ARTICLE

OBC
www.rsc.org/obc

Monosaccharide templates for *de novo* designed 4-α-helix bundle proteins: template effects in carboproteins †

Jesper Brask,[a,b] Jan M. Dideriksen,[b] John Nielsen[a] and Knud J. Jensen *[a]

[a] *Department of Chemistry, Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg, Denmark*

[b] *Department of Chemistry, Technical University of Denmark, Building 201, Kemitorvet, DK-2800 Kgs. Lyngby, Denmark*

Received 19th February 2003, Accepted 13th May 2003
First published as an Advance Article on the web 29th May 2003

De novo design and total chemical synthesis of proteins provide powerful approaches to critically test our understanding of protein folding, structure, and stability. The 4-α-helix bundle is a frequently studied structure in which four amphiphilic α-helical peptide strands form a hydrophobic core. Assembly of protein models on a template has been suggested as a way to reduce the entropy of folding. We have previously developed the concept of carbohydrates as templates in the *de novo* design of protein models termed 'carboproteins'. Here we present the chemical synthesis of three 8.1 kDa 4-α-helix bundles by oxime ligation of tetra-aminooxyacetyl functionalized D-galacto-, D-gluco-, and D-altropyranoside templates with an amphiphilic *C*-terminal hexadecapeptide aldehyde sequence. CD spectroscopy indicated that the choice of template has an effect on the overall *structure* of the carboprotein, as the altro-based carboprotein was found to be more α-helical than the corresponding galacto- and gluco-carboproteins. However, an influence on *stability* could not be detected in the present experiments, as the three carboproteins gave similar free energy of foldings (ΔG^{H2O}_F) and melting points in chemical and thermal denaturation experiments.

Introduction

It is a continuing challenge for bioorganic chemistry and structural biology to design novel molecules that mimic the three-dimensional structure and function of proteins. Protein *de novo* design offers the ultimate test of our understanding of the factors governing protein structure, folding, and stability.[1–3] Not only can complex interactions in natural proteins be studied in greater detail using smaller *de novo* designed systems, the approach also offers the prospect of access to tailor-made proteins. Among the most frequently used structural motifs for *de novo* design is the 4-α-helix bundle.[1–3] Its folding is driven by the hydrophobic collapse of amphiphilic helices and is guided by the position and nature of loop regions, by electrostatic effects, and by shape complementarity in side-chain packing.

The entropic cost in going from a 'random coil' conformation to a restrained protein structure is very significant. From the standpoint of *de novo* design, the entropic barrier imposes some restrictions and limitations, *e.g.*, the need for long sequences to provide the scaffold necessary to ensure a stable structure. To lower this barrier, Mutter and co-workers have suggested pre-organizing peptide strands on a molecular template to reduce the entropy of the construct's unfolded state.[4] Hence, the branched structures of template-assembled synthetic proteins (TASPs) facilitate folding of protein models otherwise not possible with linear sequences. Here the peptide sequences, which are secondary structure elements, are held together by the template, typically in a parallel manner. Whereas Mutter *et al.*[5–8] and Haehnel *et al.*[9–12] have mainly explored linear or cyclic peptide templates, other research groups have applied other, non-peptide templates in protein designs. These include porphyrin derivatives,[13,14] metal ions (by complexation),[15–17] a cyclohexane derivative (Kemp's triacid),[18] substituted phenyl rings,[19] and aromatic macrocycles.[20,21] However, the significance of the template geometry has been debated.[19,21] Does the template merely tie the peptide strands together, or can it with the proper choice of geometry between 'anchoring points' be used to affect the folding? Based on 4-α-helix bundle TASP structures assembled on a series of aromatic and cyclized aromatic templates, Fairlie and co-workers have concluded that at least with a sufficiently long linker between the peptide strands and the template, the template geometry is of less importance.[19] However, although these templates had different sizes, the peptide anchoring points were all in the same plane and had the same directionality within the same template.[19]

Carboproteins are protein models assembled on carbohydrate templates.[22,23] We have previously developed an efficient strategy for the synthesis of carboproteins, in which *C*-terminal peptide aldehydes are ligated by oxime bond formation to tetra-aminooxyacetyl functionalized monosaccharide templates.[24–27] The relatively rigid $^4C_1$ conformation of monosaccharides enforces well-defined axial or equatorial orientations of the hydroxyls. We have previously demonstrated that a four-stranded carboprotein has significantly higher degree of α-helicity and stability than the corresponding single-stranded carbopeptide, which has just one peptide chain attached to the monosaccharide template.[25] Also, the cooperativity in unfolding observed for carboproteins was completely absent in the carbopeptide. The question addressed in this paper is whether the structure and stability of carboproteins can be directed by the axial or equatorial orientation of hydroxyl anchoring points in the template, *i.e.*, by choice of the monosaccharide template. Such control of structure or stability could not only provide insights into the controlling effects of templates but also become a useful tool in the *de novo* design of proteins.

Materials and methods

General procedures

NovaSyn TG resin and Fmoc-amino acids for peptide synthesis were from Novabiochem (Läufelfingen, Switzerland), HBTU and HOBt were from QBiogene (Illkirch, France), while all other commercial compounds were purchased from Sigma-Aldrich (Copenhagen, Denmark). ESI-MS spectra were

† See ref. 43.

DOI: 10.1039/b301948a

obtained on a Micromass LCT instrument by direct injection of an aqueous solution of the lyophilized product. Deconvolution was performed with the MassLynx transform algorithm. ^{1}H and ^{13}C NMR spectra were recorded at 500 and 125.7 MHz, respectively, on a Varian Inova 500 spectrometer. HPLC analyses were carried out on a Waters system (600 control unit, 996 PDA detector, 717 Plus autosampler, Millennium32 control software) equipped with either a Waters Symmetry300 C_{18} 5 μm column, or a Waters Symmetry300 C_4 5 μm column, both 3.9 × 150 mm. Preparative HPLC was carried out on a similar Waters system (with a Delta 600 pump) equipped either with stack of three 40 × 100 mm column cartridges of Waters Prep Nova-Pak HR C_{18} 6 μm 60 Å, or with a single 25 × 100 mm column cartridge of Waters Delta-Pak C_4 15 μm 300 Å. Peptides and carboproteins were eluted with mixtures of acetonitrile and H_2O, both containing 0.1% TFA.

Preparation of templates 1, 2, and 3

Glc*p* and Alt*p* templates **2** and **3** were prepared as previously described for Gal*p* template **1**.[24]

Methyl 2,3,4,6-tetra-*O*-(Boc_2-Aoa)-α-D-Glc*p* was obtained in 61% yield from methyl α-D-Glc*p*. ^{1}H NMR ($CDCl_3$), δ: 5.57 (t, J = 9.8 Hz, 1H, H-3), 5.11 (t, J = 9.8 Hz, 1H, H-4), 5.04 (d, J = 3.4 Hz, 1H, H-1), 4.91 (dd, J = 10.2 Hz, J = 3.4 Hz, 1H, H-2), 4.57–4.44 (m, 8H, $COCH_2ON$–), 4.34–4.31 (m, 2H, H-6), 4.10–4.07 (m, 1H, H-5), 3.40 (s, 3H, –OCH_3), 1.54–1.53 (m, 72H, Boc CH_3). ^{13}C NMR ($CDCl_3$), δ: 167.3–167.0 (–*C*OCH_2ON–), 150.8–150.6 (Boc CO), 96.9 (C-1), 85.1–85.0 (Boc –*C*$(CH_3)_3$), 73.4–72.8 (–CO*C*H_2ON–), 71.9 (C-2/3/4), 71.1 (C-2/3/4), 70.1 (C-2/3/4), 67.4 (C-5), 63.2 (C-6), 56.2 (–OCH_3), 28.9–28.7 (Boc CH_3). Boc deprotection proceeded in quantitative yield to provide template **2**. ^{1}H NMR (CD_3OD), δ: 5.61 (t, J = 9.9 Hz, 1H, H-3), 5.27 (t, J = 9.6 Hz, 1H, H-4), 5.11 (dd, J = 9.7 Hz, J = 3.5 Hz, 1H, H-2), 5.04 (d, J = 3.1 Hz, 1H, H-1), 4.58–4.37 (m, 10H, –$COCH_2ON$– + H-6), 4.18–4.14 (m, 1H, H-5), 3.45 (s, 3H, OCH_3). ^{13}C NMR (CD_3OD), δ: 169.7–169.1 (–*C*OCH_2ON–), 97.2 (C-1), 71.8–71.5 (–CO*C*H_2ON–), 71.4 (C-2/3/4), 71.2 (C-2/3/4), 69.7 (C-2/3/4), 67.5 (C-5), 63.2 (C-6), 55.4 (–OCH_3).

Methyl 2,3,4,6-tetra-*O*-(Boc_2-Aoa)-α-D-Alt*p* was obtained in 70% yield from methyl α-D-Alt*p*. ^{1}H NMR ($CDCl_3$), δ: 5.30 (t, J = 3.3 Hz, 1H, H-3), 5.20 (dd, J = 8.8 Hz, J = 3.1 Hz, 1H, H-4), 5.10 (dd, J = 3.4 Hz, J = 0.8 Hz, 1H, H-2), 4.65 (d, J = 1.8 Hz, 1H, H-1), 4.62–4.47 (m, 8H, –$COCH_2ON$–), 4.33–4.36 (m, 3H, H-5 + H-6), 3.38 (s, 3H, –OCH_3), 1.55–1.53 (m, 72H, Boc CH_3). ^{13}C NMR ($CDCl_3$), δ: 167.4–166.4 (–*C*OCH_2ON–), 150.7–150.6 (Boc CO), 99.0 (C-1), 85.2–85.0 (Boc –*C*$(CH_3)_3$), 72.9–72.8 (–CO*C*H_2ON–), 70.1 (C-2), 68.3 (C-3/4), 66.5 (C-3/4), 65.0 (C-5/6), 64.1 (C-5/6), 56.4 (–OCH_3), 28.7 (Boc CH_3). Boc deprotection proceeded in quantitative yield to provide template **3**. ^{1}H NMR (CD_3OD), δ: 5.49 (t, J = 4.1 Hz, ^{1}H, H-3), 5.29 (dd, J = 9.2 Hz, J = 3.5 Hz, ^{1}H, H-4), 5.17 (dd, J = 3.6 Hz, J = 0.8 Hz, ^{1}H, H-2), 4.80 (d, J = 0.8 Hz, ^{1}H, H-1), 4.70–32.42 (m, 11H, –$COCH_2ON$– + H-6 + H-5), 3.45 (s, 3H, –OCH_3). ^{13}C NMR (CD_3OD), δ: 169.9–32.2 (–*C*OCH_2ON–), 99.1 (C-1), 71.4–71.2 (–CO*C*H_2ON–), 70.4 (C-2), 68.3 (C-3/4), 66.7 (C-3/4), 65.1 (C-5/6), 64.0 (C-5/6), 55.7 (–OCH_3).

Preparation of model compound 7

Template **2** (25 mg, 27 μmol, incl. 4 × TFA) was dissolved in 0.1 M NaOAc buffer, pH 4.76 (1 mL) and freshly distilled CH_3CHO (1 mL) was added. The solution was stirred for 14 h, then concentrated *in vacuo*, taken up in H_2O (5 mL), and purified by prep. C_{18} RP-HPLC. Compound **7** was isolated as a colorless oil. Yield 14 mg, 89%. ^{1}H NMR ($CDCl_3$), δ: 7.56–7.47 (m, 2H, *E*-oxime), 6.85–6.73 (m, 2H, *Z*-oxime), 5.60 (t, J = 9.9 Hz, ^{1}H, H-3), 5.18–5.12 (m, ^{1}H, H-4), 4.98–4.95 (m, 2H, H-1 + H-2), 4.67–4.47 (m, 8H, –$COCH_2ON$–), 4.32–4.25 (m, 2H, H-6), 4.06–4.04 (m, ^{1}H, H-5), 3.41 (s, 3H, –OCH_3), 1.92–1.83 (m, 12H, –ON=CHCH_3). ^{13}C NMR ($CDCl_3$), δ: 170.5–169.6 (–*C*OCH_2ON–), 149.8–149.4 (–ON=*C*HCH_3), 97.4 (C-1), 71.6 (C-2/3/4), 70.8–70.3 (–CO*C*H_2ON– + C-2/3/4), 69.6 (C-2/3/4), 67.7 (C-5), 63.1 (C-6), 56.2 (–OCH_3), 15.7–15.6 (–ON=CH*C*H_3), 12.7–12.6 (–ON=CH*C*H3). ESI-MS calcd for $C_{23}H_{34}N_4O_{14}$: 590.21 Da. Found: m/z 591.11 $[M + H]^+$, 613.1 $[M + Na]^+$, 629.1 $[M + K]^+$.

Preparation of peptide aldehyde Ac-YEELLKKLEELLKKAG-H

TG resin (5 g, 0.29 mmol g^{-1}) was placed in a polypropylene syringe equipped with a polyethylene filter. *o*-PALdehyde (819 mg, 2.9 mmol), HBTU (1.10 g, 2.9 mmol), HOBt (444 mg, 2.9 mmol), and DIPEA (993 μL, 5.8 mmol) were dissolved in DMF (20 mL), and transferred to the resin after 5 min. After 4 h, the resin was washed with DMF (3×) and DCM (3×), treated with Ac_2O–DCM (1 : 3, 20 mL) for 30 min, and washed with DCM (5×) and DMF (5×). Next, $H_2NCH_2CH(OCH_3)_2$ (1.56 mL, 14.5 mmol) and $NaBH_3CN$ (911 mg, 14.5 mmol) were dissolved in DMF–AcOH (99 : 1, 20 mL) and added to the resin. After 18 h, the resin was washed with DMF (5×) and DCM (5×), and dried *in vacuo*. A portion of the dried resin (2.0 g) was transferred to another syringe, and reacted with $(Fmoc\text{-}Ala)_2O$, formed from Fmoc-Ala-OH (1.81 g, 5.8 mmol) and DIPCDI (449 μL, 2.9 mmol) in DCM–DMF (9 : 1, 20 mL) over 15 min. After 2 h, the resin was washed with DMF (3×) and DCM (3×), and the coupling repeated. After another 2 h, and the same washing procedure, the resin was treated with Ac_2O–DCM (1 : 3, 20 mL) for 30 min, washed with DCM (5×), and dried *in vacuo*. By Fmoc deprotection of a resin sample, and measuring the absorption of the resulting piperidine-dibenzofulvene adduct at 290 nm, the resin substitution was calculated as 0.15 mmol g^{-1}. Next, the functionalized resin (1.15 g) was transferred to a MilliGen 9050 PepSynthesizer, and the remaining part of the sequence (H-YEELLKKLEELLKK-) synthesized by the FastMoc protocol, using 4 eq. of HBTU, 4 eq. of HOBt, and 8 eq. of DIPEA in each activation step. Fmoc-Tyr(*t*Bu)-OH, Fmoc-Glu(*t*Bu)-OH, Fmoc-Leu-OH, and Fmoc-Lys(Boc)-OH protected amino acids were used. Following chain elongation, the resin was treated with Ac_2O–DCM (1 : 3, 10 mL) for 30 min, washed with DCM (5×), and dried *in vacuo*. Finally the peptide was cleaved off with TFA–H_2O (19 : 1, 10 mL) for 2 h. The resin was filtered and washed with TFA (3 × 5 mL), and the combined cleavage mixture and washings were concentrated, dissolved in H_2O (10 mL), and lyophilized. The resulting off-white powder was purified by prep. C_{18} RP-HPLC to give the peptide aldehyde as a white powder after lyophilization. Yield 43 mg; 10% (incl. 4 × TFA). ESI-MS, calcd for $C_{90}H_{152}N_{20}O_{26}$: 1929.12 Da. Found: m/z 956.55 $[M - H_2O + 2H]^{2+}$, 644.29 $[M + 3H]^{3+}$, 483.63 $[M + 4H]^{4+}$.

Preparation of carboproteins 4, 5, and 6

All three carboproteins were prepared by dissolving template **1**, **2**, or **3** and peptide aldehyde (6 eq., 50% excess) in 0.1 M NaOAc buffer, pH 4.76. After 2 h, excess peptide aldehyde was removed by prep. C_4 RP-HPLC purification. All three carboproteins were obtained in quantitative yields after lyophilization (21–32 mg, incl. 4 × TFA). ESI-MS, calcd for $C_{375}H_{626}N_{84}O_{114}$: 8135.59 Da (average isotope distribution). Found: m/z 678.80 $[M + 12H]^{12+}$, 740.36 $[M + 11H]^{11+}$, 814.17 $[M + 10H]^{10+}$, 904.44 $[M + 9H]^{9+}$, 1017.25 $[M + 8H]^{8+}$, 1162.30 $[M + 7H]^{7+}$, 1355.57 $[M + 6H]^{6+}$, 1626.39 $[M + 5H]^{5+}$. The spectra were deconvoluted to the putative molecular ion, M = 8136 Da (see Fig. 2).

Analytical techniques

SEC was performed on the analytical HPLC system using an Amersham Biotech Superdex 75 HR 10/30 column, eluting with

1 mL min^{-1} 50 mM phosphate, 100 mM NaCl, pH 7.0. 100 μL carboprotein sample was injected from a 200 μM solution. The SEC column was calibrated with albumin, ovalbumin, chymotrypsinogen A, ribonuclease A, aprotinin, and vitamin B_{12}. CD spectra were recorded on a Jasco J-710 instrument. Carboprotein solutions were 10–20 μM (concentration determined from the Tyr absorption at 275 nm) in 50 mM phosphate, pH 7.0. The CD spectrum of **7** was recorded from a 100 μM solution in methanol, and the MRE calculated based on four residues per template. ANS fluorescence was measured with a SLM Aminco spectrofluorometer. Excitation was at 370 nm, ANS concentration was 20 μM, and carboprotein concentration 200 μM in 50 mM phosphate buffer, pH 7.0.

Fitting procedure

Non-linear fits were computed with Gnuplot for MS-Windows 32 bit, version 3.7. GuHCl denaturation data were fitted to the generalized equation $[\theta]_{222} = [\theta]_N f_N(1 - c_1[\mathrm{GuHCl}]) + [\theta]_D(1 - f_N)(1 - c_2[\mathrm{GuHCl}])$, in which c_1 and c_2 are constants and f_N, the fraction of folded protein, is obtained from the relationship $\Delta G_F = \Delta G_F^{H2O} + m[\mathrm{GuHCl}] = -RT\ln(f_N/(1 - f_N))$. Thermal denaturation data were fitted to the equation $[\theta]_{222} = [\theta]_N f_N + [\theta]_D(1 - f_N)$, with $\Delta G_F = -RT\ln(f_N/(1 - f_N))$ given by $\Delta G_F = \Delta H_F(1 - T/T_m) - \Delta C_p(T_m - T + T\ln(T/T_m))$, in which ΔC_p is the change in heat capacity associated with folding.

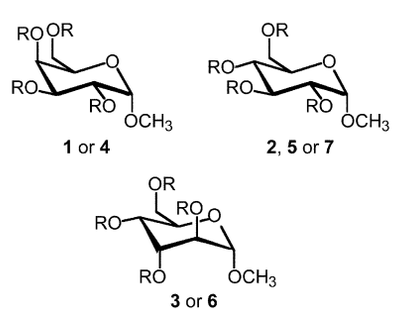
Results and discussion

Design, preparation and characterization

The previously reported carboproteins were all assembled on D-galactopyranoside (Gal*p*) templates.[22–27] For the present study, new templates based on D-glucopyranoside (Glc*p*) and D-altropyranoside (Alt*p*) were prepared and compared to the Gal*p* template (Fig. 1). The Alt*p*-template was of special interest due to the *trans* diaxial arrangement of the O-2 and O-3 hydroxyls. With MMFF-minimized structures (Maestro/Macromodel software) of α-D-Gal*p* and α-D-Alt*p*, distances were measured from O-2, O-3 and O-4 perpendicular to a plane defined by C-2, C-3 and C-5 in the monosaccharides. With a negative value indicating a position below the plane, the following distances (in Å) were measured for Gal*p*: −0.78 (O-2), +0.66 (O-3) and +2.09 (O-4). The corresponding distances for Alt*p* were: +1.34 (O-2), −1.35 (O-3) and +0.48 (O-4). Assuming that the peptide strands are aligned perpendicular to this plane, peptide strands anchored to O-2 and O-3 are offset with (0.78 + 1.34) + (0.66 + 1.35) Å = 4.13 Å in a Gal*p*-carboprotein, compared to an Alt*p*-carboprotein.

Functionalized templates **1**, **2**, and **3** were prepared from methyl α-D-Gal*p*, methyl α-D-Glc*p*, and methyl α-D-Alt*p*, respectively, by per-*O*-acylation with Boc_2NOCH_2COOH (Boc_2-Aoa-OH) followed by acidolytic deprotection, as previously described for **1**.[24] The *C*-terminal peptide aldehyde used for all three carboproteins, Ac-Tyr-Glu-Glu-Leu-Leu-Lys-Lys-Leu-Glu-Glu-Leu-Leu-Lys-Lys-Ala-Gly-H, was prepared by solid-phase peptide synthesis (SPPS) using a backbone amide linker (BAL) strategy.[28–31] Inexpensive 2,2-dimethoxyethylamine was anchored by the amino function through *o*-BAL to a solid support. Acylation of the secondary amine with $(Fmoc\text{-}Ala)_2O$ followed by standard Fmoc SPPS gave the support-bound peptide. Treatment with TFA–H_2O (19 : 1) released the peptide into solution with concomitant cleavage of the acetal to form the *C*-terminal aldehyde. The choice of sequence was based on the heptad repeat used by Mezo and Sherman to prepare 4-α-helix bundles.[21] Incorporation of Ala near the *C*-terminal was rationalized from its large helix propensity, whereas *N*-terminal Tyr allowed concentration determination from the UV-absorption, and the *C*-terminal glycinal enabled oxime ligations with no risk of racemization at the aldehyde stage. With the aminooxy functionalized templates and the peptide aldehyde in hand, chemoselective ligations by oxime formation proceeded using a 50% excess of peptide aldehyde in aqueous acetate buffer at pH 4.76. The α-D-Gal*p*-carboprotein **4**, α-D-Glc*p*-carboprotein **5**, and α-D-Alt*p*-carboprotein **6** were obtained in *quantitative* yield after prep. C_4 RP-HPLC (Fig. 1).

Analysis by less hydrophobic C_4,[32,33] rather than C_{18}, RP-HPLC gave significantly sharper peaks and together with ESI-MS (Fig. 2) showed pure compounds of the expected mass. Further, analysis by size exclusion chromatography (SEC) gave for all three carboproteins a single symmetrical peak at the same elution volume (Fig. 3), indicating monomeric structures of the same size. From calibration with a series of small globular proteins, the carboprotein elution volume was found



1 - 3: R = H_2NOCH_2CO-
4 - 6: R = Ac-YEELLKKLEELLKKA-$NHCH_2CH{=}NOCH_2CO$-
7: R = $CH_3CH{=}NOCH_2CO$-

Fig. 1 Gal*p*-structures **1** and **4**, Glc*p*-structures **2**, **5** and **7**, and Alt*p*-structures **3** and **6**. Structures **1**–**3** are templates, **4**–**6** are carboproteins, and **7** is a reference compound for CD spectroscopy.

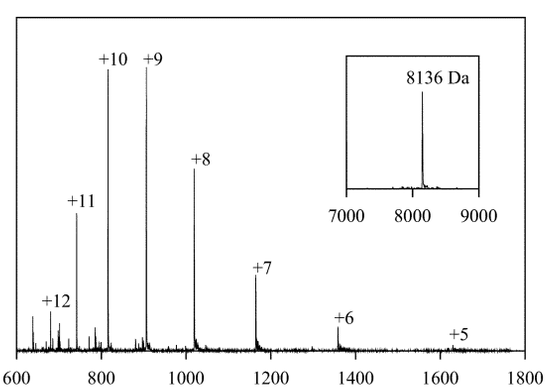


Fig. 2 ESI-MS spectrum of carboprotein **6** (calculated MW 8135.59 Da) with multiple charged species. The insert shows the deconvoluted spectrum.

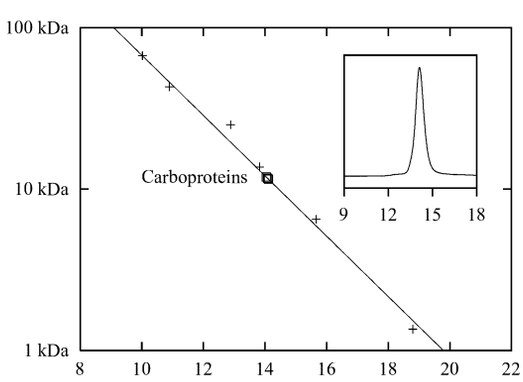


Fig. 3 SEC of carboproteins **4**, **5**, and **6**. The insert shows the elution profile of carboprotein **4**.

Table 1 Parameters obtained from denaturation curves

GuHCl denaturation			
Carboprotein	**4** (Gal*p*)	**5** (Glc*p*)	**6** (Alt*p*)
$[\theta]_N$/deg cm^2 dmol^{-1}	−21747 ± 571	−22681 ± 762	−24993 ± 721
ΔG_F^{H2O}/kcal mol^{-1}	−8.8 ± 1.9	−9.1 ± 2.5	−8.0 ± 1.9
m/kcal mol^{-1} M^{-1}	1.4 ± 0.3	1.5 ± 0.4	1.3 ± 0.4
Thermal denaturation			
Carboprotein	**4** (Gal*p*)	**5** (Glc*p*)	**6** (Alt*p*)
T_m/°C	90 ± 1	86 ± 1	85 ± 1
ΔH_F/kcal mol^{-1}	−15 ± 2	−12 ± 2	−14 ± 2

to correspond to an 11.7 kDa globular protein. The apparent size-difference to the 8.1 kDa carboproteins can be explained by a non-globular or less-compact structure of the carboproteins.

To provide a reference compound for CD measurements, template **2** was also reacted with acetaldehyde to yield structure **7** (Fig. 1). Characterization of **7** by ^{1}H NMR revealed an approx. 1 : 1 ratio of *E*- and *Z*-oxime isomers, with the *E*-isomer C*H*=N signal (multiplet) being 0.7 ppm downfield of the *Z*-isomer signal. This distribution has previously been observed for carboproteins and hence appear to be general for structures prepared by this methodology.[23,34]

Structure and stability

The CD spectra of carboproteins **4**, **5**, and **6** recorded in aqueous buffer (pH 7) showed that while all three structures contained α-helix, Alt*p*-carboprotein **6** was somewhat more α-helical than the two other structures (Fig. 4). From the mean residue ellipticity at 222 nm (MRE, $[\theta]_{222}$) the α-helicity was calculated to be 64%, 66%, and 77% for **4**, **5**, and **6**, respectively.[35] First of all, it is noteworthy that Alt*p*-carboprotein **6**, with its 'negative design', does not have a lower degree of α-helicity than carboproteins **4** and **5**. The calculated degree of α-helicity rests on accurate determination of the carboprotein concentration. For this we relied on quantification by UV-absorption, as the carboproteins each contain four Tyr residues. It should be emphasized that although the difference is small, the higher degree of α-helicity for Alt*p*-carboprotein **6** was observed consistently in all experiments. In a control experiment, the CD spectrum of reference compound **7** was recorded in methanol (Fig. 4), showing only low MRE values. We thus conclude that the template itself has only an insignificant contribution to the CD spectra of carboproteins. This leads to the suggestion that the higher helicity of carboprotein **6** can be explained by the influence of the *trans* diaxial arrangement of the O-2 and O-3 hydroxyls in Alt*p* on the alignment of the peptide strands. The observation can potentially be explained in terms of a better side-chain packing arrangement resulting from the 4.13 Å offset of the O-2 and O-3 anchored peptide chains in carboprotein **6** relative to **4** (and **5**).

However, the higher content of α-helix in carboprotein **6** was not reflected in an increased resistance towards denaturation. When subjected to denaturation by guanidinium chloride (GuHCl), the three carboproteins showed very similar behaviour, *i.e.*, high stability and cooperative unfolding (Fig. 5). Analysis by non-linear fitting to the equation of the linear extrapolation method (LEM)[36,37] provided the parameters reported in Table 1. Whereas $[\theta]_N$ ($[\theta]_{222}$ in absence of denaturant) was higher for **6**, the parameters ΔG_F^{H2O} (the extrapolated free energy of folding in H_2O) and m (related to the cooperativity of the folding) were not significantly different from those obtained for carboproteins **4** and **5**. The values of ΔG_F^{H2O} are noted to be comfortably below −5 kcal mol^{-1}, a number used as a rule of thumb for what is needed to give a fully folded structure.[1] Likewise, the three structures were all found to be very resistant toward thermal denaturation, allowing only the lower half of the denaturation curves to be recorded (Fig. 6). The reversibility of the denaturation was tested in an experiment in which carboprotein **4**, after going through stepwise heating to 90 °C, was left to obtain room temperature in the CD instrument. A new recording at 20 °C after 24 h showed 86% of the previously obtained helicity. The thermal denaturation data were fitted to an equation derived from thermodynamics,[38,39] applying the assumption that $[\theta]_N$ and $[\theta]_D$ are temperature independent, to give the T_m (melting points) and ΔH_F (enthalpies of folding) reported in Table 1. Melting points are in general high, but with only small differences between the carboproteins.

Stability towards denaturation does not necessarily imply that a protein structure is native-like.[3] Fluorescent dyes such as 8-anilino-1-naphthalenesulfonic acid (ANS) have long been used to probe structural features of proteins due to their affinity

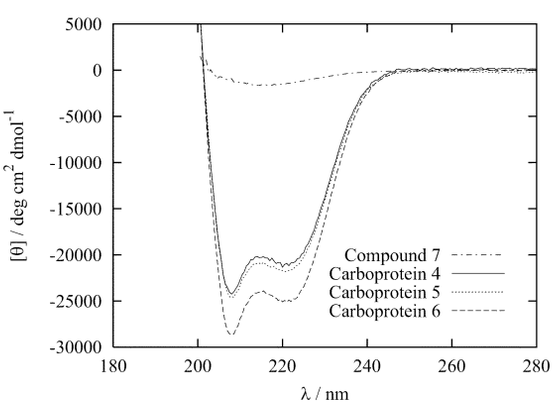


Fig. 4 CD spectra of carboproteins **4**, **5**, and **6**, and of the reference template **7**.

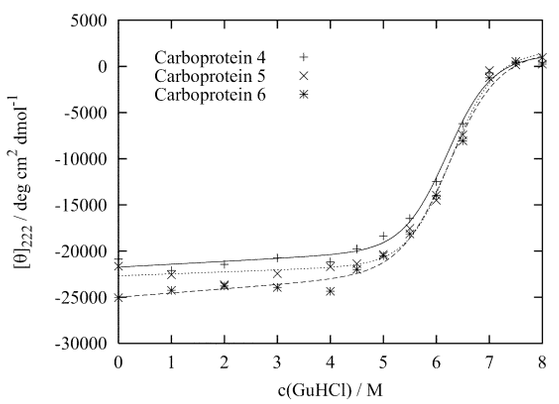


Fig. 5 GuHCl denaturation of carboproteins **4**, **5**, and **6**. Data points are shown together with fitted denaturation curves, obtained as described in Materials and methods.

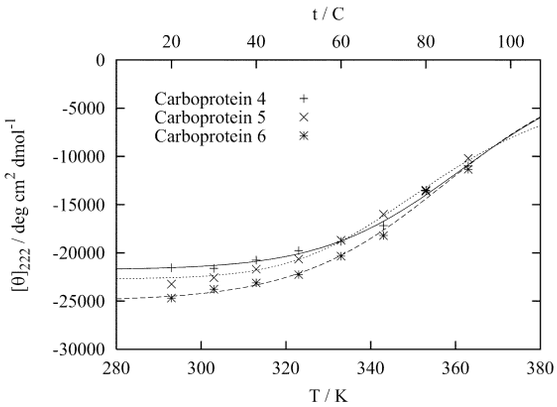


Fig. 6 Thermal denaturation of carboproteins **4**, **5**, and **6**. Data points are shown together with fitted denaturation curves, obtained as described in Materials and methods.

for hydrophobic environments, which results in an increase in fluorescence.[40] Hence, with ANS binding to molten globules but not to native-like structures, it is a convenient test of *de novo* designed proteins.[41] With 10 times excess of carboprotein relative to ANS, structures **4**, **5**, and **6** all bound ANS (Fig. 7). This could indicate presence of molten globule character, which would not be unexpected based on the relatively flat thermal denaturation curves and the simple peptide sequence with an all-Leu core.[42] However, ANS could in principle also bind to the hydrophobic 'pocket' formed by the four *N*-terminal Tyr residues in the 4-α-helix bundles. In this context it is interesting to note that the carboprotein with the highest helicity, **6**, also gave the highest ANS fluorescence intensity.

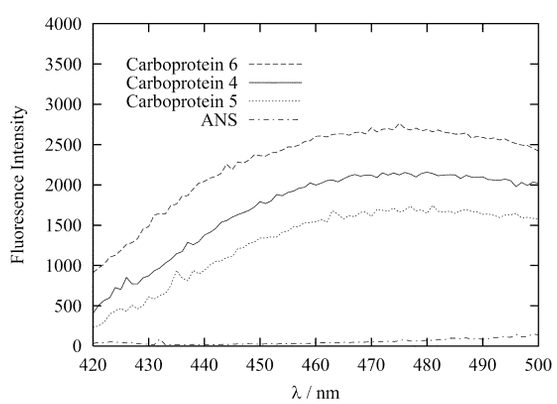


Fig. 7 ANS binding to carboproteins **4**, **5** and **6**, as determined by fluorescence spectroscopy.

Conclusion

In summary, three carboproteins based on Gal*p*, Glc*p*, and Alt*p* templates were prepared and characterized. All three carboproteins were obtained in quantitative yield following oxime ligation and work-up by prep. C_4 RP-HPLC. Characterization by ESI-MS and SEC indicated pure, monomeric structures. Interestingly, CD spectroscopy revealed a somewhat higher content of α-helix in Alt*p*-carboprotein **6** compared to Gal*p*- and Glc*p*-carboproteins **4** and **5**. This could be due to an effect of the template on the protein *structure*. However, as in the work of Fairlie and co-workers,[19] the choice of template was found to have no or only little effect on the protein *stability*. Within the experimental conditions applied here, the three carboproteins showed equally good resistance toward chemical and thermal denaturation, as revealed by the obtained thermodynamic parameters. All three carboproteins bound ANS, indicative of some molten globule character.

Abbreviations ‡

ANS, 8-anilino-1-naphthalenesulfonic acid; Aoa, aminooxyacetic acid; *o*-BAL, *ortho* backbone amide linker; DIPCDI, *N*,*N*′-diisopropylcarbodiimide; DIPEA, *N*,*N*-diisopropylethylamine; DMF, *N*,*N*-dimethylformamide; Fmoc, 9-fluorenylmethyloxycarbonyl; HBTU, *N*-[(1*H*-benzotriazol-1-yl)-(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide; HOBt, 1-hydroxybenzotriazole; LEM, linear extrapolation method; MRE, mean residue ellipticity; *o*-PALdehyde, 5-(2-formyl-3,5-dimethoxy-phenoxy)-pentanoic acid; SPPS, solid-phase peptide synthesis; TG, TentaGel; TFA, trifluoroacetic acid. Amino acid symbols denote the L-configuration unless stated otherwise. All solvent ratios are volume/volume unless stated otherwise.

Acknowledgements

We thank Ms Karen Jørgensen for recording the CD spectra, Dr Malcolm Gillies for help with the Maestro software, and Dr Elisar Barbar for valuble comments on the manuscript. We express gratitude to the Alfred Benzon Foundation (Bøje Benzon Fellowship to K. J. J.), the Villum Kann Rasmussen Foundation (J. N.) and the Technical University of Denmark (J. B.) for financially supporting this work.

‡ See ref. 44.

References

1 L. Baltzer, H. Nilsson and J. Nilsson, *Chem. Rev.*, 2001, **101**, 3153–3163.
2 W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri and A. Lombardi, *Annu. Rev. Biochem.*, 1999, **68**, 779–819.
3 R. B. Hill, D. P. Raleigh, A. Lombardi and W. F. DeGrado, *Acc. Chem. Res.*, 2000, **33**, 745–754.
4 M. Mutter and S. Vuilleumier, *Angew. Chem., Int. Ed. Engl.*, 1989, **28**, 535–554.
5 M. Mutter, G. G. Tuchscherer, C. Miller, K.-H. Altmann, R. I. Carey, D. F. Wyss, A. M. Labhardt and J. E. Rivier, *J. Am. Chem. Soc.*, 1992, **114**, 1463–1470.
6 M. Mutter, P. Dumy, P. Garrouste, C. Lehmann, M. Mathieu, C. Peggion, S. Peluso, A. Razaname and G. Tuchscherer, *Angew. Chem., Int. Ed.*, 1996, **35**, 1482–1485.
7 S. Peluso, P. Dumy, C. Nkubana, Y. Yokokawa and M. Mutter, *J. Org. Chem.*, 1999, **64**, 7114–7120.
8 M. Mutter and G. Tuchscherer, *Cell. Mol. Life Sci.*, 1997, **53**, 851–863.
9 H. K. Rau and W. Haehnel, *J. Am. Chem. Soc.*, 1998, **120**, 468–476.
10 H. K. Rau, N. DeJonge and W. Haehnel, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 11526–11531.
11 H. K. Rau, N. DeJonge and W. Haehnel, *Angew. Chem., Int. Ed.*, 2000, **39**, 250–253.
12 R. Schnepf, P. Hörth, E. Bill, K. Wieghardt, P. Hildebrandt and W. Haehnel, *J. Am. Chem. Soc.*, 2001, **123**, 2186–2195.
13 T. Sasaki and E. T. Kaiser, *J. Am. Chem. Soc.*, 1989, **111**, 380–381.
14 K. S. Åkerfeldt, R. M. Kim, D. Camac, J. T. Groves, J. D. Lear and W. F. DeGrado, *J. Am. Chem. Soc.*, 1992, **114**, 9656–9657.
15 M. R. Ghadiri, C. Soares and C. Choi, *J. Am. Chem. Soc.*, 1992, **114**, 825–831.
16 M. R. Ghadiri, C. Soares and C. Choi, *J. Am. Chem. Soc.*, 1992, **114**, 4000–4002.
17 M. W. Mutz, M. A. Case, J. F. Wishart, M. R. Ghadiri and G. L. McLendon, *J. Am. Chem. Soc.*, 1999, **121**, 858–859.
18 M. Goodman, Y. Feng, G. Melacini and J. P. Taulane, *J. Am. Chem. Soc.*, 1996, **118**, 5156–5157.
19 A. K. Wong, M. P. Jacobsen, D. J. Winzor and D. P. Fairlie, *J. Am. Chem. Soc.*, 1998, **120**, 3836–3841.
20 A. S. Causton and J. C. Sherman, *Bioorg. Med. Chem.*, 1999, **7**, 23–27.
21 A. R. Mezo and A. R. Sherman, *J. Am. Chem. Soc.*, 1999, **121**, 8983–8994.
22 K. J. Jensen and G. Barany, *J. Pept. Res.*, 2000, **56**, 3–11.
23 K. J. Jensen and J. Brask, *Cell. Mol. Life Sci.*, 2002, **59**, 859–869.

24 J. Brask and K. J. Jensen, *J. Pept. Sci.*, 2000, **6**, 290–299.
25 J. Brask and K. J. Jensen, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 697–700.
26 J. Brask, H. Wackerbarth, K. J. Jensen, J. Zhang, J. U. Nielsen, J. E. T. Andersen and J. Ulstrup, *Bioelectrochemistry*, 2002, **56**, 27–32.
27 J. Brask, H. Wackerbarth, K. J. Jensen, J. Zhang, I. Chorkendorff and J. Ulstrup, *J. Am. Chem. Soc.*, 2003, **125**, 94–104.
28 K. J. Jensen, J. Alsina, M. F. Songster, J. Vágner, F. Albericio and G. Barany, *J. Am. Chem. Soc.*, 1998, **120**, 5441–5452.
29 J. Alsina, K. J. Jensen, F. Albericio and G. Barany, *Chem. Eur. J.*, 1999, **5**, 2787–2795.
30 F. Guillaumie, J. C. Kappel, N. M. Kelly, G. Barany and K. J. Jensen, *Tetrahedron Lett.*, 2000, **41**, 6131–6135.
31 U. Boas, J. Brask, J. B. Christensen and K. J. Jensen, *J. Comb. Chem.*, 2002, **4**, 223–228.
32 V. Steiner, M. Schär, K. O. Börnsen and M. Mutter, *J. Chromatogr.*, 1991, **586**, 43–50.
33 D. L. Steer, P. E. Thompson, S. E. Blondelle, R. A. Houghten and M.-I. Aguilar, *J. Pept. Res.*, 1998, **51**, 401–412.
34 H. Miyabe, M. Torieda, K. Inoue, K. Tajiri, T. Kiguchi and T. Naito, *J. Org. Chem.*, 1998, **63**, 4397–4407.
35 Y.-H. Chen, J. T. Yang and K. H. Chau, *Biochemistry*, 1974, **13**, 3350–3359.
36 C. N. Pace, *Methods Enzymol.*, 1986, **131**, 266–280.
37 M. M. Santoro and D. W. Bolen, *Biochemistry*, 1992, **31**, 4901–4907.
38 W. J. Becktel and J. A. Schellman, *Biopolymers*, 1987, **26**, 1859–1877.
39 J. C. Waldner, S. J. Lahr, M. H. Edgell and G. J. Pielak, *Biopolymers*, 1999, **49**, 471–479.
40 C. J. Seliskar and L. Brand, *J. Am. Chem. Soc.*, 1971, **93**, 5414–5420.
41 G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev and V. N. Uversky, *Biopolymers*, 1991, **31**, 119–128.
42 T. M. Handel, S. A. Williams and W. F. DeGrado, *Science*, 1993, **261**, 879–885.
43 Portions of this work were reported in preliminary form: J. Brask, J. M. Dideriksen and K. J. Jensen, in *Peptides: The Wave of the Future, Proceedings of the 17th American Peptide Symposium*, ed. M. Lebl and R. A. Houghten, American Peptide Society, San Diego, 2001, pp. 454–455.
44 Abbreviations used for amino acids and the designations of peptides follow the rules of the IUPAC-IUB Joint Commission on Biochemical Nomenclature: IUPAC-IUB, *J. Biol. Chem.*, 1972, **247**, 977–983.