Supplemental Information

for

Deciphering the Hidden Informational Content of Protein Sequences

THE FOLDABILITY OF PROINSULIN HINGES ON A FLEXIBLE ARM THAT IS DISPENSABLE IN THE MATURE HORMONE

Ming Liu, Qing-xin Hua, Shi-Quan Hu, Wenhua Jia, Yanwu Yang, Sunil Evan Saith, Jonathan Whittaker, Peter Arvan*, and Michael A. Weiss*

Table of Contents

Purpose of Supplement

In Figure S1 are shown comparative 2D ^{1}H-NMR spectra of KP-insulin and KP-proinsulin, monomeric analogs of human insulin and human proinsulin, respectively. Spectra were observed in 10 mM deuterioacetic acid at pD 3.0. In Figure S2 are shown selected arm-related NOEs extracted from 3D ^{13}C-edited ^{1}H-^{1}H NOESY spectrum of DKP-proinsulin at pH 7.4. In Figure S3 is shown a superposition of T-state crystallographic protomers highlighting the variable positions of (and distance between) the aromatic rings of Phe^{B1} and Tyr^{A14}. ^{1}H-NMR resonance assignments of *des*-Phe^{B1}-insulin as a dimer in 10 mM deuterioacetic acid at pD 3.0 are provided in Table S1. Diagnostic main-chain ^{13}C-NMR chemical shifts and secondary shifts are given in Table S2-A in relation to mean values in canonical elements of secondry structure (α-helix and β-strand) in Table S2-B. Results of the application of a two-state thermodynamic model to CD-detected guanidine unfolding studies of insulin and des-B1-insulin are summarized in Table S3.

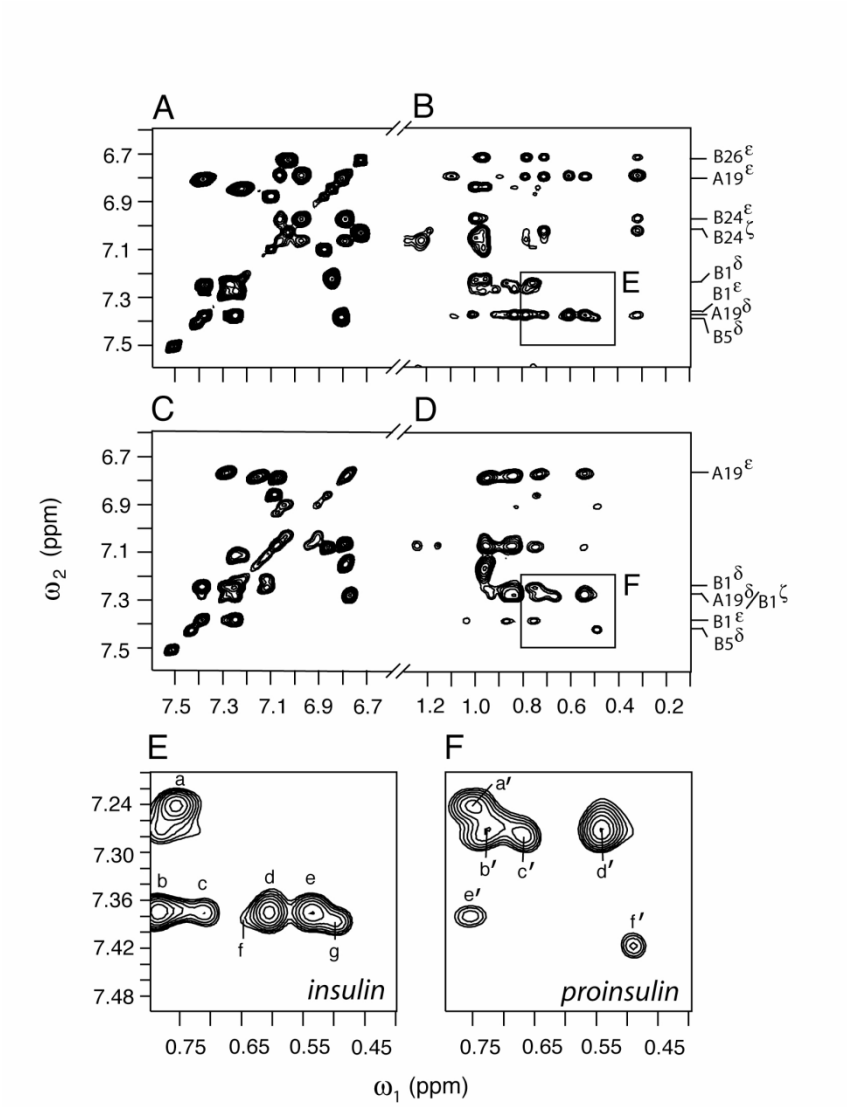

Figure S1. **N-terminal arm of proinsulin resembles T-state-specific features of insulin.** (*A* and *B*) ^{1}H-NMR spectra of KP-insulin as a monomer in 10 mM deuterioacetic acid (pH 3.0) and 25° C. The TOCSY spectrum of aromatic spin systems (*A*) enables assignment of inter-residue contacts to aliphatic resonances in NOESY spectrum (*B*). Selected aromatic assignments are given at right. (*C* and *D*) Corresponding spectra of KP-proinsulin under the same conditions. Selected aromatic assignments are given at right. (*E* and *F*) Expansion of corresponding boxes in panels *B* and *D*, respectively. TOCSY and NOESY mixing times were in each case 55 ms and 200 ms, respectively. "KP" designates substitutions Pro^{B28}→Lys and Lys^{B29}→Pro, which hinder dimerization and so enable NMR studies of monomeric proteins (se main text for references). Resonance assignments are as follows. In panel *E*: (a) B1H$_\delta$ – A13 δ-CH$_3$; (b) A19 H$_\delta$ – A16 δ-CH$_3$; (c) A19 H$_\delta$ – A2 γ'-CH$_3$; (d) overlap of

A19 H_{δ} – A2 δ-CH_3 and A19 H_{δ} – B15 δ_1-CH_3; (e) A19 H_{δ} – A2 $H_{\gamma 2}$; (f) B5 H_{δ} – A10 γ'-CH_3; and (g) B5 H_{δ} – A10 γ'-CH_3. In panel *F*: (a') B1 H_{δ} – A13 δ-CH_3; (b') A19 H_{δ} – A2 γ'-CH_3; (c') A19 HD – B15 δ_1-CH_3; (d') A19 H_{δ} – B15 δ_2-CH_3; and (e') B1 H_{δ} – A13 δ-CH_3; and (f') B5 H_{δ} – A10 δ-CH_3.

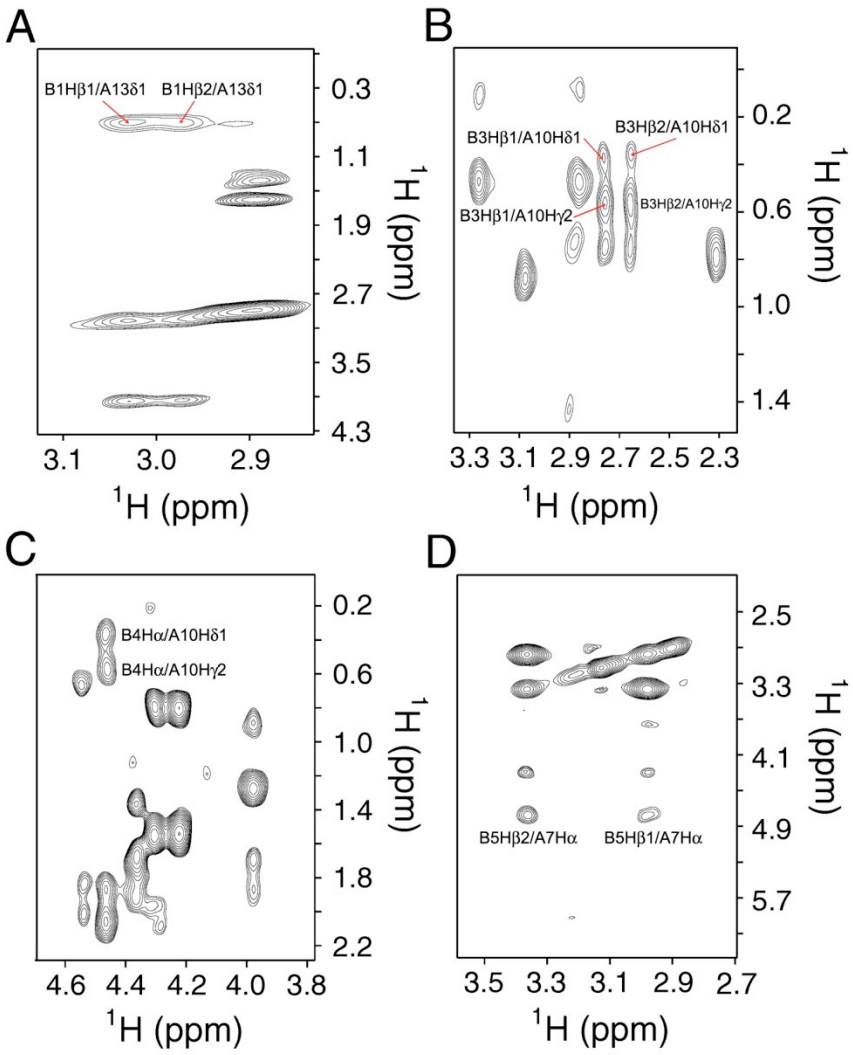

Figure S2. **^{13}C-edited NOESY spectrum of DKP-proinsulin.** Inter-residue NOEs are shown from arm-related residues: (*A*) Phe^{B1} C_β, (*B*) Asn^{B3} C_β, (*C*) Gln^{B4} C_α, and (*D*) His^{B5} C_β. Planes shown are extracted from 3D spectrum; assignments of cross peaks are as labeled. Despite such NOEs, flexibility of the distal arm is demonstrated by motional narrowing of ^{1}H-NMR resonances and patterns of ^{13}C-NMR chemical shifts (Table S2). The solution structure of DKP-proinsulin was recently described by Yang, Y. et al. (see *J. Biol. Chem.* 285, 7847-51 (2010)).

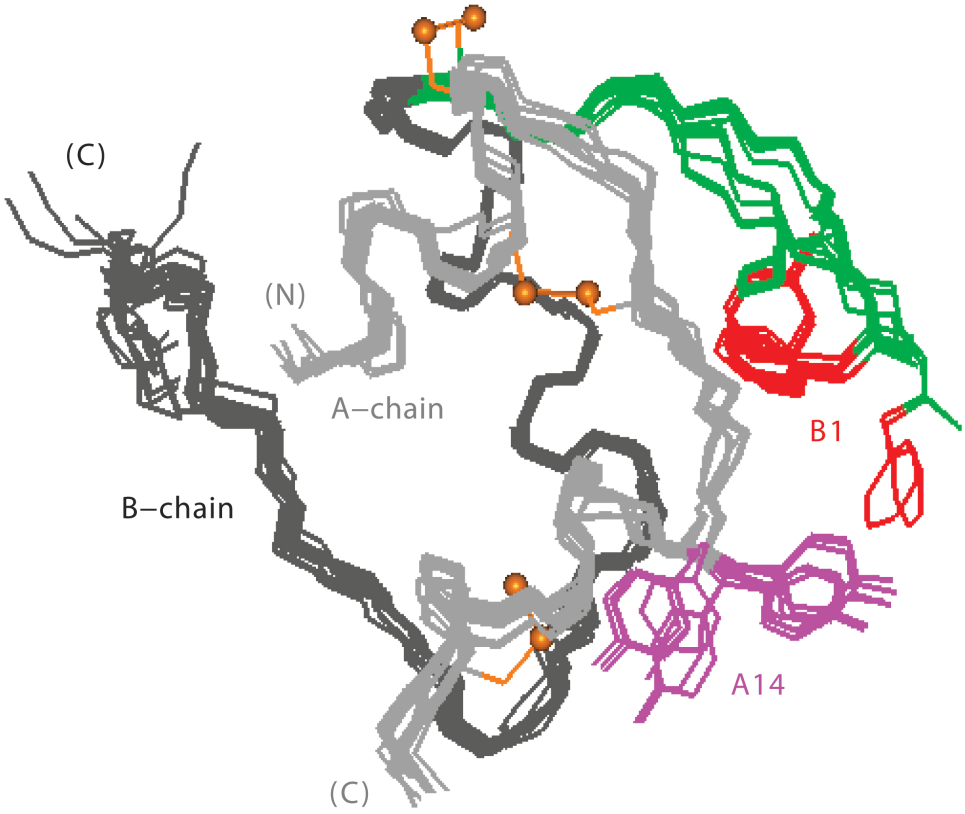

Figure S3. **Variable spatial relationships between aromatic rings of Phe^{B1} and Tyr^{A14}.** Superposition of 14 T-state protomers extracted from multiple crystal structures of T_6 and $T_3R^f_3$ zinc insulin hexamers. Whereas a few structures predict observable NOEs between these rings (including the classical 4INS; molecules 1 and 2), the majority of structures do not and in fact exhibit ring-ring distances greater than 10 Å (see PDB entries 1APH, 1BPH, 1CPH, 1DPH, 1TRZ, 1TYL, 1TYM, 2INS, 1ZNI, 1LPH, 1G7A, 1G7A). Structures were aligned according to the main-chain atoms of helical segments B9-B19, B1-A8, and A12-A19. The side chains are B1 and A14 are highlighted in red and magenta, respectively. The A-chain is shown in light gray, and the B-chain in green (B1-B7) or dark gray (B8-B30). One representative set of disulfide bridges are shown (sulfur atoms as gold spheres).

Table 1. Selected chemical shifts of ^{1}H-NMR resonances of *des*-B1- insulin and wild-type insulin (pH 3.0 and 25° C)

residue	$C_\alpha H$	$C_\beta H$	others
A. analog			
A3 Val	3.58		$C\gamma H_3$ 1.08, 0.96
A8 Thr	3.95	4.44	$C\gamma H_3$ 1.28
A10 Ile	4.49	1.47	$C\gamma H_2$ 0.99,0.48, $C\gamma' H_3$ 0.60, $C\delta H_3$ 0.24
A13 Ieu	4.03	1.61, 1.48	$C\gamma H$ 1.52, $C\delta H_3$ 0.92,0.82
A16 Leu	4.15	1.71, 1.68	$C\gamma H$ 1.68, $C\delta H_3$ 0.81, 0.74
B2 Val	3.81		$C\gamma H_3$ 1.00
B4 Gln	4.51	1.85	
B6 Leu	4.50	1.76, 0.82	$C\gamma H$ 1.54,$C\delta H_3$ 0.80, 0.70
B12 Val	3.29		$C\gamma H_3$ 0.48, 0.37
B14 Ala	4.01	$C\beta H_3$ 1.44	
B15 Leu	3.61	0.81, 0.01	$C\gamma H$ 1.21, $C\delta H_3$ 0.42, 0.01
B27 Thr	4.56	4.13	$C\gamma H_3$ 1.16
B28 Pro	4.15		$C\gamma H_2$ 1.92,1.75, $C\delta H_2$ 3.45
B29 Lys	4.40	1.93, 1.83	$C\gamma H_2$ 1.52,1.45 $C\delta H_2$ 1.69, $Ce H_2$ 2.99
B30 Thr	4.34	4.30	$C\gamma H_3$ 1.21
B. wild-type			
A3 Val	3.60		$C\gamma H_3$ 1.09, 0.98
A8 Thr	3.98	4.49	$C\gamma H_3$ 1.29
A10 Ile	4.40	1.52	$C\gamma H_2$ 0.98,0.47, $C\gamma' H_3$ 0.63, $C\delta H_3$ 0.22
A13 Ieu	3.74	1.34, 1.22	$C\gamma H$ 1.22, $C\delta H_3$ 0.74,0.74
A16 Leu	4.20	1.93, 1.61	$C\gamma H$ 1.74, $C\delta H_3$ 0.88, 0.81
B2 Val	4.09	1.85	$C\gamma H_3$ 0.86, 0.84
B4 Gln	4.51	1.95	
B6 Leu	4.50	1.82, 0.78	$C\gamma H$ 1.62,$C\delta H_3$ 0.78, 0.71
B12 Val	3.32		$C\gamma H_3$ 0.52, 0.43
B14 Ala	4.14	$C\beta H_3$ 1.47	
B15 Leu	3.63	1.01, 0.00	$C\gamma H$ 1.21, $C\delta H_3$ 0.42, 0.02
B27 Thr	4.49	4.18	$C\gamma H_3$ 1.18
B28 Pro	4.19		$C\gamma H_2$ 1.92,1.75, $C\delta H_2$ 3.46
B29 Lys	4.41	1.94, 1.83	$C\gamma H_2$ 1.54,1.45 $C\delta H_2$ 1.71, $Ce H_2$ 3.02
B30 Thr	4.34	4.29	$C\gamma H_3$ 1.23

[a]Chemical shifts are relative to 5,5-dimethylsilapentanesulfonate (DSS; presumed to be at 0 ppm). Spectra were acquired in D_2O and assigned by inspection based on published values. The proteins are dimeric under these conditions. Selected differences in chemical shift are consistent with absence of Phe^{B1} ring current with retention of native-like structure (see main text).

Table S2-A. Arm ^{13}C-chemical shifts and secondary shifts in human proinsulin

	C_α	C_β	$\delta_{C\alpha}$ (secondary shift)	$\delta_{C\beta}$ (secondary shift)
B1^{Phe}	**58.36**	**41.02**	**0.36**	**2.02**
B2^{Val}	**62.00**	**32.81**	**-0.30**	**0.71**
B3^{Asn}	**53.70**	**38.22**	**0.8**	**0.32**
B4^{Gln}	54.61	31.37	-1.49	2.97
B5^{His}	57.74	29.25	-	-
B6^{Leu}	53.67	44.90	-1.43	2.6
B7^{Cys}	53.94	-	-1.22	-
B8^{Gly}	47.50		2.4	

Bold indicates chemical shifts and secondary shifts indicative of segmental flexibility. The secondary C_α shifts of residues B1-B3 are either positive (B1 and B3) or not sufficiently negative (B2) to conform to a β-stand pattern.

The secondary C_β shifts of B2 and B3 are likewise attenuated. Secondary shifts are not provided for His^{B5} due to effects of pH-dependent side-chain protonation. The large positive secondary C_α shift of Gly^{B8} reflects its positive phi angle in a T-state-specific β-turn. Secondary shifts are defined as
the difference between observed chemical shifts and tabulated random-coil
values.

Table S2-B. Average secondary shift of canonical α-helix and β-strand in proinsulin

α-helix $\delta_{C\alpha}$ average (B9-B19)	α-helix $\delta_{C\beta}$ average (B9-B19)	β-strand $\delta_{C\alpha}$ average (B24-B27)	β-strand $\delta_{C\beta}$ average (B24-B27)
3.16	-0.65	-1.21	1.38

These values are in accordance with expected mean values based on the solution structure of
DKP-proinsulin (Yang, Y. et al. (2010) *J. Biol. Chem.* 285, 7847-51).

Table S3. Thermodynamic modeling (Guanidine Denaturation)

analog	temp (°C)	ΔG_u	$\Delta\Delta G_u$	C_{mid}	*m*

					(kcal/mol/M)
human insulin	4	4.14 ± 0.04	----------	5.1 ± 0.06	0.81 ± 0.01
des-B1-insulin	4	4.04 ± 0.07	-0.36 ± 0.16	5.3 ± 0.1	0.76 ± 0.02

Studies conducted in 50 mM potassium phosphate (pH 7.4) at 4° C. Stability is inferred from guanidine denaturation studies based on a two-state model extrapolated to zero denaturant concentration. Uncertainties shown pertain to fitting and are taken to be ± 0.1 kcal/mole in relation to additional possible sources of experimental error (guanidine concentration, temperature, buffer pH and salt concentration). ΔG_u and $\Delta\Delta G_u$ values are given in kcal/mole. C_{mid} is defined as the denaturant concentration at which the fractional unfolding is 50%. The m-value represents the slope of ΔG_u versus guanidine concentration on extrapolated to zero denaturant concentration.