Synthesis of functionalized *de novo* designed 8–16 kDa model proteins towards metal ion-binding and esterase activity†

A. Pernille Tofteng,[a] Thomas H. Hansen,[a] Jesper Brask,[b] John Nielsen,[a] Peter W. Thulstrup[a] and Knud J. Jensen*[a]

Received 19th March 2007, Accepted 29th May 2007
First published as an Advance Article on the web 19th June 2007
DOI: 10.1039/b704159d

De novo design and total chemical synthesis of proteins provides a powerful approach for biological and biophysical studies with the ability to prepare artificial proteins with tailored properties, potentially of importance for biophysical studies, material science, nanobioscience, and as molecular probes. In this paper, the previously developed concept of carbohydrates as templates is employed in the *de novo* design of model proteins (artificial helix bundles) termed 'carboproteins'. The 4-α-helix bundle is a macromolecular structure, where four amphiphilic α-helical peptide strands form a hydrophobic core. Here this structure is modified towards achieving metal ion-binding and catalytic activity. We report: (i) test of directional effects from different tetravalent carbohydrate templates, (ii) synthesis and evaluation of carboproteins functionalized with phenol, pyridyl or imidazolyl moieties as potential ligands for metal ion-binding as well as for catalysis. Our results include: (i) support of our previous 'controversial' finding that for some carboproteins the degree of α-helicity depends on the template, *i.e.*, that there is, to some extent, a controlling effect from the template, (ii) demonstration of binding of Cu(II) to tetra-functional carboproteins by electrospray ionization-time of flight-mass spectrometry (ESI-TOF-MS), UV–VIS absorption spectroscopy and size exclusion chromatography-inductively coupled plasma-mass spectrometry (SEC-ICP-MS); (iii) a kinetic investigation of the esterase activity.

Introduction

A key contribution from synthetic bioorganic chemistry to the emerging field of nanobioscience is the design and construction of novel molecules that mimic the three-dimensional structure and function of proteins. Complex interactions in natural proteins can be studied in greater detail using smaller *de novo* designed systems. The approach also offers access to tailor-made proteins, *e.g.* for self-assembly on surfaces or with novel functionalities. A frequently used structural motif in *de novo* design is the 4-α-helix bundle.[1] Its folding is driven by the hydrophobic collapse of internally hydrogen-bonded, 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 an unfolded conformation to a folded protein structure is significant. From the standpoint of *de novo* design, the entropic barrier to folding imposes some restrictions and limitations, *e.g.*, due to the long sequences which provide the scaffold to hold strands of α-helices and β-sheets together. To lower this barrier, Mutter and coworkers have suggested pre-organizing peptide strands on a molecular template to reduce the entropy of the construct's unfolded state.[2] Hence, the branched structures of template-assembled synthetic proteins (TASPs) facilitate folding of protein models otherwise not possible with a linear sequence. Here the peptide sequences, which form secondary structure elements, are held together by the template, typically in a parallel manner. Whereas Mutter *et al.*[3] and Haehnel *et al.*[4] have mainly explored linear or cyclic peptide templates, other research groups have applied other, non-peptide templates in protein designs. These include porphyrin derivatives,[5] metal ions (by complexation),[6] a cyclohexane derivative (Kemp's triacid),[7] substituted phenyl rings,[8] and aromatic macrocycles.[9] However, the significance of the template geometry has been debated.[8,10] Does the template merely tie the peptide strands together, or can it, with the proper choice of geometry between 'anchoring points', affect the folding?

Metal-binding peptides and proteins are ubiquitous, not only as functional metalloproteins, *e.g.* enzymes, but also for complex metal-transport systems, scavenging of metals in plants *etc.* For example, it has found increasing attention that Cu(II) is implicated in the formation of amyloid Aβ peptide.[11] Artificial metal ion-binding peptides and proteins, including metalloproteins, are also of increasing interest.[12] In a recent study, Williams and coworkers have reported artificial oligomers, derived from an aminoethylglycine PNA backbone with pyridyl and bipyridyl ligands for metal complexation.[13]

Carboproteins are model proteins assembled on carbohydrate templates.[14] 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.[15] The relatively rigid $^4C_1$ conformation of monosaccharides enforces well-defined axial or equatorial orientations of the hydroxyls. We have previously observed a somewhat higher degree of α-helicity in an Alt*p*-carboprotein compared to corresponding Gal*p*- and

[a]*Department of Natural Sciences, Faculty of Life Sciences, University of Copenhagen, Thorvaldsensvej 40, DK-1871, Frederiksberg C, Denmark. E-mail: kjj@life.ku.dk; Fax: +45 3528 2398; Tel: +45 3528 2430*
[b]*Novozymes A/S, 2880, Bagsvaerd, Denmark*
† The HTML version of this article has been enhanced with colour images.

Glc*p*-carboproteins, according to CD spectroscopy. This could be due to an effect of the template on the protein *structure*. However, the choice of template was found to have little or no effect on protein *stability*. Towards single-molecule nanobioscience we have designed and synthesized thiol-functionalized carbohydrates which form self-assembled monolayers on Au(111) surfaces.[16,17] Interestingly, this provided us with an estimated diameter of the thiol-linked carboprotein of 20–24 Å in an upright position in a self-assembled monolayer on a Au(111) surface.

The peptide sequences used in carboproteins so far have been purposely simplistic, using Leu, Glu, and Lys to generate four turns, as well as Tyr and Ala at the termini.‡ The four peptide strands of 'tetravalent' carboproteins are identical and parallel, as they are attached to a common template.

Our aim in this study was to design, synthesize and evaluate pyridyl and imidazoyl functionalized carboproteins as potential metal ion chelators, in order to achieve artificial metalloproteins that potentially could incorporate redox-active metals and find use in nanobioelectronics and catalysis. In the design of metallo-carboproteins, *i.e.* metal ion-binding carboproteins, we relied on a four-ligand design, with either four phenol, pyridyl or imidazoyl moieties at the N-termini of peptides attached to the template through the C-termini (Fig. 1). This strategy should position the four potential monodentate ligands in close proximity at the end of the helices, and also allow for solvent accessibility. We focused on the binding of redox-active Cu(I)–Cu(II). If successful, this would amount to *de novo* design of a metalloprotein with a non-natural metal-binding site in a 4-helix bundle. In addition, an imidazoyl functionalized carboprotein was investigated for its esterase activity, in the absence of metal ions.

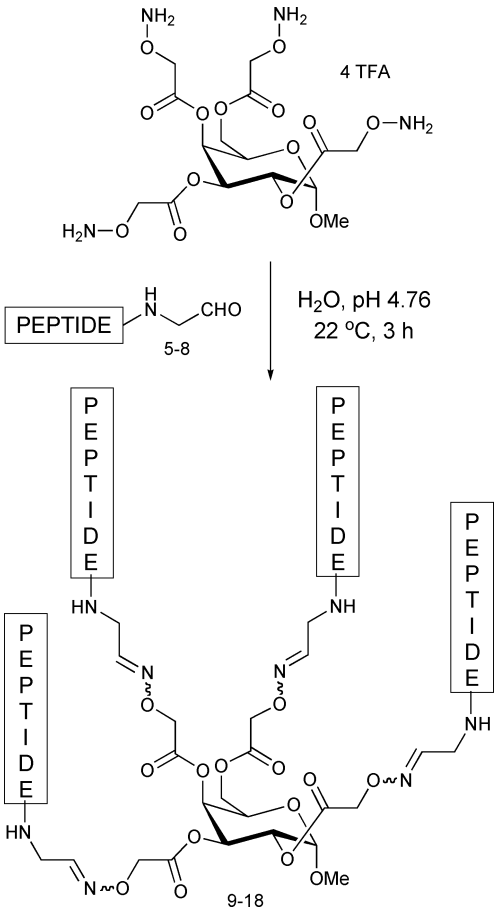


Fig. 1 Preparation of carboproteins: generalized structure exemplified with Gal*p* template.

Results and discussion

Here we present a combinatorial ensemble with 10 carboproteins, of which 7 are novel structures (Table 1). Three new peptide aldehydes, together with one previously synthesized, were combined with three templates.

Templates

Templates **1–3**, functionalized with aminooxyacetic acid (**4**) were prepared as before.[14–18]

Synthesis of peptide aldehydes

C-terminal peptide aldehydes **5**[16] and **6–8**, which were required as building blocks, were assembled using general Fmoc-chemistry on a backbone amide linker (BAL) handle with a C-terminal glycinal residue masked as an acetal (Fig. 2).[18,19] After completion of peptide chain assembly, the peptide was released from the support and the side-chain protecting groups removed by treatment with conc. TFA, which also converted the acetal to the required aldehyde.

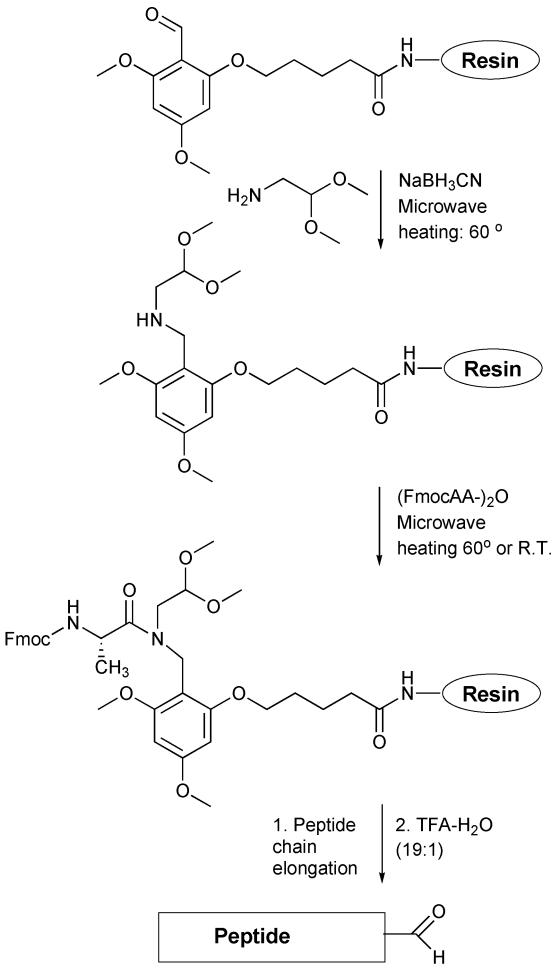
Table 1 The degree of helicity for the carboproteins 9–18 as determined from the mean residue ellipticity at 222 nm

		5	**6**	**7**	**8**
	Peptide	(Tyr)	(D-Pal)	(Pyp)	(Iza)
Template	**1, (Gal*p*)**	67% (**9**)	68% (**12**)	58% (**15**)	78% (**18**)
	2, (Alt*p*)	83% (**10**)	69% (**13**)	57% (**16**)	n.a.
	3, (Dimer Gal*p*)	48% (**11**)	39% (**14**)	41% (**17**)	n.a.

Previously,[19] the acetal building block had been attached to the PALdehyde resin by reductive amination for 2 × 18 hours. However, in the present work this was optimized using microwave heating to 60 °C to reduce the reaction time to 2 × 10 minutes.

The synthesis of histidine (His) containing peptide aldehydes by the above strategy poses a specific problem, as the most common side-chain protecting group for His is trityl (Trt) and its acidolytic removal is most conveniently performed in the presence of a silane as a hydride source to quench the rather stable trityl carbenium ion. However, aldehydes are likely to be reduced by silanes under acidic (TFA) conditions. Thus, these reducing conditions may conflict with the simultaneous generation of a redox-sensitive aldehyde moiety. The synthesis of a His-containing

‡ DeGrado and coworkers have previously observed a dynamic character of the *de novo* designed protein α2B (with Leu, Glu, and Lys in the core), most likely due to rapid motions of side chains in the bundle and the formation of multiple, interconverting topologies.[1a] It is likely that the peptide sequences used in the present study provide a dynamic character. However, coordination of the peptide 'arms' to metal ions might tighten the structure and reduce the dynamic character.



Peptide aldehyde

5: Ac-YEELLKKLEELLKKAG-H
6: Ac-D-PAL-EELLKKLEELLKKAG-H
7: Pyp-EELLKKLEELLKKAG-H
8: Iza-EELLKKLEELLKKAG-H

Fig. 2 Synthesis of peptide aldehydes.

peptide aldehyde using Fmoc-His(Boc)-OH gave hard to remove impurities and the N-terminal His proved difficult to N^{α}-acetylate, according to ninhydrin tests (data not shown) and MS.

Instead, we first relied on non-natural pyridyl moieties as ligands for the metal-binding site, as a pyridyl side-chain can be incorporated without a protecting group. We introduced two building blocks, D-3-pyridylalanine (D-Pal), which requires an acetylation of the N-terminal, and a 3-(3-pyridyl)propionic acid (Pyp), which lacks an amino group (Fig. 3). The synthesis of Pyp containing peptide aldehydes gave, however, low yields. Next, we also introduced an *N*-terminal imidazoyl moiety by incorporating 4(5)-imidazolecarboxaldehyde (Iza) as the last residue, without the need for a Trt protecting group. This eliminated the acetylation step but maintained the desired imidazoyl moiety in the N-terminal. The synthesis of D-Pal and Iza-peptide aldehydes was performed without difficulties, where the Pyp peptide aldehydes had given lower yields.

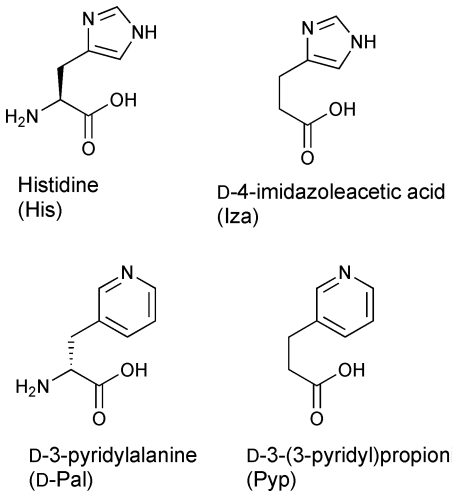
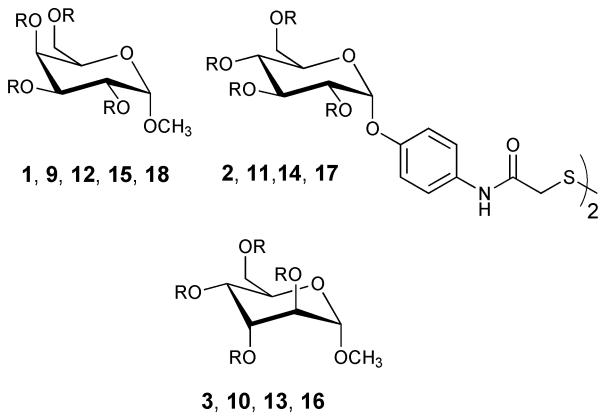


Fig. 3 Building blocks for introduction of metal-binding ligands.

Synthesis of carboproteins

The carboproteins were synthesized using our previously established protocol for chemoselective oxime ligation of peptide aldehydes to the deprotected templates **1–3** in generally good yields (Fig. 4).



 # Linker or peptidyl-linker

1-3: R = H_2NOCH_2CO-
9-11: R = Ac-YEELLKKLEELLKKA-$NHCH_2CH{=}NOCH_2CO$-
12-14: R = Ac-D-PAL-EELLKKLEELLKKA-$NHCH_2CH{=}NOCH_2CO$-
15-17: R = Pyp-EELLKKLEELLKKA-$NHCH_2CH{=}NOCH_2CO$-
18: R = Iza-EELLKKLEELLKKA-$NHCH_2CH{=}NOCH_2CO$-

Fig. 4 Templates and peptide sequences.

CD spectroscopy

The circular dichroism (CD) spectra of carboproteins **9**, **10**, **12**, **13**, **15**, **16**, and **18** all showed UV CD signals indicative of high content of helical secondary structure (Fig. 5). The remaining carboproteins **11**, **14**, and **17** were also helical (not displayed here). The degree of helicity was assessed by the ellipticity at 222 nm ($[\theta]_{222}$)[20] and the monomeric carboproteins showed a high degree of helicity ranging from ~57% to 83%, while the dimers had somewhat lower helicities (Table 1).[2] Alt*p* carboprotein **10** showed the highest degree of helicity. This validated our previous

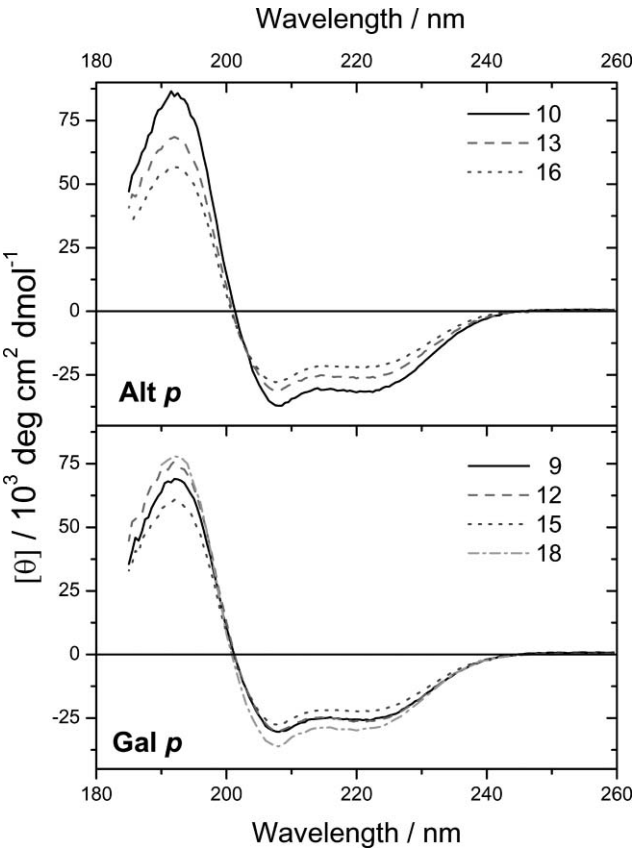


Fig. 5 UV CD spectra of carboproteins in aqueous buffer solution (10 μM carboprotein in 10 mM aqueous NaH_2PO_4, pH 7).

'controversial' finding that there is a trend for Alt*p* carboproteins to show a higher degree of α-helicity than the corresponding Gal*p* and Glc*p* carboproteins, pointing to the influence of the template for secondary structural formation. Carboproteins **9–11** were included for comparison, especially of CD spectra.

Metal ion-binding studied by MS spectrometry and UV spectroscopy

Numerous redox-active metalloproteins and enzymes rely on the Cu(I)–Cu(II) redox pair, including hemocyanin, tyrosinase, superoxide dismutase, multi-copper oxidases, and other blue-copper proteins.[21] The binding sites usually contain histidine, tyrosine, and cysteine residues, the latter in the case of blue-copper binding sites.[22] The mechanisms behind insertion of copper in proteins are not simple, in biological systems copper chaperone proteins are implemented, while in solution it is often facile *via* Cu(I) with subsequent oxidation to Cu(II).[23] This latter approach was used in this study. It is well known that all proteins and peptides bind Cu(II) under alkaline conditions, due to coordination by deprotonated amide nitrogens. This reaction is used in the biuret method for protein quantification. In order to avoid this type of unspecific binding in our studies, pH was kept in the neutral to slightly acidic range.

For the metal ion-binding studies we selected phenol-functionalized carboprotein **9** and imidazoyl-functionalized carboprotein **18** for detailed analysis by ESI-MS, SEC-ICP-MS and UV–Vis spectroscopy, where carboprotein **9** containing N-terminal Tyr moieties served as a control.

Mass spectrometry

Binding of Cu(I)–Cu(II) to several carboproteins was studied by electrospray mass spectrometry (ESI-MS). Carboproteins were, after addition of an excess of Cu(I)Cl in water, injected using a buffer solution of 50% acetonitrile–water and 0.2% formic acid. Deconvoluted spectra of the obtained data revealed the binding of one or two Cu(II) (Fig. 6 and 7). This is an interesting result, as Cu(I) disproportionates over time to Cu(II) and Cu(0). The fact that binding can be seen at this acidic pH gives a strong indication that this binding is *not* non-specific binding to the backbone, where binding of Cu(II) is likely to occur, at least in part, by metal-induced amide deprotonation.[24] Interestingly, Hou and Zagorski suggested that in the binding of Cu(II) to the amyloid Aβ peptide, the His side-chains first anchor Cu(II) followed by backbone amide deprotonation. This could mean that Cu(II) is mainly inserted at the N-termini, close to the imidazole moieties. The deconvoluted MS spectra show that the highest intensity signal is obtained from the copper-free, *i.e.* non-chelated, carboprotein. Although the mass spectrometer does not provide fully quantitative data, this could indicate a weak binding to copper or that only Cu(II) can bind and that the Cu(II) concentration is low as it arises from the disproportionation. However, the carboprotein-Cu(II) complexes may also dissociate to some degree in the mass spectrometer.

UV–Vis spectroscopy

Binding of Cu(II) to the imidazoyl Gal*p* carboprotein **18** was also studied by UV–Vis spectroscopy (Fig. 8). Addition of excess

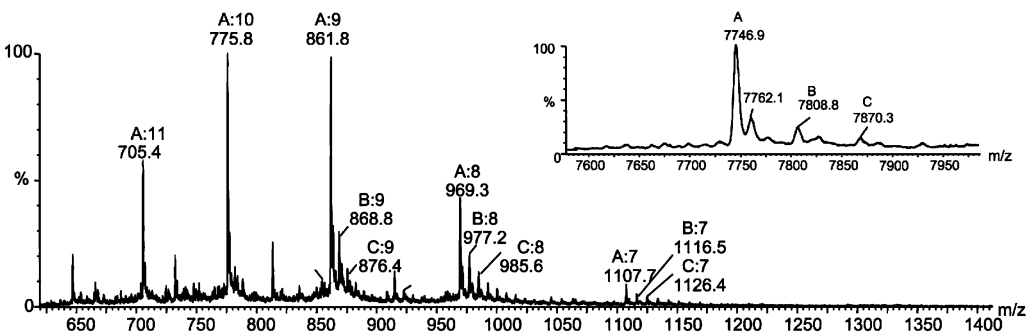


Fig. 6 ESI-MS for carboprotein **18** with Cu(II) binding where (A) is the carboprotein without Cu(II) binding, (B) is the carboprotein binding one Cu(II) and (C) is the carboprotein binding two Cu(II). The spectrum shows the protonation patterns with a deconvoluted spectrum in the right corner.

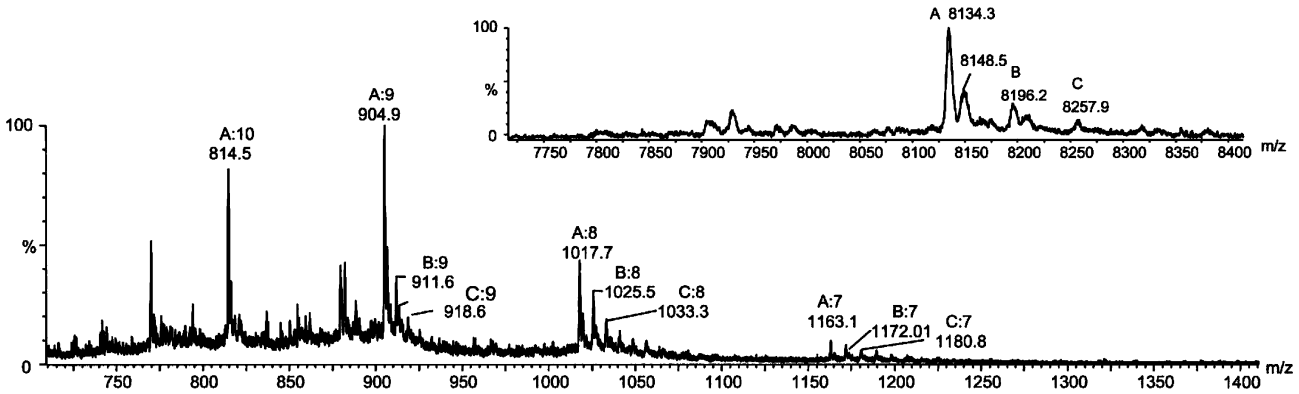


Fig. 7 ESI-MS for carboprotein **9** with Cu(II) binding where (A) is the carboprotein without Cu(II) binding, (B) is the carboprotein binding one Cu(II) and (C) is the carboprotein binding two Cu(II). The spectrum shows the protonation patterns with a deconvoluted spectrum in the right corner.

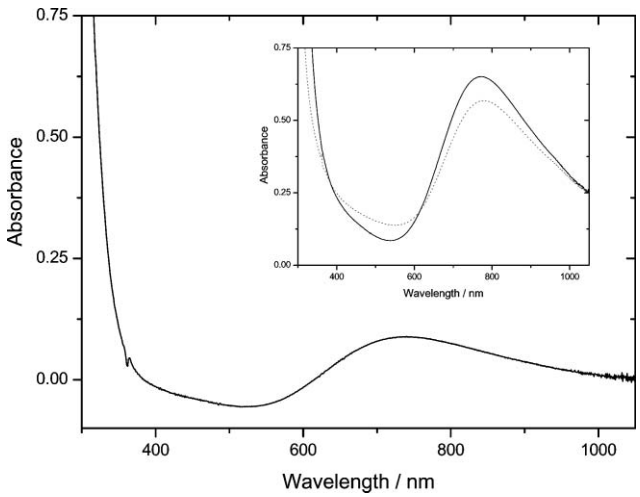


Fig. 8 UV–VIS absorption spectra of carboprotein **18** in 25 mM NaOAc buffer pH 5.5 after incubation with solid Cu(I)Cl. The main spectrum is the differential spectrum compared to the reference. The insert shows the sample (——) and reference (---) spectra, respectively.

Cu(I), followed by mixing and separation of the solid copper, yielded a light blue solution for both sample and reference, where the latter did not contain carboprotein. The color was due to formation of the tetraaqua-copper(II) ion $[Cu(H_2O)_4]^{2+}$, following the disproportionation of Cu(I) to Cu and Cu(II). However, the sample did give rise to certain spectral shifts with a new shoulder at around 300–350 nm and to a change in the wide absorption band of the tetraaqua species around 760 nm. The near-UV signals are not attributable to absorption by the carboprotein, which has an absorption band at *ca.* 265 nm. Although the signals are not intense, the spectra in Fig. 8 demonstrate binding of Cu(II) to the carboprotein. The new near-UV feature could likely be from a Cu(II)-imidazoyl charge-transition, while the change in the visible spectrum indicates an effect on the Cu(II) d–d-absorption. However, it is clearly a case where Cu(II) is only weakly bound to the carboproteins, possibly retaining several aqua ligands. Potentially, Cu(I) could occupy some of the carboprotein binding sites but this would not be detectable by absorption in the visible or near-UV region. However, neither the Cu(I)- nor the aqua-species were identified in the ESI-MS.

SEC-ICP-MS

Next, the two carboproteins **9** and **18** were evaluated for their possible chelating ability towards Cu, using chromatography on a SEC column to separate different complexes and an ICP-MS as a detector. The ICP-MS is a very sensitive detector of different elements, and was used in this study to detect binding of ^{63}Cu. Several papers have described SEC-ICP-MS for speciation analysis.[25] The method in the present study was described recently; it was shown that SEC can be used to separate even very labile metal complexes, demonstrated by the fact that it was possible to re-inject a peak from the sample without any loss of signal intensity.[25] We speculated that SEC-ICP-MS could be used to provide additional evidence for the existence of complexes between a carboprotein and Cu, as the carboprotein would be separated from the Cu ions not bound to it. Also, only complexes stable enough to survive the SEC column would be detected, every peak in the SEC chromatogram would be correlated with a possible signal for ^{63}Cu in the ICP-MS. For the SEC-ICP-MS experiments, the carboproteins **9** and **18** were dissolved in NH_4OAc buffer and mixed with excess CuCl, followed by injection into the column and monitoring of ^{63}Cu. Each chromatogram gave rise to more than one peak, demonstrating that several species were formed (Fig. 9). Using the calibration curve and the right axis it was possible to measure the size of the eluting peaks in the sample. The largest species were the ones eluting first and the Cu species eluting from 15–22 min had a size of 7–16 kDa. This indicates that both monomeric carboproteins as well as non-covalent carboprotein dimers were involved in complexation with Cu under these conditions. Also it was evident from the chromatogram that the larger carboprotein **18** was forming larger-sized complexes with Cu than the carboprotein **9**. This is seen by the fact that carboprotein **18** is starting to elute earlier than **9**. Some smaller complexes are also eluting in each run, but from the fact that the carboproteins were pure according to HPLC, these are likely to be small sulfur impurities that chelate Cu strongly and for this reason are very dominant in the chromatograms.

In summary, these data clearly indicate that the phenol-functionalized **9** and imidazoyl-functionalized Gal*p* carboprotein **18** bind copper, most likely as Cu(II); however, the site of binding could not be inferred. The fact that multiple Cu ions were observed bound to the carboprotein by mass spectrometry could indicate

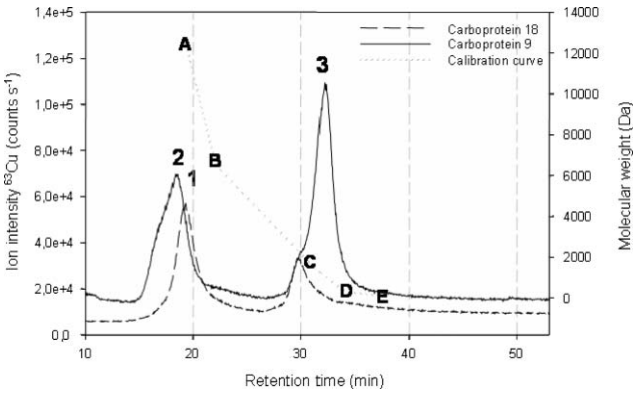


Fig. 9 SEC LC-ICP-MS chromatograms, using a 50 mM ammonium acetate eluent and monitoring the ^{63}Cu isotope show peaks for the CP **18** Cu complex (1) and the CP **9** Cu complex (2). Some non-bound Cu is seen eluting later (3). The *y*-axis left represents counts s^{-1} for the ^{63}Cu isotope and the *x*-axis represents time (min) after the injection of the sample. The right *y*-axis represents the molecular weight (Da) calibration curve made on the basis of A cytochrome c (12 500 Da), B aprotinin (6512 Da), C vitamin B12 (1355 Da), D glutathione (307 Da) and E cysteine (121 Da).

either (i) binding of copper to 1–2 imidazoyl ligands, or (2) non-specific binding, possibly involving also the backbone.

Esterase activity

The esterase activity of carboprotein **18** was evaluated using the 4-nitrophenyl butyrate (PNPB) assay. Carboprotein **18** could potentially be a catalyst for this ester hydrolysis, *via* insertion of the butyrate chain of the substrate in the centre of a helix bundle, and subsequent ester hydrolysis by the imidazoyl residues present at the solvent accessible surface. Table 2 shows the initial rate constants obtained from the spectrophotometric assay, where the formation of 4-nitrophenolate is monitored at 410 nm.

The data in Table 2 show that carboprotein **18** yields an increase in activity compared to the corresponding concentration of free imidazol, but the effect is at best a factor of 2–3. This modest effect, however, could provide a starting point for development of new catalytic, small proteins. As this requires access to a hydrophobic surface, this also relates to the question: to which extent are these small proteins molten globule-like?

Table 2 Esterase activity on 4-nitrophenyl butyrate in phosphate buffer at pH 7. Rates determined from kinetic measurements of formation of 4-nitrophenolate absorbance signal at 410 nm comparing carboprotein **18** (with four imidazoyl residues) with 4-imidazoleacetic acid hydrochloride (Iza)[a]

System	Rate/$\Delta mAU_{410\,nm}$ min^{-1}
CP **18** 0.1 mM	5.13
CP **18** 0.5 mM	9.23
Iza 0.5 mM	3.05
Iza 2.0 mM	6.39
Uncatalyzed	4.08

[a]The kinetic plots showed an initial lag phase of 3–5 minutes, and the reported rates were determined through linear regression on data taken after 12 minutes, where the course of the plots was linear.

Conclusions

We have successfully synthesized an ensemble of carboproteins carrying imidazoyl and pyridyl moieties. N-terminal imidazoyl and pyridyl functionalized peptide aldehydes were synthesized on solid phase using a BAL strategy. Microwave radiation was used to optimize the reductive amination step. Four copies of each peptide aldehyde were coupled to a tetra-aminooxy functionalized carbohydrate template by oxime formation. Carboproteins in general showed a high degree of helicity, according to CD spectroscopy. Interestingly, an Alt*p* carboprotein showed a higher degree of α-helicity than the corresponding Gal*p* and Glc*p* carboproteins. This validated our previous 'controversial' finding that there is an influence of the template on secondary structural formation. While we thus reproduced and confirmed our previous result, it now became clear that it also depends on the peptide sequence, as only the Tyr-terminated sequence showed this higher degree of α-helicity. Representative carboproteins were tested for their ability to bind Cu(I)–Cu(II). ESI-MS, SEC-IC-MS, and UV–Vis spectroscopy consistently showed that phenol-functionalized **9** and imidazoyl-functionalized Gal*p* carboprotein **18** bound copper, most likely as Cu(II), however, the site of binding could not be unequivocally determined. However, copper-binding occurred under acidic conditions, which minimizes non-specific binding. Carboprotein **18** catalyzed a modest increase in the rate of ester hydrolysis.

Experimental

General

The synthesis of Boc_2-Aoa-OH has been described in earlier publications from our laboratories.[15] However, it has now become available from NeoMPS (France). Automated peptide synthesis was carried out on an Applied Biosystems 433A peptide synthesizer. Manual peptide synthesis was performed in polypropylene syringes equipped with a polyethylene filter. Reagents for solid-phase synthesis were purchased from Sigma-Aldrich, Fluka, NovaBiochem or Iris Biotech and used without further purification. A small amount of chemicals was inherited from previous projects and used after NMR or MS analysis.

A 10 mM phosphate buffer, pH 7, was prepared by dissolving $NaH_2PO_4{\cdot}2H_2O$ (0.78 g, 5 mmol) in water (500 mL) and then adjusting to pH 7 with 1 M aq. NaOH. A 0.1 M acetate buffer, pH 4.76, was prepared by dissolving an equal amount of $NaOAc{\cdot}3H_2O$ (10 mmol) and AcOH (10 mmol) in water (200 mL). A HP 8452A Diode Array spectrophotometer was used in Fmoc quantifications. Analytical HPLC was performed on a Waters system, 600 control unit, 996 PDA detector, 717 Plus autosampler, Millenium 32 control software. For the analysis of peptides and general analysis, a Waters XTerra 300 C_{18} column (3.5 μm, 3.0 × 50 mm) was used, and for the analysis of carboproteins, a Waters Symmetry 300 C_4 column (5 μm, 3.9 × 150 mm) was used. Preparative and semi-preparative HPLC were performed on a similar Waters system with a Delta 600 pump. Preparative HPLC was performed on a stack of three 40 × 100 mm column cartridges of Waters Prep Nova-Pak HR C_{18} 6 μm 60 Å. Semi-preparative HPLC was performed on a single 25 × 10 mm column cartridge of Waters Delta-Pak C_4 15 μm 300 Å or a single 40 × 10 mm

column cartridge of Waters Prep Nova Pak HR C_{18} 6 μm 60 Å. The solvents used for HPLC were (A) H_2O with 0.1% TFA, (B) CH_3CN with 0.1% TFA, (C) H_2O, (D) CH_3CN.

CD spectra were recorded on a JASCO J-710 spectropolarimeter using cylindrical Hellma quartz cells. Carboprotein solutions were approximately 10 μM in 10 mM NaH_2PO_4, pH 7.0. The absolute concentration was determined spectroscopically (tyrosine absorption at 274 nm, using $\varepsilon = 1420\ cm^{-1}\ M^{-1}$, pyridyl absorption at 264 nm, using $\varepsilon = 1600\ cm^{-1}\ M^{-1}$, imidazole absorption at 214 nm, using $\varepsilon = 6500\ cm^{-1}\ M^{-1}$). The MRE was calculated based on one UV active residue per peptide chain and according to ref. 20. The number *n* is the number of residues not counting the C-terminal glycinal oximes. The helicity was calculated according to a formula by Yang *et al.* [20] NMR spectra were recorded on a Bruker Avance 300 spectrometer equipped with a BBO probe. UV–Vis absorption for metal-binding and kinetics studies were carried out on an HP-8453 diode-array spectrophotometer with a thermostated cell at 25 °C in 10 mm quartz cells (Hellma).

Masses were determined using an ESI-TOF-MS connected to a Waters 2795 HPLC equipped with a Waters 996 PDA detector. The ESI-TOF-MS was a Micromass LCT apparatus (Waters, USA) equipped with an ESI probe and spectra were acquired in positive mode. The ESI-TOF-MS was calibrated using the standard calibration procedure of the manufacturer (mixture of PEGs). The total ESI-TOF-MS spectra were acquired in the range from 100–1500 Da and deconvolution was performed using the transform algorithm provided by Masslynx 4.0 software. LC-ICP-MS experiments was performed on a HPLC (Agilent 1100 Series, Agilent Technologies, UK) coupled to Diode Array Detector (DAD) and an ICP-MS (Agilent 7500c, Agilent Technologies, UK) equipped with a PFA micro-flow nebulizer. All connections were PEEK tubings (0.17 mm id). All LC-ICP-MS chromatographic data were processed using Plasma Chromatographic Software v. B-02-04 (Agilent Technologies, UK). The size exclusion column (SEC) used in these experiments was a Superdex Peptide 10/300 GL (glass, 10 × 300 mm, 13 μm crosslinked agarose–dextran, Amersham Biosciences, USA), with an optimum separation range between 100 and 7000 Da. This column was evaluated in an earlier study.[25]

Preparation of the aminooxy-functionalized template

The synthesis of functionalized α-D-galactopyranoside and α-D-altropyranoside templates have been described in earlier publications.[14–18]

Solid phase synthesis of peptide aldehydes Ac-D-Pal-EELLKKLEELLKKAG-H (6) and Pyp-EELLKKLEELLKKAG-H (7)

The *o*-PALdehyde-TG resin (0.5 mmol) was prepared as previously described.[15] Next was a reductive amination which used aminoacetaldehyde dimethyl acetal (5.0 mmol, 10 eqv.) and $NaBH_3CN$ (5.0 mmol, 10 eqv.) dissolved in MeOH–AcOH (99 : 1). This mixture was added to the *o*-PALdehyde-TG resin (2.5 g, 0.26 mmol g^{-1}) before the resin was placed in a microwave oven and heated for 10 min at 60 °C. The resin was washed with DMF (×5) and CH_2Cl_2 (×5) and the reductive amination step was repeated. Then Fmoc-Ala-OH (5.0 mmol, 10 eqv.) was dissolved in CH_2Cl_2–DMF (9 : 1, 15 ml), DIPCDI (387 μl, 2.5 mmol) was added, and after preactivation for 15 min, transferred to the resin (0.5 mmol). The reaction mixture was placed on a shaker for 2 h (or alternatively heated in a microwave oven to 60° for 2 × 10 min) and then washed with DMF (×5) and CH_2Cl_2 (×5). The coupling procedure and the wash were repeated, using 10 eqv. of the Fmoc-amino acid and 5 eqv. of DIPCDI. Next, the resin was capped with CH_2Cl_2–Ac_2O (1 : 3) for 30 min and washed again with DMF (×3) and CH_2Cl_2 (×3) and dried *in vacuo*. An Fmoc quantification was performed to estimate the overall loss in loading (0.19–0.22 mmol g^{-1}). Then the resin was divided into two portions (of 0.25 mmol) and the middle part (-EELLKKLEELLKK-) of sequence **6** and sequence **7** were synthesized individually using an ABI 433 peptide synthesizer. Fmoc-D-Pal-OH (4 eqv.) was manually coupled to the side-chain protected EELLKKLEELLKK-N[$CH_2CH(OCH_3)_2$]-*o*-BAL-TG resin using standard HBTU–HOBt–DIEA (3.8 : 4 : 7.8 eqv.) coupling in DMF (10 mL, 45 min). After a wash with DMF (×3) and DCM (×3), the peptide was deprotected with piperidine–DMF (1 : 4), washed accordingly and acetylated with CH_2Cl_2–Ac_2O (1 : 3) for 30 min. 3-Pyridinepropionic acid (Pyp) was manually coupled to side-chain protected EELLKKLEELLKK-N[$CH_2CH(OCH_3)_2$]-*o*-BAL-TG resin using standard HBTU–HOBt–DIEA (3.8 : 4 : 7.8 eqv.) coupling in DMF (10 mL, 2 hours). Both peptidyl-resins were dried and treated with TFA–H_2O (95 : 5) (10 mL each) for 4 h and washed with TFA (5 mL × 2) to release (cleave) peptide aldehydes **6** and **7**, respectively. TFA was removed *in vacuo* and the peptide aldehydes were redissolved in CH_3CN–H_2O (1 : 1) (4 ml) and purified by prep C_{18} RP-HPLC. The fractions containing the product were pooled and concentrated *in vacuo* to remove CH_3CN and lyophilized.

Peptide aldehyde 6

Yield 154 mg, 34% (calculated according to the loading of dipeptide being 77% of original loading and incl. 4 × TFA). ESI-MS, calcd. for $C_{89}H_{151}N_{21}O_{25}$: 1915.28 Da. Found: *m/z* 949.54 $[M - H_2O + 2H]^{2+}$, 639.36 $[M + 3H]^{3+}$ 633.35 $[M - H_2O + 3H]^{3+}$, 479.78 $[M + 4H]^{4+}$, 475.27 $[M - H_2O + 4H]^{4+}$.

Peptide aldehyde 7

Yield 140 mg, 36% (calculated according to the loading of dipeptide being 85% of original loading and incl. 4 × TFA). ESI-MS, calcd. for $C_{87}H_{148}N_{20}O_{24}$: 1858.23 Da. Found: *m/z* 921.04 $[M - H_2O + 2H]^{2+}$, 621.02 $[M + 3H]^{3+}$, 614.34 $[M - H_2O + 3H]^{3+}$, 465.56 $[M + 4H]^{4+}$.

Peptide aldehyde 8, Iza-EELLKKLEELLKKAG-H

The sequence EELLKKLEELLKKA-N[$CH_2CH(OCH_3)_2$]-*o*-BAL-TG resin (2 × 0.25 mmol, loading 0.19 of dipeptide) was synthesized as described above. 4-Imidazoleacetic acid (Iza) hydrochloride was manually coupled to the sequence using HBTU-HOBt-DIEA (3.8, 4, and 7.8 eqv., respectively) coupling in DMF (10 mL). Then the resin was dried and cleaved with TFA–H_2O (95 : 5) (10 mL) for 4 hours, followed by a wash with TFA (2 × 5 mL). TFA was removed *in vacuo* and the peptide aldehyde was redissolved in CH_3CN–H_2O (1 : 1, 4 ml) and purified by prep C_{18} RP-HPLC. The fractions containing the product were pooled and concentrated *in vacuo* to remove CH_3CN and then lyophilized.

Peptide aldehyde 8

Yield 89 mg, 21% (calculated based on loading of dipeptide being 72% of original loading and incl. 4 × TFA). ESI-MS, calcd for $C_{84}H_{145}N_{21}O_{24}$: 1833.18 Da. Found: *m/z* 908.95 $[M - H_2O + 2H]^{2+}$, 612.33 $[M + 3H]^{3+}$, 605.97 $[M - H_2O + 3H]^{3+}$, 454.74 $[M - H_2O + 4H]^{4+}$.

General procedure for the preparation of 64 AA carboproteins on Gal*p* and Alt*p* templates

The deprotected templates (as specified below) were dissolved in an aqueous buffer 0.1 M NaOAc, pH 4.76. Peptide aldehyde (6 eqv.) was added and after 2 h the reaction was analyzed by HPLC and purified by semi-preparative C_4 RP-HPLC.

General procedure for the preparation of 128 AA carboproteins assembled onto a disulfide functionalized Gal*p* template

The deprotected template **2** (1.5 mg) was dissolved in DMSO (100 μL) and an aqueous buffer 0.1 M NaOAc, pH 4.76 (4 mL). The peptide aldehyde (10 eqv.) was added and after 2 h the reaction was analyzed by HPLC and purified by semi-preparative C_4 RP-HPLC.

Four-stranded 64 AA carboprotein with Tyr on Gal*p* (9)

Yield 9 mg, 86% (incl. 16 × TFA). ESI-MS, calcd. for $C_{375}H_{626}N_{84}O_{114}$: 8135.48 Da. Found: *m/z* 1628.01 $[M + 5H]^{5+}$, 1356.98 $[M + 6H]^{6+}$, 1163.23 $[M + 7H]^{7+}$, 1017.94 $[M + 8H]^{8+}$, 904.94 $[M + 9H]^{9+}$, 814.54 $[M + 10H]^{10+}$, 740.59 $[M + 11H]^{11+}$. Deconvoluted spectra see Fig. 7.

Four-stranded 64 AA carboprotein with Tyr on Alt*p* (10)

Yield 13 mg, 80% (incl. 16 × TFA). ESI-MS, calcd. for $C_{375}H_{626}N_{84}O_{114}$: 8135.48 Da. Found: *m/z* 1356.95 $[M + 6H]^{6+}$, 1163.21 $[M + 7H]^{7+}$, 1017.93 $[M + 8H]^{8+}$, 904.93 $[M + 9H]^{9+}$, 814.54 $[M + 10H]^{10+}$, 740.57 $[M + 11H]^{11+}$.

Eight-stranded 128 AA carboprotein with Tyr (11)

Yield 9 mg, 79% (incl. 32 × TFA). ESI-MS, calcd. for $C_{764}H_{1260}N_{170}O_{230}S_2$: 16561.16 Da. Found: *m/z* 1658.25 $[M + 10H]^{10+}$, 1507.60 $[M + 11H]^{11+}$, 1382.00 $[M + 12H]^{12+}$, 1275.75 $[M + 13H]^{13+}$, 1184.69 $[M + 14H]^{14+}$, 1036.74 $[M + 16H]^{16+}$.

Four-stranded 64 AA carboprotein with D-Pal on Gal*p* (12)

Yield 20 mg, 96% (incl. 16 × TFA). ESI-MS, calcd. for $C_{371}H_{622}N_{88}O_{110}$: 8075.44 Da. Found: *m/z* 1616.10 $[M + 5H]^{5+}$, 1346.89 $[M + 6H]^{6+}$, 1154.61 $[M + 7H]^{7+}$, 1010.41 $[M + 8H]^{8+}$, 898.26 $[M + 9H]^{9+}$, 808.54 $[M + 10H]^{10+}$.

Four-stranded 64 AA carboprotein with D-Pal on Alt*p* (13)

Yield 16 mg, 85% (incl. 16 × TFA). ESI-MS, calcd. for $C_{371}H_{622}N_{88}O_{110}$: 8075.44 Da. Found: *m/z* 1616.39 $[M + 5H]^{5+}$, 1346.86 $[M + 6H]^{6+}$, 1154.61 $[M + 7H]^{7+}$, 1010.41 $[M + 8H]^{8+}$, 898.25 $[M + 9H]^{9+}$, 808.55 $[M + 10H]^{10+}$.

Eight-stranded 128 AA carboprotein with D-Pal (14)

Yield 10 mg, 89% (incl. 32 × TFA). ESI-MS, calcd. for $C_{756}H_{1252}N_{178}O_{222}S_2$: 16441.16 Da. Found: *m/z* 1646.14 $[M + 10H]^{10+}$, 1496.55 $[M + 11H]^{11+}$, 1371.91 $[M + 12H]^{12+}$, 1266.47 $[M + 13H]^{13+}$.

Four-stranded 64 AA carboprotein with Pyp on Gal*p* (15)

Yield 9 mg, 88% (incl. 16 × TFA). ESI-MS, calcd. for $C_{363}H_{610}N_{84}O_{106}$: 7847.23 Da. Found: *m/z* 1570.37 $[M + 5H]^{5+}$, 1308.84 $[M + 6H]^{6+}$, 1122.00 $[M + 7H]^{7+}$, 981.88 $[M + 8H]^{8+}$, 872.90 $[M + 9H]^{9+}$.

Four-stranded 64 AA carboprotein with Pyp on Alt*p* (16)

Yield 12 mg, 79% (incl. 16 × TFA). ESI-MS, calcd. for $C_{363}H_{610}N_{84}O_{106}$: 7847.23 Da. Found: *m/z* 1570.63 $[M + 5H]^{5+}$, 1308.81 $[M + 6H]^{6+}$, 1122.00 $[M + 7H]^{7+}$, 981.89 $[M + 8H]^{8+}$, 872.92 $[M + 9H]^{9+}$.

Eight-stranded 128 AA carboprotein with Pyp (17)

Yield 9 mg, 82% (incl. 32 × TFA). ESI–MS, calcd. for $C_{740}H_{1228}N_{170}O_{214}S_2$: 15984.99 Da. Found: *m/z* 1777.94 $[M + 9H]^{9+}$, 1600.60 $[M + 10H]^{10+}$, 1455.15 $[M + 11H]^{11+}$, 1333.96 $[M + 12H]^{12+}$, 1143.58 $[M + 14H]^{14+}$, 1000.81 $[M + 16H]^{16+}$.

Four-stranded 64 AA carboprotein with Iza on Gal*p* (18)

Yield 13 mg, 91% (incl. 16 × TFA). ESI-MS, calcd. for $C_{351}H_{598}N_{88}O_{106}$: 7747.03 Da. Found: *m/z* 1546.35 $[M - H_2O + 5H]^{5+}$, 1291.05 $[M + 6H]^{6+}$, 1107.44 $[M + 7H]^{7+}$, 969.34 $[M + 8H]^{8+}$, 861.79 $[M + 9H]^{9+}$, 775.72 $[M + 10H]^{10+}$. Deconvoluted spectra see Fig. 6.

Sample preparation for spectroscopic methods

SEC-ICP-MS. For the SEC-ICP-MS experiments, 0.1 mg of the carboproteins **9** and **18** were dissolved in 1 ml 50 mM NH_4OAc and mixed with excess Cu(I)Cl (Merck, 98%). 100 μl was injected into the column and ^{63}Cu was monitored. Each chromatogram gave rise to more than one peak, revealing that several complexes were formed. Using the calibration curve and the right axis, it was possible to measure the size of the eluting peaks in the sample. The largest species were the ones eluting first and the Cu species eluting from 15–22 min had a size between 7–16 kDa.

UV–Vis. Measurements on carboprotein **18** were carried out in 25 mM NaOAc buffer at pH 5.5. A reference and a sample with *ca.* 1 mM carboprotein were separately incubated with solid Cu(I)Cl as in SEC-ICP-MS sample preparations. Following thorough mixing, both the reference and sample solutions turned light blue.

Esterase assay. The esterase activity was evaluated using the 4-nitrophenyl butyrate (*p*-NP-b) assay where the formation of 4-nitrophenolate is monitored at 410 nm. A stock solution of *p*-NP-b in isopropanol was prepared (9.0 μL *p*NP-butyrate in

500 μL iPrOH), and stored in the dark at 4 °C. Immediately prior to measurements, the substrate was diluted 100 times in 0.5 M phosphate buffer, pH 7, and mixed with either pure buffer, solutions of carboprotein **18**, or the corresponding free 4-imidazoleacetic acid hydrochloride (Iza).

Acknowledgements

We thank Dr Per-Ola Norrby, Göteborg University, Dr Morten J. Bjerrrum and Christian Bukh, University of Copenhagen, Faculty of Life Sciences, for valuable discussions, and Karen Jørgensen and Solveig Kallesøe, University of Copenhagen, Faculty of Natural Sciences, for technical assistance with recording CD spectra.

Notes and references

1 (*a*) W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri and A. Lombardi, *Annu. Rev. Biochem.*, 1999, **68**, 779–819; (*b*) L. Baltzer, H. Nilsson and J. Nilsson, *Chem. Rev.*, 2001, **101**, 3153–3163; (*c*) R. B. Hill, D. P. Raleigh, A. Lombardi and W. F. DeGrado, *Acc. Chem. Res.*, 2000, **33**, 745–754.

2 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.

3 (*a*) M. Mutter, P. Dumy, P. Garrouste, C. Lehmann, M. Mathieu, C. Peggion, S. Peluso, A. Razaname and G. Tuchscherer, *Angew. Chem., Int. Ed. Engl.*, 1996, **35**, 1482–1485; (*b*) S. Peluso, P. Dumy, C. Nkubana, Y. Yokokawa and M. Mutter, *J. Org. Chem.*, 1999, **64**, 7114–7120; (*c*) M. Mutter and G. Tuchscherer, *Cell. Mol. Life Sci.*, 1997, **53**, 851–863.

4 (*a*) H. K. Rau and W. Haehnel, *J. Am. Chem. Soc.*, 1998, **120**, 468–476; (*b*) H. K. Rau, N. DeJonge and W. Haehnel, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 11526–11531; (*c*) H. K. Rau, N. DeJonge and W. Haehnel, *Angew. Chem., Int. Ed.*, 2000, **39**, 250–253; (*d*) R. Schnepf, P. Hörth, E. Bill, K. Wieghardt, P. Hildebrandt and W. Haehnel, *J. Am. Chem. Soc.*, 2001, **123**, 2186–2195.

5 (*a*) T. Sasaki and E. T. Kaiser, *J. Am. Chem. Soc.*, 1989, **111**, 380–381; (*b*) 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.

6 (*a*) M. R. Ghadiri, C. Soares and C. Choi, *J. Am. Chem. Soc.*, 1992, **114**, 825–831; (*b*) M. R. Ghadiri, C. Soares and C. Choi, *J. Am. Chem. Soc.*, 1992, **114**, 4000–4002; (*c*) M. W. Mutz, M. A. Case, J. F. Wishart, M. R. Ghadiri and G. L. McLendon, *J. Am. Chem. Soc.*, 1999, **121**, 858–859.

7 M. Goodman, Y. Feng, G. Melacini and J. P. Taulane, *J. Am. Chem. Soc.*, 1996, **118**, 5156–5157.

8 A. K. Wong, M. P. Jacobsen, D. J. Winzor and D. P. Fairlie, *J. Am. Chem. Soc.*, 1998, **120**, 3836–3841.

9 (*a*) A. S. Causton and J. C. Sherman, *Bioorg. Med. Chem.*, 1999, **7**, 23–27; (*b*) A. R. Mezo and A. R. Sherman, *J. Am. Chem. Soc.*, 1999, **121**, 8983–8994.

10 A. R. Mezo and A. R. Sherman, *J. Am. Chem. Soc.*, 1999, **121**, 8983–8994.

11 L. Hou and M. G. Zagorski, *J. Am. Chem. Soc.*, 2006, **128**, 9260–9261.

12 W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri and A. Lombardi, *Annu. Rev. Biochem.*, 1999, **68**, 779–819.

13 (*a*) B. P. Gilmartin, K. Ohr, R. L. McLaughlin, R. Koerner and M. E. Williams, *J. Am. Chem. Soc.*, 2005, **127**, 9546–9555; (*b*) K. Ohr, B. P. Gilmartin and M. E. Williams, *Inorg. Chem.*, 2005, **44**, 7876–7885.

14 (*a*) K. J. Jensen and G. Barany, *J. Pept. Res.*, 2000, **56**, 3–11; (*b*) K. J. Jensen and J. Brask, *Cell. Mol. Life Sci.*, 2002, **59**, 859–869.

15 (*a*) J. Brask and K. J. Jensen, *J. Pept. Sci.*, 2000, **6**, 290–299; (*b*) J. Brask and K. J. Jensen, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 697–700; (*c*) J. Brask, H. Wackerbarth, K. J. Jensen, J. Zhang, J. U. Nielsen, J. E. T. Andersen and J. Ulstrup, *Bioelectrochemistry*, 2002, **56**, 27–32.

16 J. Brask, H. Wackerbarth, K. J. Jensen, J. Zhang, I. Chorkendorff and J. Ulstrup, *J. Am. Chem. Soc.*, 2003, **125**, 94–104.

17 H. Wackerbarth, A. P. Tofteng, K. J. Jensen, I. Chorkendorff and J. Ulstrup, *Langmuir*, 2006, **22**, 6661–6671.

18 J. Brask and K. J. Jensen, *J. Pept. Sci.*, 2000, **6**, 290–299.

19 F. Guillaumie, J. C. Kappel, N. M. Kelly, G. Barany and K. J. Jensen, *Tetrahedron Lett.*, 2000, **41**, 6131–6135.

20 Y.-H. Chen, J. T. Yang and K. H. Chau, *Biochemistry*, 1974, **13**, 3350–3359.

21 W. Kaim, B. Schwederski, *Bioinorganic Chemistry*, Wiley, New York, 1991.

22 E. I. Solomon, *Inorg. Chem.*, 2006, **45**, 8012–8025.

23 P. M. Hanna, D. R. McMillin, M. Pasenkiewicz-Gierula, W. E. Antholine and B. Reinhammar, *Biochem. J.*, 1988, **253**, 561–568.

24 L. Hou and M. G. Zagorski, *J. Am. Chem. Soc.*, 2006, **128**, 9260–9261.

25 D. P. Persson, T. H. Hansen, P. E. Holm, J. K. Schjoerring, H. C. B. Hansen, J. Nielsen, I. Cakmak and S. Husted, *J. Anal. At. Spectrom.*, 2006, **21**, 996–1005.