Published in final edited form as:
Chem Commun (Camb). 2010 November 21; 46(43): . doi:10.1039/c0cc03141k.

Total chemical synthesis of human proinsulin†,‡

Samuel Luisier[§,a], **Michal Avital-Shmilovici**[§,a], **Michael A. Weiss**[b], and **Stephen B. H. Kent**[a]

Michael A. Weiss: michael.weiss@case.edu; Stephen B. H. Kent: skent@uchicago.edu

[a]Department of Chemistry, Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637, USA

[b]Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106, USA

Abstract

A convergent synthetic strategy based on modern chemical ligation methods was used to make human proinsulin. The synthetic protein was characterized by LCMS, CD spectroscopy, and by 1D- and 2D-NMR spectroscopy. Synthetic human proinsulin had full biochemical activity in a receptor-binding assay.

The human proinsulin molecule is the key to efficient biosynthesis of human insulin. Proinsulin is biosynthesized as an 86 amino acid residue polypeptide chain, and after folding and formation of three native disulfide bonds is processed by prohormone convertases in the pancreatic β-cell to form insulin. Proinsulin is composed of the B and A peptides of insulin linked together by the 35-amino acid C domain.[1,2] Recent solution NMR structural studies of a monomeric proinsulin indicate that the C domain is largely unstructured.[3] The primary role of the C domain of intact proinsulin is to favor the formation of the correct disulfides, as found in mature insulin.[4] Several insulin gene mutations are thought to cause diabetes by impairment of proinsulin folding leading to unremitting endoreticular stress and beta-cell apoptosis.

The NMR structure of a monomeric analogue of human proinsulin was recently reported.[3] Although bovine proinsulin has been crystallized[5] and proinsulin was co-crystallized with insulin,[6,7] no X-ray structure of human proinsulin has been obtained. We anticipate that the use of a racemic protein mixture made up of L-proinsulin and D-proinsulin could lead to the formation of highly ordered centrosymmetric crystals, which can be used for X-ray crystal structure determination.[8] The mirror image protein D-proinsulin can only be prepared by total chemical synthesis. Here, we report the efficient total chemical synthesis of human proinsulin using modern chemical ligation methods.[9]

The polypeptide chain of proinsulin has 86 amino acids containing six cysteine residues (Fig. 1a), and the folded protein contains three intramolecular disulfide bonds. The presence of six Cys residues makes proinsulin a good target for total chemical synthesis by native chemical ligation, which involves the thioester-mediated amide-forming covalent condensation of unprotected synthetic peptides.[10,11] Our synthetic strategy is shown in Fig. 1b. The two peptide-thioesters, L-proinsulin(Thz^{19}-Cys^{71})-$^{\alpha}$COSR (R = $-CH_2CH_2CO$–Ala–COOH) and L-proinsulin(Phe^{1}-Val^{18})-$^{\alpha}$COSR, and the Cys-peptide L-proinsulin(Cys^{72}-

‡Electronic supplementary information (ESI) available: Peptide synthesis; HPLC and LC-MS analysis; chemical synthesis of D-proinsulin; CD spectra of D- and L-proinsulin; 1D- and 2D-NMR spectra of D- and L-proinsulin. See DOI: 10.1039/c0cc03141k

†This article is part of the 'Enzymes and Proteins' web-theme issue for ChemComm.

Correspondence to: Stephen B. H. Kent, skent@uchicago.edu.

§These authors contributed equally to this work.

Asn^{86}) building blocks were prepared by manual stepwise Boc-chemistry "*in situ* neutralization" solid phase peptide synthesis[12] (see ESI‡). The synthesis of the full length 86-amino acid polypeptide chain of the target L-proinsulin was achieved using two native chemical ligations, starting from the C-terminal segment.[13] Data for the ligations are shown in Fig. 2.

L-proinsulin(Thz^{19}-Cys^{71})-$^{\alpha}$COSR (25 mg, 4.3 μmol) and L-proinsulin(Cys^{72}-Asn^{86}) (8.3 mg, 4.7 μmol) were dissolved to a concentration of 4.3 mMand 4.7 mM, respectively, in 6M guanidine·HCl, buffered with 0.2 M sodium phosphate, pH 6.9, and containing 100 mM (4-carboxymethyl)thiophenol (MPAA) and 20 mM TCEP·HCl. The reaction mixture was purged with helium for 5 min and sealed. Upon complete reaction (~4.5 h), the N-terminal Thz- of the product was converted to Cys- by addition of 0.4 M methoxylamine·HCl and overnight reaction at pH 4.0. The crude product L-proinsulin(Cys^{19}-Asn^{86}) was separated from salts and low MW contaminants on a solid phase extraction C18 cartridge and lyophilized. The second ligation between L-proinsulin(Phe^{1}-Val^{18})-$^{\alpha}$COSR (5.1 mg, 2.3 μmol) and L-proinsulin(Cys^{19}-Asn^{86}) (17 mg, 2.3 μmol) was performed under the same conditions at a peptide concentration of 5.5 mM.

Upon completion of the second ligation (~46.5 h), the full-length polypeptide L-proinsulin(Phe^{1}-Asn^{86}) was recovered by precipitation from water, and the precipitated peptide redissolved and lyophilized. The crude linear (1–86) peptide was then folded in a buffer solution containing 50 mM glycine/NaOH (pH 10.5), 1 mM EDTA, 1 mM reduced glutathione (GSH), 1 mM oxidized glutathione (GSSG) at peptide concentration of 0.1 mg mL^{-1} [14] (Fig. 3). The folding solution was purged, sealed, and incubated overnight at 4 °C. The reaction mixture was then acidified to pH 2 with dil. HCl and purified by preparative reverse-phase HPLC to give pure L-proinsulin (5.45 mg, 0.58 μmol, 26% for the whole synthesis, based on limiting peptide L-proinsulin(Thz^{19}-Cys^{71})-$^{\alpha}$COSR). The synthetic protein was characterized by LC-MS.

Chemical synthesis of the mirror image protein molecule D-proinsulin was performed following the same procedure as described for the L-proinsulin (see ESI‡). Both the L- and the D-proinsulin were characterized by CD spectroscopy, receptorbinding, and by 1D- and 2D-NMR spectroscopy. CD spectra were recorded on an Aviv spectrometer at 25 °C. Protein samples were dissolved in phosphate-buffered saline (pH 7.4). The CD spectra of chemically synthesized and biosynthesized (recombinant) L-proinsulin showed essentially identical curves (Fig. 4a). As expected, the CD spectrum of D-proinsulin was of equal magnitude, but inverted, compared to the spectrum of L-proinsulin (see ESI‡). In an insulin receptor-binding assay, the results were compatible with expectations: the biosynthetic *insulin* was the most active in binding (K_d = 0.06 ± 0.01 nM); the synthetic and the Lilly biosynthetic L-proinsulin showed equal activity (K_d = 2.7 ± 0.4 nM, K_d = 2.8 ± 0.4 nM, respectively), *i.e.* lower binding than *insulin* by the expected amount; and, finally, the enantiomer D-proinsulin was inactive in the binding assay (its K_d was too great to measure, more than 100-fold greater than that of L-proinsulin) (Fig. 4b). We also analyzed L-proinsulin and D-proinsulin by 1D and 2D ^{1}H-NMR (for NMR data, see ESI‡). As enantiomeric forms, the L- and D-proinsulin had the same NMR spectra. The 1D NMR results showed clearly that both L- and D- synthetic proinsulin are competent to self-assemble into R_6 zinc/phenol hexamers.

With the enantiomers L- and D-proinsulin in hand, we attempted to crystallize proinsulin by employing the racemic crystallization method.[8] However, human proinsulin was refractory to crystallization even under these conditions, most likely because of its tendency to form hexamers and aggregates. Molecular models of classical crystal lattices observed in studies

of *insulin* suggest that for proinsulin an analogous mode of hexamer–hexamer assembly would be precluded by the C domain.

In conclusion, we have developed an efficient total chemical synthesis of human proinsulin by the chemical ligation of three synthetic peptide segments. The synthetic 86-residue polypeptide folds efficiently with the formation of three native disulfide bonds. The resulting synthetic L-protein is indistinguishable from authentic recombinant human proinsulin. The D-proinsulin protein molecule exhibits a mirror image CD spectrum and is inactive in receptor binding. NMR analysis showed that synthetic proinsulin formed native-like R6 hexamers. In order to simplify crystallographic analysis using the racemic crystallization method, we are now using the same synthetic strategy described here to make monomeric analogue DKP human proinsulin enantiomers (DKP = HisB10 → Asp, ProB28 → Lys and Lys29 → Pro).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge support from the University of Chicago Diabetes Research and Training Center (DRTC); and the NIH (grant DK0697674 to MAW at Case Western Research University). SL was supported by the Swiss National Science Foundation (grant no. PBBEP2-125612). We also thank Wenhua Jia for the CD spectra, Jonathan and Linda Whittaker for the binding assays, and Qing-xin Hua and Yanwu Yang for NMR studies.

Notes and references

1. Steiner DF, Oyer PE. Proc Natl Acad Sci U S A. 1967; 57:473. [PubMed: 16591494]
2. Oyer PE, Sooja C, Peterson JD, Steiner DF. J Biol Chem. 1971; 246(5):1375. [PubMed: 5101771]
3. Yang Y, Hua Q, Liu J, Shimizu EH, Choquette MH, Mackin RB, Weiss MA. J Biol Chem. 2010; 285(11):7847. [PubMed: 20106974]
4. Johansson J, Ekberg K, Shafqat J, Henriksson M, Chibalin A, Wahren J, Jornvall H. Biochem Biophys Res Commun. 2002; 295:1035. [PubMed: 12135597]
5. Rosen LS, Fullerton WW, Low BB. Arch Biochem Biophys. 1972; 152:569. [PubMed: 4673814]
6. Steiner DF. Nature. 1973; 243:528. [PubMed: 4550045]
7. Low BW, Fullerton WW, Rosen LS. Nature. 1974; 248:339. [PubMed: 4819640]
8. Pentelute BL, Gates ZP, Tereshko V, Dashnau JL, Vanderkooi JM, Kossiakoff AA, Kent SB. J Am Chem Soc. 2008; 130:9695. [PubMed: 18598029]
9. Kent SBH. Chem Soc Rev. 2009; 38:338–51. [PubMed: 19169452]
10. Dawson PE, Muir TW, Clarklewis I, Kent SBH. Science. 1994; 266:776. [PubMed: 7973629]
11. Dawson PE, Kent SBH. Annu Rev Biochem. 2000; 69:923. [PubMed: 10966479]
12. Schnolzer M, Alewood P, Jones A, Alewood D, Kent SBH. Int J Pept Res Ther. 2007; 13:31.
13. Bang D, Kent SBH. Angew Chem Int Ed. 2004; 43:2534.
14. Mackin RB, Choquette MH. Protein Expression Purif. 2003; 27:210.

(a) FVNQHLCGSH[10] LVEALYLVCG[20] ERGFFYTPKT[30]
RREAEDLQVG[40] QVELGGGPGA[50] GSLQPLALEG[60]
SLQKRGIVEQ[70] CCTSICSLYQ[80] LENYCN[86]

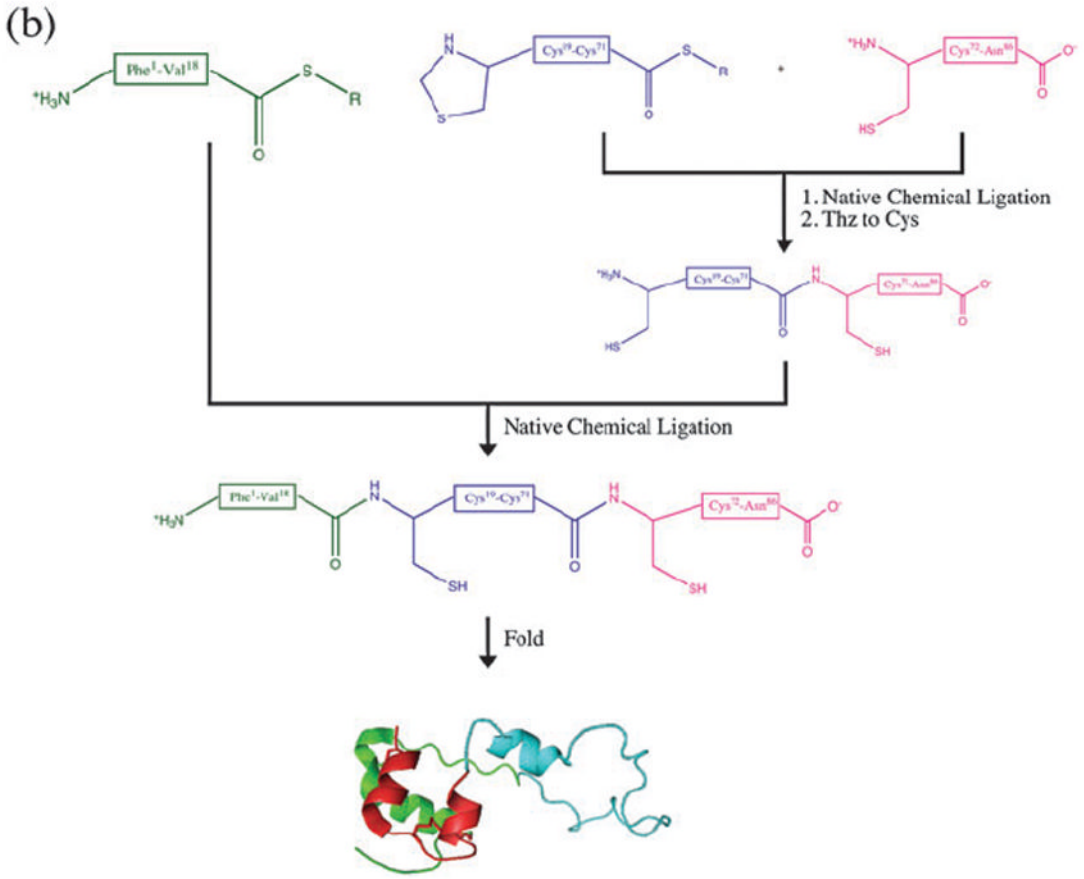

Fig. 1.

(a) Amino acid sequence of human proinsulin. Color coding corresponds to the peptide segments used in synthesis. (b) Synthetic strategy used for the total chemical synthesis of human proinsulin by native chemical ligation. R = $-CH_2CH_2CO–Ala–COOH$.

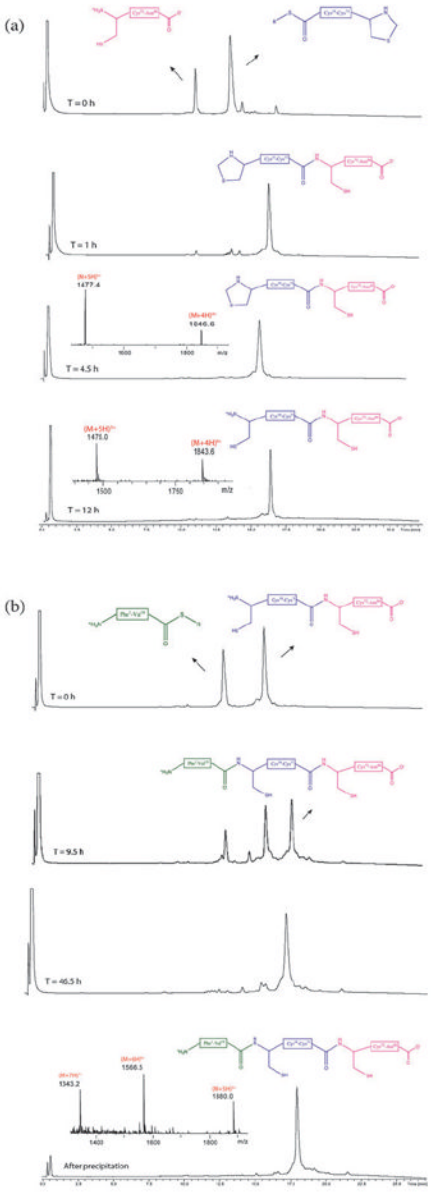

Fig. 2.

Synthesis of the L-proinsulin polypeptide chain. (a) LC-MS analytical data (λ = 210 nm) for the first ligation between L-proinsulin-(Thz^{19}-Cys^{71})-$^{\alpha}$COSR and L-proinsulin(Cys^{72}-Asn^{86}), followed by conversion of Thz- to Cys-. (b) LC-MS analytical data for the second ligation between L-proinsulin(Phe^{1}-Val^{18})-$^{\alpha}$COSR and L-proinsulin(Cys^{19}-Asn^{86}); the times shown refer to overall elapsed times for the synthesis. The chromatographic separations were performed using a linear gradient (9%–53%) of buffer B in buffer A over 22 min (buffer A = 0.1% TFA in water, buffer B = 0.08% TFA in acetonitrile).

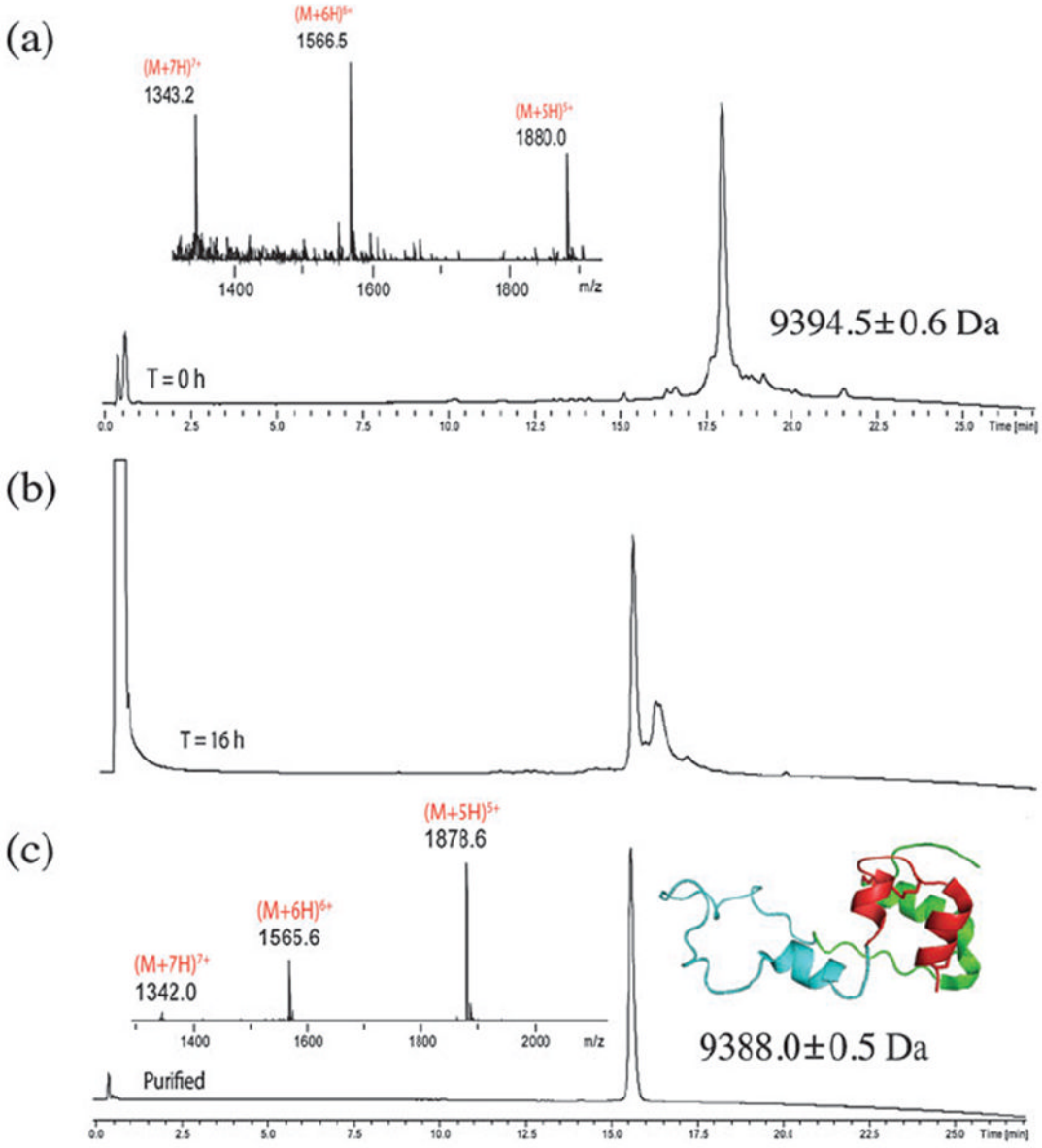

Fig. 3.

Analytical LC-MS analysis (λ = 210 nm) of L-proinsulin's folding. (a) Crude linear polypeptide 1–86 (mass: obs. = 9394.5 ± 0.6 Da, calc. 9394.6 Da, av isotopes); (b) folding reaction after 16 h; folding conditions are described in the text; (c) folded human proinsulin, after prep HPLC purification (mass: obs. = 9388.0 ± 0.5 Da, calc. 9388.5 Da). The chromatographic separations were performed under the same conditions used for Fig. 2.

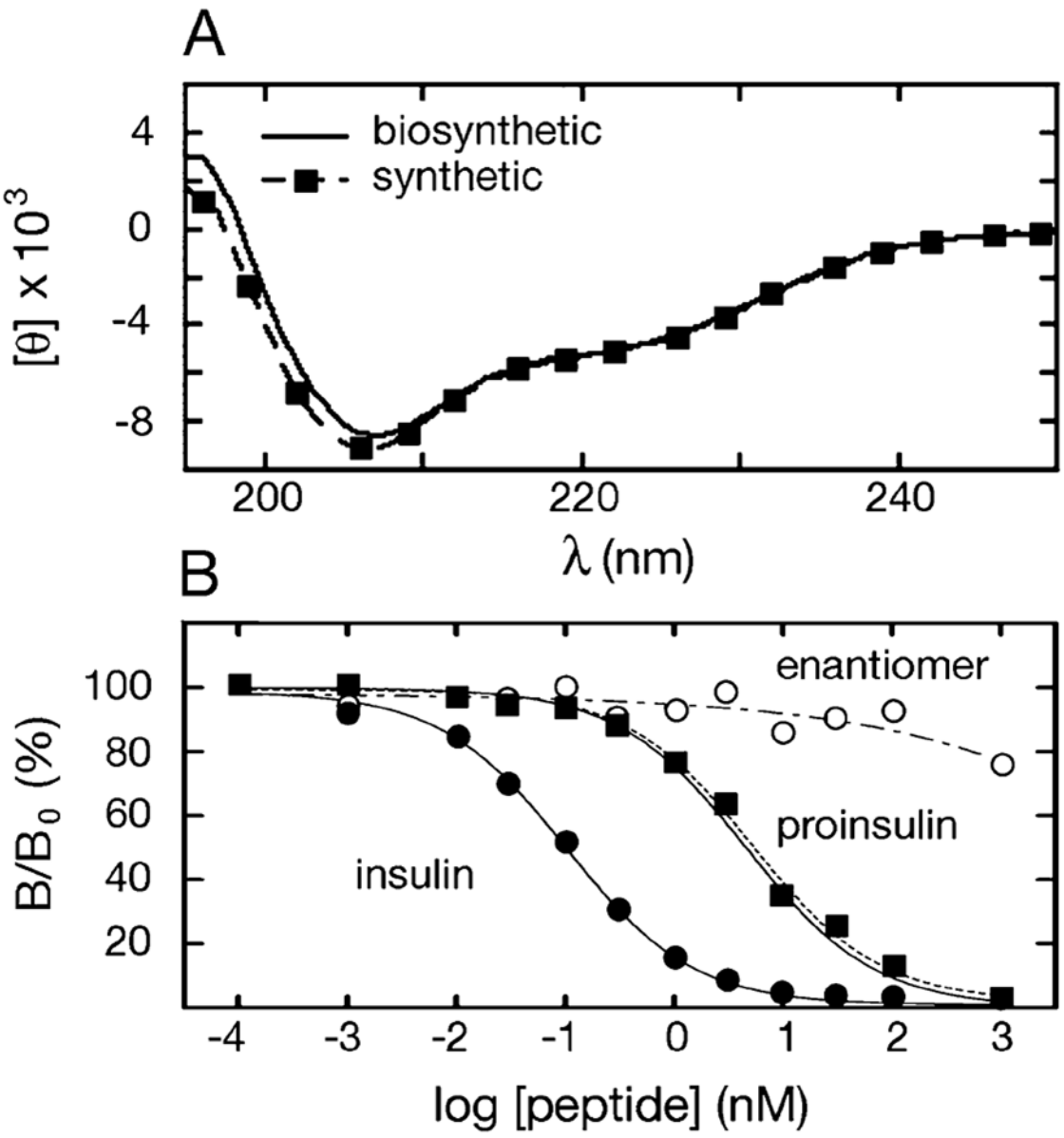

Fig. 4.

(a) CD spectra of biosynthetic *vs.* synthetic L-proinsulin. (b) Insulin receptor-binding assays for biosynthetic *insulin* (●), biosynthetic L-proinsulin (··■··), synthetic L-proinsulin (■), synthetic D-proinsulin (–·○–·).