus 0.1 M imidazole-HCl (pH 8.0). Crystals (~50 μm) were then cryo-cooled by plunging directly into liquid nitrogen without cryo-protectant. Data were collected from a single crystal at the MX2 beamline at the Australian Synchrotron and processed to 1.30 Å resolution using XDS (26). Molecular replacement was conducted with PHASER (38) with a porcine insulin monomer (Protein Data Bank entry 1B2A) as a search model. A single copy was located within the crystallographic asymmetric unit. The structure was then refined against all data to 1.40 Å resolution using PHENIX v1.8.2_1309 (39) iterated with model improvement in COOT (30). Data collection and refinement statistics are in **Table S5**; Ramachandran plot statistics are: 95.7% of residues lie the preferred region and 4.3 % in the allowed region; region definitions are those implemented in COOT v0.7.1 (30).

D. Spectroscopy and Molecular Modeling.

Circular dichroism (CD) and thermodynamic stabilities. Far-ultraviolet CD spectra were obtained using an Aviv spectropolarimeter as described (18). Thermodynamic stabilities were probed at 25 °C by CD-monitored guanidine denaturation at helix-sensitive wavelength 222 nm as described (18). Estimates of the free energy of unfolding and other fitting parameters were obtained by application of a two-state model as described (40).

Heteronuclear NMR studies of μIR complexes. Spectra were obtained at 900 MHz and 25 °C. Two μIR samples were prepared, each containing a 1:1:1 complex of the L1-CR fragment IR310.T, the KP-insulin analog and αCT peptide. In the first sample ("complex 1") both the KP-insulin analog and αCT peptide contained selective ^{13}C and ^{15}N labels; in the second sample ("complex 2") only the αCT peptide was labeled. In each case the unlabeled L1-CR glycosylated fragment (51 kDa; the same fragment as employed in complex 1) was exchanged to 25 mM sodium phosphate in D_2O (pD 7.2, direct meter reading).

Complex 1 was obtained on addition of Ala^{B30}-KP-insulin ([^{13}C, ^{15}N]-labeled at Gly^{B23}, Phe^{B24}, Tyr^{B26} and Ala^{B30}; see above) and αCT peptide R-KTFED$\underline{\text{YL}}$HNV$\underline{\text{VF}}$VPR-K as described above and [^{13}C, ^{15}N]-labeled at Tyr^{708}, Leu^{709}, Val^{713} and Phe^{714} (*underlined*). *Complex 2* was obtained on addition of unlabeled KP-insulin and the same labeled αCT peptide; comparison of the spectra of the two complexes immediately distinguished αCT resonances from insulin-related resonances. In each case the samples were then incubated for 4 hrs at 4 °C and concentrated for NMR studies to a final concentration of 0.2 mM or 0.1 mM (Complex 2). Samples of the free labeled αCT and of free labeled Ala^{B30}-KP-insulin were also prepared for baseline NMR studies of the unbound components.

The low concentration of the complexes restricted NMR studies to analysis of resonance chemical-shift fingerprints as provided by TROSY-enhanced HSQC spectroscopy (TROSY: transverse relaxation-optimized NMR spectroscopy; HSQC: heteronuclear single-quantum coherence); such spectra provide cross peaks between ^{1}H-^{13}C spin pairs (41). Conventional [^{1}H, ^{13}C]-HSQC spectra and [^{1}H, ^{13}C]-TROSY-enhanced spectra of the complexes were acquired on a Bruker 900 MHz Avance spectrometer at the Cleveland Center for Membrane and Structural Biology. [^{1}H, ^{13}C]-HSQC spectra of respective aliphatic- and aromatic regions were acquired using standard pulse sequences. [^{1}H, ^{13}C]-TROSY spectra of aromatic spin systems were acquired as described by Wüthrich and colleagues (42); acquisition times were typically 12 hrs per HSQC spectrum and 15 hrs per TROSY spectrum as extensive signal averaging was required to obtain adequate signal-to-noise ratios. Although the labeled moieties also contained ^{15}N at selective sites, [^{1}H-^{15}N]-TROSY spectra in H_2O solution were not obtained due to the low concentration of the complexes. [^{1}H, ^{13}C]-HSQC spectra of free labeled αCT and free labeled Ala^{B30}-KP insulin were acquired on Bruker 700 MHz Avance spectrometer at 25 °C; acquisition time were 3 hrs. All data sets were processed using NMRPipe (43), and the spectra were analyzed using PIPP (44).

Homonuclear ^{1}H-NMR spectroscopy of insulin analogs. NOE spectroscopy (NOESY), total correlation spectroscopy (TOCSY), and double-quantum-filtered correlation spectroscopy (DQF-COSY) spectra of KP-insulin, DKP-insulin, [D-Ala^{B20}, D-Ala^{B23}]-DKP-insulin, ΔPhe^{B24}-KP-insulin, ΔPhe^{B25}-DKP-insulin, and Gly^{B24}-DKP-insulin were acquired at 600, 700 and/or 800 MHz and at 25 °C or 32 °C in (a) aqueous solution at pH 7.0-8.0, (b) 10 mM deuterioacetic acid at pH 3.0, and in the case of the ΔPhe^{B24} and Gly^{B24} analogs, (c) 20% deuterioacetic acid (pH 1.9) as described (45). Matched spectra of Gly^{B24}-DKP-insulin and its parent DKP-insulin were also obtained at 600 and 800 MHz at a series of NOESY mixing times from 50-300 ms. The solution structure of Gly^{B24}-DKP-insulin employed long-range NOEs observed in aqueous solution at pH 7.0-8.0 (which in the case of residues in the C-terminal segment of the B chain were not observed in 20% deuterioacetic acid).

Heteronuclear multidimensional NMR spectroscopy. A labeled sample of ΔPhe^{B25}-DKP-insulin (in which the *des*-octapeptide[B23-B30] fragment was uniformly enriched with ^{13}C and ^{15}N; see above) was prepared in nitrogen-purged H_2O (7% D_2O) at pH 7.0 in a 300-μl Shigemi NMR tube with *ca* 0.5 mM concentration. All the experiments were performed at 25 °C using Bruker Avance 700 MHz spectrometer equipped with a triple resonance cryo-probe as described (18). Spectra were processed with NMRPipe (43) and analyzed with PIPP (44). Assignments were obtained from 3D HNCACB, CBCA(CO)NH, C(CO)NH, H(CCO)NH and HCCH-TOCSY spectra and extended by analyses of 3D ^{13}C- and ^{15}N-separated NOESY-HSQC spectra (41). Three kinds of 4D-NOESY spectra based on 4D time-shared NOESY experiment (46), 4D ^{15}N, ^{15}N-edited NOESY, 4D ^{15}N, ^{13}C-edited NOESY and 4D ^{13}C, ^{13}C-edited NOESY, were used to derive inter-proton distance constraints between labeled residues; inter-proton distance constrains between labeled residues and unlabeled residues were derived from 3D ^{13}C-seperated NOESY; and distance constrains between unlabeled residues were derived from 2D homonuclear NOESY spectra.

Interpretation of NOEs in Analog 4. To enable interpretation of inter-residue NOEs in flexible polypeptide segments *versus* domains of stable structure, a series of NOESY spectra of DKP-insulin and Gly^{B24}-DKP-insulin at successive mixing times from 50-300 ms. Long-range NOEs between the aliphatic resonances of Ile^{A2} and the aromatic ring of Tyr^{A19} (similar in DKP-insulin and Gly^{B24}-DKP-insulin) were employed as a reference standard for the build-up of long-range NOEs *within a stably-maintained structural domain*. Whereas in DKP-insulin in D_2O (pD 6.6-7.6 at 25 °C or 32 °C) a similar pattern of mixing-time-dependent long-range NOEs was observed involving the side chains of Phe^{B24} and Tyr^{B26}, in the Gly^{B24} analog long-range NOEs involving Phe^{B25} and Tyr^{B26} were absent at mixing times shorter than 200 ms and weak at longer mixing times. The latter pattern was interpreted as transient contacts by a flexible segment. This interpretation was supported by back-calculation of ensemble-averaged simulated NOESY spectra. Distance geometry (DG) / restrained molecular dyanmics (RMD) models were cross-validated by aromatic ^{1}H-NMR ring-current shifts with attention to the chemical shifts of the Leu^{B15} spin system as a function of the position and orientation of residues Phe^{B24}, Phe^{B25}, and Tyr^{B26} in DKP-insulin and as a function of the position and orientation of Phe^{B25} and Tyr^{B26} in Gly^{B24}-DKP-insulin (47). Characteristic upfield secondary ^{1}H chemical shifts of Leu^{B15} aliphatic resonances in DKP-insulin, due to the neighboring Phe^{B24} ring current, are markedly attenuated in spectra of Gly^{B24}-DKP-insulin, implying that the Phe^{B25} aromatic ring does not stably re-engage the B-chain α-helix in a corresponding manner.

NMR structure calculations. The solution structure of ΔPhe^{B25}-DKP-insulin was calculated by distance geometry (DG), simulated annealing (SA), and restrained molecular dynamics (RMD) using XPLOR-NIH (48, 49) as follows. The initial structure was calculated by employing only unambiguous NOEs, and then the calculations were iteratively performed with successive NOE analysis. All assigned NOEs were classified into strong, medium, weak, and very weak categories corresponding to inter-proton distance restraints of 1.8-2.7 Å (1.8-2.9 Å for NOEs involving NH protons), 1.8-3.3 Å (1.8-3.5 Å for NOEs involving NH protons), 1.8-5.0 Å, and 1.8-6.0 Å, respectively (50, 51). Backbone ϕ and ψ torsion angle restraints were generated using TALOS+ (52, 53), in which minimum error ranges were set to ±20-30° from the average predicted value (54, 55). Root-mean-square deviations (RMSDs) were calculated with respect to both mean coordinates. The solution structure of ΔPhe^{B25}-DKP-insulin was calculated based on 931 NOE distance restraints and 86 dihedral-angle restraints (**Table S6**). A Ramachandran plot of the final structures was obtained (56). Control ensembles were calculated that included explicit constraints on these angles to the three Ramachandran regions preferred by ΔPhe in prior peptide studies (in each case ±40º) (57-61). Lennard-Jones (L-J) potential energy functions were calculated with empirical energy parameters as implemented in CHARMM19 (62). Simulated NOESY spectra based on structure models were calculated by back-calculation using X-PLOR as described (47); a rotational correlation time of 2.5 ns was assumed. In the solution structure of analog 4, helix-related hydrogen bonds were inferred from the pattern of protected amide resonances as observed in D_2O solution at pH 2.5 or in a co-solvent containing 20% deuterioacetic acid, conditions that favor observation of main-chain amide resonances.

Analysis of NMR chemical shifts. The ^{1}H and ^{13}C chemical shifts of random-coil residues and patterns of secondary-structure-associated shifts were obtained from Sykes and coworkers (63, 64). ^{1}H-NMR aromatic ring-current shifts were calculated based on crystal structures of insulin or NMR-derived ensembles as described (85).

Molecular modeling of insulin analogs. Models of ΔPhe^{B24}-DKP-insulin were generated using a protocol similar to that employed in the NMR-derived models ΔPhe^{B25}-DKP-insulin based on the following assumptions. (i) Distance constraints for the α-helical domain of ΔPhe^{B24}-DKP-insulin helix core region (residues A1-A21 and B1-B22) were assumed to be identical to those determined by NMR analysis of the ΔPhe^{B25} analog; (ii) the B23-B30 segment was assumed not to be in stable contact with the α-helical moiety of the insulin, such that only sequential and intra-residue distance constraints for the B23-B30 segment were retained (*i.e.*, all constraints between the B23-B30 segment and helical core domain were removed); and (iii) main-chain dihedral angles of the ΔPhe residue were assumed to lie within one or another of the allowed regions of the Ramachandran plot (±40°) based on prior crystallographic studies of peptides containing ΔPhe.

Molecular modeling of μIR complexes. Comparative modeling of the complex of IR310.T, αCT and insulin was performed using MODELLER (65) employing as molecular templates the structure of the insulin receptor fragments from PDB entry 3W11 (24) and the structure of native insulin (PDB entry 2G4M (66)). The models included residues disordered in the crystal structure of Complex A presented here, *viz.* Cys159-Asn168 and Lys265-Gln276 of insulin receptor and Asp^{B28}-Thr^{B30} of insulin. An initial fifty models were created for each complex and the structure with the lowest energy was selected for molecular dynamics simulations. IR310.T contains N-linked glycosylation sites at residues Asn16, Asn25, Asn111, Asn215 and Asn255 (67); *N*-acetyl glucosamine carbohydrate was thus attached at each of these sites. Molecular dynamics (MD) calculations were performed using the

GROMACS (v4.5.5) suite (68) employing the OPLS-aa force field (69, 70). Each complex was solvated in a box (85 × 85 × 85 Å) of TIP4P water (71). Ionizable residues and protein termini were assumed to be in their charged state. Sodium and chloride ions were added to neutralize the system and provide a final ionic strength of 0.10 M. Protein and solvent (including ions) were coupled separately to a thermal bath at 300 K employing velocity rescaling (72) applied with a coupling time of 1.0 ps. The pressure was maintained at 1 bar by coupling to a Berendsen barostat (73) with a coupling constant of 5.0 ps and compressibility of 4.5×10^{-5} bar. The time step used was 2 fs. Simulations were performed with a single non-bonded cutoff of 10 Å and applying a neighbor-list update frequency of 10 steps (20 fs). The particle-mesh Ewald method was used to account for long-range electrostatics (74), applying a grid width of 1.2 Å and a fourth-order spline interpolation. Bond lengths were constrained using the LINCS algorithm (75). All simulations consisted of an initial minimization of water molecules, followed by 100 ps of MD with the protein restrained. Following positional restraints of MD, all restraints on the protein were removed and MD continued for a further 20 ns. Partial atomic charges for ΔPhe were obtained from the electrostatic potential calculated at the HF/6-311+G(d,p)//HF/6-31G(d) level using the GAUSSIAN-09 program (76). The structure of (Z)-2-formamido-3-phenylacrylamide was optimized with both the amide moieties restrained to have a planar geometry. The electrostatic potential was calculated at this optimized geometry using the CHELPG approach. Partial atomic charges were determined from this potential using the RESP method (77). Partial atomic charges on the aromatic carbon and hydrogen atoms, and the amide atoms were constrained to those values used throughout the OPLS-aa force field. The resulting partial atomic charges on the ΔPhe residue were q_{N}=-0.30, q_{NH}=0.22, $q_{\mathrm{C}\alpha}$=0.03, q_{C}=0.5, q_{O}=-0.5, $q_{\mathrm{C}\beta}$=-0.055, $q_{\mathrm{H}\beta}$=0.03, $q_{\mathrm{C}\gamma}$=0.075, $q_{\mathrm{Ca}_{\rho}}$=-0.055, $q_{\mathrm{Ha}_{\rho}}$=0.055.

Ab initio *quantum mechanical (QM) calculations.* Electron density and molecular electrostatic potential (MEP) were calculated using B3LYP method and 6-31G(d) basis set using the GAUSSIAN-09 utility *Cubegen* (76); isosurface electrostatic maps were generated using Jmol software (78). *Ab initio* QM calculations were performed on molecular coordinates of isolated amino acids obtained from coordinates derived from the PDB without any further optimization.

Molecular Graphics. Graphics were generated using POVRAY (79), DINO (80) and /or CHIMERA (29).

REFERENCES

1. Brems DN*, et al.* (1992) Altering the association properties of insulin by amino acid replacement. *Protein Eng.* 5:527-533.
2. Weiss MA, Hua QX, Lynch CS, Frank BH, & Shoelson SE (1991) Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective ^{2}H and ^{13}C labeling with application to protein design. *Biochemistry* 30:7373-7389.
3. Mirmira RG & Tager HS (1989) 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.
4. Merrifield RB, Vizioli LD, & Boman HG (1982) Synthesis of the antibacterial peptide cecropin A (1-33). *Biochemistry* 21:5020-5031.
5. Hu SQ*, et al.* (1993) Steric requirements at position B12 for high biological activity in insulin. *Biochemistry* 32:2631-2635.
6. Hua QX*, et al.* (2006) A conserved histidine in insulin is required for the foldability of human proinsulin. Structure and function of an Ala^{B5} analog. *J. Biol. Chem.* 281:24889-24899.
7. Barany G & Merrifield RB (1980) *The Peptides*, eds Gross E & Meienhofer J (Academic Press, New York), pp 273-284.
8. Jain RM, Rajashankar KR, Ramakumar S, & Chauhan VS (1997) First observation of left-handed helical conformation in a dehydro peptide containing two L-Val residues. Crystal and solution structure of Boc-L-Val-ΔPhe-ΔPhe-ΔPhe-L-Val-OMe. *J. Am. Chem. Soc.* 119:3205-3211.
9. Gupta M*, et al.* (2008) Dehydrophenylalanine (ΔPhe) as a β breaker: extended structure terminated by a Δ-Phe-induced turn in the pentapeptide Boc-Phe1-Ala2-Ile3-ΔPhe4-Ala5-OMe. *ChemBioChem.* 9:1375-1378.
10. Thim L*, et al.* (1986) Secretion and processing of insulin precursors in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 83:6766-6770.
11. Kjeldsen T (2000) Yeast secretory expression of insulin precursors. *Appl. Microbiol. Biotechnol.* 54:277-286.
12. Wang Y*, et al.* (2001) Human insulin from a precursor overexpressed in the methylotrophic yeast *Pichia pastoris* and a simple procedure for purifying the expression product. *Biotechnol. Bioeng.* 73:74-79.
13. Whittaker J*, et al.* (2001) Alanine scanning mutagenesis of a type 1 insulin-like growth factor receptor ligand binding site. *J. Biol. Chem.* 276:43980-43986.
14. Whittaker J & Whittaker L (2005) Characterization of the functional insulin binding epitopes of the full-length insulin receptor. *J. Biol. Chem.* 280:20932-20936.

15. Sohma Y*, et al.* (2010) Contribution of residue B5 to the folding and function of insulin and IGF-I: constraints and fine-tuning in the evolution of a protein family. *J. Biol. Chem.* 285:5040-5055.

16. Whittaker J*, et al.* (2012) α-Helical element at the hormone-binding surface of the insulin receptor functions as a signaling element to activate its tyrosine kinase. *Proc. Natl. Acad. Sci. U. S. A.* 109:11166-11171.

17. Denley A*, et al.* (2007) Differential activation of insulin receptor substrates 1 and 2 by insulin-like growth factor-activated insulin receptors. *Mol. Cell. Biol.* 27:2569-2577.

18. Yang Y*, et al.* (2010) An Achilles' Heel in an amyloidogenic protein and its repair. Insulin dynamics, misfolding, and therapeutic design. *J. Biol. Chem.* 285:10806-10821.

19. Saker F*, et al.* (1998) Glycemia-lowering effect of cobalt chloride in the diabetic rat: Role of decreased gluconeogenesis. *Am. J. Physiol.* 274:E984-E991.

20. Hedo JA, Harrison LC, & Roth J (1981) Binding of insulin receptors to lectins: evidence for common carbohydrate determinants on several membrane receptors. *Biochemistry* 20:3385-3393.

21. Wang ZX (1995) An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule *FEBS Lett.* 360:111-114.

22. Soos MA*, et al.* (1986) Monoclonal antibodies reacting with multiple epitopes on the human insulin receptor. *Biochem. J.* 235:199-208.

23. Sell C*, et al.* (1994) Effect of a null mutation of the insulin-like growth factor I receptor gene on growth and transformation of mouse embryo fibroblasts. *Mol. Cell. Biol.* 14:3604-3612.

24. Menting JG*, et al.* (2013) How insulin engages its primary binding site on the insulin receptor. *Nature* 493:241-245.

25. McPhillips TM*, et al.* (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron. Radiat.* 9:401-406.

26. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D. Biol. Crystallogr.* 66:133-144.

27. Karplus PA & Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336:1030-1033.

28. Bricogne G*, et al.* (2012) BUSTER, version 2.11.4 (Global Phasing Ltd, Cambridge, United Kingdom).

29. Pettersen EF*, et al.* (2004) UCSF Chimera- a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605-1612.

30. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* 60:2126-2132.

31. Brunger AT, DeLaBarre B, Davies JM, & Weis WI (2009) X-ray structure determination at low resolution. *Acta Crystallogr. D Biol. Crystallogr.* 65:128-133.

32. Smart OS*, et al.* (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D. Biol. Crystallogr.* 68:368-380.

33. Lou M*, et al.* (2006) The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. *Proc. Natl. Acad. Sci. U. S. A.* 103:12429-12434.

34. Berry MB*, et al.* (2001) Structure of an anti-HIV monoclonal Fab antibody fragment specific to a gp120 C-4 region peptide. *Proteins* 45:281-282.

35. Celikel R, Ruggeri ZM, & Varughese KI (2000) von Willebrand factor conformation and adhesive function is modulated by an internalized water molecule. *Nat. Struct. Biol.* 7:881-884.

36. Roversi P, Blanc E, Vonrhein C, Evans G, & Bricogne G (2000) Modelling prior distributions of atoms for macromolecular refinement and completion. *Acta Crystallogr. D. Biol. Crystallogr.* 56:1316-1323.

37. Blanc E*, et al.* (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D. Biol. Crystallogr.* 60:2210-2221.

38. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J. Appl. Cryst.* 40:658-674.

39. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66:213-221.

40. Sosnick TR, Fang X, & Shelton VM (2000) Application of circular dichroism to study RNA folding transitions. *Methods Enzymol.* 317:393-409.

41. Bax A (1994) Multidimensional nuclear magnetic resonance methods for protein studies. *Curr. Opin. Struct. Biol.* 4:738-744.

42. Pervushin K, Riek R, Wider G, & Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.* 94:12366-12371.

43. Delaglio F*, et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6:277-293.

44. Garrett DS, Powers R, Gronenborn AM, & Clore GM (1991) A common sense approach to peak picking in two- three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* 95:214-220.

45. Hua QX*, et al.* (1996) Mapping the functional surface of insulin by design: structure and function of a novel A-chain analogue. *J. Mol. Biol.* 264:390-403.

46. Xu Y, Long D, & Yang D (2007) Rapid data collection for protein structure determination by NMR spectroscopy. *J. Am. Chem. Soc.* 129:7722-7723.

47. Jacoby E, Hua QX, Stern AS, Frank BH, & Weiss MA (1996) Structure and dynamics of a protein assembly. 1H-NMR studies of the 36 kDa R6 insulin hexamer. *J. Mol. Biol.* 258:136-157.

48. Schwieters CD, Kuszewski J, Tjandra N, & Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 160:65-73.

49. Schwieters CD, Kuszewski JJ, & Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog. NMR Spec.* 48:47-62.

50. Williamson MP, Havel TF, & Wüthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* 182:295-315.

51. Clore GM*, et al.* (1986) The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* 5:2729-2735.

52. Cornilescu G, Delaglio F, & Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 13:289-302.

53. Shen Y, Delaglio F, Cornilescu G, & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44:213-223.

54. Caffrey M*, et al.* (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J.* 17:4572-4584.

55. Cai ML*, et al.* (2001) Solution structure of the constant region of nuclear envelope protein LAP2 reveals two LEM-domain structures: one binds BAF and the other binds DNA. *EMBO J.* 20:4399-4407.

56. Laskowski RA, Macarthur MW, Moss DS, & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283-291.

57. Ramagopal UA, Ramakumar S, Joshi RM, & Chauhan VS (1998) Crystal structure of Boc-LAla-ΔPhe-ΔPhe-ΔPhe-ΔPhe-NHMe: a left-handed helical peptide. *J. Pept. Res.* 52:208-215.

58. Vijayaraghavan R, Kumar P, Dey S, & Singh T (1998) Design of peptides with aβ-dehydro residues: synthesis, crystal structure and molecular conformation of N-Boc-L-Ile-ΔPhe-L-Trp-OCH_3. *J. Pept. Res.* 52:89-94.

59. Mitra SN, Dey S, Karthikeyan S, & Singh TP (1997) Design of specific structures using α, β-dehydro-phenylalanine residues: synthesis, crystal structure, and molecular conformation of Boc-L-Val-Δ Phe-Δ Phe-L-Val-Δ Phe-Δ Phe-L-Val-OCH_3, a 3_{10}-helical heptapeptide. *Biopolymers* 41:97-105.

60. Ramakumar S*, et al.* (2004) De novo design and characterization of a helical hairpin eicosapeptide: emergence of an anion receptor in the linker region. *Structure* 12:389-396.

61. Gupta M & Chauhan VS (2010) De novo design of α,β-didehydrophenylalanine containing peptides: From models to applications. *Biopolymers* 95:161-173.

62. Brooks BR*, et al.* (2009) CHARMM: The biomolecular simulation program. *J. Comp. Chem.* 30:1545-1614.

63. Wishart DS, Bigam CG, Holm A, Hodges RS, & Sykes BD (1995) ^{1}H, ^{13}C, and ^{15}N random coil NMR chemical-shifts of the common amino-acids. 1. Investigation of nearest-neighbor effects. *J. Biomol. NMR* 5:67-81.

64. Wishart DS*, et al.* (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6:135-140.

65. Fiser A & Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* 374:461-491.

66. Mueller-Dieckmann C*, et al.* (2007) On the routine use of soft X-rays in macromolecular crystallography. Part IV. Efficient determination of anomalous substructures in biomacromolecules using longer X-ray wavelengths. *Acta Crystallogr. D Biol. Crystallogr.* 63:366-380.

67. Elleman TC*, et al.* (2000) Mutational analysis of the N-linked glycosylation sites of the human insulin receptor. *Biochem. J.* 347:771-779.

68. Hess B, Kutzner C, Van Der Spoel D, & Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435-447.

69. Jorgensen WL & Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657-1666.

70. Damm W, Frontera A, Tirado–Rives J, & Jorgensen WL (1997) OPLS all□atom force field for carbohydrates. *J. Comp. Chem.* 18:1955-1970.

71. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926-935.

72. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.

73. Berendsen HJ, Postma JPM, van Gunsteren WF, DiNola A, & Haak J (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684-3690.

74. Essmann U*, et al.* (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577-8593.

75. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4:116-122.

76. Frisch M*, et al.* (2009) Gaussian 09, revision A. 02; Gaussian, Inc., Wallingford CT.

77. Bayly CI, Cieplak P, Cornell W, & Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97:10269-10280.

78. Willighagen E & Howard M (2007) Fast and Scriptable Molecular Graphics in Web Browsers without Java3D. http://dx.doi.org/10.1038/npre.2007.50.1.

79. Anonymous (2004) Povray: The persistence of vision raytracer. http://www.povray.org.

80. Philippsen A (2003) DINO: Visualizing structural biology. http://www.dino3d.org.

81. Huang Y, Liang Z, & Feng Y (2001) The relationship between the connecting peptide of recombined single chain insulin and its biological function. *Sci. China C Life Sci.* 44:593-600.

82. Hua QX, Shoelson SE, Kochoyan M, & Weiss MA (1991) Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* 354:238-241.

83. Ludvigsen S, Olsen HB, & Kaarsholm NC (1998) A structural switch in a mutant insulin exposes key residues for receptor binding. *Journal of molecular biology* 279:1-7.

84. Derewenda U*, et al.* (1991) X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. *J. Mol. Biol.* 220:425-433.

85. Hua QX*, et al.* (1998) Mini-proinsulin and mini-IGF-I: homologous protein sequences encoding non-homologous structures. *J. Mol. Biol.* 277:103-118.

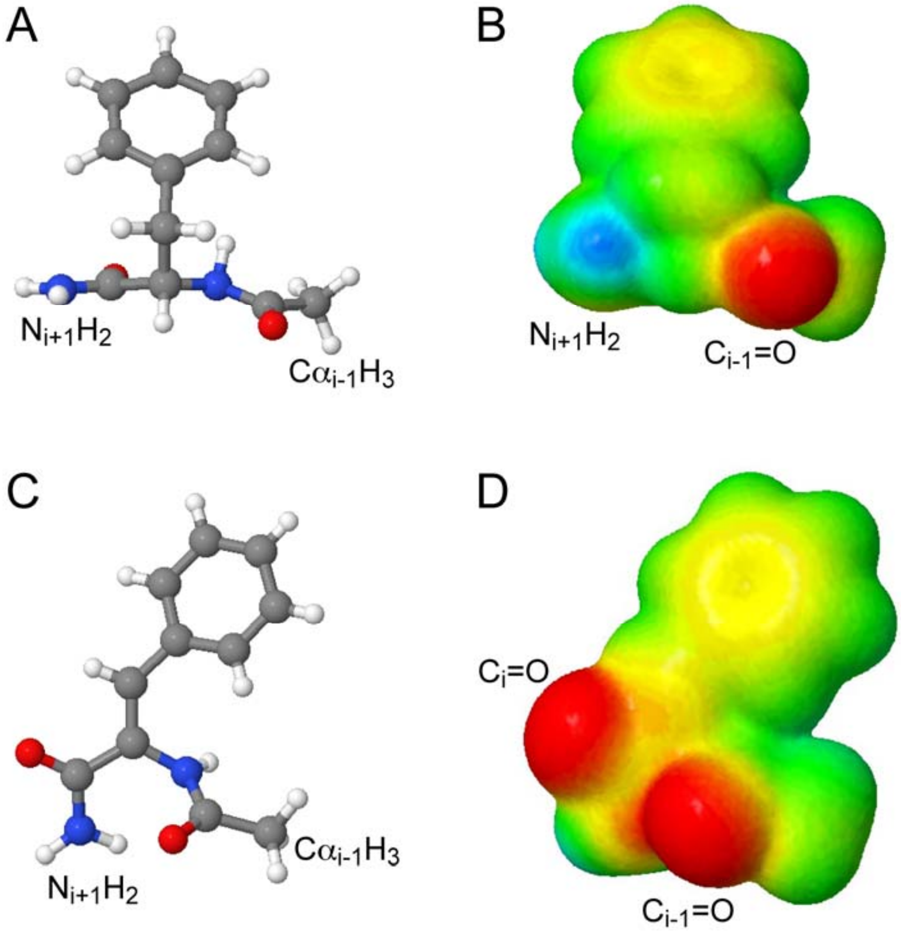

Figure S1. Structure of Phe *versus* (α,β)-dehydro-Phe. (*A* and *B*) Canonical phenylalanine. (*A*) Stick model of Phe as an isolated amino acid capped by an N-terminal acetyl group and C-terminal amide group. (*B*) *Ab initio* QM model of the electrostatic surface of the capped Phe molecule. (*C* and *D*) *Z* isomer of (α,β)-dehydro-phenylalanine (designated ΔPhe; (9)). (*C*) Stick model of ΔPhe as an isolated amino acid capped by an N-terminal acetyl group and C-terminal amide group. (*D*) *Ab initio* QM model of the electrostatic surface of the capped ΔPhe molecule. Extended conjugation of aromatic π electrons with the double bonds Cα=Cβ and carbonyl C=O enforces the near-planarity of this non-standard residue. Isosurface representation of electron density and MEP maps were calculated as described in ***Supplementary Text*** above.

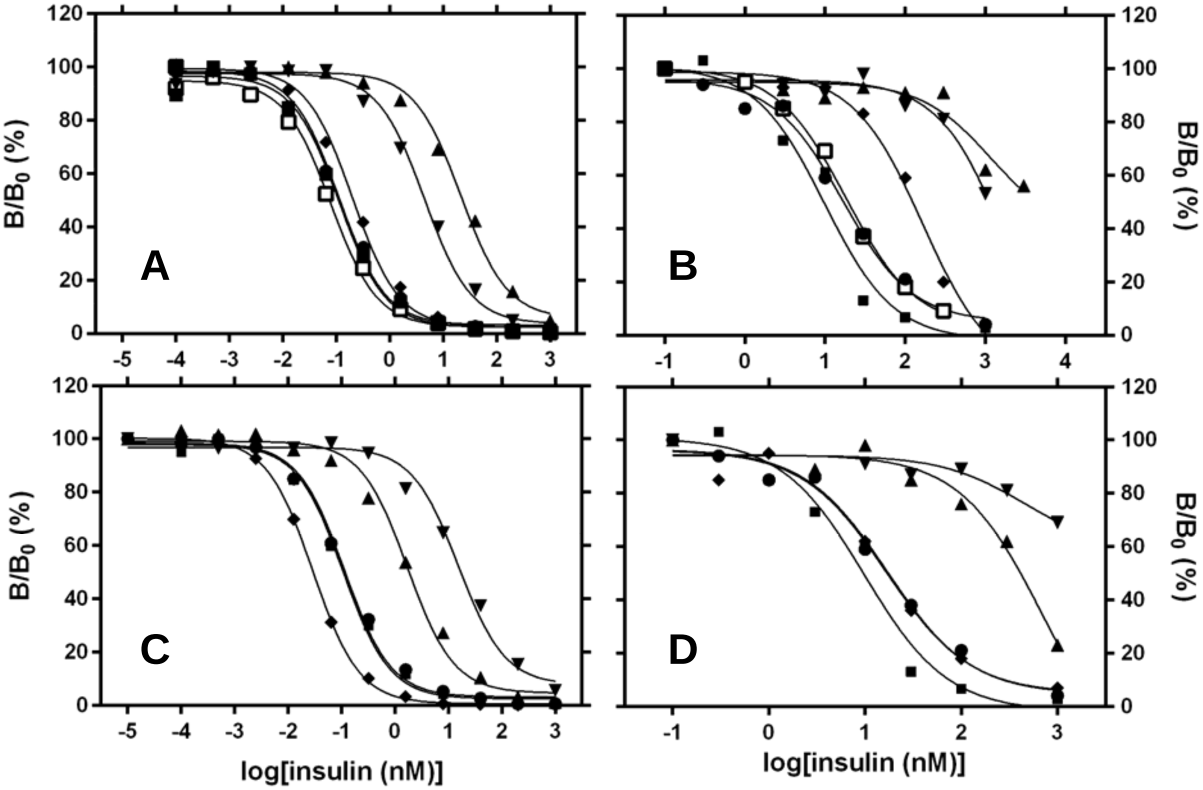

Figure S2. Comparative receptor-binding studies of insulin analogs. (*A*) Holoreceptor assays. Comparison of native insulin (■; K_d=0.07±0.008 nM), POT-insulin (□; K_d=0.05±0.007 nM) , and KP-insulin (●; K_d=0.07±0.01 nM) to Ala-scanning analogs of POT-insulin¶; Ala^{B24} (▲; K_d=14±2 nM), and Ala^{B25} (▼; K_d=3±0.4 nM), and Ala^{B26} (◆; K_d=0.13±0.02 nM). (*B*) µIR assembly assays with similar trends in relative stabilities (same symbols). These data provide evidence that Phe^{B24}, Phe^{B25}, and to a lesser extent, Tyr^{B26} contribute to the stability of the µIR ternary complex in accordance with their respective roles in the binding of insulin to holoreceptor. (*C*) Holoreceptor assays. Comparison of native insulin (■; K_d=0.07±0.008 nM) and KP-insulin (●; K_d=0.07±0.01 nM) with high-affinity Asp^{B10}-related analog DKP-insulin (◆; K_d=0.02±0.003 nM) *versus* low-affinity analogs Ala^{B12} (▲; K_d=0.8±0.11 nM) and single-chain 53-residue porcine insulin precursor (PIP) (▼; K_d=9.2±1.5 nM). (*D*) µIR assays with similar trends among the low-affinity analogs but lack of Asp^{B10}-related enhancement of stability (same symbols). These data provide evidence that (i) Val^{B12} contributes to the stability of the µIR ternary complex in accordance with its role in the binding of insulin to holoreceptor, (ii) substitution His^{B10}→Asp (putative Site 2 residue) stabilizes holoreceptor but has no effect on µIR, and (iii) tethering B30 to A1 *via* a two-residue linker (sequence AK) in PIP markedly impairs µIR assembly in accordance with impaired binding of the single-chain insulin analog to holoreceptor.

¶POT designates Pro^{B28}-Orn^{B29}-Thr^{B30} wherein the wild-type lysine at position B29 of wild-type human insulin is substituted by ornithine (human insulin sequence Pro^{B28}-Lys^{B29}-Thr^{B30}). The purpose of this modification is to eliminate a tryptic site and so facilitate semi-synthesis (see ***Supplementary Text****). The Orn^{B29} substitution does not affect activity.*

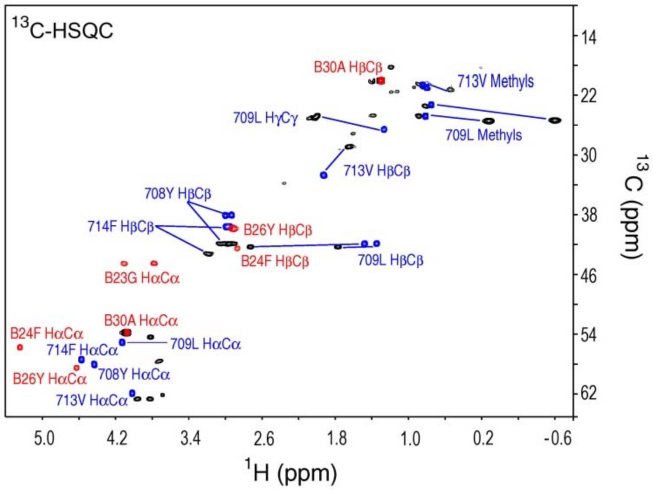

Figure S3. NMR analysis of unlabeled IR310.T complexed with labeled KP-porcine insulin and labeled αCT peptide ("complex 1"). [^{1}H-^{13}C] HSQC spectrum of selected aliphatic resonances in the μIR complex (*black*) is shown relative to [^{1}H-^{13}C] HSQC spectra of KP-porcine insulin (*red*) and free αCT (*blue*). Labeled aromatic sites in the insulin analog were Phe^{B24} and Tyr^{B26}; labeled aromatic sites in αCT were Tyr708 and Phe714 as indicated. Note marked upfield complexation shifts of Leu709 resonances due to neighboring aromatic ring currents in the μIR complex. As expected, the complexation shifts of Ala^{B30} are negligible.

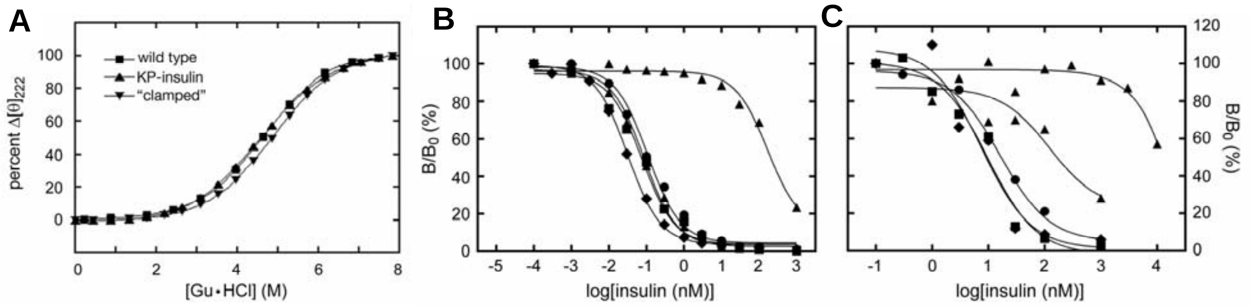

Figure S4. Characterization of insulin analogs. (*A*) Protein denaturation was monitored by CD at 222 nM as a function of the concentration of guanidine hydrochloride: (■) native human insulin, (▲) KP-insulin, and (▼) [D-Ala^{B20}, D-Ala^{B23}]-KP-insulin in which the D-amino acids were inserted as a chiral "clamp" within the B20-B23 β-turn. Inferred free energies of unfolding (ΔG_u) at 25 °C are (native insulin) 3.3±0.1 kcal mol^{-1}, (KP-insulin) 2.8±0.1 kcal mol^{-1}, and ([D-Ala^{B20}, D-Ala^{B23}]-KP-insulin) 3.8±0.1 kcal mol^{-1}. The paired D-Ala substitutions in the β-turn thus augment stability by 1(±0.2) kcal mol^{-1}. (*B*) Holoreceptor assays. Comparison of native insulin (■; K_d=0.07±0.008 nM) and KP-insulin (●; K_d=0.07±0.01 nM) with ΔPhe^{B24}-KP-insulin (▲; K_d=0.05±0.008 nM), ΔPhe^{B25}-KP-insulin (▼; K_d=124.5±28.6 nM), and [D-Ala^{B20}, D-Ala^{B23}]-insulin (◆; K_d=0.03±0.006 nM). (*C*) μIR assays exhibit similar trends to holoreceptor (*B*) except that ΔPhe^{B24} is associated with impaired assembly (but to a lesser extent than ΔPhe^{B25}). The D-Ala^{B20}, D-Ala^{B23} analog exhibits enhanced binding in to holoreceptor but resembles native insulin in relation to μIR assembly. The symbols in (*C*) are the same as in **(*B*)**. These data demonstrate that the paired substitutions [D-Ala^{B20}, D-Ala^{B23}] (designed as a "chiral lock" of the B20-B23 β-turn; see ***main text***) stabilizes the free hormone in accordance with the enhanced affinity of [D-Ala^{B20}, D-Ala^{B23}]-insulin for holoreceptor (see **Table S2**). ΔPhe^{B25} impairs assembly of the μIR ternary complex in accordance with its destabilization of the holoreceptor complex.

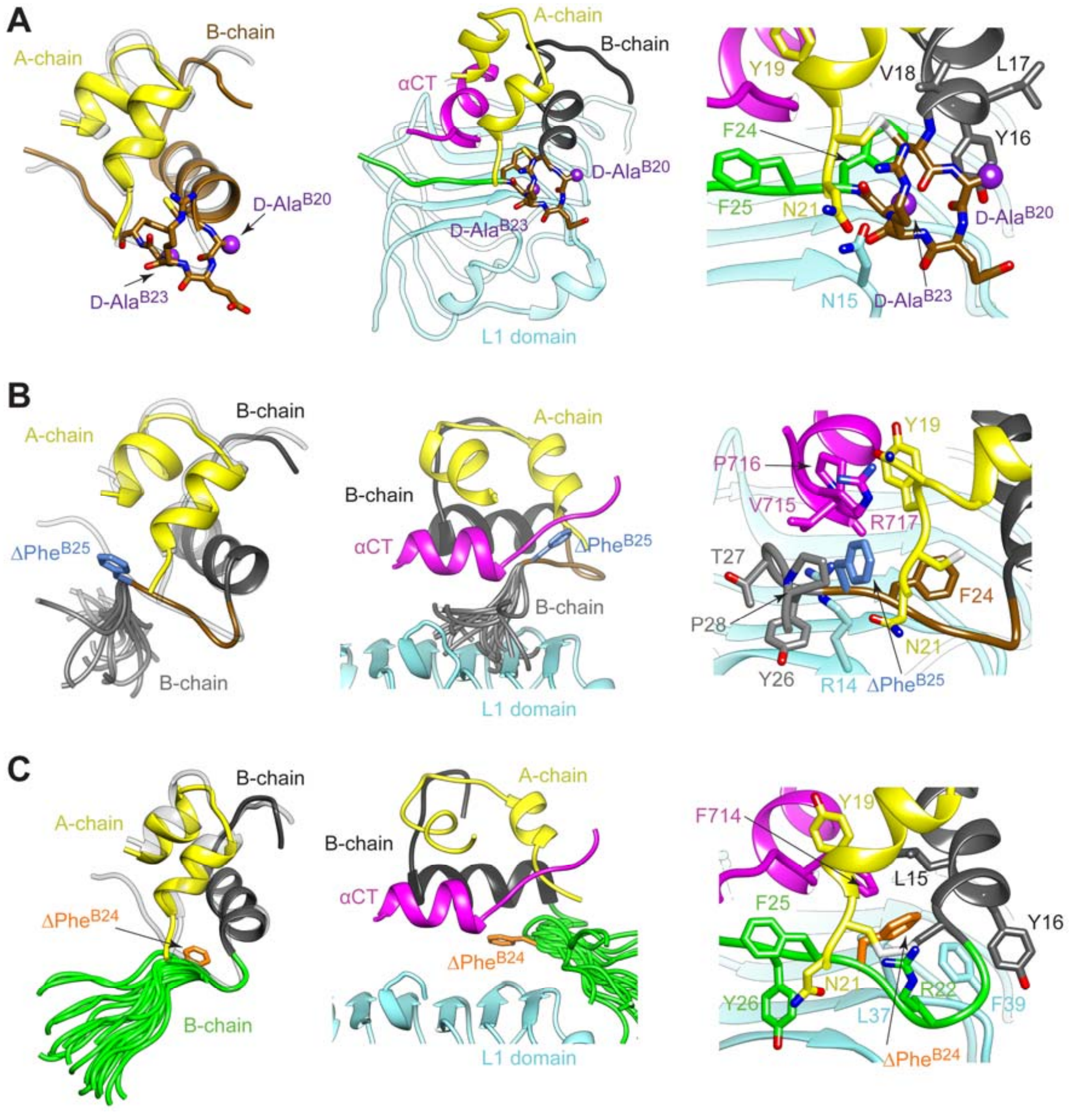

Figure S5. Structure-function relationships of non-standard insulin analogs. (*A*) Analog 1: [D-Ala^{B20}, D-Ala^{B23}]-insulin. *Left:* Crystal structure of free analog overlaid on wild-type insulin (*light gray*). Analog segment B19-B24 is shown in stick representation; respective C$_{\beta}$ atoms of D-Ala^{B20} and D-Ala^{B23} are shown as *purple* spheres. *Center:* model of the analog bound to Site 1. *Right:* environment of the variant B20-B23 segment within the μIR model. (*B*) Analog 2: ΔPhe^{B25}-insulin. *Left:* Solution structure relative to wild-type insulin (*light gray*). *Center:* models of the analog bound to Site 1, indicating that the ΔPhe^{B25} substitution results in residues B26-B27 (and consequently B28-B30) being directed towards the L1-β$_{2}$ surface. *Right:* detail of the environment of ΔPhe^{B25} within the μIR model. (*C*) Analog 3: ΔPhe^{B24}-insulin. *Left:* predicted structure of the analog relative to native insulin (*light gray*). *Center:* models of analog bound to the μIR, indicating its ability to adopt a conformation similar to that of native insulin bound to Site 1. *Right:* environment of ΔPhe^{B24} within the μIR model. Models in center and right panels of *(A-C)* are derived from the insulin-complexed μIR structure presented in **Fig. 4**; structure of native insulin is that of PDB entry 4INS.

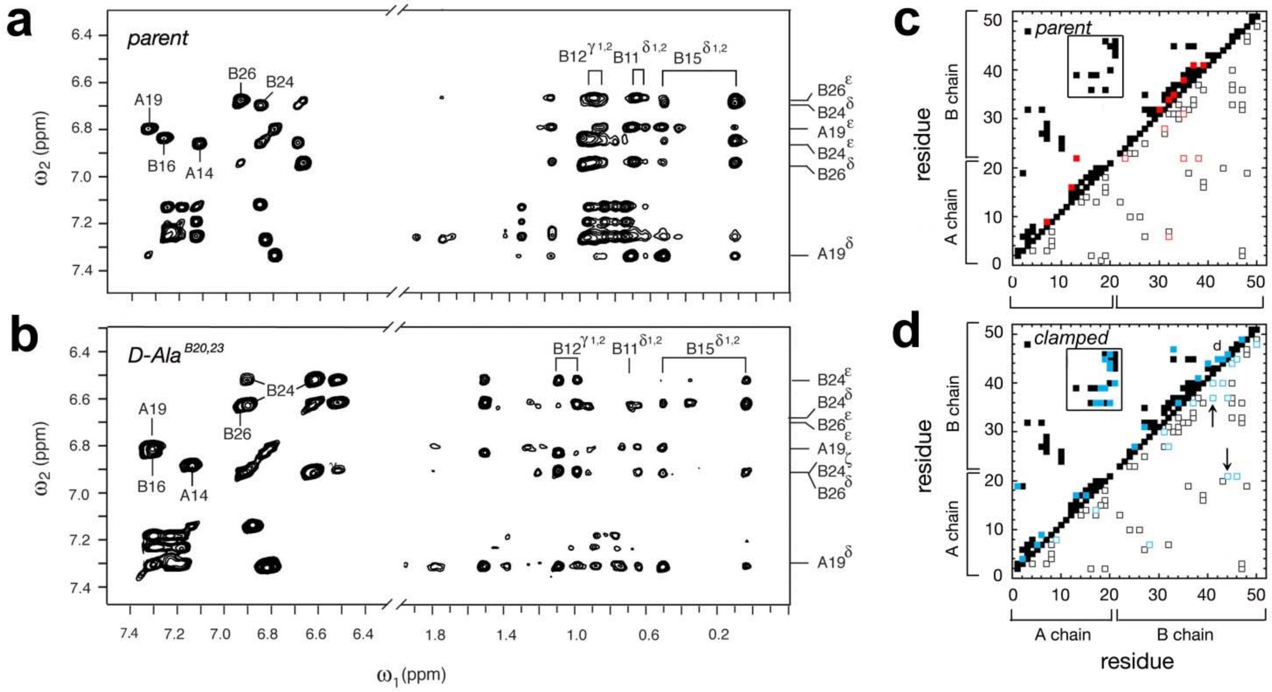

Figure S6. NMR characterization of [D-Ala^{B20}, D-Ala^{B23}]-DKP-insulin. (*A*) and (*B*): ^{1}H-NMR NOESY spectra of DKP-insulin (parent) and [D-Ala^{B20}, D-Ala^{B23}]-DKP-insulin, respectively. Aromatic spin systems (TOCSY spectrum at *left* in each panel) and NOEs involving aliphatic resonances (NOESY spectrum at *right* in each panel). Spectra of parent and [D-Ala^{B20}, D-Ala^{B23}]-DKP-insulin are similar, with subtle enhancement of chemical-shift dispersion in the D-Ala-stabilized analog; this is especially marked in the aromatic region. Spectra were acquired in D$_2$O at pD 7.6 (direct meter reading) and 32 °C; TOCSY and NOESY mixing times were 55 ms and 200 ms, respectively. (*C*) and (*D*): Pattern of inter-residue NOE contacts in DKP-insulin (parent) and [D-Ala^{B20}, D-Ala^{B23}]-DKP-insulin, respectively. Diagonal plots summarize at *upper left* main-chain/main-chain and main-chain/side-chain NOEs (*filled squares*); and at *lower right*, side-chain/side-chain NOEs (*open squares*). *Red squares* in left diagonal plot indicate contacts in parent spectrum not observed in analog; *blue squares* in right diagonal plot indicate contacts in analog not observed in parent. Boxes regions (*upper left* in each panel) contain contacts between C-terminal α-helix of A-chain (residues A12-A20; horizontal axis) and α-turn-β super-secondary structure of the B-chain (residues B9-B28; vertical axis). Arrows indicate positions B20 and B23.

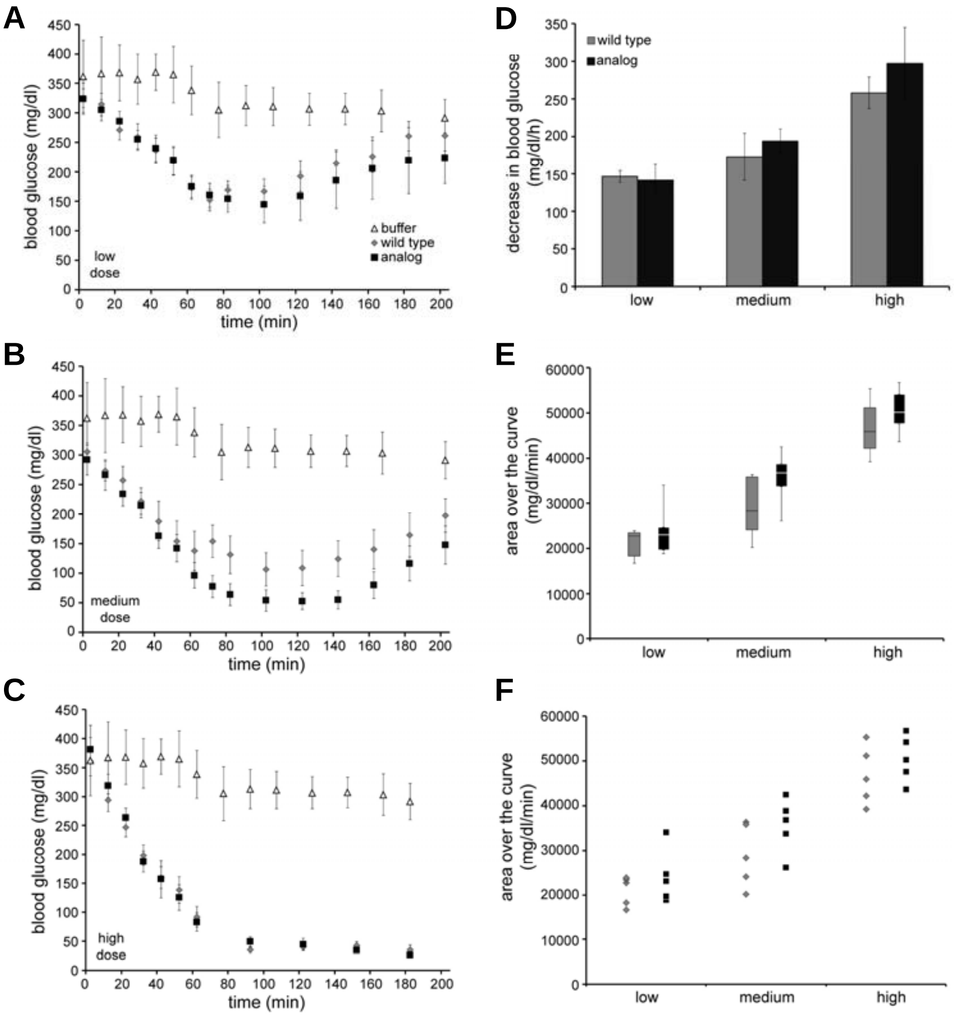

Figure S7. Rodent studies of [D-Ala^{B20}, D-Ala^{B23}]-KP-insulin in the treatment of diabetes mellitus. (*A*-*C*) Time course of blood glucose concentration following subcutaneous injection of native human insulin (◆) or the turn-stabilized analog ([D-Ala^{B20}, D-Ala^{B23}]-KP-insulin; ■) at three doses: 6.7 μg/rat (*A*), 20 μg/rat (*B*) and 36 μg/rat (*C*). Mean initial values of blood glucose (*N*=5 for each group) were similar in each set of rats; error bars represent standard deviations. Results of control injection of diluent are also shown (Δ) as performed at a different date using rats of similar body mass and degree of hyperglycemia. (*D*) Histogram showing initial rate of decline and standard error of the mean among the two groups of diabetic rats at three doses: 6.7 μg ("low dose"), 20 μg ("medium dose"), and 36 μg ("high dose"). *Gray* and *black bars* pertain to native insulin and the [D-Ala^{B20}, D-Ala^{B23}] analog, respectively. At medium and high doses the analog exhibited a trend toward greater initial rate of fall in blood glucose concentration but differences were not significant at *p*=0.05. (*E,F*) Overall potencies as measured by integrated areas above the curves shown in (*A*-*C*) relative to a horizontal line at the initial blood glucose concentration: (*E*) box plot and (*F*) actual data points (5 per analog at each dose). In box plots shaded rectangles delimit estimated upper and lower quartiles; bar indicates median; and thin lines designated minimum and maximum data points. Differences in potency at medium dose are near the threshold of statistical significance (*p*=0.07).

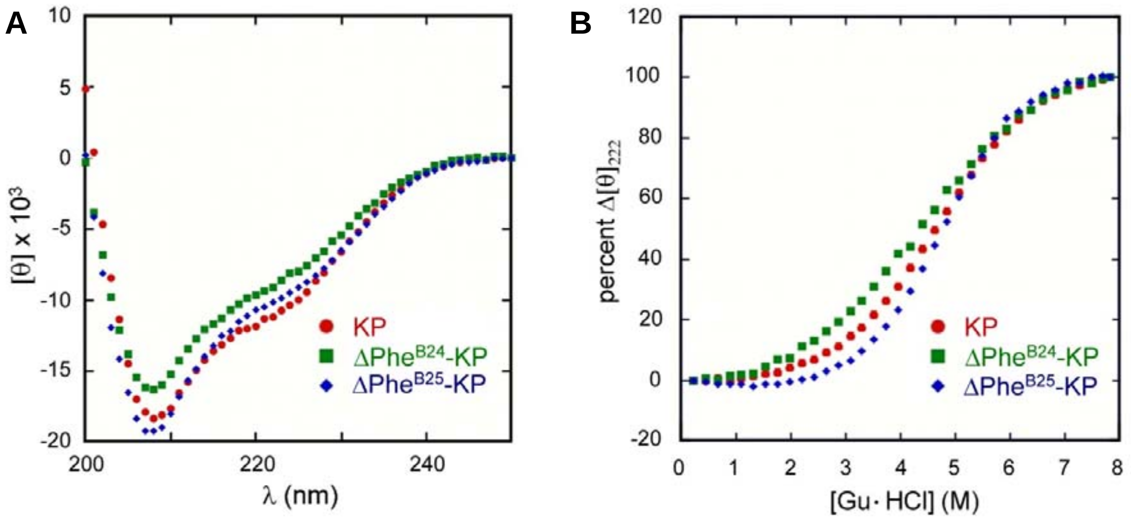

Figure S8. Biophysical characterization of ΔPhe-containing insulin analogs. (*A*) Far UV-CD spectra of ΔPhe-KP-insulin analogs. Spectra are of (●) KP-insulin, (■) ΔPhe^{B24}-KP-insulin, and (◆) ΔPhe^{B25}-KP-insulin. Spectra acquired at 25 °C are normalized to provide mean residue ellipticities ([θ]). (*B*) Thermodynamic stabilities of ΔPhe-KP-insulin analogs *vs* KP-insulin. Unfolding transitions are of (●) KP-insulin, (■) ΔPhe^{B24}-KP-insulin, and (◆) ΔPhe^{B25}-KP-insulin. CD-detected unfolding transitions induced by guanidine hydrochloride are plotted as percent change in mean residue ellipticity at helix-sensitive wavelength 222 nM. Data were acquired at 25 °C. Inferred thermodynamic parameters and other two-state model parameters are given in **Table S4**.

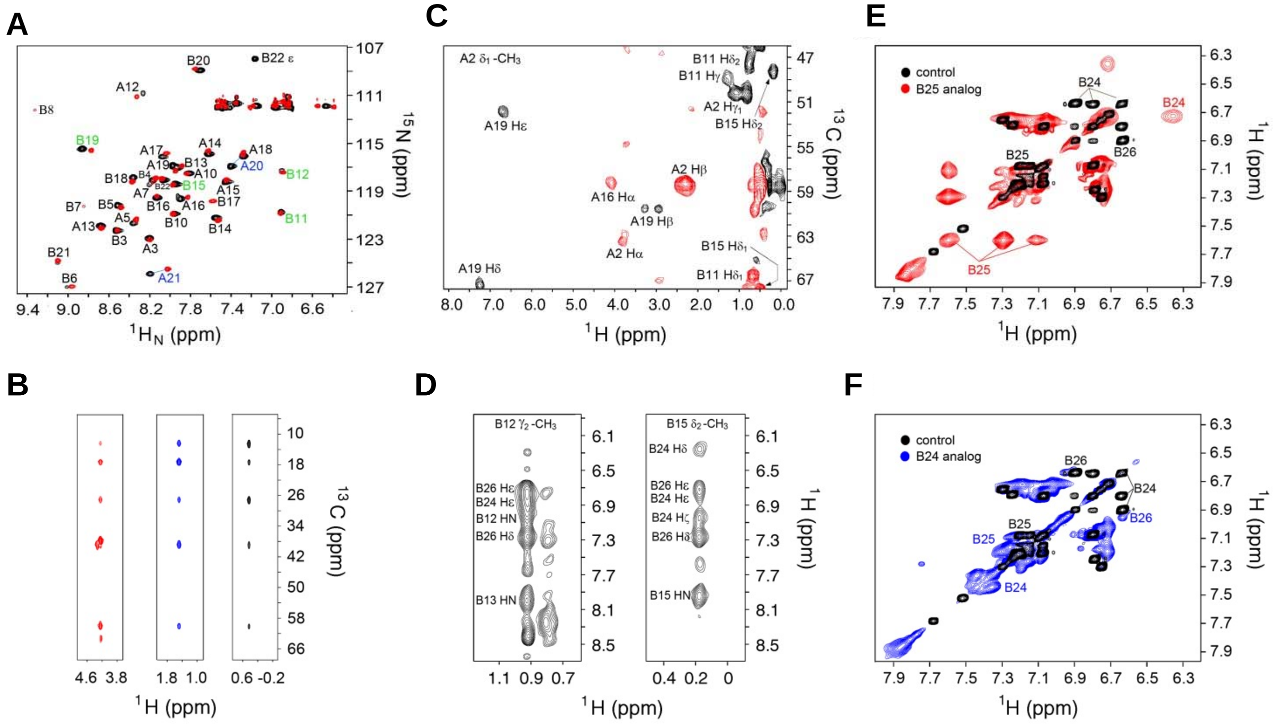

Figure S9. NMR studies of ΔPhe^{B25}-DKP-insulin. (*A*) ^{1}H,^{15}N-HSQC "fingerprint" spectra. Cross peaks reflect uniform isotopic labeling of ΔPhe^{B25}-DKP-insulin (*red*) and DKP-insulin (*black*) in residues A1-A21 and B1-B22 (*i.e.*, excepting C-terminal segment B23-B30; see ***SI Text***). Highlighted in *blue* and *green* are sites close to the B24/B26 aromatic rings; *blue* indicates large chemical-shift perturbation. (*B*) Heteronuclear spin-system identification in ΔPhe^{B25}-DKP-insulin. Strip plot of HCCH-TOCSY (HCC) derived from Ile^{A2} in ΔPhe^{B25} -DKP-insulin. Red indicates cross peaks from the Cβ of Ile^{A2} whereas *blue* indicates cross peaks from the Cβ of Ile^{A2}, and *black* indicates cross peaks from the $\delta 2$ methyl group of Ile^{A2}. (*C*) NOEs between ^{13}C-labeled residues and ^{13}C-labeled residues derived from 4D [^{13}C, ^{13}C]-edited NOESY spectrum for Ile^{A2}. (*D*) NOEs between ^{13}C-labeled methyl groups of Val^{B12} and unlabeled aromatic rings in the B24-B26 segment (left panel); NOEs between ^{13}C-labeled methyl groups of Leu^{B15} methyl group and these unlabeled aromatic rings (right panel). (*E*) Overlay of the spectra of ΔPhe^{B25}-KP-insulin (*red*) and KP-insulin (*black*). (*F*) Overlay of the spectra of ΔPhe^{B24}-KP-insulin (*blue*) and KP-insulin (*black*). The spectrum of the active B24 analog is broader than that of KP-insulin or the inactive B25 analog with more limited chemical-shift dispersion.

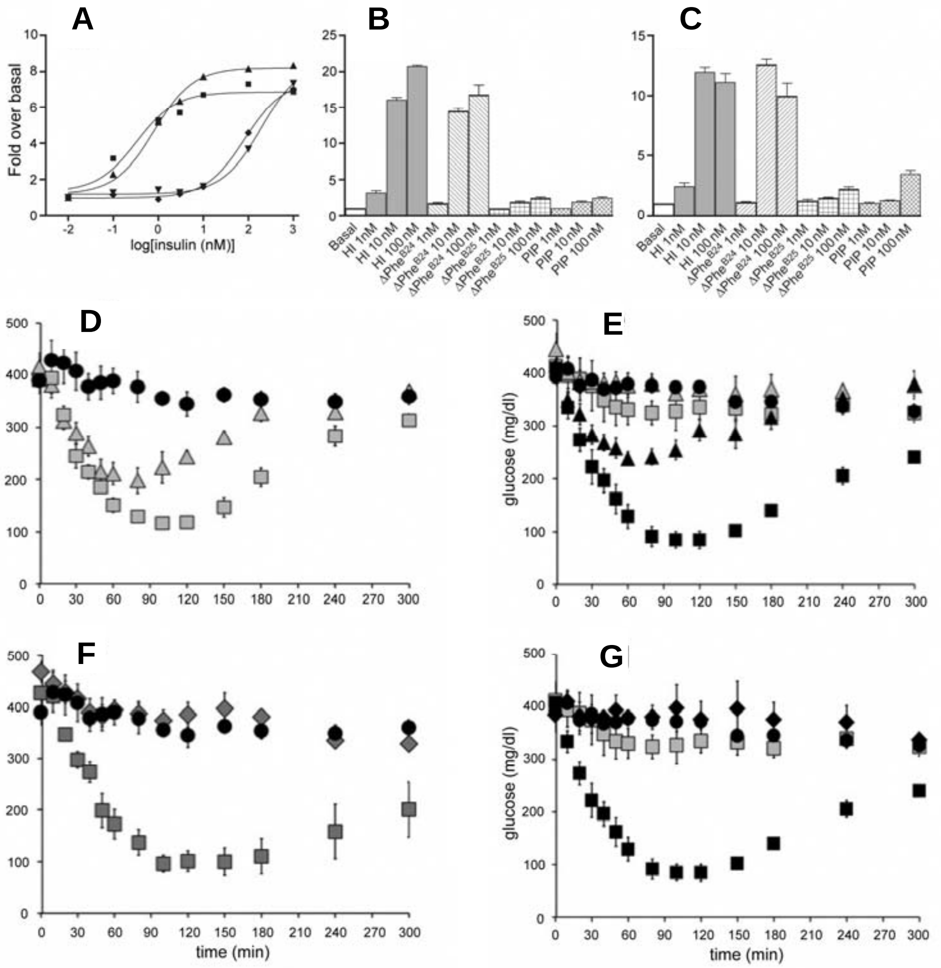

Figure S10. Analogs 2 and 3: tyrosine kinase activation, cellular studies and biological activity in rats. (*A*) Fold-increase in autophosphorylation of the B isoform of the insulin receptor (IR-B) on binding of (■) native human insulin, (▲) ΔPhe^{B24}-KP-insulin, (▼) ΔPhe^{B25}-KP-insulin, or (◆) porcine insulin precursor PIP. A shift to the right indicates impaired TK activation. (*B*) Hormone-stimulated autophosphorylation of IR-B in mouse fibroblasts engineered to express IR-B but not IR-A or IGF-1R. (*C*) Hormone-stimulated phosphorylation of AKT in mouse fibroblasts engineered to express IR-B but not IR-A or IGF-1R. Concentrations of ligands in (*B*) and (*C*) are as indicated; PIP indicates inactive single-chain analog of porcine insulin (53 residues with linker Ala-Lys; see ref. (81)). Time course of blood glucose concentration following (*D* and *F*) subcutaneous (SQ) injection of insulin analogs or (*E* and *G*) intravenous (IV) injection of insulin analogs. (*D*) Effects of SQ ΔPhe^{B24}-KP-insulin (▲,▲) *versus* KP-insulin (■,■) at two doses: 5 μg/rat (*gray symbols*) and 50 μg/rat (*black symbols*). Error bars represent standard deviations. (*E*) Effects of ΔPhe^{B24}-KP-insulin (▲) and KP-insulin (■) at a single dose of 10 μg/rat given IV bolus. (*F*) Effects of SQ ΔPhe^{B25}-KP-insulin (◆; high-dose only) *vs* KP-insulin at high- and low doses (■,■), defined as 5 μg/rat (*gray symbol*) and 50 μg/rat (*black symbols*). (*G*) Effects of ΔPhe^{B25}-KP-insulin (◆) and KP-insulin (■) at a single dose of 20 μg/rat given IV bolus. In each study, mean initial values of blood glucose (*N*=6 for each group) were similar in each set of rats; error bars represent standard deviations and a control administration of buffer (Lilly diluent) is shown by filled circles (●).

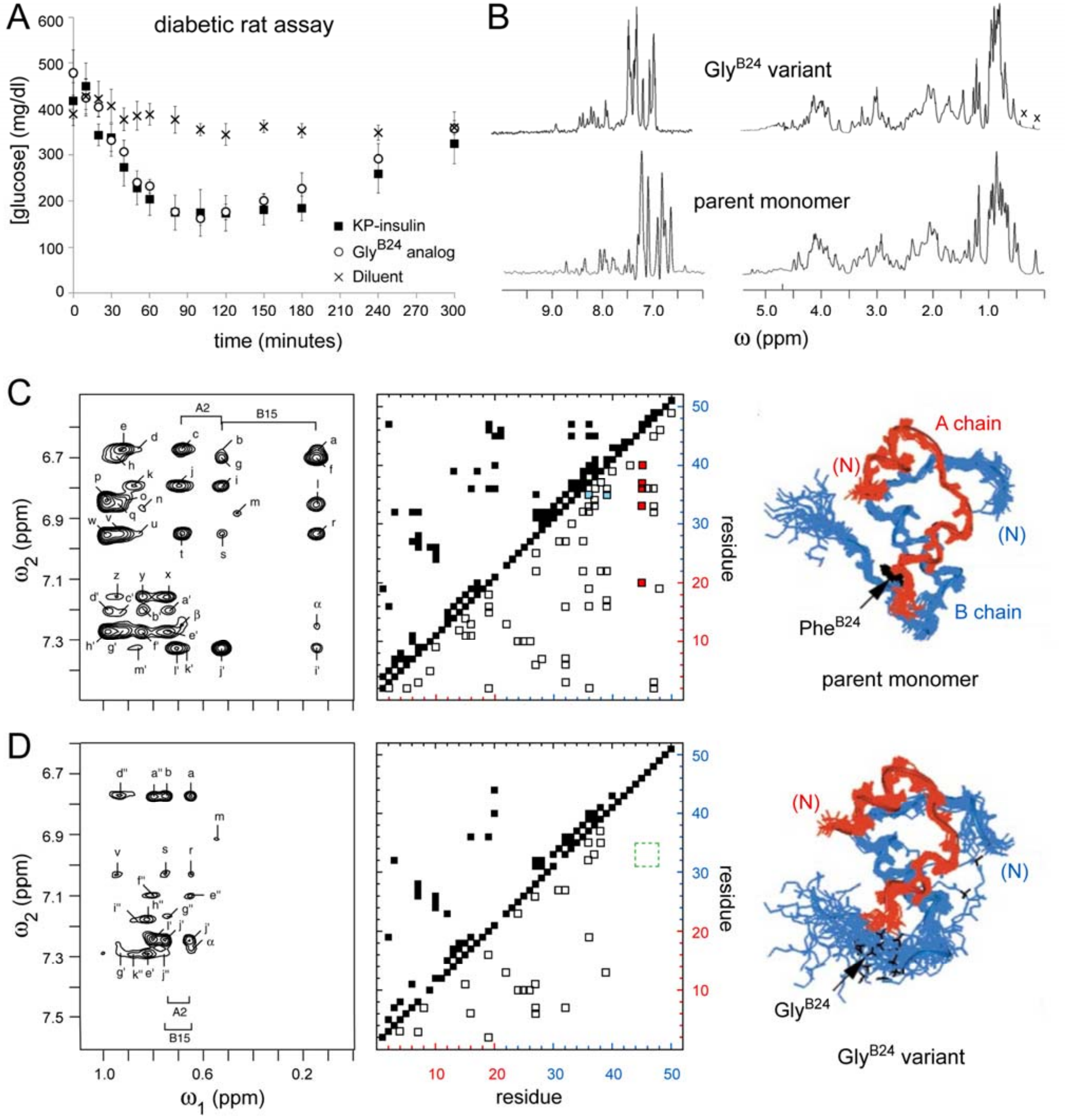

Figure S11. Studies of Gly^{B24} insulin analogs. (*A*) Biological activity of a Gly^{B24} insulin analog in a rat model of DM (18),is similar to that of the parent Phe^{B24} analog; the studies employed male Sprague-Dawley rats rendered diabetic by streptozotocin fasted overnight and then injected with 20 μg of the insulin analogs into the tail vein at *t*=0 min. Symbols: (■) KP-insulin (also designated insulin *lispro*), (**O**) Gly^{B24}-Orn^{B29}-insulin, and (**✕**) buffer control employing Lilly protein-free diluent. (*B*) Comparison of ^{1}H-NMR spectra of the engineered insulin monomer DKP-insulin (45) and variant Gly^{B24}-DKP-insulin at 25 ^{o}C in aqueous D_2O solution at pD 7.0 (direct meter reading); spectra were obtained at 600 MHz with mixing time 200 ms. The aromatic region of the spectrum is shown at *left,* and aliphatic region at *right.* Spectrum of the Gly^{B24} variant exhibits a marked reduction in chemical-shift dispersion, including absence of upfield-shifted Leu^{B15} spin system. Non-protein resonances at right are indicated by “x”. (*C* and *D*) Comparative 2D-NMR structural analyses of DKP-insulin and Gly^{B24}-DKP-insulin. Each panel contains three parts: *left,* portion of NOESY spectrum in D$_2$O containing long-range contacts between aromatic protons (vertical axis; ω_2) and aliphatic protons (horizontal axis; ω_1); *center,* summary of inter-residue NOE contacts by diagonal plot (residues 1-21 represent the A chain (*red* scales), and residues 22-51 the B chain (*blue* scales); and *right,* DG/RMD ensembles showing the A chain in red and B chain in blue. In the diagonal plots, contacts between main-chain protons are shown at upper left whereas contacts between side chains or between side-chain and main-chain protons are shown at lower right. In (*C*), side-chain contacts involving Leu^{B15} and Phe^{B24} are highlighted in blue and red, respectively; in (*D*) the absence of key long-range NOEs is indicated by empty box (*green dashed* borders; inset at lower right). The control ensemble of DKP-insulin was calculated based on corresponding homonuclear ^{1}H-NMR spectra; see **Table S7** for restraint information and statistical parameters. Small differences between diagonal plots in panel *C* (*right*) versus **Fig. S6*C*** are due to our efforts to match spectra of respective variants versus parent in each case in relation to experimental conditions (protein concentration, pH, temperature, and magnetic field strength). The NOESY spectra shown were acquired at 600 MHz at pD 7.6 and 32 ^{o}C with mixing time 200 ms.

The solution structure of Gly^{B24}-DKP-insulin in aqueous solution at neutral pH is less disordered than that of GlyB24-insulin as characterized in 20% deuterioacetic acid (pH 1.9) but nonetheless exhibits marked destabilization of the C-terminal B-chain β-turn and β-strand (82). Inter-residue NOEs were not observed in the spectrum of the Gly^{B24} variant indicative of a reported structural switch in which Phe^{B25} takes the place formerly occupied by Phe^{B24} in the native hydrophobic core (83), presumably due to the alternative (and less active) monomeric template employed in the prior study (Gly^{B24}-Glu^{B16}-insulin). The PDB accession code for the solution structure of Gly^{B24}-DKP-insulin is 2MPI.

Assignments in NOESY spectra. Panel *C (left)*: (a) B26-Hε/B15-δ_2-CH$_3$ (b) B26-Hε/B15-δ_1-CH$_3$ (c) B26-Hε/B11-δ_1-CH$_3$ (d) B26-Hε/A3-γ_2-CH$_3$ (e) B26-Hε/A3-γ_1-CH$_3$ — B26-Hε/B12-γ-CH$_3$ (f) B24-Hδ/B15-δ_2-CH$_3$ (g) B24-Hδ/B15-δ_1-CH$_3$ (h) B24-Hδ/B15-Hβ2 (i) A19-Hε/A2-δ-CH$_3$ (j) A19-Hε/A2-γ'-CH$_3$ (k) A19-Hε/A3-γ_2-CH$_3$ — A19-Hε/A2-Hγ_2 (l) B24-Hε/B15-δ_2-CH$_3$ (m) B5-Hδ/A10-δ-CH$_3$ (o) B16-Hε/B17-δ_2-CH$_3$ (q) B24-Hε/B15-β_2-CH$_3$ (p) B16-Hε/B17-δ_1-CH$_3$ — B5-Hδ/A10-H$\gamma_{1,2}$ (s) B26-Hδ/B15-δ_1-CH$_3$ (t) B26-Hδ/B11-δ_1-CH$_3$ (u) B26-Hδ/A3-γ_2-CH$_3$ (w) B26-Hδ/B12-γ_1-CH$_3$ (x) B1-Hδ/A13-δ_2-CH$_3$ (y) B1-Hδ/A13-δ_1-CH$_3$ (i') A19-Hδ/B15-δ_2-CH$_3$ (j') A19-Hδ/B15-δ_1-CH$_3$ — A19-Hδ/A2-δ-CH$_3$ (l') A19-Hδ/A16-δ_1-CH$_3$ (g') B1-Hε/B17-δ_2-CH$_3$ (h') B1-Hε/B18-γ_1-CH$_3$ — A19-Hδ/A2-Hγ_2. Panel *D* (*left*): (a) B26-Hε/B15-δ_2-CH$_3$ — A19-Hε/B15-δ_2-CH$_3$ (b) B26-Hε/B15-δ_1-CH$_3$ — A19-Hε/B15-δ_1-CH$_3$ — A19-Hε/A2-γ'-CH$_3$ (a") A19-Hε/A16-δ-CH$_3$ (d") B26-Hε/B12-γ-CH$_3$ — A19-Hε/A2-Hγ (m) B5-Hδ/A10-δ-CH$_3$ (r) B26-Hδ/B15-δ_2-CH$_3$ (s) B26-Hδ/B15-δ_1-CH$_3$ (v) B26-Hδ/B12-γ-CH$_3$ (e") B25-Hδ/B15-δ_2-CH$_3$ (f") A14-Hε/A16-δ-CH$_3$ (g") B5-Hδ/B6-δ_2-CH$_3$ (h") B1-Hδ/B2-γ-CH$_3$ (i") B1-Hδ/B6-Hβ_2 (j') A19-Hδ/A2-δ-CH$_3$ (j') A19-Hδ/B15-δ_1-CH$_3$ (l') A19-Hδ/A16-δ-CH$_3$ (α) B25-Hε/B15-δ_2-CH$_3$ (j") B1-Hε/B6-δ-CH$_3$ (e') B1-Hε/B2-γ-CH$_3$ — B1-Hε/A13-δ-CH$_3$ (k") B1-Hε/B6-Hβ_2 (g') B1-Hε/B17-δ-CH$_3$.

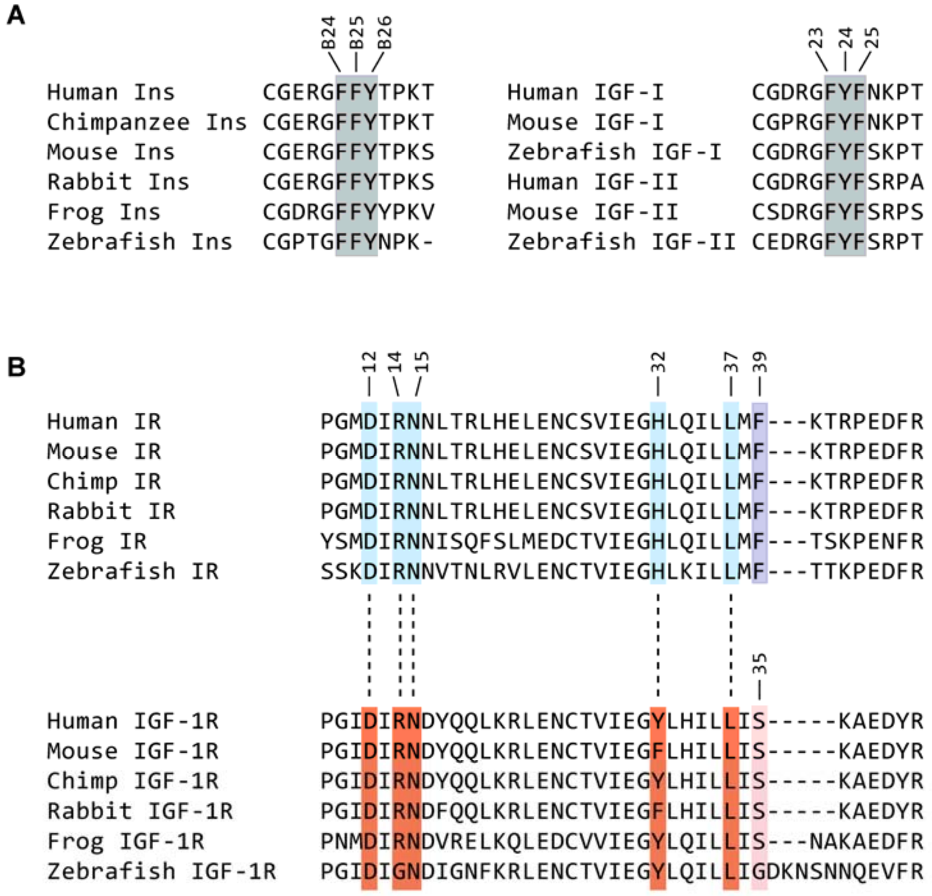

Figure S12. Sequence and structural comparison of insulins, IGFs and their receptors. (*A*) Alignment of representative insulin and IGF sequences, showing the conservation of the aromatic triplet Phe^{B24}-Phe^{B25}-Tyr^{B26} and their counterpart in IGFs. (*B*) Alignment of representative sequences of insulin receptor and IGF-1R, showing the conservation of residues involved in interaction with the aromatic triplet as observed in the structure of the μIR complex presented in the ***main text***. In both (*A*) and (*B*), residues N- and C-terminal to the aligned sequences have been omitted for brevity; the numbering above each alignment block is that of the top sequence.

Table S1. Receptor-binding studies

Analog	**Holoreceptor**		**μIR**	
	K_d (nM)	rel. affinity (%)	K_d (nM)	rel. affinity (%)
native insulin	0.08 ± 0.01	100	7.5 ± 1.6	100
KP-insulin[*]	0.07 ± 0.01	*ca* 100	11.4 ± 3.0	*ca* 70
DKP-insulin[†]	0.02 ± 0.01	*ca* 400	5.7 ± 1.5	*ca* 130
Orn^{B29}-insulin[‡]	0.05 ± 0.01	*ca* 160	4.0 ± 1.6	*ca* 190
Ala^{B24}-Orn^{B29}-insulin	13.9 ± 2.3	*ca* 1	> 10^2 (*n.f.*)[§]	< 0.1
Ala^{B25}-Orn^{B29}-insulin	3.0 ± 0.4	*ca* 3	> 10^2 (*n.f.*)	< 0.1
Ala^{B26}-Orn^{B29}-insulin	0.13 ± 0.02	*ca* 60	79 ± 21	*ca* 10
single-chain ins. (PIP)[¶]	9.2 ± 1.5	< 0.1	> 10^2 (*n.f.*)	< 0.1
Ala^{B12}-DKP-insulin	0.8 ± 0.1	*ca* 10	303 ± 8	*ca* 2
D-Ala^{B20}, D-Ala^{B23}-insulin	0.04 ± 0.01	*ca* 220	4.8 ± 1.2	*ca* 160
ΔPhe^{B24}-KP-insulin[‖]	0.05 ± 0.01	*ca* 150	33 ± 8	*ca* 20
ΔPhe^{B25}-KP-insulin	125 ± 29	< 0.01	>10^2 (*n.f.*)	< 0.1

[*] KP-insulin contains substitutions Pro^{B28}→ Lys and Lys^{B29}→Pro.

[†] DKP-insulin contains substitutions His^{B10}→Asp, Pro^{B28}→ Lys, and Lys^{B29}→Pro.

[‡] Orn^{B29}-insulin contains ornithine instead of lysine at position B29 to eliminate a tryptic site and thereby facilitate trypsin-mediated semi-synthesis using unprotected fragments.

[§] *n.f.* : not fitted

[¶] PIP (porcine insulin precursor) is a 53-residue single-chain insulin containing substitution Thr^{B29}→Ala (as in porcine insulin and two-residue "mini-C-domain" with sequence Ala-Lys. Whereas PIP exhibits a 100-fold decrease in receptor-binding affinity relative to porcine or human insulin, a 50-residue SCI in which Lys^{B29} is connected to Gly^{A1} by a peptide bond (with deletion of residue B30) exhibits no detectable activity (84). Inclusion of Thr^{B30} leads to a 51-residue SCI whose receptor-binding affinity is decreased by 1500-fold relative to its two-chain parent (85).

[‖] ΔPhe designates the *Z* isomer.

Table S2. ^{1}H-NMR chemical shifts in monomeric insulin in free and μIR-bound form

	Free form	**μIR-bound form**	**Random coil***
Insulin			
Gly^{B23}	4.09/44.44, 3.76/44.44	3.96/45.51	
Phe^{B24}	5.22/55.70	4.54/57.98	
	2.86/42.44	2.98/39.45	
	6.64/132.87, 6.81/130.60	7.37/131.80, 7.25/129.62	6.80/131.35, 6.85/130.52
	6.81/131.03	7.33/130.35	6.84/129.03
Tyr^{B26}	4.60/59.45	4.52/57.97	
	2.90/39.78	2.92/38.95	
	6.91/132.70, 6.62/118.01	7.14/133.43, 6.84/118.48	6.82/132.23, 6.58/117.30
Ala^{B30}	4.05/53.70	4.08/53.78	4.26/52.84
	1.29/19.95	1.30/20.06	1.38/19.06
αCT			
Y708	4.44/58.20	4.52/57.97	
	2.94/38.13, 3.01/38.13	2.96/41.91, 3.05/41.91	2.92/38.95
	7.05/133.14, 6.77/118.19	7.06/133.19, 6.77/118.18	6.82/132.23, 6.58/117.30
L709	4.13/55.24	4.36/54.92	
	1.35/41.96, 1.49/41.96	1.77/42.31, 2.73/42.31	1.62/42.38
	1.27/26.62	2.01/24.84	1.53/26.76
	0.76/23.30, 0.82/24.82	-0.60/25.37, 0.12/25.45	0.76/24.67, 0.77/24.27
V713	4.02/62.07	4.12/62.06	
	1.93/32.75	1.65/28.91	1.97/32.71
	0.85/20.63, 0.81/20.99	0.54/21.27, 1.19/18.27	0.82/21.11, 0.83/21.31
F714	4.58/57.55	4.54/57.98	
	2.99/39.70, 2.96/39.70	3.18/43.27	2.98/39.45
	7.17/131.39, 7.29/131.51,	7.18/131.35, 7.23/131.65	6.80/131.35, 6.85/130.52
	7.26/129.76	7.31/131.62	6.84/129.03

* The ^{1}H and ^{13}C chemical shifts of random coil residues were obtained from Sykes and coworkers (63, 64).

TABLE S3 Diffraction data and crystallographic refinement statistics for μIR complex

Data set		**Crystal 1 (=D_1)**	**Crystal 2**	**Crystal 3**	**D_1 + Crystal 2 (=D_2)**	**D_2 + Crystal 3 (=D_3)**
Space group		P23	P23	P23	P23	P23
a (Å)		169.04	169.87	169.08	169.04	169.04
Resolution (Å)		20.0-3.5 (3.6-3.5)[*]	20.0-3.6 (3.7-3.6)	20.0-3.8 (3.9-3.8)	20.0-3.5 (3.6-3.5)	20.0-3.5 (3.6-3.5)
Redundancy		17.1 (16.7)	14.7 (14.4)	10.9 (11.2)	30.6 (16.7)	39.2 (16.7)
R_{merge}		0.17 (14.1)	0.14 (10.0)	0.16 (15.9)	0.22 (14.8)	0.21 (14.8)
$\langle I/\sigma(I) \rangle$	3.9-3.8 Å	0.59	0.60	0.27	0.86	0.91
	3.8-3.7 Å	0.47	0.35	n/a[†]	0.55	0.55
	3.7-3.6 Å	0.48	0.33	n/a	0.54	0.54
	3.6-3.5 Å	0.21	n/a	n/a	0.17	0.17
	Overall	7.26	14.8	10.6	12.3	12.1
$CC_{1/2}$ (%)[‡]	3.9-3.8 Å	33.9	29.0	12.9	49.7	53.3
	3.8-3.7 Å	27.5	17.4	n/a	35.5	34.7
	3.7-3.6 Å	20.3	12.3	n/a	28.1	30.8
	3.6-3.5 Å	12.7	n/a	n/a	11.6	12.3
	Overall	99.8	100.0	99.9	99.9	99.9

Refinement	
Resolution (Å)	20.0-3.5
No. reflections	19092
R_{work} / R_{free} [§]	0.264 / 0.284
No. protein atoms	4552
No. carbohydrate atoms	141
$\langle B \rangle$ protein (Å^2)	237.
$\langle B \rangle$ carbohydrate (Å^2)	270.
σ_{bonds} (Å) / σ_{angles} (°)	0.010 / 1.3

[*] Numbers in parentheses refer to completeness in the outer resolution shell.

[†] n/a: Not applicable, as beyond the resolution cut-off of the relevant data set.

[‡] $CC_{1/2}$ is the Pearson correlation coefficient between the pairwise intensities for each reflection calculated from two random halves of the data set. All values listed are significant at at least the p=0.001 level of probability.

[§] Free set: 5% of reflections. R-factors are computed using the expectation of the model structure factor amplitude (37).

Table S4. Thermodynamic stabilities of insulin analogs[*]

Analog	**ΔG_u** (kcal mol^{-1})	**C_{mid}** (M)	**m** (kcal mol^{-1} M^{-1})
[D-Ala^{B20}, D-Ala^{B23}]-KP	3.8±0.1	5.2±0.1	0.78±0.02
KP-insulin	3.0±0.1	4.7±0.1	0.47±0.04
∆PheB^{B24}-KP	2.2±0.1	4.6±0.3	0.59±0.03
∆PheB^{B25}-KP	2.9±0.1	4.5±0.2	0.64±0.01
GlyB^{B24}-KP	2.2±0.1[†]	4.2±0.2	0.53±0.03

[*] Stabilities were inferred from CD-detected guanidine denaturation studies at 25 °C as described (40).

[†] Matched control guanidine titration of KP-insulin on this date yielded an estimate of ΔG_u of 2.9±0.1 kcal mol^{-1}, and so the ***main text*** cites a $\Delta\Delta G_u$ of 0.7±0.2 kcal mol^{-1}.

Table S5. Diffraction data processing and crystallographic refinement statistics for [D-Ala^{B20},D-Ala^{B23}]-insulin

Data Processing	
Wavelength (Å)	0.9537
Resolution range (Å)	18.39-1.30 (1.35-1.30)*
Space group	$I2_13$
a (Å)	78.03
Redundancy	11.3 (7.2)
Completeness (%)	99.8 (99.3)
R_{merge}	0.10 (3.34)
$<I/\sigma(I)>$	13.9 (0.5)
$CC_{1/2}$ †	0.999 (0.170)
Refinement	
Resolution range (Å)	18.39-1.40 (1.44-1.40)
No. reflections	15722 (1439)
R_{work} / R_{free}‡	0.141 / 0.163
No. atoms protein	793
No. atoms solvent	83
<B> protein (Å^2)	22.0
<B> solvent (Å^2)	36.0
σ_{bonds} (Å) / σ_{angles} (°)	0.009 / 1.07

* Numbers in parentheses refer to completeness in the outer resolution shell.

† $CC_{1/2}$ is the Pearson correlation coefficient between the pairwise intensities for each reflection calculated from two random halves of the data set. All values listed are significant the *p*=0.001 level of probability.

‡ Free set: 5% of reflections.

Table S6. NMR-derived restraints and structure calculations: ΔPhe^{B25}-DKP-insulin

Parameter	**SA ensemble**[*]
RMSD from experimental distance restraints (Å)	
all (931)	0.056 ± 0.012
intraresidue, *i*=*j* (294)	0.065 ± 0.015
sequential, |*i*-*j*|=1 (217)	0.048 ± 0.011
medium-range, 1<|*i*-*j*|<5 (192)	0.049 ± 0.011
long-range, |*i*-*j*|≥5 (228)	0.040 ± 0.009
Max. distance restraint violation (Å)	0.288
RMSD from dihedral-angle restraints (°)	
Total (86)	0.669 ± 0.153
Φ (39)	0.445 ± 0.107
Ψ (39)	0.678± 0.152
Max. dihedral angle violation (°)	3.454
RMSD from idealized covalent geometry	
bonds (Å)	0.010 ± 0.002
angles (°)	1.061 ± 0.237
Improper dihedral angles (°)	2.356 ± 0.527
Average pairwise r.m.s. deviation[†] (Å)	
RMSD of all heavy atoms	0.63 ± 0.11
RMSD of backbone atoms	0.29 ± 0.10

[*] Mean ± standard error where applicable; "SA" denotes *simulated annealing*.

[†] Residues A1-A21, B1-B27.

Table S7. Homonuclear ^{1}H-NMR-derived structures: Gly^{B24}-DKP-insulin versus DKP-insulin

	Gly^{B24}	**Phe^{B24}**
Restraint Information		
Distance restraints		
all	548	749
Intra-residue, $i=j$	212	194
sequential, $\|i-j\|=1$	158	153
medium-range, $1<\|i-j\|<5$	100	153
long-range, $\|i-j\|\geq 5$	78	184
Dihedral-angle restraints		
total	42	48
ϕ	24	30
ψ	18	18
Hydrogen-bond related restraints	19	24
Atomic RMSDs (Å)		
Main chain[*]		
α-helices[†]	0.43	0.31
A chain	0.49	0.32
B chain	0.44	0.36
Side chains		
α-helices[†]	0.90	0.65
A chain	0.96	0.67
B chain	0.85	0.75
Restraint Violations[‡]		
NOE violations (Å)	0.08	0.02
dihededral-angle violations (°)	0.97	0.23
Deviations from Ideal Geometry		
bond length (Å)	0.008	0.006
bond angle (°)	0.98	0.93
Empirical Energy Functions		
NOE restraint energy (kcal mol^{-1})	126±10	14±2
van der Waals (kcal mol^{-1})	13±6	10±2
improper dihedral angles (kcal mol^{-1})	7±1	9±2

[*] RMSD values were calculated according to Cα atoms of residues A2-A19 and B3-B19.

[†] Helical segments are defined as residues A2-A7, A13-A19 and B9-B19.

[‡] NOE and dihedral force constants were 40 kcal-Å^{-2} and 40 kcal-radian^{-2}, respectively (1 kcal = 4.18 kJ).