DOI: 10.1002/cbic.200800263

3- Instead of 4-Helix Formation in a De Novo Designed Protein in Solution Revealed by Small-Angle X-ray Scattering

Rasmus Høiberg-Nielsen,*[a] A. Pernille Tofteng Shelton,[a] Kasper K. Sørensen,[a] Manfred Roessle,[b] Dimitri I. Svergun,[b, c] Peter W. Thulstrup,[a] Knud J. Jensen,*[a] and Lise Arleth*[a]

De novo design and chemical synthesis of proteins and their mimics are central approaches for understanding protein folding and accessing proteins with novel functions. We have previously described carbohydrates as templates for the assembly of artificial proteins, so-called carboproteins. Here, we describe the preparation and structural studies of three α-helical bundle carboproteins, which were assembled from three different carbohydrate templates and one amphiphilic hexadecapeptide sequence. This heptad repeat peptide sequence has been reported to lead to 4-α-helix formation. The low resolution solution structures of the three carboproteins were analyzed by means of small-angle X-ray scattering (SAXS) and synchrotron radiation circular dichroism (SRCD). The ab initio SAXS data analysis revealed that all three carboproteins adopted an unexpected 3+1-helix folding topology in solution, while the free peptide formed a 3-helix bundle. This finding is consistent with the calculated α-helicities based on the SRCD data, which are 72 and 68% for two of the carboproteins. The choice of template did not affect the overall folding topology (that is for the 3+1 helix bundle) the template did have a noticeable impact on the solution structure. This was particularly evident when comparing 4-helix carboprotein monomers with the 2×2-helix carboprotein dimer as the latter adopted a more compact conformation. Furthermore, the clear conformational differences observed between the two 4-helix (3+1) carboproteins based on D-altropyranoside and D-galactopyranoside support the notion that folding is affected by the template, and subtle variations in template distance-geometry design may be exploited to control the solution fold. In addition, the SRCD data show that template assembly significantly increases thermostability.

Introduction

De novo design and total chemical synthesis of artificial protein mimics is a powerful approach to test our understanding of protein folding. By designing protein mimics with reduced complexity, specific sequence-to-folding questions that are difficult to answer in more complex systems can be addressed. Assembly of peptide chains on a template, for example, a cyclic peptide, to form so-called "template assembled synthetic proteins" (TASP) has been suggested as a method for testing protein folding. TASP molecules allow one to mimic an isolated folding unit, for example, a subdomain, and at the same time take into account the fixating and stabilizing force imposed by the rest of the protein.

Since the introduction of the TASP concept by Mutter and co-workers, several groups[1–3] have explored this strategy. Most of this work has focused on designing 4-α-helix bundles by attaching short amphiphilic peptides to templates of various rigidity, for example, linear[4–6] and cyclic peptides,[7–10] porphyrins,[1,11–15] cavitands,[16,17] and phenyl based structures.[18] Despite this endeavor, it still remains elusive whether the template acts by directing the amphiphilic peptides into a predetermined folding topology as proposed by Mutter and co-workers or if the template can be regarded as a neutral scaffold merely holding the peptides together as suggested by Fairlie and co-workers.[18] The reason for the rather poorly understood effect of the template most likely has to do with the experimental difficulties related to structural characterization. First of all, characterization by NMR is made difficult by the structural economy used in de novo design, in which several copies of the same peptide sequence are used. Furthermore, since most de novo designed proteins fail to crystallize, no crystal structure for any TASP structure has been reported, and additional options for structural characterization are limited. Consequently structural elucidation has so far primarily been based on CD spectroscopy, simple chemical shift dispersion in ^{1}H NMR or non-covalent bound fluorescent dyes with low

[a] Dr. R. Høiberg-Nielsen, Dr. A. P. Tofteng Shelton, K. K. Sørensen, Prof. P. W. Thulstrup, Prof. K. J. Jensen, Prof. L. Arleth
Department of Natural Sciences, Faculty of Life Sciences
University of Copenhagen, 1871 Frederiksberg C (Denmark)
Fax: (+45) 35-33-23-98
E-mail: rhn@life.ku.dk
kjj@life.ku.dk
lie@life.ku.dk

[b] Dr. M. Roessle, Dr. D. I. Svergun
European Molecular Biology Laboratory, Hamburg Outstation
22603 Hamburg (Germany)

[c] Dr. D. I. Svergun
Institute of Crystallography, Russian Academy of Sciences
Leninsky pr. 59, 117333 Moscow (Russia)

binding specificity, which provide even less information. However, the information from this kind of structural characterization is limited, since molten-globule-like structures may contain residual, secondary structure close to that expected for the native protein. Likewise, the use of fluorescent dyes is not very informative nor reliable in this case since many native proteins also bind the dyes.[19] Therefore, these techniques should be complemented with a structural technique, preferably one which allows fast data acquisition and does not depend on a protein's ability to crystallize. Small-angle X-ray scattering (SAXS),[20] including recently developed techniques for ab initio data analysis[21] is in particular suited for de novo protein research since the measurements are fast and are performed on solutions of proteins. Here we present a comparative SAXS study of three α-helical bundle carboproteins with different carbohydrate templates to test the effect of the peptide sequence and the template on their global folding.

Results

Design

We have previously described carbohydrates as templates for the design of artificial protein or protein mimics, which we termed carboproteins.[22] Four copies of a peptide sequence reported to favor 4-helix bundles[17] were attached to monosaccharide templates with a parallel orientation. These carboproteins were characterized using mainly CD spectrometry, including denaturation studies, and also through H–D exchange of amide NH's by 1D NMR, which revealed a NH exchange protected core. We observed a reproducible effect of the template (Galp vs. Altp) on the degree of α-helicity of the carboprotein; such a directing effect from the template had not been reported before. A diameter for carboproteins of about 23–27 Å, which is consistent with 4-helix bundle formation in the self-assembled monolayer (SAM), has been proposed from studies of SAMs on Au(111) surfaces.[23,24]

Carboproteins **1** and **2** were synthesized as previously described,[25–27] while novel carboprotein **3** (with a cyclo-dithioerythritol or cDTE template) was prepared by extension of our recently reported synthesis of a related cyclo-dithiothreitol (cyclo-DTT) structure.[24] The four peptide units of carboproteins **1** and **2** have a parallel orientation when a 4-helix bundle is formed. However, the Altp template in **2** has a 1,2-*trans* diaxial orientation at O-2 and O-3. The 2-helix carboprotein **3** was designed to dimerize, which could occur in parallel, antiparallel, "bisecting U" topology or "cross-linked" (extended) between several monomers. Template **3** (cDTE) is a *meso* compound, and hence stereochemically pure. One aim of this study was to analyze the effects of peptide sequence (for example, side-chain packing) on the folding and compare them with directing effects of the templates (Figure 1).

Small-angle X-ray scattering

The recorded scattering profiles of carboproteins **1**–**3** and the free peptide **4** in Figure 2A show remarkably similar features, although differences are seen in the finer details. Next, we focused on Fourier-transformed data,[28] which provides information on the intraparticle distances, that is, to the possible connection lines within the particle. The differences between carboproteins **1**–**3** and the complex of the nontemplated peptide **4** are more easily observed in the indirect Fourier transformed data (Figure 2B), in which the obtained pair distance distribution functions ($p(r)$-functions) of carboproteins **1**–**3** clearly differ from each other and from the $p(r)$ of the free peptide. The nearly bell shaped $p(r)$-function of nontemplated peptide **4** with a D_{max} of 36 Å indicates a rather compact spherical structure. However, for all carboproteins (**1**--**3**) the shape of the $p(r)$-functions agrees with the structure typically found for two-domain proteins. The peak at 19 Å is the most representative intramolecular distance in the protein, and the shoulder at about 42 Å corresponds to the distance between the center of mass of each domain.[29–31] Furthermore, the

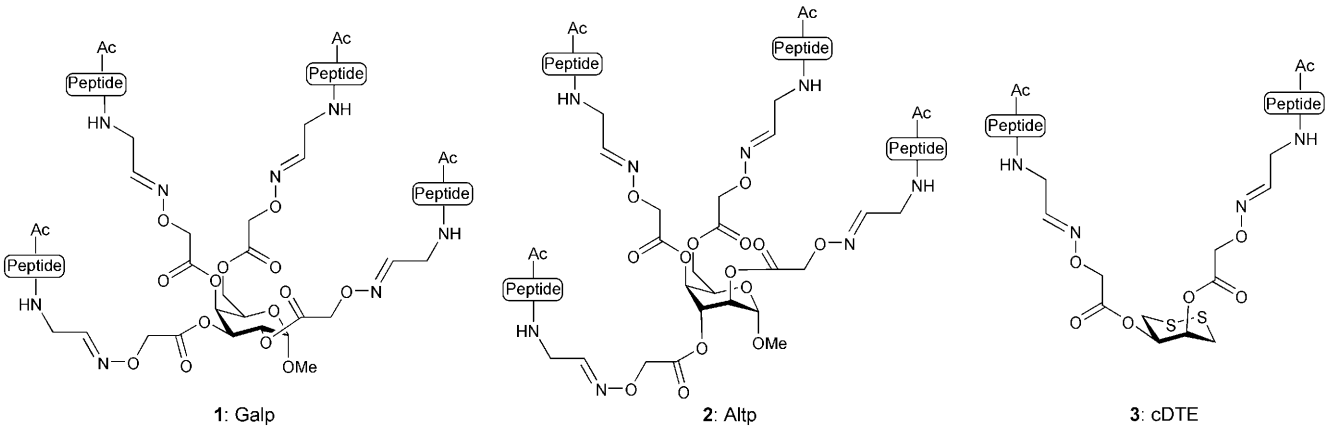

Figure 1. Structure of carboproteins **1**–**3** assembled on methyl α-galactopyranoside (Galp), methyl α-altropyranoside (Altp), and cyclo-dithioerythritol (cDTE). Peptide sequence: Ac-YEELLKKLEELLKKAG-H. Carboproteins **1** and **2** are structural isomers (**2** has a *trans* diaxial arrangement of the O-2 and O-3 hydroxyls). The 2-helix cDTE carboprotein **3** was designed as a 'hemi' 4-helix bundle, which would dimerize to a 2×2-helix bundle. For comparison, the corresponding nontemplated peptide Ac-YEELLKKLEELLKKAG-NH_2 (**4**) was included in this study.

Table 1. List of parameters obtained from indirect Fourier transformation.

	20 °C				74 °C			
Name	$I(0)$ [10^{-3} cm^{-1}]	M_W [kD]	R_g [Å]	D_{max} [Å]	$I(0)$ [10^{-3} cm^{-1}]	M_W [kD]	R_g [Å]	D_{max} [Å]
carboprotein 1	4.20 ± 0.08	6.32	18.45 ± 0.56	59.41 ± 2.60	2.95 ± 0.04	4.44	16.86 ± 0.41	57.03 ± 1.90
carboprotein 2	4.69 ± 0.09	7.06	24.10 ± 0.73	78.67 ± 2.85	2.65 ± 0.05	4.00	15.33 ± 0.43	49.13 ± 2.25
carboprotein 3	4.37 ± 0.06	6.58	16.88 ± 0.38	53.79 ± 1.89	3.18 ± 0.03	4.79	15.60 ± 0.22	49.22 ± 1.53
peptide 4	3.19 ± 0.04	4.80	13.32 ± 0.1	36.05 ± 0.64	–		–	–

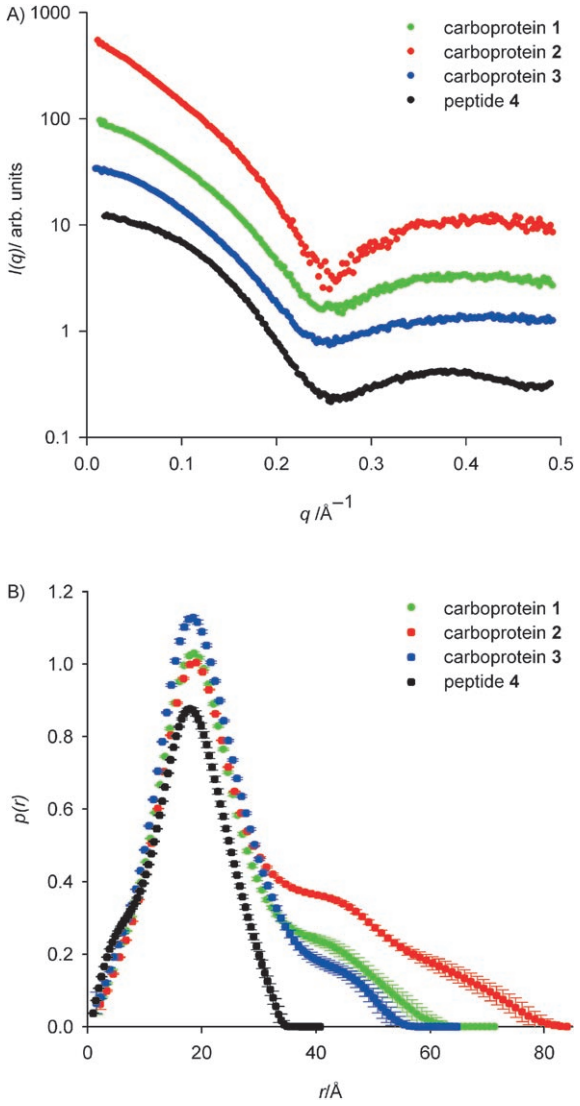

Figure 2. A) Background subtracted raw data that has been scaled. B) Pair distance distribution function obtained by an indirect Fourier transformation of the data.

slowly decreasing tail region following the shoulder reveals that one of the two domains may have an elongated form.

The radius of gyration, R_g, and the maximal diameter, D_{max}, as determined by indirect Fourier transformation (See Table 1) support the view that carboprotein **2** adopts a solution structure significantly more extended than that of carboproteins **1** and **3**, as the dimensions of protein **2** are considerably larger. However, when heated these differences diminish, and in all three cases R_g as well as D_{max} are reduced. The decrease in forward scattering for all carboproteins when heated is most likely simply an effect of the change in mass density and hydration pattern; this occurs as a result of exposure of the hydrophobic surface.

Ab initio modeling using GASBOR

An ab initio method based on chain-compatible dummy residues developed by Svergun and co-workers[32] was used to restore the 3D solution shape from the experimental data in the q range $0.008 < q < 0.496$ Å^{-1}. This method employs a simulated annealing procedure to find the spatial distribution of dummy residues that gives the best fit against the experimental data. Furthermore, a constraint is imposed on the models to ensure that they have a spatial distribution of dummy residues corresponding to that of a normal protein. Since the number of dummy residues exceeds the actual information content of the scattering curve, this method does not give unique solutions.[33–35] To take into account the non-unique nature of the method, 15 ab initio models were generated from each dataset, and subsequently an average structure was found using the programs SUPCOMB and DAMAVER.[36] The average χ values of the restored models against the raw data were 2.02, 2.72, 2.21 and 2.65 for carboproteins **1**, **2, 3** and the nontemplated peptide **4**, respectively. Representative model fits are displayed in Figure 6. Moreover, the individual solutions of the restored envelopes seemed to be robust since the normalized spatial discrepancy (NSD) values on average were reasonably low (0.836, 0.981, 0.854 and 0.783 for carboproteins **1**, **2**, **3** and peptide **4**, respectively; NSD values below unity indicate that models are similar).[36]

The SAXS reconstructed molecular envelopes of carboproteins **1**–**3** presented in Figure 3 have clear similarities. Apparently all carboproteins have two components: a compact nearly spherical part (~30–40 Å) and a thinner elongated part (~20–50 Å). Furthermore, carboproteins **1**–**3** as well as the complex of the nontemplated peptide **4** all have a small, but noticeable cavity inside (~5–9 Å). Besides these common features, several individual features can be seen. The solution structure of carboprotein **3** reveals that the compact part of the protein is wider, and the thin elongated part is accordingly shorter. The bulky tail-region of carboprotein **2** is also clearly different form that of carboproteins **1** and **3**. In the case of carboprotein **2**, all the independent models included in the structural averaging had a cluster of dummy residues situated near the end of the tail part; however, this part of carboprotein **2**

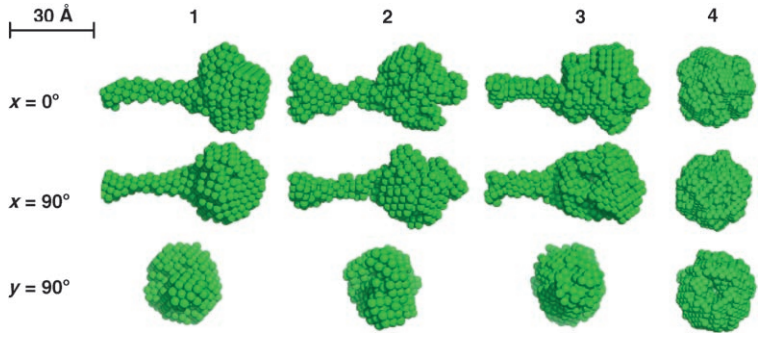

Figure 3. Ab initio reconstructed molecular envelopes of the solution shape of the three de novo designed carboproteins **1**–**3** and the free peptide **4** viewed from different angles. The envelopes of all three carboproteins seem to be composed of two different parts: a compact globular part and a thinner tail-like part. Note that the size and shape of the compact part resembles that of the complex of peptide **4**. The reconstructed envelopes of the carboproteins as well as the free peptide complex have a cavity inside (not visible in the figure).

had a tendency to be slightly more structurally diversified than the remaining part of carboprotein **2**, which gave raise to the marginally higher NSD-value for carboprotein **2**.

Rigid body modeling

Rigid body modeling was employed as a supplementary tool to provide additional support for the results obtained by ab initio modeling. The program SASREF[37] was used to refine the tertiary conformation of an ensemble of α-helical model peptides against the experimental SAXS-data. The α-helical model peptide was generated as a standard α-helix using the program PyMol, and no further structural optimization was done. This approach allowed us to carefully examine the oligomeric state of peptide **4**. As can be seen in Figure 4, the fits obtained when using monomeric or dimeric helices are significantly worse than those obtained when using trimeric or tetrameric helices. The difference between the model fit obtained with trimer and tetramer helixes are somewhat smaller, but nonetheless, the fact that the χ value for the 4-helix bundle is more than twice that of the 3-helix bundle clearly suggests that the peptide **4** self-assembles into a 3-helix bundle. Furthermore, as judged form the scattering pattern of peptide **4** in Figure 4, the formation of this 3-helix bundle is insensitive to a eightfold dilution; this suggests that the formation of this structure is rather specific. Monomeric and dimeric states are not promoted upon dilution. This finding is consistent with previous observations of carboproteins that appeared to have only one peak in their size exclusion chromatography profiles.

Based on the similarities in size and shape of the compact part of carboproteins **1**--**3** on the one hand, and the complex formed by self-association of peptide **4** on the other, we analyzed carboproteins **1**–**3** by rigid body modeling of only two components, namely, an α-helical peptide already used to model the nontemplated peptide **4** complex and the final 3-helix bundle model obtained by modeling the peptide **4** complex. When doing so, the obtained models were basically in agreement with the ab initio models. Thus, the fourth helix was not in close proximity to the 3-helix bundle, but was rather protruding away from the center (results not shown). However, especially in the case of carboproteins **1** and **2**, the two components of the model were clearly disconnected despite the fact that SASREF penalizes disconnected models. This suggests that the fourth peptide strand has greater dimensions than what can be accounted for by an α-helix. Moreover, when using this approach the χ values were unsatisfactory high (4.02, 4.44 and 4.06 for carboproteins **1**, **2** and **3**, respectively).

In order to increase the dimensions of models and improve the low-q part of the fits, the three helix

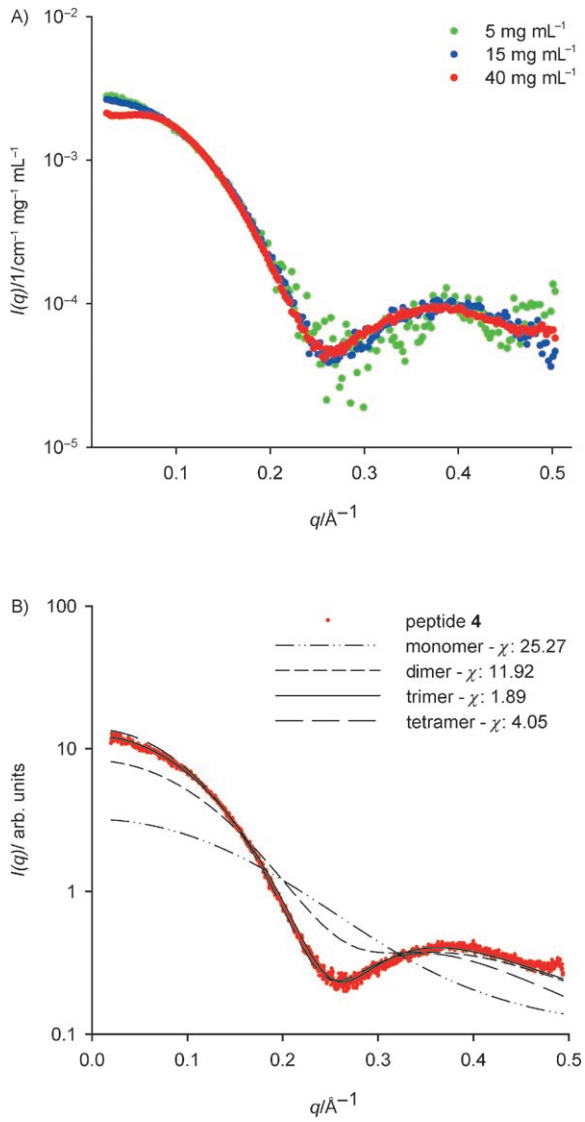

Figure 4. A) SAXS-data of the complex of peptide **4** recorded at three different concentrations. The depression of the scattering intensity at low-*q* and 40 mg mL^{-1} is a result of interparticle interference. B) Rigid body modeling of atomic models of the amphiphilic α-helix hexadecapeptide. The graph shows the fits obtained by rigid body modeling of monomeric, dimeric, trimeric and tetrameric α-helix peptides against the experimental SAXS-data from peptide **4**. As can be seen visually as well as from the χ-values displayed in the legend, the best fit was obtained by modeling the data as a 3-helix bundle.

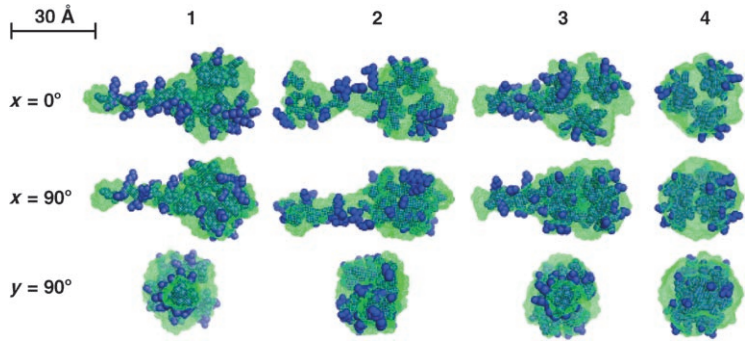

Figure 5. Results of rigid body modeling. The Figure shows the models obtained by rigid body modeling (blue spheres) overlaid with the envelopes obtained by ab initio modeling (green mesh). The rigid body model of the nontemplated peptide **4** complex has been obtained by rigid body modeling of three α-helical model peptides against the experimental SAXS data. The resulting 3-helix bundle model in combination with an additional peptide strand was used to model carboproteins **1**–**3**. Prior to the process of tertiary structure refinement the fourth α-helical peptide strand was individually stretched to fit each of the envelopes of the ab initio models of carboprotein **1**–**3**. The rigid body models and the ab initio envelopes were superimposed with the program SUBCOMB. A high degree of similarity was observed, which supports the ab initio results.

bundle portion of the model was conserved but the fourth hexadecapeptide was stretched manually for each of the carboproteins, thus spoiling its α-helical structure. This was done by taking an atomic model of α-helical hexadecapeptide strand and stretching it to fit into the thin elongated part of each ab initio envelope, as shown in Figure 5. The two components were aligned using the program PyMol. The α-helical peptide strand was stretched until it fitted into the elongated part of the ab initio envelope. The rigid body models obtained by this procedure are exhibited in Figure 5. As can be seen, the models are in very good agreement with the ab initio models. Moreover, the χ values for the theoretical scattering intensities of these models were significantly improved in comparison to the experimental data (3.06, 3.36 and 3.58 for carboproteins **1**, **2** and **3**, respectively). This improvement was mainly achieved in the lower-*q* region, thus reflecting that these models describe the overall shape and dimensions of the molecules more accurately. Furthermore, this suggests that the fourth helix of carboproteins **1**–**3** does not adapt an α-helical structure. However, despite the fact that these models give an accurate fit to the experimental data at low-*q*, carboproteins **1**–**3** do not give a good agreement at high-*q*. Whereas these rigid body models may be in good agreement with overall shape and topology of the molecules they do accurately describe the finer details above ~0.30 Å, which are much more precisely represented in the retrieved ab initio models. The rigid body model of the complex formed by peptide **4**, on the other hand, fits the experimental data surprisingly well throughout the entire *q*-range, strongly suggesting that the complex of peptide **4** is, in fact, a 3-α-helix bundle.

Synchrotron radiation circular dichroism (SRCD)

In order to provide further spectroscopic information to investigate the differences in the SAXS solution structure of carboproteins **2** and **3**, these carboproteins were analyzed by synchroton radiation CD spectroscopy. As can be seen from Figure 7, both structures have a pronounced maximum at 190 nm and a minimum at 220 nm, which is the fingerprint of an α-helical structure. Thus, both spectra show signals indicative of high content of helical secondary structure. Based on $[\theta]_{220}$, the degree of α-helicity was calculated according to the formula pro-

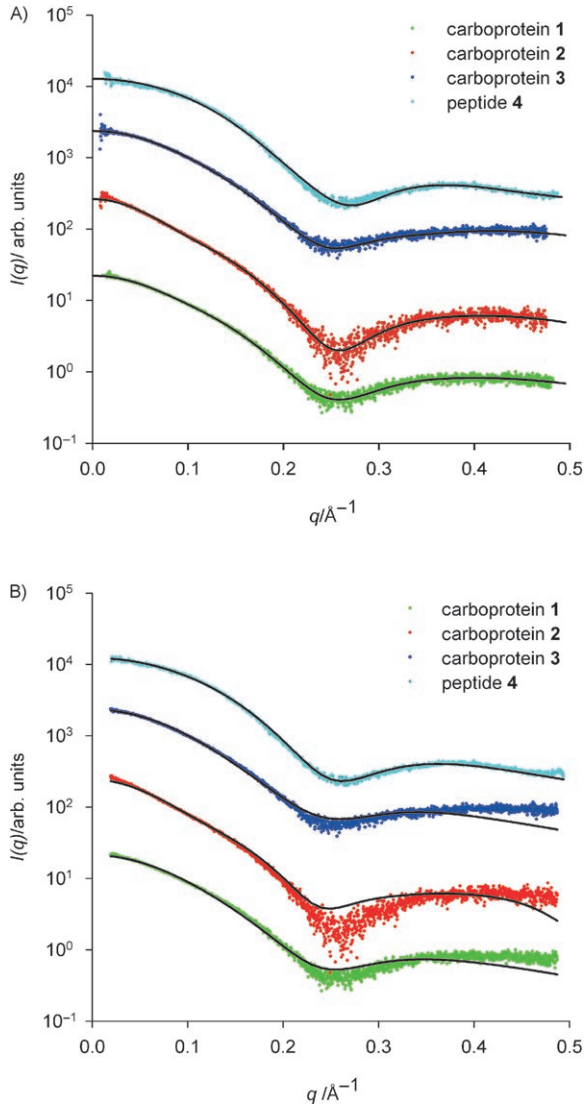

Figure 6. A) Representative model fits of retrieved ab initio models of carboproteins **1**–**3** and the complex of peptide **4**. B) Rigid body model fits. The complex of peptide **4** was modeled with three α-helical peptide strands. The final model of this complex was used in combination with a fourth non-helical peptide strand to model the structure of carboproteins **1**–**3**. Whereas the 3-helix rigid body model fits the experimental data surprisingly well throughout the entire *q*-range, the obtained models for carboproteins **1**–**3** do not fit the finer structure well above ~0.30 Å.

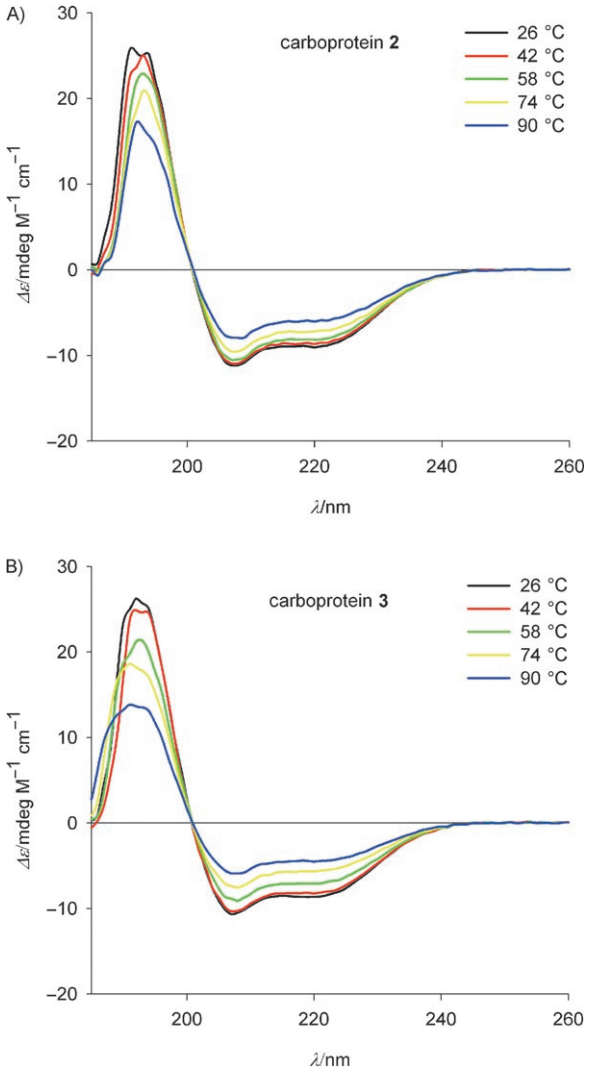

Figure 7. A) Synchrotron radiation circular dichroism of carboprotein **2**. Note the maxima at 190 nm and the minima at ~208 nm and ~222 nm, which indicate an α-helical structure. B) Synchrotron radiation circular dichroism of carboprotein **3**.

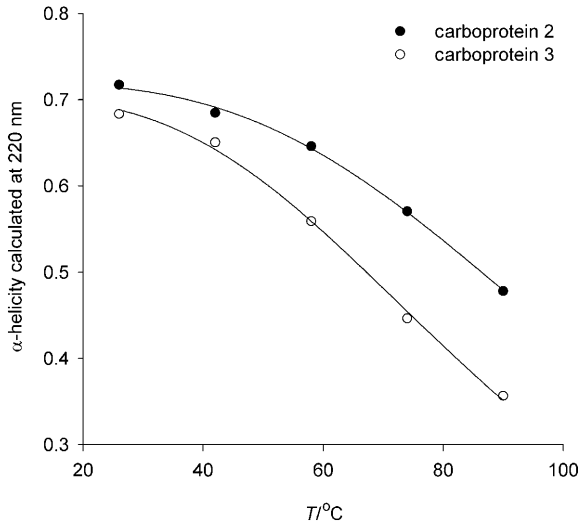

Figure 8. α-Helicity of 4-helix carboprotein **2** and of 2×2-helix **3** as function of temperature. The α-helicity was calculated based on $[\theta]220$ as described by Chen and co-workers[38] Data were fitted to a sigmoidal (logistical) function. The steeper decline in the curve of the 2×2 helix dimer reveals increased stability of carboproteins gained by attaching the helixes to a common template.

posed by Chen and co-workers[38] to 72 and 68% for carboproteins **2** and **3**, respectively. These results agree with the 67% α-helicity previously reported for carboproteins.[25] The differences observed at $[\theta]_{220}$ were found throughout the spectra. However, since the spectra have the characteristics of α-helical structure, these differences were especially pronounced at 190 and 220 nm.

For the purpose of comparison, the α-helicity was calculated based on $[\theta]_{220}$ at five different temperatures. As can be seen from the decrease in α-helicity as a function of rising temperature in Figure 8, carboproteins **2** and **3** obviously have different degrees of thermostability, with the template Altp assembled carboprotein **2** being the most stable.

Discussion

The presented results unequivocally show that carboproteins 1–3 do not form 4-helix bundles. First of all, the indirect Fourier transformation data clearly excludes the possibility of 4-helix bundle formation, since a 4-helix bundle of these structures would have a D_{max} of less than 40 Å. The molecular length of the α-helical form of this peptide is ~24 Å assuming 3.6 residues per turn and a helical repeating unit of 5.4 Å. In addition, the template and linkers are likely to extend these structures by 5–7 Å; however, even if the Gly-Aoa linkers were fully extended, a 4-helix bundle of these structures could not extend beyond 40 Å. Similarly, the molecular width of a 4-helix bundle of this peptide is expected to be ~30 Å and could not exceed 40 Å as this would disconnect the helixes.

Thus, the question arises: which alternative solution folds are adopted by carboproteins **1**–**3**? The distinct two-domain shape of the *p*-(*r*) functions of carboproteins **1**–**3** suggests that not all of the four helices are bundled together. Indeed, this is also what must be concluded from the reconstructed ab initio envelopes as well as the rigid body models. These clearly show that carboproteins **1**–**3** adopt a 3+1 helix fold, whereas nontemplated peptide **4**, forms a 3-helix bundle. Moreover, the complex formed from the nontemplated peptide **4** showed no tendency for formation of higher oligomeric states even at 40 mg mL^{-1} and it was insensitive to an eightfold dilution; this suggests that the 3-helix formation of this peptide is rather specific. The notion of 3-helix bundle formation is, furthermore, consistent with the α-helicities calculated based on the CD data. The calculated α-helicities were 72% and 68% for carboproteins **2** and **3**, respectively, which are again compatible with a structure consisting of three helical and one random coiled-like peptide.

The ab initio models of the carboproteins as well as trimer of peptide **4** all turned out to have a small cavity in the middle of the bundle. This is supported by the rigid body modeling of the trimer complex formed by peptide **4**. The resulting spatial

arrangement of the helices shows that the helices are unexpectedly far apart; this suggests a rather loose packing of the hydrophobic amino acid side-chains. One explanation for this packing behavior could be related to the simplistic design of the sequence. Since the all-Leu design of the hydrophobic core might be incompatible with the typical knobs-into-holes type of side-chain packing that is formed from the alternating packing of small and large hydrophobic side-chains, this could provide a more dynamic core.

The length of the compact part of the envelope of carboproteins **1** and **2** is ~30 Å, which is close to that expected of a 16 amino acid helix (~24 Å) plus the linker and template (~5-7 Å). Carboprotein **3**, on the other hand, has a more elongated body that extends to ~35 Å, but at the same time the length of the excluded helix is accordingly shorter. The additional flexibility of this 2×2 helix carboprotein seems to enable it to adopt a conformation in which the fourth helix is less excluded from the rest of the protein. Therefore this carboprotein probably has less exposure of hydrophobic surface. Furthermore, as a result of this additional flexibility, the compact part of carboprotein **3** adopts a solution structure closer to that of the bundle formed by self-association of peptide **4**. In contrast, the conformational flexibilities of carboproteins **1** and **2** appear to be restricted by the template since the fourth amphipatic "helix" has a more extended and consequently a more solvent exposed conformation. Thus, the SAXS-data suggest that the template restricts the "fourth" helix to adopt a thermodynamically less favorable conformation. In other words, there is a directing effect of the template.

Moreover, when comparing the $p(r)$-functions of carboproteins **1** and **2**, it is clear that template stoichiometry also impacts the solution structure. This is especially evident when looking at interatomic distances above 35 Å, as these distances are much more populated in carboprotein **2** than in carboprotein **1**. This means that the D-altropyranoside template somehow facilitates the folding of the excluded peptide strand into a less extended and more compact conformation. The pronounced differences between $p(r)$-functions at intermediate to high distances show that the D-galactopyranoside template forces the construct to adopt a fundamentally different solution fold than that of carboproteins **1** and **3**.

The question arises as to whether specific heptad repeat sequences form coiled-coils or 3-, 4-, 5-helix bundles or even higher order structures. In this work, we focused on formation of 3- versus 4-helix structures. The de novo protein coil-Ser which has an all-Leu core with Leu in positions *a* and *d*, forms an antiparallel trimer.[39] In contrast, there are only few examples of *parallel* triple helical structures. However, the de novo sequence coil-V^aL^d forms a parallel trimeric coiled coil (VEALEKKVAALESK-VQALEKKVEALEHG).[40] The α_1A (Ac-GELEELLKKLKEELKG-OH) and α_1B (Ac-GELEELLKKLKELLKG-OH) sequences from the DeGrado group were reported to form tetrameric assemblies as 4-helix bundles. Interestingly, the formation of disulfide linked 5-helix assembly rather than expected 4-helix bundle has been described.[41] The sequence used in the present study was Ac-$Y^aE^bE^cL^dL^eK^fK^gL^aE^bE^cL^dL^eK^fK^gA^aG$-H, N-acetylated and coupled as the C-terminal aldehyde in which a, b, c, d, e, f, and g indicate positions in the heptad repeat. It thus allows for four 'turns' in an α-helix, sufficient for helix bundle formation, and includes three Leu and one Tyr (N-terminal) and one Ala (C-terminal) in a and d positions. Sherman and coworkers[17] have used related sequences to design 3- and 4-helix bundles[42] and have found that their 4-helix structures are more stable than the corresponding 3-helix structures and that the trimer is in a monomer-dimer equilibrium. They speculated that they could 'force' formation of a 3-helix or 4-helix by placing the corresponding number of helices on their cavitand templates. The results we present here contradict the view implicit in reports from several groups working with TASPs, for which no X-ray structures have been reported, that the number of α-helical sequences attached to a central template determines the oligomeric state of the formed helical bundle.

Based on sedimentation equilibrium experiments and CD spectroscopy on 4-α-helix TASPs, Fairlie and coworkers, on the other hand, concluded that template shape, size and directionality did *not* influence helix-bundle formation given sufficiently long linkers.[18] This conclusion somewhat contradicts the finding that carboproteins have a substantial increase in α-helicity compared to the unattached peptide.[25,26] Similarly, it has been shown that β-sheet formation can be induced well below the critical chain length onset of β-sheet structure by fixing the peptide to rigid template molecules.[43–45] The work presented in this paper supports the view that template assembly impacts on structure formation; however, to our knowledge this is the first time it has been thoroughly demonstrated that template assembly influences tertiary structure formation in addition to inducing secondary structure.

It was somewhat surprising that the dimensions of all carboproteins decreased upon heating. For the vast majority of naturally occurring proteins, R_g actually increases upon denaturation. Furthermore, in the case of disulfide-free proteins without prosthetic groups, R_g scales with the chain contour length according to the power low relationship expected for a Gaussian random coil.[46] However, similar observations have been reported before for other de novo designed triple helix bundles.[47] Thus, in these cases the elongated native structures have greater dimensions than their random coil-like denatured state.

Finally, it should be emphasized that the thermal stabilities of carboprotein **2** and **3** are very impressive; however, these high stabilities are not unusual for an all leucine α-helix bundle, which typically is highly resistant to elevated temperatures[48–51] as well as chemical denaturation.[25,42,52,53] In fact, the burial of hydrophobic side groups is considered to be the main packing force, and hence the key parameter determining the stability of helix bundles.[54,55] More interestingly though, is the noticeably higher thermal stability of carboprotein **2** compared to **3**, which reveals the impact of the template on stability. In light of the cited reports on highly stable α-helix bundles, this finding may appear trivial. Nonetheless, more recent reports have shown that conformational specificity in redesigned proteins often is achieved at the expense of decreased stability.[48,56] For instance, structural specificity in de novo designed proteins may be obtained by increasing the specificity of inter-

helical contact surfaces. This has been achieved through substitutions of nonpolar, aliphatic side-chain groups with aromatic or conformationally restricted β-branched ones (for example, substitution of Leu with Tyr).[19] Therefore, the stability gained by the 'template assembly' may be exploited to maintain a sufficient degree of stability which would allow increase of the conformational specificity, and consequently decreasing the stability, without compromising the fold.

Conclusions

We have designed and synthesized three carboproteins to investigate the effects of the template assembly on stability and structure. By means of SAXS and ab initio data analysis we have shown that these carboproteins adopt an unexpected 3+1 helix folding in solution. The template stoichiometry design did *not* control the overall folding topology. That is, the template carrying four potentially α-helical peptide strands did not form a 4-helix bundle, but a 3-helix bundle; however, the template did actually have a noticeable influence on the conformation. The template effect was especially evident when comparing the solution structure of the 2×2-stranded carboprotein on one hand and the 4-stranded carboproteins on the other. Furthermore, the reconstructed solution structures of carboproteins **1** and **2** revealed conformational differences. Thus, subtle variations in template distance-geometry design may be used to control the solution structure. In addition, our results show that the 'template assembly' utilized in the design of the presented carboproteins contributes to a significant increase in thermostability, which may be exploited in the pursuit of creating conformational specific de novo designed proteins.

Experimental Section

General: The synthesis of Boc_2-Aoa-OH has been described in an earlier publication from our laboratories.[57] It is now available from NeoMPS (Strasbourg, 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. Most chemicals were purchased from Sigma–Aldrich, Fluka, NovaBiochem (Schwalbach, Germany) or Iris Biotech (Marktredwitz, Germany) and used without further purification.

A phosphate buffer (10 mM, pH 7) was prepared by dissolving $NaH_2PO_4{\cdot}2H_2O$ (0.78 g, 5 mmol) in water (500 mL) and then titrating to pH 7 with NaOH (1 M aq). An acetate buffer (0.1 M, 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 to quantify the amount of Fmoc cleaved. Analytical HPLC was performed on a Waters system (Milford, MA, USA), with a 600 control unit, a 996 PDA detector, a 717 Plus autosampler, and Millenium 32 control software. For the analysis of peptides a Waters XTerra 300 C_{18} column (3.5 μm, 3.0×50 mm) was used. For the analysis of carboproteins a Waters Symmetry 300 C_4 column (5 μm, 3.9×150 mm) was used. Preparative and semipreparative 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. Semipreparative HPLC was performed on a single C_4 25×10 mm 15 μm 300 Å column cartridge (FeF Chemicals, Køge, Denmark) or a single C_{18} 40×10 mm HR 6 μm 60 Å column cartridge of Waters Prep Nova Pak. 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.

NMR spectra were recorded on a Bruker Avance 300 spectrometer. 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 from Waters equipped with an ESI probe and spectra were acquired in positive mode.

Synthesis of carboproteins. (4*R*,5*S*)-1,2-dithiane-4,5-diyl bis(2-(bis(*tert*-butoxycarbonyl)aminooxy)acetate): (*S*,*R*)-cyclo-dithioerythritol (39 mg, 0.26 mmol) and *N*,*N*-di-Boc-aminooxyacetic acid (251.4 mg, 0.86 mmol) were dissolved in pyridine-CH_2Cl_2 (1:1, 6 mL). The reaction was stirred with molecular sieves (4 Å) for 1 h before diisopropylcarbodiimide (DIC, 109 mg, 0.86 mmol) and 4-dimethylaminopyridine (DMAP, 0.13 mL, 0.17 mmol) were added. After 2 h additional DIC (0.17 mL, 1.38 mmol) was added and the reaction was stirred for an additional 30 min. Finally, the molecular sieves were removed and the reaction mixture was concentrated in vacuo. The solution was dissolved in CH_3CN, centrifuged and purified by preparative C_{18} RP-HPLC. Yield 129 mg, 72%. ^1H NMR ($CDCl_3$, 300 MHz), δ: 1.41 (s, 36H), 2.4–3.41 (m, 4H), 4.45 (s, 4H), 5.10–5.22 (m, 2H); ^{13}C NMR (300 MHz, $CDCl_3$,): $\delta = 28.04$, 72.17, 84.51, 150.07, 166.26. ESI-MS, calculated for $C_{28}H_{46}N_2O_{14}S_2$: 698.2391 Da. Found *m/z* 717.2574 $[M+H_2O+H]^+$.

Carboprotein 3, (Ac-Tyr-Glu-Glu-Leu-Leu-Lys-Lys-Leu-Glu-Glu-Leu-Leu-Lys-Lys-Ala-Gly)2-cyclo-DTE: (4*R*,5*S*)-1,2-dithiane-4,5-diyl bis(2-(aminooxy)acetate) (4*R*,5*S*)-1,2-dithiane-4,5-diyl bis(2-(bis(*tert*-butoxycarbonyl)aminooxy)acetate) (35 mg, 0.05 mmol) was dissolved in CH_2Cl_2/TFA (1:1, 5 mL) and stirred for 1 h. The solution was then concentrated in vacuo and lyophilized. Yield 15 mg, >99%.

Carboprotein **3** (4*R*,5*S*)-1,2-dithiane-4,5-diyl bis(2-(aminooxy)acetate) (5 mg, 16.8 μmol) and peptide aldehyde (100 mg, 50.4 μmol) were dissolved in 4 mL of a 2:1 solution of CH_3CN and NaOAc buffer (0.1 M, pH 4.76). The reaction was stirred for 6 h and purified by prep C_{18} RP-HPLC; yield 12 mg, 84%. ESI-MS, *m/z*: 1374.57 $[M+3H]^{3+}$, 1031.38 $[M+4H]^{4+}$, 825.34 $[M+5H]^{5+}$, 688.12 $[M+6H]^{6+}$, (calcd for $C_{188}H_{314}N_{42}O_{56}S_2$: 4122.89 Da).

Peptide 4, Ac-Tyr-Glu-Glu-Leu-Leu-Lys-Lys-Leu-Glu-Glu-Leu-Leu-Lys-Lys-Ala-Gly-NH2: Peptide 4 was synthesized using standard automated Fmoc-chemistry on a Novasyn TGR 0.27 $mmol\,g^{-1}$ resin and purified by prep C_4 RP-HPLC; yield 132 mg, 68%. ESI-MS, *m/z*: 984.69 $[M+Na+H]^{2+}$, 973.70 $[M+2H]^{2+}$, 649.47 $[M+3H]^{3+}$ (calcd for $C_{90}H_{153}N_{21}O_{26}$ 1944, 13 Da).

Sample preparation and initial characterization: Samples ($10\,mg\,mL^{-1}$) were prepared by dissolving lyophilized sample (1 mg) into buffer (100 μL, 50 mM NaOAc, pH 5.5). The concentrations were controlled by measuring absorption at 280 nm on a Nanodrop spectrophotometer (NanoDrop Technologies). Prior to SAXS-measurements the carboproteins were subjected to size exclusion chromatography. Judged from the retention profiles, monodisperse conditions were found for carboproteins **1**–**2** up to $10\,mg\,mL^{-1}$. Carboprotein **3**, on the other hand, formed a monodisperse dimer as designed.

Small-angle X-ray scattering measurements and data processing: SAXS-measurements were performed on the X33 small-angle X-ray scattering beamline of the European Molecular Biology Labo-

ratory (EMBL) at the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) using standard procedures. Data were collected on a MAR345 image plate detector covering a range of $0.01 < q < 0.496$ Å^{-1} ($q = 4\ \pi\ \sin\ \theta/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength). The scattering intensities of buffer backgrounds were measured both before and after the sample and the averaged background scattering were subtracted from the scattering of the sample according to standard procedures. Reference solutions of bovine serum albumin (BSA) of known concentration (~5 mg mL^{-1}) were used for absolute calibration. As an independent check, absolute calibration was performed on water; the deviation of the two methods was found to be within 2%. The data were converted to into a direct space representation in terms of the pair distance distribution function, $p(r)$, by means of a indirect Fourier transformation[58] using a Bayesian analysis method.[28] In the case of homogenous particles, the $p(r)$-function represents the probability of finding a point within the particle at the distance r starting form an arbitrary point inside the particle. Thus for proteins, which almost have a uniform electron density, the $p(r)$ function can be considered as a histogram of inter-particle distances. The $p(r)$ is related to the scattering intensity $I(q)$, through the Fourier Transform [Equation (1)]:

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty rqI(q) \sin(rq)\, dq \qquad (1)$$

Ab initio shape determination: The solution shapes of carboproteins **1**–**3** were reconstructed from the experimental data using the ab initio method GASBOR.[32] The method allows for the retrieval of structural information down to a resolution of ~0.5 nm.[32] In this method, simulated annealing and nonlinear minimization routines are employed to minimize the discrepancy between a chain-like ensemble of dummy residues and the experimental data. At each calculation step, one dummy residue is moved and the resulting scattering curve is calculated using the Debye formula. Traditionally, SAXS ab initio methods have avoided contribution from internal structure by neglecting the higher q-portion of the scattering curve. GASBOR, on the other hand, has a build-in constraint that allows it to exploit the entire q-range and thus obtain a higher resolution. This is achieved through a so-called histogram penalty weight constraint that forces the model to have a distribution of adjacent dummy residues corresponding to the distribution of α-atoms typically found in proteins. Thus, besides optimization of the conformation of the dummy residues to the experimental data this method tries to obey certain physically meaningful constraints. The basic function to be minimized using this method is [Eq. (2)]:

$$E(r) = \chi^2 + \alpha P(r) \qquad (2)$$

Where χ^2 is the reduced chi-square of the experimental data against the model and $P(r)$ is a penalty term which is added via the Lagrange multiplier α. The $P(r)$ consists of three independent penalty terms: The first term imposes a natural spatial distribution of the dummy residues by demanding that a spatial distribution in the model corresponds to that typically found in proteins. The second term imposes interconnectivity in the model, such that interconnected models are preferred over unconnected models. The last term enforces a kind of inertia in the center of mass of the model such that the center of mass is not moved unnecessarily during the process of ab initio fitting. Even though the histogram penalty weight forces the model to have a predefined distribution of adjacent dummy residues it does not require the chain to be unbranched. For this reason GASBOR works equally well with branched molecules such as TASPs. By using this method it was possible to exploit the experimental resolution of 13 Å ($2\pi/q_{max}$) in the data analysis.

All carboprotein ab initio models were obtained using 68 dummy residues to account for both the peptide part (64 dummy residues) and the carbohydrate part (4 dummy residues). To take into account the difference in mean electron density between the peptide and carbohydrate part, the number of carbohydrate dummy residues was calculated as [Eq. (3)]:

$$N_{residues} = \frac{M_{W\,temp} \cdot \nu_{temp}\, \rho_{temp} - \rho_{buffer}}{M_{W\,avAA} \cdot \nu_{avAA}\, \rho_{avAA} - \rho_{buffer}} \qquad (3)$$

Here $M_{W\,temp}$(~632 g mol^{-1}) and $M_{W\,avAA}$ (135 g mol^{-1}) are the molecular weights of template and the peptide part of the protein, respectively. ν_{temp}(0.62 cm^3 g^{-1}) and ν_{avAA} (0.73 cm^3 g^{-1}) are, respectively, the partial specific volume of the template and the peptide component. ρ_{buffer}(335 e nm^{-3}) is the mean electron density of the buffer and ρ_{temp} (510 e nm^{-3}) and $\rho avAA$(447 e nm^{-3}) are, respectively, the mean electron density of the template and the peptides. Using these values, the scattering contributions of the template corresponds approximately to four dummy residues.

The ab initio models were obtained by using 100 annealing steps with 40 000 iterations at each step. However, a maximum of 4000 successes (configurations improving the fit) at each step were allowed. Besides the above-mentioned adjustments, only default values were used. Fifteen independent ab initio models were generated for each of the three carboproteins and an average model was found by using the programs DAMAVER and SUPCOMB[36] in batch mode; for more detailed information see refs. [59] and [60].

Rigid body modeling: The tertiary structure carboproteins **1**–**3** as well as peptide 4 was elucidated using the rigid body modeling program SASREF.[37] This program employs simulated annealing to optimize the spatial arrangement of an ensemble of subunits in a complex against experimental SAXS data. This is done by randomly selecting one subunit and moving it randomly in an arbitrarily direction followed by a random rotation. After each step the scattering amplitudes of the shifted subunit are recalculated and the scattering intensity of the entire complex is computed via the Debye equation and compared to the experimental data. Besides optimizing the scattering intensity of the model complex to that of the experimental data, SASREF is constrained to avoid steric clashes and to be interconnected.

Synchrotron radiation circular dichroism (SRCD): Far-UV CD spectra were obtained using the CD instrument at the UV1 beamline at the ASTRID synchrotron at the Institute for Storage Ring Facilities (ISA), University of Aarhus, Denmark. This UV absorption and CD beamline has been described in detail previously.[61] It has been shown that at UV1, protein samples do not deteriorate upon sequential scanning, despite the high UV photon flux of the beamline.[62] CD spectra of carboproteins **1**, **2**, and **3** were measured in an aqueous solution of sodium acetate buffer (50 mM, pH 5.5). All protein concentrations were 5.0 mg mL^{-1}. For all samples a 0.1 mm cylindrical SUPRASIL quartz cell (Hellma GmbH, Germany) was used for both sample and reference spectra. Temperature was controlled with a sample holder attached to a Peltier element, and spectra were recorded sequentially at the following temperatures: 26, 34, 42, 50, 58, 66, 74, 82 and 90 °C. Reference spectra were only recorded at 26 °C, and a reference scan preceded and followed every temperature series. A standard 10 mg mL^{-1} camphor sulfonic acid solution was used for obtaining a calibration factor for the signal

intensities. The CD spectra were obtained by averaging two scans with one point per nm and 45 s. collection time per point in the 260–180 nm interval. The degree of α-helicity was assessed based solely on the mean residue ellipticity at 220 nm ($[\theta]_{220} = -40\,000/(1-k)$, with $k = 2.6$).

Acknowledgements

We gratefully acknowledge the SAXS beam time provided by European Molecular Biology Laboratory at the X33 beamline of the Deutsches Elektron Synchrotron, Hamburg, Germany and the DANSYNC organization for funding our experiments. The Faculty of Life Sciences is acknowledged for a Ph.D. stipend to R.H.N., and FNU for a grant to K.J.J.

Keywords: ab initio calculations · carboproteins · protein design · protein structures · small-angle X-ray scattering

[1] K. S. Akerfeldt, R. M. Kim, D. Camac, J. T. Groves, J. D. Lear, W. F. DeGrado, *J. Am. Chem. Soc.* **1992**, *114*, 9656.
[2] M. R. Ghadiri, C. Soares, C. Choi, *J. Am. Chem. Soc.* **1992**, *114*, 4000.
[3] M. Goodman, Y. B. Feng, G. Melacini, J. P. Taulane, *J. Am. Chem. Soc.* **1996**, *118*, 5156.
[4] M. Mutter, K. H. Altmann, G. Tuchscherer, S. Vuilleumier, *Tetrahedron* **1988**, *44*, 771.
[5] M. Mutter, R. Hersperger, K. Gubernator, K. Muller, *Protein-Struct. Funct. Gen.* **1989**, *5*, 13.
[6] P. E. Dawson, S. B. H. Kent, *J. Am. Chem. Soc.* **1993**, *115*, 7263.
[7] I. Ernest, S. Vuilleumier, H. Fritz, M. Mutter, *Tetrahedron Lett.* **1990**, *31*, 4015.
[8] M. Mutter, G. G. Tuchscherer, C. Miller, K. H. Altmann, R. I. Carey, D. F. Wyss, A. M. Labhardt, J. E. Rivier, *J. Am. Chem. Soc.* **1992**, *114*, 1463.
[9] A. Grove, M. Mutter, J. E. Rivier, M. Montal, *J. Am. Chem. Soc.* **1993**, *115*, 5919.
[10] T. Arai, Y. Ide, Y. Tanaka, T. Fujimoto, N. Nishino, *Chem. Lett.* **1995**, 381.
[11] T. Sasaki, E. T. Kaiser, *Biopolymers* **1990**, *29*, 79.
[12] T. Sasaki, E. T. Kaiser, *J. Am. Chem. Soc.* **1989**, *111*, 380.
[13] T. Arai, K. Kobata, H. Mihara, T. Fujimoto, N. Nishino, *Bull. Chem. Soc. Jap.* **1995**, *68*, 1989.
[14] H. Mihara, N. Nishino, R. Hasegawa, T. Fujimoto, *Chem. Lett.* **1992**, 1805.
[15] H. Mihara, N. Nishino, R. Hasegawa, T. Fujimoto, S. Usui, H. Ishida, K. Ohkubo, *Chem. Lett.* **1992**, 1813.
[16] B. C. Gibb, A. R. Mezo, A. S. Causton, J. R. Fraser, F. C. S. Tsai, J. C. Sherman, *Tetrahedron* **1995**, *51*, 8719.
[17] A. R. Mezo, J. C. Sherman, *J. Am. Chem. Soc.* **1999**, *121*, 8983.
[18] A. K. Wong, M. P. Jacobsen, D. J. Winzor, D. P. Fairlie, *J. Am. Chem. Soc.* **1998**, *120*, 3836.
[19] S. F. Betz, D. P. Raleigh, W. F. DeGrado, *Curr. Opin. Struct. Biol.* **1993**, *3*, 601.
[20] P. Linder, T. Zemb in *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, Elsevier, Amsterdam, **2008**.
[21] M. V. Petoukhov, D. I. Svergun, *Curr. Opin. Struct. Biol.* **2007**, *17*, 562.
[22] K. J. Jensen, J. Brask, *Cell. Mol. Life Sci.* **2002**, *59*, 859.
[23] J. Brask, H. Wackerbarth, K. J. Jensen, J. D. Zhang, J. U. Nielsen, J. E. T. Andersen, J. Ulstrup, *Bioelectrochemistry* **2002**, *56*, 27.
[24] H. Wackerbarth, A. P. Tofteng, K. J. Jensen, I. Chorkendorff, J. Ulstrup, *Langmuir* **2006**, *22*, 6661.
[25] J. Brask, K. J. Jensen, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 697.
[26] J. Brask, J. M. Dideriksen, J. Nielsen, K. J. Jensen, *Org. Biomol. Chem.* **2003**, *1*, 2247.
[27] A. P. Tofteng, T. H. Hansen, J. Brask, J. Nielsen, P. W. Thulstrup, K. J. Jensen, *Org. Biomol. Chem.* **2007**, *5*, 2225.
[28] S. Hansen, *J. Appl. Crystallogr.* **2000**, *33*, 1415.
[29] D. I. Svergun, *J. Appl. Crystallogr.* **1992**, *25*, 495.
[30] L. Shi, M. Kataoka, A. L. Fink, *Biochemistry* **1996**, *35*, 3297.
[31] C. Birck, M. Malfois, D. Svergun, J. P. Samama, *J. Mol. Biol.* **2002**, *321*, 447.
[32] D. I. Svergun, M. V. Petoukhov, M. H. J. Koch, *Biophys. J.* **2001**, *80*, 2946.
[33] C. E. Shannon, *Bell Syst. Tech. J.* **1948**, *27*, 379.
[34] P. B. Moore, *J. Appl. Crystallogr.* **1980**, *13*, 168.
[35] D. Taupin, V. Luzzati, *J. Appl. Crystallogr.* **1982**, *15*, 289.
[36] V. V. Volkov, D. I. Svergun, *J. Appl. Crystallogr.* **2003**, *36*, 860.
[37] M. Nöllmann, W. M. Stark, O. Byron, *J. Appl. Crystallogr.* **2005**, *38*, 874.
[38] Y. H. Chen, J. T. Yang, K. H. Chau, *Biochemistry* **1974**, *13*, 3350.
[39] B. Lovejoy, S. Choe, D. Cascio, D. K. Mcrorie, W. F. DeGrado, D. Eisenberg, *Science* **1993**, *259*, 1288.
[40] J. A. Boice, G. R. Dieckmann, W. F. DeGrado, R. Fairman, *Biochemistry* **1996**, *35*, 14480.
[41] C. Garcia-Echeverria, *J. Am. Chem. Soc.* **1994**, *116*, 6031.
[42] A. S. Causton, J. C. Sherman, *J. Pept. Sci.* **2002**, *8*, 275.
[43] H. Diaz, J. R. Espina, J. W. Kelly, *J. Am. Chem. Soc.* **1992**, *114*, 8316.
[44] M. Feigel, *Liebigs Ann. Chem.* **1989**, 459.
[45] D. S. Kemp, *Trends Biotechnol.* **1990**, *8*, 249.
[46] J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, K. W. Plaxco, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12491.
[47] F. W. Kotch, R. T. Raines, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3028.
[48] M. Munson, R. O'Brien, J. M. Sturtevant, L. Regan, *Protein Sci.* **1994**, *3*, 2015.
[49] X. Jiang, E. J. Bishop, R. S. Farid, *J. Am. Chem. Soc.* **1997**, *119*, 838.
[50] R. Lutgring, J. Chmielewski, *J. Am. Chem. Soc.* **1994**, *116*, 6451.
[51] S. Y. M. Lau, A. K. Taneja, R. S. Hodges, *J. Biol. Chem.* **1984**, *259*, 13253.
[52] A. S. Causton, J. C. Sherman, *Bioorg. Med. Chem.* **1999**, *7*, 23.
[53] O. D. Monera, F. D. Sonnichsen, L. Hicks, C. M. Kay, R. S. Hodges, *Protein Eng.* **1996**, *9*, 353.
[54] K. A. Dill, H. S. Chan, *Biochemistry* **1990**, *29*, 7133.
[55] B. W. Matthews, *Annu. Review Biochem.* **1993**, *62*, 139.
[56] R. B. Hill, D. P. Raleigh, A. Lombardi, N. F. Degrado, *Acc. Chem. Res.* **2000**, *33*, 745.
[57] J. Brask, K. J. Jensen, *J. Pept. Sci.* **2000**, *6*, 290.
[58] O. Glatter, *J. Appl. Crystallogr.* **1977**, *10*, 415.
[59] M. H. J. Koch, P. Vachette, D. I. Svergun, *Quart. Rev. Biophys.* **2003**, *36*, 147.
[60] D. I. Svergun, M. H. J. Koch, *Rep. Prog. Phys.* **2003**, *66*, 1735.
[61] S. Eden, P. Limao-Vieira, S. V. Hoffmann, N. J. Mason, *Chem. Phys.* **2006**, *323*, 313.
[62] F. Wien, A. J. Miles, J. G. Lees, S. V. Hoffmann, B. A. Wallace, *J. Synchrotron Radiat.* **2005**, *12*, 517.

Received: April 17, 2008
Published online on October 10, 2008

