DOI: 10.1002/cbic.200800430

Total Synthesis of desB30 Insulin Analogues by Biomimetic Folding of Single-Chain Precursors

A. Pernille Tofteng,[b] Knud J. Jensen,*[b] Lauge Schäffer,[a] and Thomas Hoeg-Jensen*[a]

Insulin is a peptide hormone consisting of 51 amino acids in two chains with three disulfide bridges. Human insulin and various analogues are used for the treatment of diabetes and are produced recombinantly at ton scale. Herein, we report the chemical synthesis of insulin by the step-wise, Fmoc-based, solid-phase synthesis of single-chain precursors with solubilising extensions, which under redox conditions, spontaneously fold with the correct pairing of the three disulfide bridges. The folded, single-chain, insulin precursors can be transformed into bioactive two-chain desB30 insulin by the simultaneous removal of the solubilising extension (4–5 residues) and the chain-bridging C-peptide (3–5 residues) by employing Achromobacter lyticus *protease—a process well-known from the yeast-based recombinant production of insulin. The overall yields of synthetic insulins were as much as 6%, and the synthetic process was straightforward and not labour intensive.*

Introduction

Insulin consists of two peptide chains of 21 and 30 amino acid residues, with each chain folded into secondary structures and cross-linked by three disulfide bridges (two interchain and one intrachain). Since the discovery of insulin around 1920,[1] insulin drugs have been the cornerstone of diabetes treatment. The goal of modern diabetes treatment is to approximate the native, healthy secretion of insulin, which occurs in two major patterns. Slowly released basal insulin should be delivered to the body around the clock on a day-to-day, reproducible fashion, and fast-acting insulin should be supplied with meals, with the timing and dosing of the drug optimally adjusted to the character of each individual meal. It is generally agreed that the better the native insulin delivery can be approximated, the lower is the risk of long-term complications from diabetes, such as blindness and kidney and cardiovascular damage.[2,3]

In order to achieve ideal insulin delivery, various insulin analogues have reached clinical practice in recent years.[4] For many years, insulin was collected from animal glands, and attempts at optimising insulin delivery consisted of various formulation approaches, such as different crystal forms. However, modern recombinant technologies allow for the manufacture of insulin mutants and enable diverse modulations of the pharmacological properties of insulin.

Although recombinant methods have been very successful, they suffer from one major drawback; generally, only the 20 proteogenic amino acids are allowed.[5] Modern chemical peptide synthesis, on the other hand, allows for the introduction of an almost unlimited number of unnatural amino acids and other building blocks.

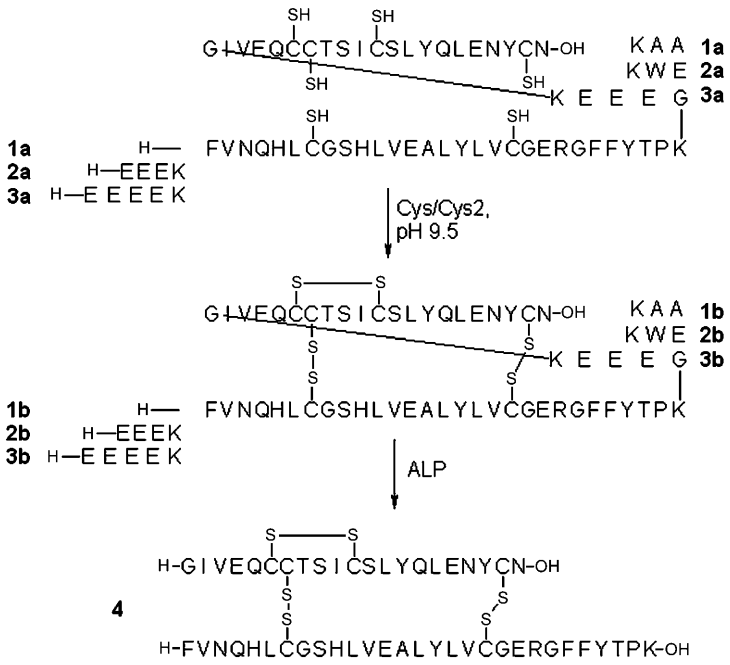
The total chemical synthesis of insulin has been reported on several occasions, but the methods are difficult and have only been practised in a few specialized laboratories.[6–9] The difficulties involve the relatively long sequences, which are difficult to solubilise, and the correct folding and pairing of the three disulfide bridges. Early methods involved the synthesis of the two separate chains,[10,11] followed by random disulfide pairing. More recently, methods for controlled disulfide pairings, by the use of triple orthogonal cysteine protection schemes, have been described.[12,13] Insulin analogues have also been prepared by native chemical ligation from three fragments,[14] followed by folding to single-chain insulin analogues or insulin-like growth factors.[15] However, two-chain insulins have not been prepared by the native ligation route. Furthermore, the ligation procedures seem to work only with special insulins containing solubilising mutations, such as the immunogenic A8 and mitogenic B10 mutations, which render the compounds questionable as drugs.[16,17] All of the described methods involve labour-intensive procedures, and overall yields are low.

In the pancreas, native insulin is produced from a single-chain precursor, which spontaneously folds to give an inactive single-chain proinsulin with the correct disulfide pairing.[18] Only upon enzymatic removal of the native 35-residue C-peptide is the bioactive two-chain form of insulin produced. The single-chain approach has been briefly exploited by chemical linkers[19] and broadly adopted in the recombinant production of insulin, for which much shorter artificial C-peptides have been developed, such as AAK or EWK.[20–22] These artificial C-peptides can be removed enzymatically to provide two-chain insulins by the utilization of the lysine-specific enzyme *Achromobacter lyticus* protease (ALP).[23]

[a] *L. Schäffer, Dr. T. Hoeg-Jensen*
Diabetes Protein and Peptide Chemistry, Novo Nordisk A/S
Novo Nordisk Park D6.1.142, 2760 Maaloev (Denmark)
Fax: (+45) 44-44-42-56
E-mail: tshj@novonordisk.com

[b] *Dr. A. P. Tofteng, Prof. K. J. Jensen*
Faculty of Life Sciences, University of Copenhagen
Thorvaldsensvej 40, 1871 Frederiksberg (Denmark)
Fax: (+45) 35-33-23-98
E-mail: kjj@life.ku.dk

Inspired by the single-chain recombinant methods, we now report the Fmoc-based total chemical synthesis of insulin as single-chain precursors, which are folded and enzymatically cleaved to provide bioactive two-chain insulin. Solid-phase synthesis normally delivers peptides in their reduced form (free Cys), and the single-chain precursors of simple constructions containing the C-peptides AAK or EWK, as known from recombinant methods, were insoluble in simple aqueous buffers. However, by supplementing the single-chain propeptides with presequences EEEK or EEEEK at the N terminus of the B-chain, the precursors became soluble at slightly basic pH values and could be folded in simple buffers. Notably, the folding of synthetic insulin proceeded even better with GEEEK as the C-peptide than with AAK or EWK, as optimised for recombinant methods. During the final processing, the C-peptide and the EEE(E)K presequences were removed simultaneously by ALP treatment to provide bioactive two-chain insulin. The B2Aib desB30 human insulin was prepared as an example of an insulin analogue containing an unnatural amino acid.



Scheme 1. Folding to give single-chain precursors **1b**, **2b** and **3b** and enzymatic ALP transformation to give two-chain desB30 human insulin **4**. Compound **1a**: single-chain insulin precursor with no solubilising extension and AAK as the bridging C-peptide. Compound **2a**: solubilising extensions EEEK and EWK as the C-peptide. Compound **3a**: extensions EEEEK and GEEEK as the C-peptide.

Results and Discussion

Peptide synthesis with AAK as the C-peptide

From the peptide N terminus, native proinsulin consists of the B-chain, the C-peptide and the A-chain. For the solid-phase synthesis of peptide acids, the C-terminal amino acid is generally immobilised on a hydroxyl resin. Since the C-terminal residue of the insulin A-chain is asparagine, which contains a side-chain amide, we chose to anchor the first amino acid by its side-chain to a Rink amide linker resin. This strategy circumvents potential problems with a C-terminal ester, such as racemisation. We conducted the chemical peptide synthesis by amino acid activation with HBTU and HOBt/HOAt (80:20) and Fmoc removal with 20% piperidine in NMP on a Tentagel resin on a standard automated peptide synthesizer. We used the AAK sequence as the C-peptide (**1a**; Scheme 1). For analysis, we cleaved intermediary peptides from the resin at various times during synthesis, and these data showed the synthesis to proceed with the desired sequence as the main product (Figure 1).

We cleaved the crude 53-mer single-chain intermediate **1a** from the resin with TFA/H_2O/TES (95:2.5:2.5) and precipitated it from diethyl ether. Unfortunately, the full-length product, which had six reduced Cys residues, was insoluble in simple aqueous buffers, even at pH values above 12 or below 2. As a control experiment, we reduced the identical recombinant precursor **1b** (MI3) with either excess thiol (MESNa) or phosphine (TCEP) to provide **1a**. These experiments also produced precipitates of reduced 53-mer insulin **1a**. Because of the poor solubility of reduced insulin **1a**, we were not able to purify the peptide prior to folding.

We dissolved the reduced synthetic **1a** in 6 M guanidinium chloride (GuHCl), but analysis by LC-MS showed a broad peak with no distinct mass peak for the expected product. We folded **1a** by either dialyzing a solution of it in 6 M GuHCl against a buffer with dilute GuHCl or by simply diluting it into buffer, as detailed below. Such procedures resulted in the appearance of a minor peak in the LC-MS with a smaller retention time and the expected mass signals for folded product **1b**. However, the handling of the peptide in GuHCl with dialysis filters provided very low yields on repeated purification attempts. For this reason, we sought a precursor that was soluble in a simple buffer.

Peptide synthesis with EEEK or EEEEK as N-terminal extensions and EWK or GEEEK as C-peptides

Insulin is an acidic peptide, so the molecule attains an overall negative charge at neutral and basic pH. The isoelectric point of insulin is 5.5. Accordingly, providing insulin with additional negative charges seemed like a viable approach to solving the solubility problems. For this reason, we placed an extension at the N terminus of the B-chain (EEEK), which could be removed with enzymatic ALP-based removal of the C-peptide (cleavage at K, **2a**; Scheme 1). We also included a negative charge in the C-peptide. The EWK C-peptide has been shown to provide better yields in recombinant insulin production in yeast due to better folding.[22]

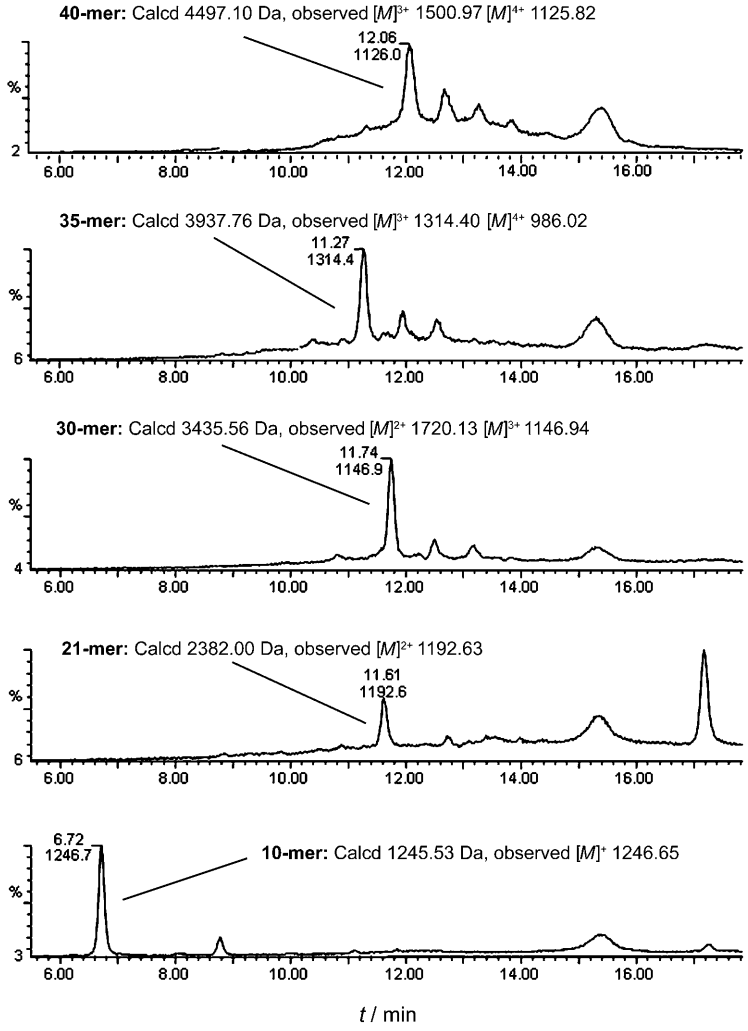


Figure 1. LC-MS data of peptide intermediaries cleaved from the resin after 10, 21, 30, 35 and 40 couplings. All peptide samples were dissolved in MeCN/H_2O (1:1) prior to analysis except for the 21-mer, which could be only solubilised in a 20 mM phosphate buffer containing 6 M GuHCl, pH 7.5.

We synthesized the 57-mer peptide **2a** as described above with a Rink ChemMatrix resin by using an automated peptide synthesizer and microwave heating to decrease the coupling times and improve the purity.[24] We cleaved the crude single-chain intermediate **2a** from the resin with TFA/TIPS/H_2O/DMB/DTT (89:5:2:2:2) and precipitated it from diethyl ether. Contrary to intermediate **1a** above, the intermediate **2a** could be solubilised in aqueous buffer at pH 8–10 (>5 mg mL^{-1}). However, LC-MS analysis of crude **2a** showed a broad peak with no clear mass peak. There is precedence for such complex HPLC profiles of unfolded reduced peptides.[25,26] Since HPLC gave a broad peak with no mass corresponding to **2a**, we did not purify the crude product prior to folding. However, upon folding crude **2a** in redox buffers, as detailed below, a new LC-MS peak appeared at an earlier retention time. This peak displayed a well-defined LC-MS mass spectrum corresponding to the desired product **2b** as both $[M]^{4+}$ and $[M]^{5+}$ ions (Figure 2).

In order to improve the solubility and perhaps improve the folding even further, we synthesized another sequence with the microwave peptide synthesizer, this time with EEEEK as the presequence and GEEEK as the C-peptide. The resulting peptide **3a** was highly soluble at neutral and basic pH, but its reduced form also eluted as a broad peak in the LC-MS. Contrary to **1a** and **2a**, **3a** gave the mass signal corresponding to the reduced peptide.

In order to clarify the quality of the peptide synthesis without complications from reduced, reactive cysteines, we also synthesized the single-chain intermediate **2a** with the cysteines blocked by Acm protection, in order to prevent Cys-based side reactions, which could include alkylations, random folding and oligomer formations. The LC-MS of the Acm-blocked peptide showed the expected mass from the largest eluted peak ($t_R = 4.17$ min). Accordingly, the peptide synthesis proceeded quite well considering the long sequence. However, when multiple reduced cysteines were present in the sequence, the material eluted as a very broad peak, which was very difficult to characterize by MS.

We attempted to fold the Acm-protected version of **2a**, with either iodine in dilute acetic acid or TMSCl in DMSO. However, the solubility of Acm$_6$-**2a** was poor under such conditions, and we could not detect any folded material by LC-MS.

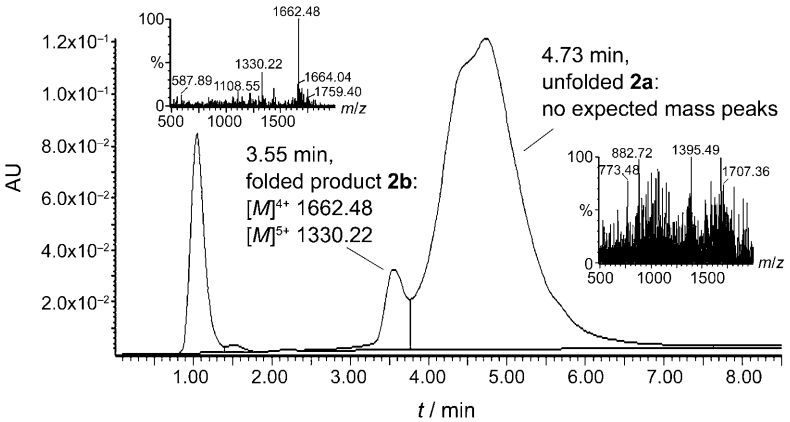


Figure 2. C-peptide EWK: crude reduced precursor **2a** at 4.73 min showed no well-defined mass peaks, and folded precursor **2b** at 3.55 min (5–6%) displayed the expected product masses, $[M]^{4+}$ and $[M]^{5+}$.

Folding experiments

As mentioned above, purified recombinant **1b** could be reduced with excess thiol or phosphine to produce **1a**, but the material precipitated when simple buffers were used. However, by carrying out the reductions in 6 M GuHCl, the reduced intermediate **1a** could be kept in solution. When we refolded under redox conditions, which was adopted from proinsulin work,[27] folded single-chain insulin **1b** could be regained from recombinant **1b** in yields of approximately 50–70%, as quantified by HPLC. In other words, the chemical reduction and refolding of a pure single-chain insulin precursor can proceed in quite high yields.[28,29]

In extensive folding studies of the synthetic sequence **1a** (C-peptide AAK), we kept the peptide in solution with 6 M GuHCl buffer and folded it using three different methods. Initially we dialyzed a protein solution of approximately 0.2 mM (in 6 M GuHCl buffer, pH 9) against a refolding buffer containing 1 mM Cys and 8 mM cystine at pH 9, and a decreasing amount of GuHCl in order to keep precipitation at a minimum. The final yield of purified product was very low. The second folding method involved the simple addition of refolding buffer containing 1 mM Cys and 8 mM cystine at pH 9.5 directly to the reduced peptide in 6 M GuHCl (protein concentration of 0.6 mM). We incubated the mixture for 48 h at room temperature, diluted it to a protein concentration of 0.1 mM, and then removed GuHCl by dialysis. Precipitation occurred during the dialysis, which complicated purification and gave a very low final yield. The third folding method started with the conversion of reduced peptide **1a** to its corresponding S-sulfonate.[30,31] The reduced peptide **1a** required 6 M GuHCl to solubilise it. We accomplished the folding of the corresponding S-sulfonates by adding 350 mM Cys to a solution of the S-sulfonated peptide dissolved in NH_4HCO_3. The overall purified yield was low. Although it was possible to obtain the precursor **1b** from the folding of synthetic **1a**, the amount of precipitation during folding decreased the purified yields to approximately 1–2% and made the process very time consuming. Therefore, we carried out the remaining experiments with EEE(E)K extensions at the N terminus of the B-chain in order to eliminate the precipitation problems.

The synthetic, EEEK-solubilised, reduced intermediate **2a** (C-peptide EWK) eluted as a broad peak by LC-MS, so the reduced intermediate could not be purified. Instead, we treated crude **2a** (0.5 mM) with aqueous buffers in the pH range of 7.5–11. Notably, the material did not dissolve fully below pH 8. We conducted the refolding in buffers containing Cys (1 mM), cystine (8 mM), Tris (10 mM), glycine (10 mM) and EDTA (1 mM)[27] at room temperature in an open vial. We analyzed the reaction mixtures by LC-MS and observed the expected product **2b** as an earlier eluting peak in yields of up to 6% (Figure 2). We observed the best yields from **2a** after folding it at pH 9.5 for 2–3 days. We based the quantifications on the LC-MS UV trace at 280 nm compared to a recombinant insulin standard, with extinction coefficients corrected for the number of Tyr and Trp residues. We could not identify any conditions that increased the folding yield higher than approximately 6%. Higher peptide concentrations (5 mM) resulted in poorer folding yields (1–2%) and lower concentrations (50 μM) did not improve the yields. Phosphate buffer or no buffer gave results similar to those in Tris buffer (5–6% folding yield). The addition of protein disulfide isomerase (0.1%, *w*/*w*) did not improve folding yields (2–3%), and similar yields were obtained with classical folding methods, such as lauroyl–sarcosine buffer in combination with cyclodextrin (4–5% yield) or charcoal as a surface catalyst (2% yield). Additionally, experimenting with the temperature or with oxygen-free conditions did not improve folding yields.

We isolated the folded intermediate **2b** by preparative HPLC; this gave **2b** in a yield of 3%, which was calculated from the crude peptide material. During the HPLC, we collected the broad residual peak of **2a**, and attempted to refold this "second-crop **2a**". However, these experiments gave very little folded product **2b** (<0.5% yield) even when the material was first fully reduced with excess thiol and/or denatured with 6 M GuHCl. Accordingly, crude **2a** seemed to contain a smaller portion of material that could be folded correctly as well as a larger portion that was unfoldable and overall failed characterization.

Notably, attempts at reducing the amount of solvent after folding by partial evaporation or by freeze-drying followed by redissolution, before samples were loaded on preparative HPLC, lowered the isolated yields. Also, any freezing and thawing of the folding mixture before preparative HPLC lowered the isolated yields.

By introducing EEEEK as the presequence and GEEEK as the C-peptide (60-mer peptide **3a**), we observed even better folding. With Cys (1 mM), cystine (8 mM), Tris (10 mM), glycine (10 mM) and EDTA (1 mM) at room temperature in an open vial at pH 9.5, we obtained **3b** in crude folding yields of 20–25% as quantified by LC-MS (Figure 3). For **3a**, we obtained the best folding yields in as little as 2–4 h (contrary to **2a**, for which folding was optimal after 2–3 days). The better folding of synthetic precursor **3a** compared to that of **2a** might have resulted from the more polar nature of the C-peptide (GEEEK vs. EWK), which might have led to the better exposure of this part of the sequence to solvent, and hence, to the tighter folding of the hydrophobic parts of the sequence towards the core of the structure, and thus, to better overall yield. Upon purification by HPLC, we isolated the folded precursor **3b** in yields of 10–12%, as calculated from the crude peptide material.

Careful LC-MS analysis of the residual unfolded material corresponding to the broad LC-MS peak from **3a** revealed the presence of several masses corresponding to (partially) folded product + multiples of 56 Da (Figure 4). These signals likely resulted from a number of cysteines that were irreversibly alkylated by *t*Bu cation during the TFA-based peptide deprotection, despite the presence of scavengers. Cys *tert*-butylation is a well-known side reaction during peptide deprotection, and we were not surprised that this was a significant problem in the present case considering the presence of six Cys and 20 *t*Bu-protected residues in the crude peptide **3a**. Notably, Cys(*t*Bu) can be deprotected with HF,[32] so it is possible that the

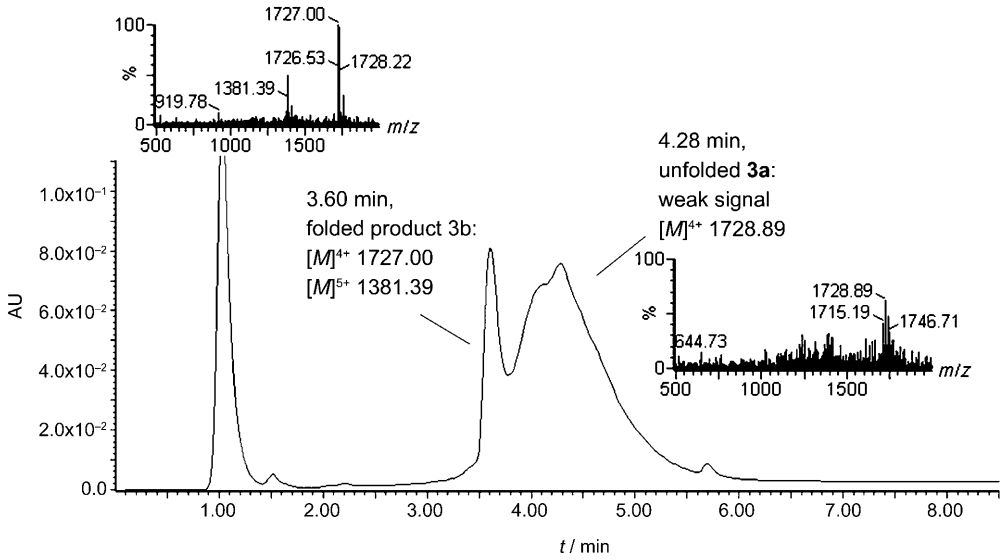


Figure 3. Crude reduced precursor **3a** (C-peptide GEEEK) showed trace MS peaks of the unfolded material at 4.28 min and folded precursor **3b** at 3.60 min (20–25%) with the expected product masses, $[M]^{4+}$ and $[M]^{5+}$.

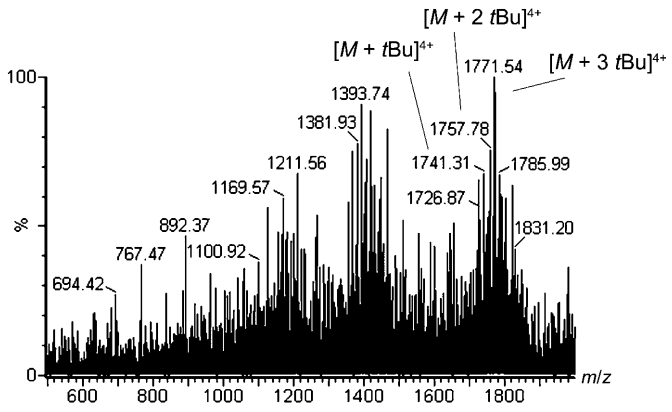


Figure 4. LC-MS of the unfolded material corresponding to the broad peak from peptide **3a** revealed the presence of partially folded *tert*-butylated side products as $[M+n56]^{4+}$ ions.

method can be improved with HF-compatible equipment, which were not available to us.

Enzymatic removal of the C-peptide and solubilising extension

We dissolved the purified single-chain peptides **2b** and **3b** separately (10 mg mL^{-1}) in carbonate buffer (pH 9.5), treated them with ALP (0.1%, *w*/*w*) and analysed the products by LC-MS. The single-chain peptides were quantitatively transformed into the two-chain insulin final product **4** (Figure 5). Notably, ALP processed the **2b** precursor faster than it did the **3b** precursor (1–2 h compared to 6–12 h), but in both cases the transformation proceeded quantitatively. The final crude products showed no significant byproducts by LC-MS analysis, although we identified product **4** with the C-peptide GEEEK still attached as a passing intermediate in the LC-MS analysis. The final purifications by HPLC proceeded in yields of approximately 50% and provided desB30 human insulin **4** in overall yield of 1% from **2b** (C-peptide EWK) and 6% from **3b** (C-peptide GEEEK), with yields calculated from the crude reduced peptides **2a**/**2b**. The purity of **4** was, in both cases, >98% in both acidic and neutral pH HPLC systems.

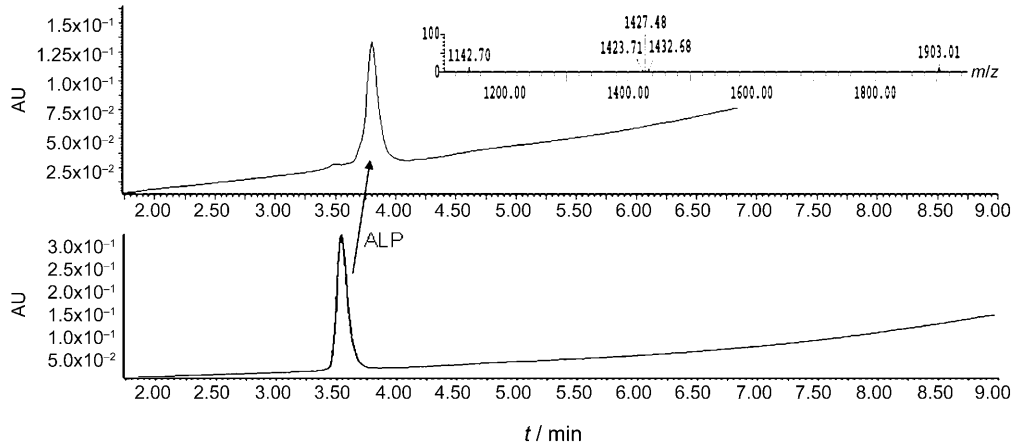


Figure 5. LC-MS of the ALP-catalysed transformation of folded single-chain precursor **3b** (3.55 min) into two-chain insulin **4** (3.80 min).

Notably, human insulin contains B30-threonine, and porcine insulin contains B30-alanine. The lysine used in the described ALP transformation was positioned at B29. Accordingly, the prepared products were desB30 insulins. However, the B30 position has little or no impact on the biological activities and biophysical properties of insulin. In fact, some insulin products used clinically are desB30 analogues.[33,34] If desired, a B30 residue can be coupled enzymatically to the B29 lysine with ALP and excess amino acid ester in a partial organic solvent, as already exploited for the transformation of porcine insulin to human insulin.[35]

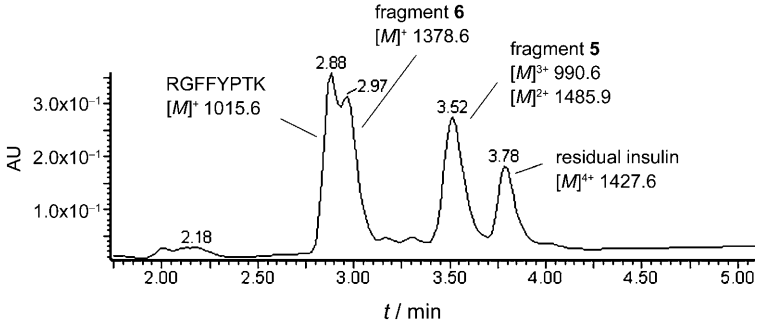


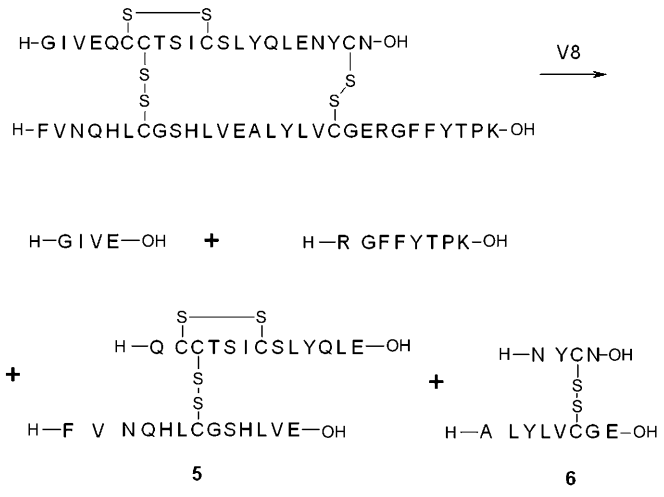
Figure 6. LC-MS showing the expected disulfide pattern with fragments **5** and **6** from V8 treatment of synthetic desB30 human insulin **4**.

Bioactivity and disulfide bridge characterization

To test whether the disulfide bonds were formed correctly we used two approaches. First, when the final product **4** was tested in an insulin receptor binding assay, it showed an affinity of 82% compared to human insulin. The high value indicates that the product was correctly folded. For comparison, recombinant desB30 human insulin **4** showed an affinity of 85% in the same assay. Insulin analogues containing mispaired disulfides display much lower insulin receptor affinities.[36]

Second, we treated product **4** with V8 protease, which is known to cleave peptides to the C terminus of their glutamyl residues. If the correct disulfide bridge pattern was in place, the fragments shown in Scheme 2 should theoretically be obtained from V8-treated **4**; the MS data shown in Figure 6 display the expected signals from fragments **5** and **6**. In theory, other disulfide bridge patterns could give rise to the same fragmentation pattern, but in combination with the high insulin receptor affinity, the results indicated that the native disulfide bridge pattern was present in the synthetic desB30 human insulin **4**.

In a demonstration of the scope of this new insulin synthesis strategy, we incorporated the nonproteinogenic amino acid Aib into synthetic insulin.[37] The B2Aib desB30 insulin analogue **7** was prepared by the described methods and showed an affinity of 80% compared to human insulin in the insulin receptor binding assay.



Scheme 2. Treatment of desB30 human insulin with V8 protease to provide fragments **5** and **6**, as detected by LC-MS.

Conclusions

In conclusion, we prepared desB30 human insulin and B2Aib desB30 human insulin chemically by the step-wise, Fmoc-based, solid-phase synthesis of single-chain precursors, preferably with EEEEK as the presequence and GEEEK as the chain-bridging C-peptide. The reduced intermediates were difficult to characterize but folded spontaneously under redox conditions to provide insulin with the correct pairing of the three disulfide bridges in folding yields of up to 25% (12% after purification). We identified *tert*-butylated side products, which likely contained Cys(*t*Bu), in the material that resisted folding. We transformed the purified single-chain precursors into bioactive two-chain insulin by a clean enzymatic transformation with ALP. We obtained overall yields of up to 6% from this synthesis strategy, and the method was robust, easy and not labour intensive. This provides a new route to the synthesis of insulin variants containing unnatural substitutions. Whether or not the method is applicable to insulin analogues with structure-disrupting substitutions, remains to be investigated.

Experimental Section

General procedures: Automated peptide synthesis was carried out by using an Applied Biosystems 433A peptide synthesizer or a Liberty microwave peptide synthesizer. Protection groups used during the peptide synthesis include Fmoc-Arg(Pbf), Asn(Trt), Cys(Trt), Gln(Trt), Glu(O*t*Bu), His(Trt), Lys(Boc), Ser(*t*Bu), Thr(*t*Bu), Trp(Boc) and Tyr(*t*Bu).

Chemicals were purchased from Sigma–Aldrich, Fluka, NovaBiochem or Iris Biotech and were used without further purification.

The LC-MS of **1a** and **1b** was performed on a Micromass LCT apparatus connected to a Waters HPLC by direct injection. The preparative HPLC of compound **1b** was performed by using a Waters system with a 600 control unit, 996 PDA detector, 717 Plus autosampler, Millenium 32 control software, a Delta 600 pump and no autosampler on a C4 RP column.

The LC-MS of compounds **2**–**7** was performed on a Waters Alliance 2795 instrument with an MS module ZQ/ZMD 4000 and MassLynx software. The preparative HPLC of compounds **2b**, **3b**, **4** and **7** was performed by using a Gilson unit with Trilution software and a C4 RP column. The analytical HPLC of compounds **2b**, **3b**, **4** and **7** was performed on a Dionex Summit instrument with Chromeleon software and C8 RP columns, with either an acidic or a neutral eluent system (0.1% TFA in H_2O to 80% CH_3CN over 20 min or 10 mM Tris, 15 mM $(NH_4)_2SO_4$ in H_2O to 80% CH_3CN over 20 min, pH 7.2).

Solid-phase peptide synthesis (SPPS) of 1a: The peptide was synthesized by using standard Fmoc chemistry. Amino acids (4 equiv) were activated with HBTU (3.8 equiv), DIPEA (7.8 equiv) and HOBt/HOAt (4 equiv, 80:20) in NMP. Rink–Tentagel resin (1 g, 0.25 mmol, 0.25–0.27 mmol g^{-1} approximate loading) was used, and the first residue was Fmoc-Asp-O*t*Bu (loaded by its side-chain to provide C-terminal Asn). The Fmoc-protected amino acids were side-chain protected when needed, and the chain elongation was carried out on an ABI 433A peptide synthesizer. All amino acids were double coupled for 45 min. Intermediate peptides and the final 53-mer precursor were cleaved from the resin with TFA/H_2O/TES (95:2.5:2.5) for 3 h followed by the partial removal of TFA in vacuo and product precipitation from diethyl ether. The analytical samples of intermediate peptides after 10, 21, 30, 35 and 40 couplings were collected, dissolved in CH_3CN/H_2O (1:1) and analyzed by LC-MS, except for the 21-mer intermediate, which required GuHCl (6 M) buffer solution for solubilisation before analysis by LC-MS. The crude 53-mer peptide was dried and used in folding studies. The yield of crude peptide, which was calculated from the initial loading of the resin, was approximately 445 mg (30%) with a protein concentration of ≈100% according to UV at 280 nm with $\varepsilon =$ 5840 L mol^{-1} cm^{-1}.

SPPS of EEEK-EWK-precursor 2a and EEEEK—GEEEK–precursor 3a: The peptides were synthesized with standard Fmoc chemistry on a Liberty microwave-assisted peptide synthesizer (CEM Corporation, North Caroline, USA). Rink-ChemMatrix (0.5 g, 0.5 mmol g^{-1} approximate loading, Matrix Innovation, Montreal, Canada) was used. Couplings were performed with a sixfold excess of amino acid, activated in situ with DIC/HOAt, for 5 min at up to 70 °C. Deprotection was performed with 5% piperidine in NMP for up to 3 min at 70 °C. After synthesis, the peptide was cleaved from the resin by treatment with TFA/TIPS/H_2O/DMB/DTT (89:5:2:2:2, 20 mL) for 2 h and precipitated by the addition of the cleavage mixture to diethyl ether (70 mL).

Folding of 1a to give 1b

Method I: Peptide **1a** (160 mg, 27 μmol) was dissolved in a Tris buffer (200 mM Tris, 6 M GuHCl, pH 7.5; 40 mL, 0.6 mM protein concentration). The crude peptide **1a** contained traces of TFA from the cleavage step, and the pH was readjusted to 7.5 before the mixture was incubated at 37 °C for 3 h to provide full dissolution. The sample was diluted to a total volume of 120 mL with GuHCl buffer (6 M, pH 9) to a total protein concentration of 0.2 mM. The peptide solution was then dialyzed against Tris buffer (20 mM, pH 9) containing decreasing amounts of GuHCl. Precipitation was observed at GuHCl concentrations below 2 M. Finally, the solution was dialysed against H_2O, lyophilized, and the crude solid was purified by RP preparative HPLC on a C4 column to provide ≈3 mg of product (2%).

Method II: Peptide **1a** (70 mg, 12 μmol) was dissolved in a Tris buffer (10 mM Tris, 6 M GuHCl, 1 mM EDTA, 0.5 mM Cys, 4.5 mM cystine, pH 9.5; 200 mL, 0.6 mM final peptide concentration). The crude peptide **1a** contained traces of TFA from the cleavage step, and the pH was readjusted to 9.5 before the mixture was incubated for 48 h. The solution was dialyzed against phosphate buffer (10 mM), eluted through a desalting column and purified by RP preparative HPLC on a C4 column to provide ≈1 mg of product (1.5%).

Method III: Peptide **1a** (160 mg, 27 μmol) was dissolved in a Tris buffer (200 mM Tris, 6 M GuHCl, pH 7.5, 8 mL). The crude peptide **1a** contained traces of TFA from the cleavage step, and the pH was readjusted to 7.5 before the mixture was incubated at 37 °C. Na_2SO_3 (128 mg, 1.02 mmol) was added, the mixture was stirred for 10 min, $Na_2S_4O_6$ (144 mg, 0.53 mmol) was added, and the mixture was incubated for 60 min at 37 °C before the $Na_2SO_3/Na_2S_4O_6$ addition was repeated. The mixture was incubated for an additional 2 h at room temperature in the dark, filtered, dialyzed extensively against H_2O for approximately 2 days and lyophilized. Sulfonated **1a** (93 mg, 52%) was dissolved in freshly made NH_4HCO_3 buffer (50 mM NH_4HCO_3, pH 9.5, 400 mL). Cys (20 mg, 0.17 mmol) was added, and the mixture was incubated at 4 °C, overnight. The solution was dialyzed against H_2O for 2 h and against NH_4HCO_3 (50 mM, pH 9.5) overnight, dialyzed extensively against H_2O for up to 2 days and lyophilized. The product was purified by RP preparative HPLC on a C4 column to provide approximately 2.5 mg (1.5%).

Folding of 3a to give 3b: Crude peptide **3a** (338 mg, 49 μmol) was suspended in a Tris buffer (10 mM Tris, 1 mM Cys, 8 mM cystine, 10 mM glycine, 1 mM EDTA, pH 10.5, 101 mL, 3.3 mg mL^{-1}, 0.5 mM peptide concentration). The pH value was readjusted to 10.5 with NaOH (1 M), and the mixture was shaken occasionally to give full dissolution within 15–30 min, at which time the pH had decreased to 9.5. The folding was monitored by LC-MS. After 4 h (3 days in the case of **2a**), the folding mixture was divided into five portions, and each was purified by preparative HPLC on a C4 column (2×25 cm) with buffers A (0.1% TFA in H_2O) and B (0.1% TFA in CH_3CN) in a gradient of 22–85% B over 40 min with a flowrate of 9 mL min^{-1}. The product **3b** was isolated by freeze-drying (40 mg, 12%). The purity was >98% by analytical HPLC with either TFA (0.1%) or Tris (10 mM, pH 7.2) buffer.

ALP treatment to give 4: Folded precursor **3b** (27 mg, 4 μmol) was dissolved in a bicarbonate buffer (50 mM $NaHCO_3$, pH 9.5, 2.7 mL, 1%, *w/w* peptide/buffer). ALP was added from a stock solution (5.4 mg mL^{-1}, dissolved in 50 mM sodium acetate, 160 mM NaCl, 50 μL, 1% *w/w* enzyme/peptide). The reaction was monitored by LC-MS. The reaction was complete after 6–8 h (1–2 h in the case of **2b**). The product was purified by HPLC on a C4 column (2× 25 cm) with buffers A (0.1% TFA in H_2O) and B (0.1% TFA in CH_3CN) in a gradient of 23–50% B over 30 min with a flowrate of 5 mL min^{-1}. Product **4** was isolated by freeze-drying (11 mg, 50% for this step and 6% overall from the crude peptide **3a**). The purity was >98% by analytical HPLC with either TFA (0.1%) or Tris (10 mM, pH 7.2) buffer.

V8 treatment to analyse the disulfide bridge pattern of 4: Product **4** (1 mg, 0.18 μmol) in phosphate buffer (0.1 M, pH 7.5, 300 μL) was treated with V8 glutamyl endopeptidase (6 μL, 1 mg mL^{-1}, 2%, *w/w*). The mixture was monitored by LC-MS, which documented the formation of products **5** and **6** (Figure 6); these support the presence of the native disulfide bridge pattern.

Receptor-affinity assay: This was a scintillation proximity assay in which partially purified insulin receptors were used. SPA PVT anti-mouse beads (Amersham Biosciences, UK) were incubated with IR antibody 83-7 and insulin receptors for 5 h at room temperature. The SPA beads were washed twice with buffer to remove any re-

ceptors not bound to the SPA beads, and ^{125}I-human insulin was added. A twofold dilution series of human insulin or insulin analogue (starting from 100 nM), was prepared in a Packard Optiplate 96, and the SPA beads were added. The final concentration of ^{125}I-human insulin was 5000 cpm per 200 μL, and the buffer composition was HEPES (100 mM), pH 7.8, NaCl (100 mM), $MgCl_2$ (10 mM) and Tween-20 (0.025%). The plate was shaken gently for 18 h at room temperature, centrifuged and counted in a Topcounter. The EC_{50} values were calculated by nonlinear regression analysis.

Abbreviations: Acm: acetamidomethyl, Aib: α-aminoisobutyric acid, ALP: *Achromobacter lyticus* protease, Boc: *tert*-butyloxycarbonyl, DIC: *N*,*N*′-diisopropylcarbodiimide, DMB: 3,5-dimethoxybenzene, DMSO: dimethylsulfoxide, DTT: 1,4-dithiothreitol, EDTA: ethylenediaminetetraacetic acid, Fmoc: 9-fluorenylmethoxycarbonyl, HBTU: *N*-[(1H-benzotriazol-1-yl)(dimethyl amino)methylidene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide, HEPES: *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid, HOAt: 1-hydroxy-7-azabenzotriazole, HOBt: 1-hydroxybenzotriazole, GuHCl: guanidine hydrochloride, LC-MS: liquid chromatography mass spectroscopy, MESNa: sodium 2-mercaptoethanesulphonate, NMP: *N*-methyl-2-pyrrolidone, Pbf: 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl, RP: reverse phase, SPA: scintillation proximity assay, *t*Bu: *tert*-butyl, TCEP: Tris-(2-carboxyethyl)phosphine, TES: triethylsilane, TIPS: triisopropylsilane, TFA: trifluoroacetic acid, TMS: trimethylsilyl, Trt: trityl, UV: ultraviolet spectroscopy, V8: glutamyl endopeptidase.

Acknowledgements

Support from the Ministry of Science, Technology and Innovation, and Novo Nordisk Corporate Research Affairs for an industrial Ph.D. stipend to A.P.T. is gratefully acknowledged.

Keywords: *Achromobacter lyticus* protease · enzyme catalysis · insulin · protein folding · solid-phase synthesis

[1] F. Banting, C. Best, J. Collip, W. Campbell, A. Fletcher, *Can. Med. Assoc. J.* **1922**, *12*, 141–146.
[2] The Diabetes Control and Complications Trial Research Group, *N. Engl. J. Med.* **1993**, *329*, 977–986.
[3] The UK-PDS Study Group, *Lancet* **1998**, *352*, 837–853.
[4] A. Lindholm, *Best Pract. Res. Clin. Gastroenterol.* **2002**, *16*, 475–492.
[5] W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239–244.
[6] P. G. Katsoyannis, *Science* **1966**, *154*, 1509–1514.
[7] F. Maerki, G. M. De, K. Eisler, B. Kamber, B. Riniker, W. Rittel, P. Sieber, *Hoppe-Seyler's Z. Physiol. Chem.* **1979**, *360*, 1619–1632.
[8] J. P. Mayer, F. Zhang, R. D. DiMarchi, *Peptide Sci.* **2007**, *88*, 687–713.
[9] Q. X. Hua, S. Q. Hu, W. H. Jia, Y. C. Chu, G. T. Burke, S. H. Wang, R. Y. Wang, P. G. Katsoyannis, M. A. Weiss, *J. Mol. Biol.* **1998**, *277*, 103–118.
[10] J. Meienhofer, E. Schnabel, H. Bremer, O. Brinkhoff, R. Zabel, W. Sroka, H. Klostermayer, D. Brandenbrug, T. Okuda, H. Zahn, *Z. Naturforschung* **1963**, *18B*, 1120–1121.
[11] A. Marglin, R. B. Merrifield, *J. Am. Chem. Soc.* **1966**, *88*, 5051–5052.
[12] K. Akaji, K. Fujino, T. Tatsumi, Y. Kiso, *J. Am. Chem. Soc.* **1993**, *115*, 11384–11392.
[13] C. S. Samuel, F. Lin, M. A. Hossain, C. Zhao, T. Ferraro, R. A. D. Bathgate, G. W. Tregear, J. D. Wade, *Biochemistry* **2007**, *46*, 5374–5381.
[14] Q. X. Hua, S. H. Nakagawa, W. Jia, K. Huang, N. B. Phillips, S. Q. Hu, M. A. Weiss, *J. Biol. Chem.* **2008**, *283*, 14703–14716.
[15] Y. Sohma, B. L. Pentelute, J. Whittaker, Q. X. Hua, L. J. Whittaker, M. A. Weiss, S. B. Kent, *Angew. Chem.* **2008**, *120*, 1118–1122; *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 1102–1106.
[16] G. Milazzo, L. Sciacca, V. Papa, I. D. Goldfine, R. Vigneri, *Mol. Carcinog.* **1997**, *18*, 19–25.
[17] J. L. Ottesen, P. Nilsson, J. Jami, D. Weilguny, M. Duhrkop, D. Bucchini, S. Havelund, J. M. Fogh, *Diabetologia* **1994**, *37*, 1178–1185.
[18] D. F. Steiner, *N. Engl. J. Med.* **1969**, *280*, 1106–1113.
[19] A. Marglin, *Peptide Sci.* **2008**, *90*, 200–202.
[20] J. Markussen, K. H. Jørgensen, A. R. Sørensen, L. Thim, *Int. J. Pept. Protein Res.* **1985**, *26*, 70–77.
[21] S.-G. Chang, D.-Y. Kim, K.-D. Choi, J.-M. Shin, H.-C. Shin, *Biochem. J.* **1998**, *329*, 631–635.
[22] T. Kjeldsen, S. Ludvigsen, I. Diers, P. Balschmidt, A. R. Sørensen, N. C. Kaarsholm, *J. Biol. Chem.* **2002**, *277*, 18245–18248.
[23] T. Masaki, M. Tanabe, K. Nakamura, M. Soejima, *Biochim. Biophys. Acta Enzymol.* **1981**, *660*, 44–50.
[24] S. A. Palasek, Z. J. Cox, J. M. Collins, *J. Pept. Sci.* **2007**, *13*, 143–148.
[25] E. C. B. Johnson, E. Malito, Y. Q. Shen, D. Rich, W. J. Tang, S. B. H. Kent, *J. Am. Chem. Soc.* **2007**, *129*, 11480–11490.
[26] E. C. B. Johnson, S. B. H. Kent, *Tetrahedron Lett.* **2007**, *48*, 1795–1799.
[27] J. Winter, H. Lilie, R. Rudolph, *Anal. Biochem.* **2002**, *310*, 148–155.
[28] W. D. Busse, S. R. Hansen, F. H. Carpenter, *J. Am. Chem. Soc.* **1974**, *96*, 5949–5950.
[29] W. D. Busse, F. H. Carpenter, *J. Am. Chem. Soc.* **1974**, *96*, 5947–5949.
[30] S. H. Wang, S. Q. Hu, G. T. Burke, P. G. Katsoyannis, *J. Protein Chem.* **1991**, *10*, 313–324.
[31] C. C. Wang, C. L. Tsou, *Trends Biochem. Sci.* **1991**, *16*, 279–281.
[32] L. Moroder, H. J. Musiol, M. Gotz, C. Renner, *Biopolymers Pept. Sci.* **2005**, *80*, 85–97.
[33] P. Home, P. Bartley, D. Russell-Jones, H. Hanaire-Broutin, J.-E. Heeg, P. Abrams, M. Landin-Olsson, B. Hylleberg, H. Lang, E. Draeger, *Diabetes Care* **2004**, *27*, 1081–1087.
[34] I. Jonassen, S. Havelund, U. Ribel, A. Plum, M. Loftager, T. Hoeg-Jensen, A. Volund, J. Markussen, *Pharm. Res.* **2006**, *23*, 49–55.
[35] J. Markussen, U. Damgaard, M. Pingel, L. Snel, A. R. Sorensen, E. Sorensen, *Diabetes Care* **1983**, *6 Suppl. 1*, 4–8.
[36] P. Sieber, K. Eisler, B. Kamber, B. Riniker, W. Rittel, F. Märki, G. M. De, *Hoppe-Seyler's Z. Physiol. Chem.* **1978**, *359*, 113–123.
[37] S. H. Hu, G. T. Burke, G. P. Schwartz, N. Ferderigos, J. B. A. Ross, P. G. Katsoyannis, *Biochemistry* **1993**, *32*, 2631–2635.

Received: June 25, 2008
Published online on November 26, 2008

