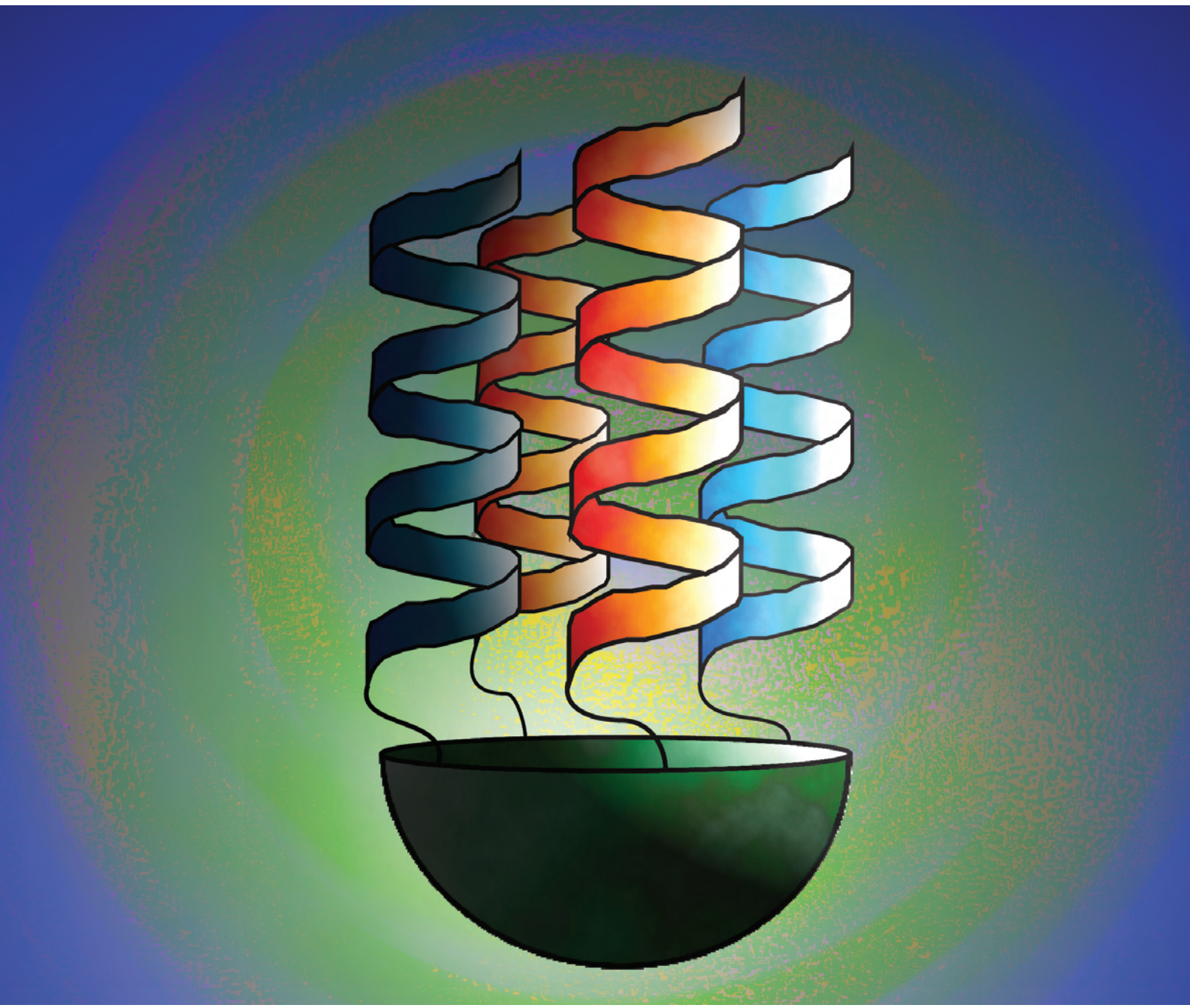
# Organic & Biomolecular Chemistry

www.rsc.org/obc

Volume 5 | Number 22 | 21 November 2007 | Pages 3557–3720



ISSN 1477-0520



RSC Publishing

# The design, synthesis, and characterization of the first cavitand-based *de novo* hetero-template-assembled synthetic proteins (Hetero-TASPs)†

**Heidi E. K. Huttunen-Hennelly‡ and John C. Sherman***

***Received 2nd August 2007, Accepted 19th September 2007***
***First published as an Advance Article on the web 12th October 2007***
**DOI: 10.1039/b711869d**

The design, synthesis, and characterization of novel cavitand-based hetero-TASPs, TASPs having different peptide sequences within one bundle, are described. Three families of hetero-TASPs were designed: the LG3/LG2 family (different linker lengths), LG3/AG3 family (altering helix hydrophobicity), and the LG3/LG2C family (anti-parallel caviteins). These first generation hetero-TASPs were found to be α-helical, stable towards guanidine hydrochloride, and monomeric in solution. The LG3/LG2 caviteins exhibited primarily native-like properties. The remaining hetero-TASP families were found to exhibit less dispersion and broader signals in the amide regions of their $^1$H NMR spectra than their respective reference caviteins. The success in the design of the LG3/LG2 hetero-TASPs suggests that subsequent hetero-TASPs may have potential to manifest superior native-like structure compared with homo-TASPs, and refinement of the linker and peptide sequences may accomplish this goal.

## Introduction

It was Anfinsen's early studies on ribonuclease that concluded that a protein's tertiary structure is encoded in its linear sequence of amino acids.[1] From this research transpired one of the greatest unanswered questions of the life sciences to which much effort has been expended over the past 25 years. How do linear polypeptides that carry all the necessary information for folding take on native three-dimensional shapes? The details of this folding process remain unclear, and its study is referred to as the "protein-folding problem". A long-term goal in this area of research has been to deduce the parameters controlling the relationship between sequence, tertiary structure, and function. Solving the protein-folding problem will have practical consequences in biology, drug development, and even medicine.

One approach used to illustrate the interactions that govern the relationship between protein sequence and structure, is through the design and characterization of *de novo* proteins.[2] *De novo* proteins are small proteins designed from first principles, retaining the crucial interactions of natural proteins, but otherwise lacking their complexity. Using *de novo* design, a peptide sequence that is predicted to adopt a particular fold can be synthesized, and the tertiary structure of the resulting protein can be readily characterized. *De novo* design affords the potential to use an incremental approach and to make individual amino acid modifications in the design to evaluate contributions of specific residues to the overall three-dimensional structure of a protein.[3] Furthermore, this discipline focuses on not only obtaining models for understanding the challenges of the folding process but to create novel molecules with native-like structures and useful functional activities for medical or industrial applications.[4]

The *de novo* concept can be expanded to include the use of templates in the synthesis of template-assembled synthetic proteins (TASPs).[5] Some examples of the molecular scaffolds used to date in the synthesis of TASPs include peptides,[5] carbohydrates,[6] porphyrins,[7] steroids,[8] calixarenes,[9] transition metals,[10] and cavitands.[11] In the method of template assembly, potential units of secondary structure are attached to a rigid scaffold in order to pre-organize the peptides and minimize the unfavorable loss of entropy upon folding. Another advantage of attaching peptides to a template is that it eliminates the need for designing loop regions between the peptide strands.

*De novo* design and template assembly have provided a means to synthesize diverse protein structures. Due to its relative simplicity, one of the most common topologies encountered amongst the examples of *de novo* proteins is the four-helix bundle.[12] An immense challenge in the design of these *de novo* four-helix bundle proteins is the creation of native-like tertiary structures, exhibiting well-packed hydrophobic cores. Examples of *de novo* proteins with well-defined tertiary structures are limited.[13] Our group has reported the synthesis of *de novo* four-helix bundle proteins using a cavitand as template, and termed the resulting proteins "*caviteins*" (*cavi*tand + pro*tein*) (Fig. 1).[14] Some of these caviteins have manifested considerable native-like properties.[14a,14d]

The TASPs synthesized in our lab to date have resulted from the simultaneous ligation of peptides strands to appropriate moieties on a cavitand (and a cyclotribenzylene, CTB) template, and have thus been limited to having only one type of peptide sequence attached within one bundle (*e.g.* three or more identical helices).[15] Having the ability to synthesize hetero-TASPs, TASPs having different peptide sequences within one bundle, would enable the creation of novel *de novo* caviteins, including an anti-parallel four-helix bundle cavitein. This article focuses on the design, synthesis, and characterization of a first generation of hetero-TASPs.

*Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, V6T 1Z1, Canada. E-mail: sherman@chem.ubc.ca; Fax: +1 (604) 822 2847*

† Electronic supplementary information (ESI) available: Additional figures. See DOI: 10.1039/b711869d

‡ Present address: Department of Chemistry, Thompson Rivers University, 900 McGill Road, PO Box 3010, Kamloops, BC, V2C 5N3, Canada. E-mail: hhuttunen@tru.ca; Fax: +1 (250) 828 5450

**Table 1** Complete sequences from N- to C-termini using one-letter-abbreviated amino acids including modified termini for peptides **1**, **2**, **3**, and **4**

| Peptide number | Peptide name | Peptide sequence |
|---|---|---|
| **1** | lg3 | $ClCH_2CO$-NH-[GGG-EELLKKLEELLKKG]-CO-$NH_2$ |
| **2** | lg2 | $ClCH_2CO$-NH-[GG-EELLKKLEELLKKG]-CO-$NH_2$ |
| **3** | ag3 | $ClCH_2CO$-NH-[GGG-EEAAKKAEEAAKKG]-CO-$NH_2$ |
| **4**[a] | lg2c | $CH_3CO$-NH-[GEELLKKLEELLKKGGC]-Spy |

[a] lg2c will be attached to the cavitand *via* a C-terminal cysteine residue; Spy is the S-pyridyl group.

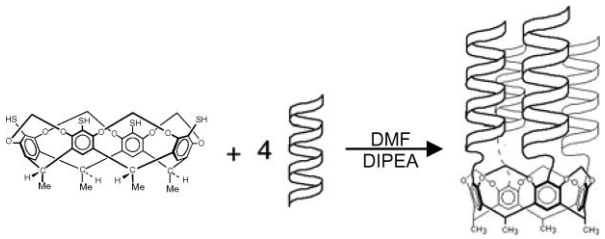



**Fig. 1** Cavitein synthesis.

Three hetero-TASP families were designed and synthesized using the lg3 peptide sequence (see Table 1) as the invariable sequence. The LG3 cavitein (bundle contained four lg3 peptide helices) had previously been studied and was found to exhibit significant native-like characteristics.[14a] In this way, LG3 served as a useful starting point for designing a hetero-TASP, and as a suitable reference cavitein. The LG3/LG2 hetero-TASPs were designed with peptides having different linker lengths; the LG3/AG3 family included two peptides with the same linker lengths but differing hydrophobic residues in the peptide sequences; and the LG3/LG2C family included anti-parallel four-helix bundle caviteins.

It is generally accepted that hydrophobic interactions are the major forces involved in initializing protein folding and stabilizing the tertiary structures of proteins.[15] Natural globular proteins have well-defined structure where specific hydrophobic interactions of the amino acid side chains pack like "knobs-into-holes".[16] For *de novo* proteins, manifestation of such structure is thus referred to as having native-like character. Although our previous caviteins have possessed several native-like characteristics,[14a,14d] the extent of side-chain packing specificity is not clear.

The LG3/LG2 family was designed to model a potential "knobs-into-holes" packing within the hydrophobic core by using two peptides sequences with different linker lengths (see Fig. 2).

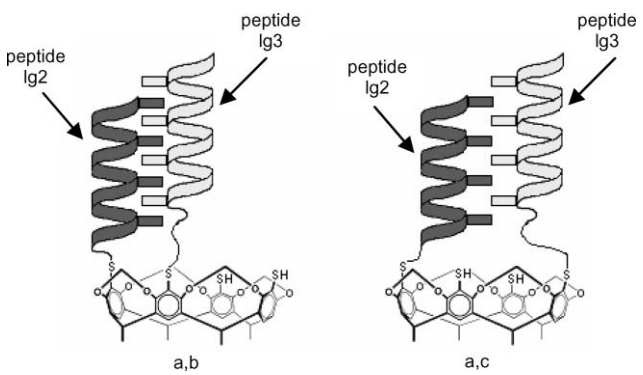



**Fig. 2** Schematic representation highlighting the a,b and a,c LG3/LG2 Hetero-TASP (note only two peptides are shown for clarity; see Table 1 for peptide labels).

Such a design would potentially enable the side chains of adjacent peptide helices to be "out of register", which might increase the efficacy of the hydrophobic packing within the four-helix bundle core.

The LG3/AG3 family was created to examine the effects of decreasing the hydrophobicity of subsequent peptide helices within a bundle. Previously, Causton synthesized an all-alanine peptide sequence linked *via* disulfide bonds to a benzylthiol cavitand template.[17] He found that his all alanine-based cavitein (four peptides had Ala in place of Leu) exhibited highly molten globule-like characteristics. If our previously designed leucine-based (lg3) caviteins were not entirely native-like due to an over-packed hydrophobic core, and the all-alanine-based (ag3) caviteins were highly under-packed, then designing a hetero-TASP composed of two ag3 peptides and two lg3 peptides could result in a more native-like cavitein.

The LG3/LG2C hetero-TASPs represent the first anti-parallel caviteins to be synthesized, which was one main goal in itself. Furthermore, perhaps the LG3/LG2C caviteins could accommodate a stable "knobs-into-holes" packing arrangement, and result in highly native-like caviteins. An anti-parallel design further permitted the evaluation of whether an increase in structural stability could be diagnosed when the peptide helices are oriented with opposing helix macrodipoles with respect to one another. We report the design and synthesis of the first anti-parallel cavitein, although other examples of anti-parallel *de novo* proteins and a few anti-parallel hetero-TASPs exist in the literature,[18] and it also is a frequent topology found in natural proteins.[19]

## Results and discussion

### Design

The peptide sequences were designed by following a minimalist approach.[12c] This included designing peptides with either four or five different amino acid residues. The peptides were also designed to be amphiphilic (*i.e.* the helix would contain a hydrophobic core and hydrophilic surfaces when folded) (see Fig. 3). Apart from Gly and Cys, the design included only amino acids with high helical propensities to promote a stable α-helix when folded. Two or three glycine residues were added to the N- or C-termini to serve as a linker between the template and the α-helical segment of the peptide. The effects of the glycine linkers on the cavitein structure and stability have been evaluated previously.[14a,14d] A Gly was incorporated at the N- or C-termini to serve as a helix-capping residue.[20] For peptides lg2, lg3, and ag3 the C-terminus was amidated to help reduce the effect of the helix macrodipole.[21] Furthermore, oppositely charged residues Lys and Glu were placed three and four amino acids apart to encourage intrahelical salt-bridges.

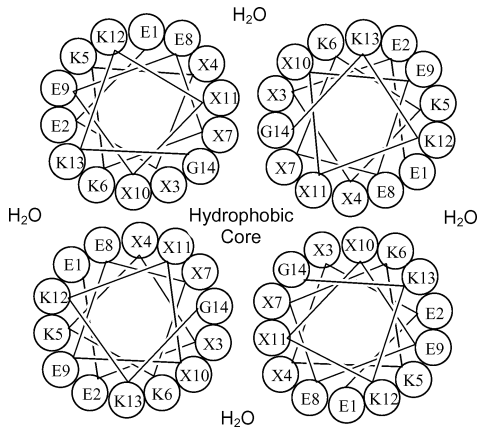



**Fig. 3** Helical wheel diagram of four strands of peptide **1** with the helices oriented in parallel (reader is looking down the helical axes from C- to N-termini).

An efficient template for TASP synthesis should pre-organize the peptides into a desired folding pattern and be easily integrated into the design. The template used for the formation of the caviteins was a cavitand, a rigid organic macrocycle with an enforced cavity.[22] The rigidity of the cavitand macrocycle limits the degrees of freedom of the constituent peptides, and thus, stabilizes the overall tertiary structure. Some of the problems associated with other templates in the synthesis of native-like four-helix bundles has been their inherent flexibility.[23] The thiol functional groups located on the upper rim of the cavitand are approximately 7 Å apart,[24] which are fitting for four-helix bundle proteins where the interhelical distances extend between 7–14 Å.[25] The thiol functionalities on the cavitand are nucleophilic, and thus serve as viable moieties for peptide attachment to efficiently afford a template-assembled synthetic protein (TASP). Furthermore, the cavitand template promotes the peptides strands into a desired orientation (*e.g.* all parallel), and the size of the cavitand (*i.e.* 4-, 5-, or 6-cavitand) can be modified to accommodate different sized bundle structures.

## Synthesis

Peptides **1**, **2**, and **3** were synthesized using standard solid-phase techniques.[21] Manual chloroacetylation of the free N-termini of the peptides was achieved by treatment with chloroacetyl chloride, resulting in an activated form of the peptide. The individual peptides were then manually treated with 95% trifluoroacetic acid (TFA), which cleaved the peptide from the resin in addition to removing the side-chain protecting groups simultaneously.

The N-terminus of peptide **4** was acetylated since future attachment of the peptide to the cavitand template would take place from the C-terminal cysteine residue *via* a disulfide bond to cavitand template. The C-terminal cysteine was activated using 2,2′-dipyridyl disulfide (DPDS) to afford the activated form of the peptide ready for incorporation into the caviteins *via* its C-terminus.

The nomenclature denotes using small letters for the peptide sequences (Table 1) and the name is then capitalized to refer to the corresponding four-helix bundle caviteins (Table 2). The first letter in either the peptide or cavitein name refers to the hydrophobic residue in the helical sequence, following by the number of glycine linkers between the peptide helices and cavitand template. Consider LG3, the capitals signify that it is a cavitein, L = leucine (the hydrophobic residue in the peptide sequence), and G3 = three glycine linkers between the peptide helices and the cavitand. For peptide **4**, the attachment to the cavitand template is *via* the C-terminus, which is indicated by the letter "c" at the end of the peptide name. For example cavitein LG2C, L = leucine, G2 = two glycine linkers, and C = peptides linked to the cavitand *via* their C-termini.

Caviteins **5–8** were synthesized by following literature procedures.[11a,14a] Hetero-TASPs **9–16** were synthesized *via* two different methodologies, one using limited peptide equivalents and second employing protecting groups. The first approach is described here (see Fig. 4), and the second approach is explained in the ESI†.

The first step entailed reacting 2.5 equivalents of peptide **1** with the cavitand template in the presence of excess DIPEA base, and DMF solvent. The subsequent peptide mixture was purified using RP-HPLC. The isolated products included two different disubstituted intermediates where the two peptides attached were either at the a,b (not shown in Fig. 4) or a,c positions on the template, respectively, in addition to mono-, tri-, and tetrasubstituted products (not shown in Fig. 4). The identities of the a,b and a,c disubsituted caviteins were assigned based on the 2 : 1 (a,b : a,c) peak area ratio in the HPLC trace (see

**Table 2** Names and **5–16**

| | Cavitein number | Cavitein name | Cavitein sequence |
|---|---|---|---|
| | **5** | LG3 | Cavitand–([peptide lg3])$_4$ |
| | **6** | LG2 | Cavitand–([peptide lg2])$_4$ |
| | **7** | AG3 | Cavitand–([peptide ag3])$_4$ |
| | **8**[a] | LG2C | Cavitand–([peptide lg2c])$_4$ |
| | **9** | 2LG3·2LG2_ab | Cavitand–([peptide lg3]$_2$·[peptide lg2]$_2$) |
| | **10** | 2LG3·2LG2_ac | Cavitand–([peptide lg3]$_2$·[peptide lg2]$_2$) |
| | **11** | 3LG3·1AG3 | Cavitand–([peptide lg3]$_3$·[peptide ag3]$_1$) |
| | **12** | 2LG3·2AG3_ab | Cavitand–([peptide lg3]$_2$·[peptide ag3]$_2$) |
| | **13** | 2LG3·2AG3_ac | Cavitand–([peptide lg3]$_2$·[peptide ag3]$_2$) |
| | **14** | 1LG3·3AG3 | Cavitand–([peptide lg3]$_1$·[peptide ag3]$_3$) |
| | **15**[a] | 2LG3·2LG2C_ab | Cavitand–([peptide lg3]$_2$·[peptide lg2c]$_2$) |
| | **16**[a] | 2LG3·2LG2C_ac | Cavitand–([peptide lg3]$_2$·[peptide lg2c]$_2$) |

[a] lg2c peptides are attached *via* their C-termini.

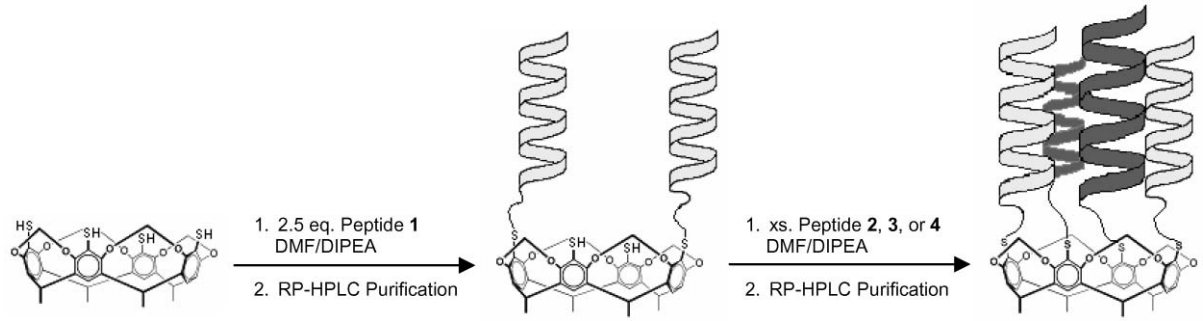



**Fig. 4** Schematic representation outlining the synthesis for an a,c Hetero-TASP.

ESI†), and by the splitting of the cavitand signal in the $^1$H NMR spectrum of 2LG3·2LG2_ab (see Fig. 17). The masses of the products were confirmed by MALDI-mass spectrometry. The product distribution contained mainly the a,b disubstituted product (57%) followed by the a,c disubstituted product (24%) and minimal amounts of the other three possible products. The purified and separated mono-, di-, trisubstituted products were then individually subjected to peptide **2**, **3**, or **4** in the presence of DIPEA base, and DMF solvent, to yield the corresponding four-helix hetero-TASPs.

## Characterization

The secondary and tertiary structures of the caviteins were analyzed using circular dichroism (CD) spectroscopy, guanidine hydrochloride denaturation, analytical ultracentrifugation (AUC), 1D $^1$H NMR spectroscopy, N-H/D exchange experiments, and 1-anilinonaphthalene-8-sulfonate (ANS) binding studies.

**Far-UV/CD Spectroscopy.** Far-UV (190–250 nm) CD spectroscopy is regularly used to quantify the extent of secondary structure present in peptides and proteins.[26] All the caviteins here are α-helical in structure, with two characteristic negative bands near 222 and 208 nm, and one positive band near 195 nm (see Fig. 5–7). An exact helical content for the caviteins was not determined since aromatic contributions by the cavitand template near 220–260 nm are known to influence the absorbance at 222 nm.[14a] For all of the caviteins, the CD spectra at concentrations of 4 and 40 μM agreed within experimental error, respectively, and therefore the 4 μM curves for each of the caviteins are not shown. This concentration-independence is consistent with monomeric caviteins.

A qualitative analysis of the CD spectra suggests that the hetero-TASPs signals approximate an average of the signals observed for their parent derivatives. Furthermore, it is apparent in Fig. 4 that the α-helicities decrease and the CD curves shift to lower wavelengths as the lg3 peptide helices are replaced with ag3 peptide helices (notice the minimum at ~203 nm for AG3, compared to ~208 nm for LG3). This shift in the CD curve toward lower wavelengths is indicative of random coil contributions, and has been observed by Kwok and Hodges in studying the effects of changing the hydrophobicity in two-stranded coiled coils.[27] Lastly, the 2LG3·2AG3_ac hetero-TASP appears to be more helical than its a,b counterpart. This difference is not observed for the other two hetero-TASP families.

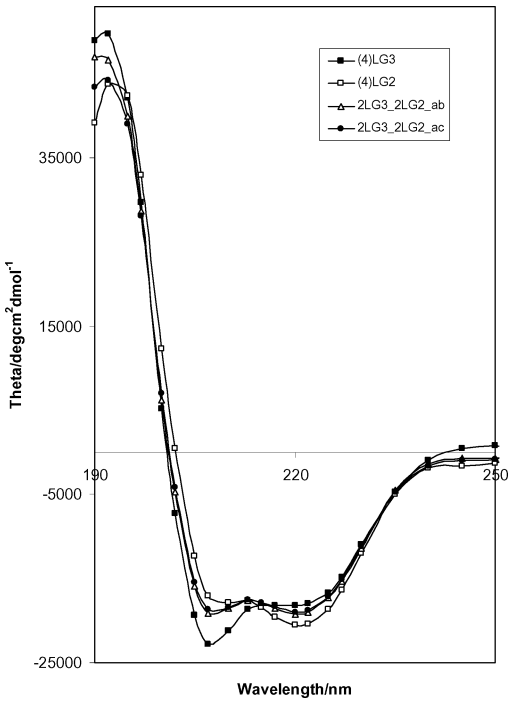



**Fig. 5** Far-UV CD spectra for the LG3/LG2 substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

**Near-UV/CD Spectroscopy.** In addition to determining the extent of α-helical secondary structure by interpreting the far-UV region of a CD spectrum, information about protein tertiary structure can be gathered by examining the near-UV (250–350 nm) spectral region.[26b] In this range, the aromatic amino acids and disulfide bonds are the typical chromophores, and the CD signals they produce are sensitive to the overall tertiary structure of the protein.[28] The cavitand template is the chromophore responsible for the absorptions in the near-UV region for caviteins. Proteins lacking well-defined three-dimensional structures (*e.g.* a molten globule or misfolded protein) produce little or no signal in the near-UV spectral region due to the time-averaged fluctuating structures.[29] On the other hand, enhanced near-UV signals are indicative of a well-defined protein structure due to the asymmetric environments of their aromatic chromophores.[30] Fig. 8–10 show the near-UV CD spectra for the caviteins.

The enhanced near-UV CD signals for the LG3/LG2 caviteins support native-like structures. The replacement of lg3 peptide helices with lg2 peptide helices had little effect on the near-UV CD

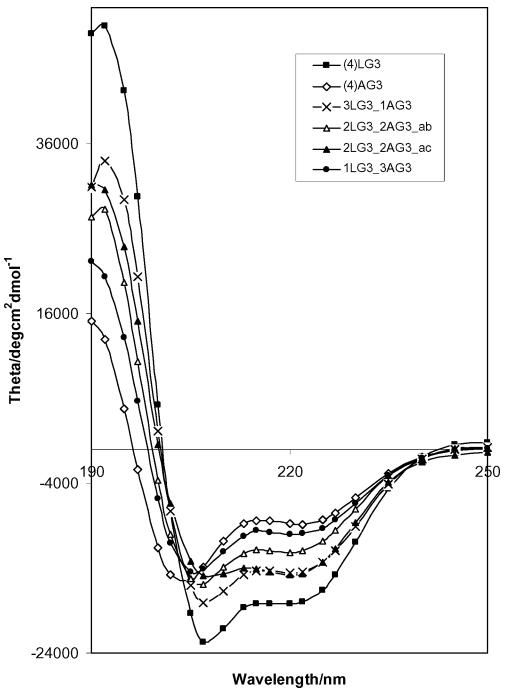



**Fig. 6** Far-UV CD spectra for the LG3/AG3 substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

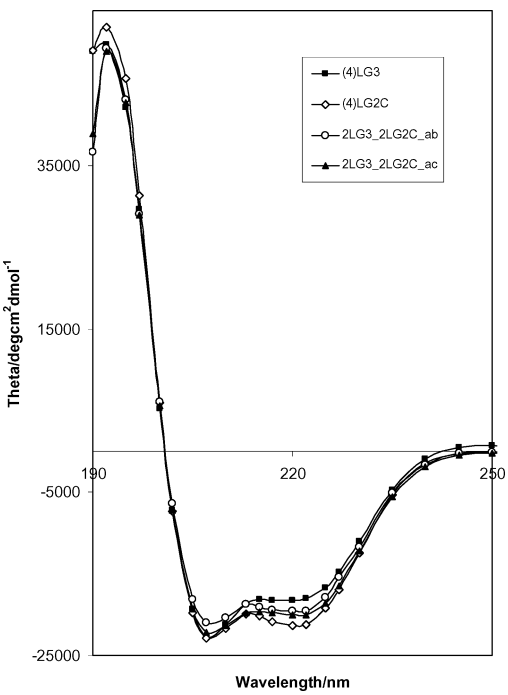



**Fig. 7** Far-UV CD spectra for the LG3/LG2C substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

signal when compared to the LG2 cavitein. The near-UV signals for 2LG3·2LG2_ab and 2LG3·2LG2_ac lie in between the curves observed for LG3 and LG2. However, the signs for the absorptions of 2LG3·2LG2_ab and 2LG3·2LG2_ac correspond to the signs of LG2, and are opposite to LG3. This suggests that the LG2 and 2LG3·2LG2_ab and 2LG3·2LG2_ac caviteins may be supercoiled

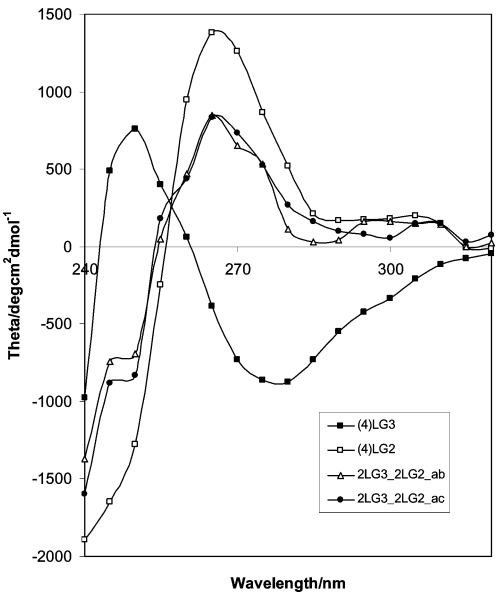



**Fig. 8** Near-UV CD spectra for the LG3/LG2 substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

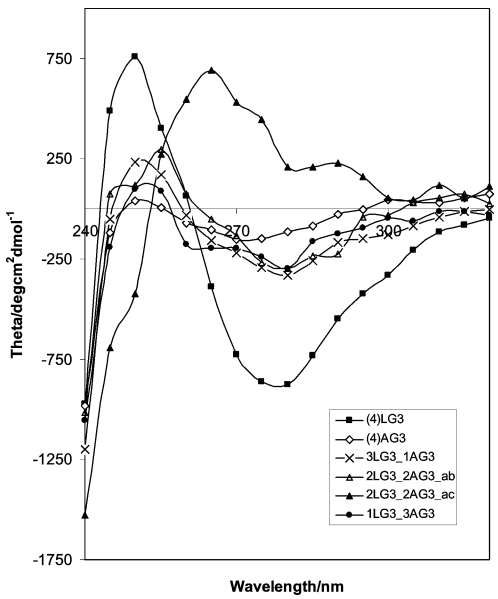



**Fig. 9** Near-UV CD spectra for the LG3/AG3 substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

in the same direction, while LG3 may be supercoiled in the reverse direction.

The near-UV spectral region displays relatively strong signals for both the LG3 and the 2LG3·2AG3_ac caviteins. The near-UV signal for 2LG3·2AG3_ac is more enhanced than the other variants of the LG3/AG3 family. On the other hand, 2LG3·2AG3_ab exhibits a reduced signal compared to 2LG3·2AG3_ac, which supports the notion that the tertiary structure of the a,c derivative is more efficiently packed than the a,b derivative. A clear distinction between the properties of an a,b and a,c derivative was not observed in the other hetero-TASPs.

The near-UV region for the anti-parallel caviteins and the LG2C reference cavitein have very little signal in this spectral region.

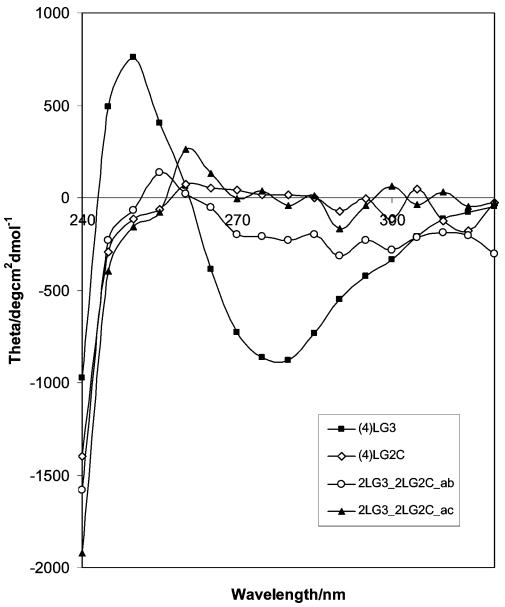



**Fig. 10** Near-UV CD spectra for the LG3/LG2C substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

This suggests little tertiary structural specificity, and could be attributable to the disulfide linkages between the template and peptide helices. Similar disulfide-linked caviteins have previously shown molten globule-like characteristics.[14c]

**Guanidine hydrochloride (GuHCl) denaturation.** GuHCl is thought to preferentially bind to protein surfaces and disrupt hydrogen bonding, hydrophobic packing, and electrostatic interactions, although the exact mechanism of denaturation is still not known.[31] The stabilities of the caviteins in the presence of GuHCl were determined at different concentrations. Fig. 11–13 display the unfolding curves of caviteins monitored at 222 nm in the presence of 0–8.0 M GuHCl at concentrations of 40 μM. An additional denaturation experiment was also carried out for each

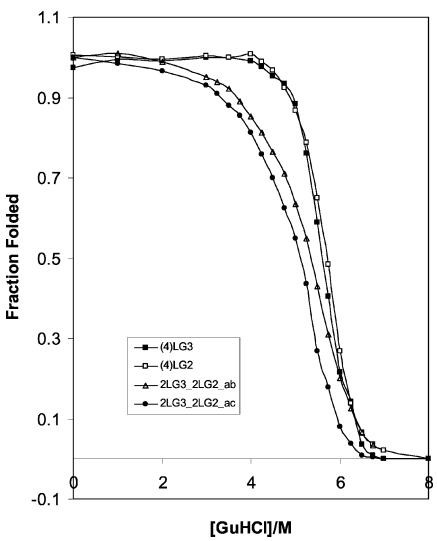



**Fig. 11** Effect of GuHCl on the fraction folded, as determined by the helicity ($[\theta]_{222}$), of the LG3/LG2 substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

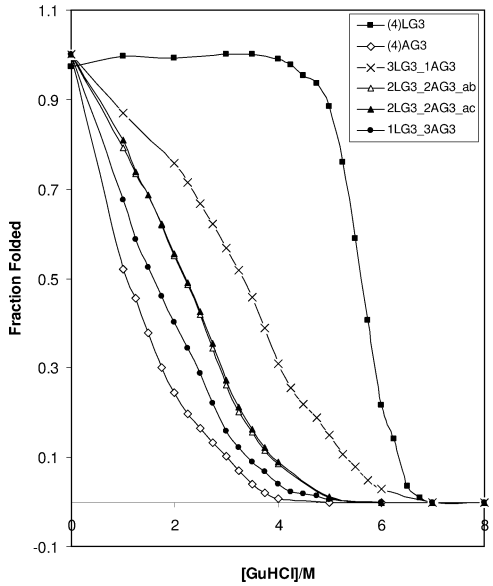



**Fig. 12** Effect of GuHCl on the fraction folded, as determined by the helicity ($[\theta]_{222}$), of the LG3/AG3 substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

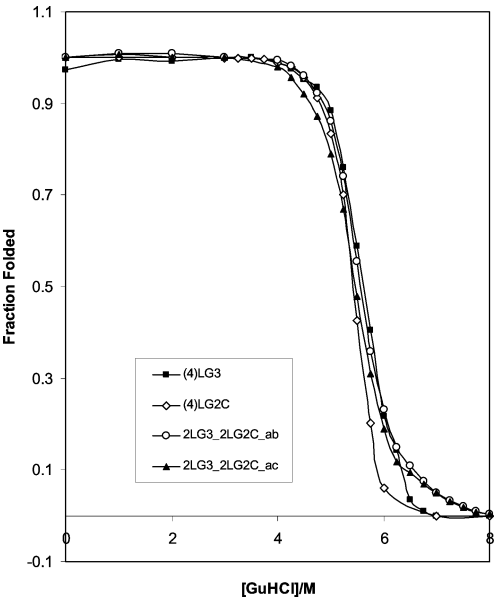



**Fig. 13** Effect of GuHCl on the fraction folded, as determined by the helicity ($[\theta]_{222}$), of the LG3/LG2C substituted caviteins at 40 μM in 50 mM phosphate buffer, pH 7.0 at 20 °C.

of the caviteins at a concentration of 4 μM (data not shown). The caviteins at the two concentrations were indistinguishable, which suggests the presence of monomers.

The cooperativity of the unfolding process can be measured qualitatively by the width and shape of the unfolding transition. A highly cooperative unfolding curve indicates that the protein existed originally as a compact, well-folded structure. A very gradual, non-cooperative unfolding transition indicates that the protein existed initially as a partially folded (*e.g.* molten globule) protein. All of the LG3/LG2 substituted caviteins exhibited cooperative two-state unfolding transitions. However, it is clear that LG3 and LG2 exhibit a more cooperative unfolding transition

than do the corresponding hetero-TASPs, which have noticeably less steep unfolding curves.

When comparing the stabilities of our caviteins towards GuHCl one can take the concentration of GuHCl required to unfold the cavitein halfway ($[GuHCl]_{1/2}$) as a rough estimate of their stabilities. A more accurate method for determining the stabilities of proteins is *via* a "linear extrapolation method", in which the free energy of unfolding is assumed to vary linearly with GuHCl concentration.[32] To determine the free energy of unfolding of the caviteins, we used the expanded method described by Santoro and Bolen which assumes the unfolding transition to be a reversible, two-state process.[33] This method uses a non-linear least squares fitting of data to derive at a $\Delta G^{0}_{H_2O}$ value. The extrapolated values for $\Delta G^{0}_{H_2O}$, $m$, and the denaturation midpoint for each of the caviteins are listed in Tables 3–5.

The folded conformations of natural proteins are typically 5–10 kcal mol$^{-1}$ more stable than the unfolded state.[32b] By altering the four-helix bundles of LG3 or LG2 to their hetero-TASP counterparts, a loss in structural stability of 3–4 kcal mol$^{-1}$ is encountered, however, the stabilities of all the LG3/LG2 caviteins still fall in the range of natural proteins. Larger $m$ values correspond to a more cooperative unfolding transition, and can therefore be used as an estimate of native-like structure.[34] The $m$ values for the four caviteins were comparable to the $m$ values found in native proteins.[35]

The unfolding transitions for the LG3/AG3 substituted caviteins were non-cooperative except for the reference cavitein LG3 (Fig. 12). It is clear from the denaturation curves that the replacement of a single lg3 helix with an ag3 helix results in a large loss in structural stability. It appears that the side chains of the alanine residues of a single helix within the four-helix bundle (3LG3·1AG3) are too small and result in an under-packed hydrophobic core.

The data in Table 5 suggests that the anti-parallel caviteins (LG3·LG2C_ab and LG3·LG2C_ac) were less stable by 2–3 kcal mol$^{-1}$ toward the chemical denaturant as compared to their reference caviteins (LG3 and LG2C). Although all of the denaturation curves look similar, the regions near 8.0 M GuHCl are less steep for the hetero-TASPs, resulting in the lower calculated stabilities. The unfolding curve for the LG2C reference cavitein was slightly steeper than the other denaturation curves, and resulted in the greatest value for the $\Delta G^{0}_{H_2O}$ values of unfolding. All caviteins of this family, however, were still of similar stability as native proteins.[32b]

**Table 3** Guanidine hydrochloride-induced denaturation data calculated for the LG3/LG2 substituted caviteins

| Cavitein | $[GuHCl]_{1/2}$/M | $m$/kcal mol$^{-1}$ M$^{-1}$ | $\Delta G^{0}_{H_2O}$/kcal mol$^{-1}$ |
|---|---|---|---|
| LG3 | 5.6 ± 0.1 | −1.9 ± 0.1 | −10.8 ± 0.4 |
| LG2 | 5.7 ± 0.1 | −1.8 ± 0.1 | −10.4 ± 0.3 |
| 2LG3·2LG2_ab | 5.4 ± 0.1 | −1.2 ± 0.1 | −6.6 ± 0.3 |
| 2LG3·2LG2_ac | 5.2 ± 0.1 | −1.4 ± 0.1 | −7.4 ± 0.5 |

**Table 4** Guanidine hydrochloride-induced denaturation data calculated for the LG3/AG3 substituted caviteins

| Cavitein | $[GuHCl]_{1/2}$/M | $m$/kcal mol$^{-1}$ M$^{-1}$ | $\Delta G^{0}_{H_2O}$/kcal mol$^{-1}$ |
|---|---|---|---|
| LG3 | 5.6 ± 0.1 | −1.9 ± 0.1 | −10.8 ± 0.4 |
| AG3 | 1.1 ± 0.1 | — | — |
| 3LG3·1AG3 | 3.3 ± 0.1 | −0.8 ± 0.1 | −3.7 ± 0.4 |
| 2LG3·2AG3_ab | 2.1 ± 0.1 | — | — |
| 2LG3·2AG3_ac | 2.2 ± 0.1 | — | — |
| 1LG3·3AG3 | 1.6 ± 0.1 | — | — |

**Table 5** Guanidine hydrochloride-induced denaturation data calculated for the LG3/LG2C substituted caviteins

| Cavitein | $[GuHCl]_{1/2}$/M | $m$/kcal mol$^{-1}$ M$^{-1}$ | $\Delta G^{0}_{H_2O}$/kcal mol$^{-1}$ |
|---|---|---|---|
| LG3 | 5.6 ± 0.1 | −1.9 ± 0.1 | −10.8 ± 0.4 |
| LG2C | 5.4 ± 0.1 | −2.3 ± 0.1 | −11.8 ± 0.4 |
| 2LG3·2LG2C_ab | 5.5 ± 0.1 | −1.6 ± 0.1 | −8.8 ± 0.4 |
| 2LG3·2LG2C_ac | 5.4 ± 0.1 | −1.5 ± 0.1 | −8.4 ± 0.4 |

**Analytical ultracentrifugation (AUC).** Sedimentation equilibrium can provide thermodynamic information of molecules including their molar mass, association constants, and stoichiometries.[36] Such studies were performed on all of the caviteins to accurately determine their oligomeric states and molecular weights in solution. The caviteins were analyzed at three different concentrations and rotor speeds. The sedimentation equilibrium data were analyzed using NONLIN,[37] and the exponential plot of absorbance *versus* radius for LG3 is shown in Fig. 14. The

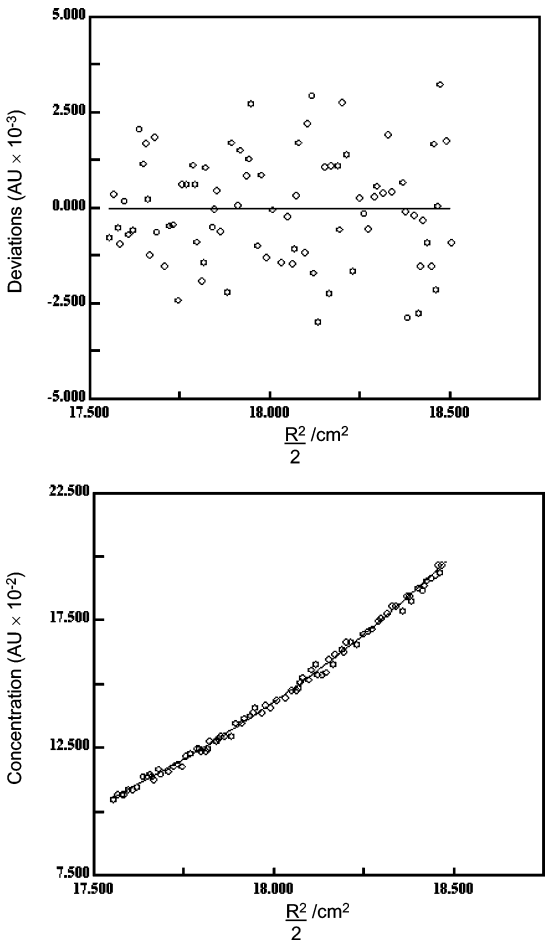



**Fig. 14** Sedimentation equilibrium concentration distributions of LG3 at a rotor speed of 27 000 rpm in 50 mM phosphate buffer, pH 7.0, 20 °C at 10 μM. The upper panel represents the residuals for the fit. In the lower panel the solid line represents a theoretical fit to a monomer equilibrium.

**Table 6** Molecular weights ($M_w$) determined by sedimentation equilibrium for the Hetero-TASPs at 20 °C in 50 mM pH 7.0 phosphate buffer at concentrations of 10, 50, and 80 μM with rotor speeds of 27 000, 35 000 and 40 000 rpm

| Cavitein | Experimentally determined $M_w$ (Da) | Calculated $M_w$ (Da) | Predominant species |
|---|---|---|---|
| LG2 | 8500 ± 600 | 8016 | Monomer |
| LG3 | 8000 ± 300 | 8240 | Monomer |
| AG3 | 7500 ± 200 | 7400 | Monomer |
| LG2C | 8700 ± 600 | 8424 | Monomer |
| 2LG3·2LG2_ab | 7800 ± 500 | 8128 | Monomer |
| 2LG3·2LG2_ac | 7900 ± 400 | 8128 | Monomer |
| 2LG3·2AG3_ab | 8000 ± 400 | 7820 | Monomer |
| 2LG3·2AG3_ac | 7900 ± 300 | 7820 | Monomer |
| 2LG3·2LG2C_ab | 8800 ± 500 | 8332 | Monomer |
| 2LG3·2LG2C_ac | 8700 ± 400 | 8332 | Monomer |

plots for the other caviteins were similar and are therefore not included. All of the caviteins were found to exist as monomers in solution and the experimental and calculated molecular weights corresponded (Table 6).

**$^1$H NMR spectroscopy.** One-dimensional $^1$H NMR spectroscopy is a simple diagnostic method used to differentiate between native-like and molten globule-like structures. Natural proteins exhibit considerable chemical shift dispersion, which is indicative of native-like structure,[38] whereas the looser packing arrangement of molten globule-like structures results in broader, less disperse signals.[39] The spectral region between 7 and 11 ppm is typically examined because distinct sharp and disperse signals are observed for native-like proteins corresponding to the slowly exchanging amide protons. A stacked plot of the amide regions of the $^1$H NMR spectra for hetero-TASP families are shown in Fig. 15–17, respectively.

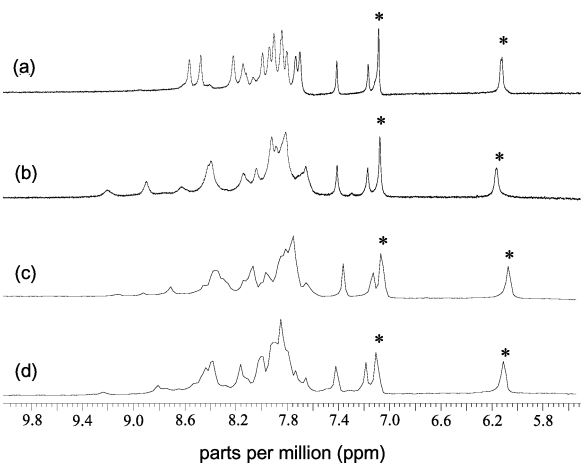



**Fig. 15** Expansions of the amide regions of 500 MHz $^1$H NMR spectra of the LG3/LG2 substituted caviteins at ~1.5 mM in 10% $D_2O$, 45 mM phosphate buffer, pH 7.0 at 20 °C. (a) LG3 (b) LG2 (c) 2LG3·2LG2_ab (d) 2LG3·2LG2_ac (* = cavitand signals).

The $^1$H NMR spectra of LG2 and LG3 each show ~13 distinguishable dispersed amide signals indicative of a well-defined amide backbone with a high content of tertiary structure. The presence of only 13 amide signals for the LG2 and LG3 caviteins suggests that many of the amino acid residues are in a degenerate environment and therefore indistinguishable from each other, likely due to the four-fold symmetry of the cavitein.

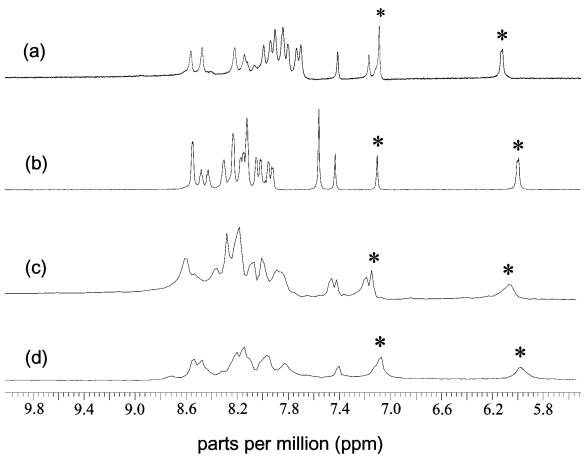



**Fig. 16** Expansions of the amide regions of 500 MHz $^1$H NMR spectra of the LG3/AG3 substituted caviteins at 1.5 mM in 10% $D_2O$, 45 mM phosphate buffer, pH 7.0 at 20 °C. (a) LG3 (b) AG3 (c) 2LG3·2AG3_ab (d) 2LG3·2AG3_ac (* = cavitand signals).

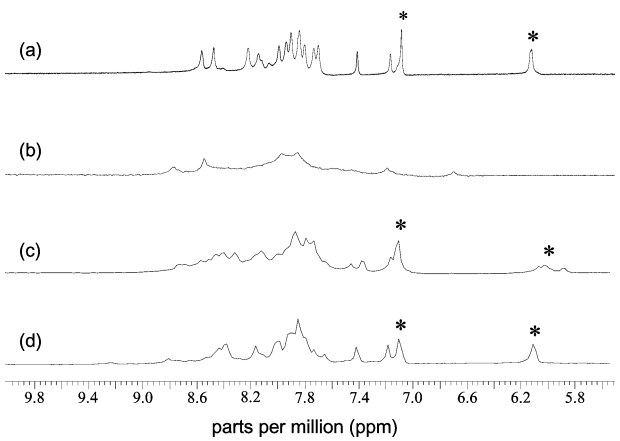



**Fig. 17** Expansions of the amide regions of 500 MHz $^1$H NMR spectra of the LG3/LG2C substituted caviteins at 1.5 mM in 10% $D_2O$, 45 mM phosphate buffer, pH 7.0 at 20 °C. (a) LG3 (b) LG2C (c) 2LG3·2LG2C_ab (d) 2LG3·2LG2C_ac (* = cavitand signals).

The amide regions of the LG3/LG2 caviteins exhibit sharp and disperse signals characteristic of native-like proteins. Additionally, the $H_{out}$ cavitand signal at ~6.1 ppm for 2LG3·2LG2_ab and 2LG3·2LG2_ac is comparable to the $H_{out}$ signal for the parent caviteins. The $H_{out}$ signal should split into a doublet since it is coupled to the $H_{in}$ proton on the cavitand template. The sharpness

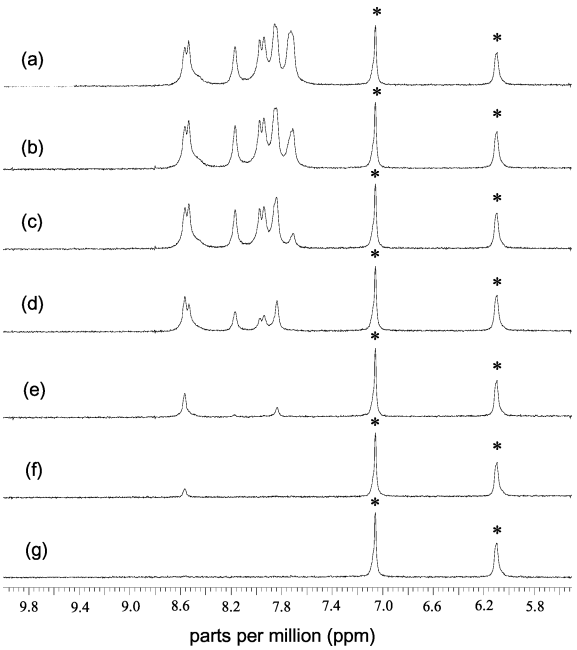



**Fig. 18** Stack plot of the 500 MHz $^1$H NMR spectra illustrating the time-dependent amide H/D exchange of LG3 at ~1.5 mM in 50 mM deuterated acetate buffer, pD 5.02 at 20 °C. (a) 6 min (b) 20 min (c) 1 h 5 min (d) 6 h 10 min (e) 23 h 19 min (f) 72 h 53 min (g) 9 d 1 h 21 min (* = cavitand signals).

of the $H_{out}$ signal for all of the caviteins of this family suggests that these caviteins are primarily populating a single conformation, supportive of native-like structure.

The amide region of LG3 and AG3 are considerably sharper than those of 2LG3·2AG3_ab or 2LG3·2AG3_ac. The amide region for AG3 however, is less dispersed than the amide region of LG3. The broad spectra for 2LG3·2AG3_ab and 2LG3·2AG3_ac suggest that these caviteins are less native-like than their parent reference caviteins, and exhibit some molten globule-like characteristics. The broad signals in the amide region of the $^1$H NMR spectra of 2LG3·2AG3_ab and 2LG3·2AG3_ac complement the non-cooperative unfolding curves observed in the GuHCl denaturation experiments, again supporting the tertiary structures of 2LG3·2AG3_ab and 2LG3·2AG3_ac as being molten globule-like. A similar argument could not be made for AG3, which has very sharp, although poorly dispersed, amide signals in its $^1$H NMR spectrum, characteristic of native-like structure, whereas a non-cooperative unfolding curve was observed in the GuHCl denaturation experiment.

The $^1$H NMR spectrum of LG3 is considerably sharper than LG2C or the anti-parallel caviteins. Unfortunately, the spectrum of the LG2C cavitein is of poor quality due to a lack of sufficient protein sample (synthesis hampered by an S-pyridyl impurity making purification difficult). Looking at the spectrum of 2LG3·2LG2C_ac, the amide signals are slightly sharper than the amide signals in the 2LG3·2LG2C_ab cavitein. It is difficult to make any strong conclusions about the tertiary structures of the anti-parallel caviteins, although they do exhibit less native-like character than LG3. One interesting observation in the $^1$H NMR spectrum for the 2LG3·2LG2C_ab (Fig. 17c) cavitein is the splitting of the cavitand signal ($H_{out}$) at ~6.1 ppm into what appears to be three signals in a ~1 : 2 : 1 ratio. This splitting of the $H_{out}$ cavitand signal supports the assignment of this cavitein as being a,b substituted.

**N-H/D exchange.** Hydrogen–deuterium exchange is a valuable diagnostic technique in the study of protein tertiary structures.[40] The rate of exchange of the amide protons in native-like proteins can be studied by NMR and has been found to be much slower when compared to such exchange in molten globule-like structures. The observed rate of exchange of an amide proton in a structured environment ($k_{obs}$) can be divided by a calculated intrinsic rate of exchange for an amide in an unstructured environment at a given temperature and pH ($k_{in}$) to give a number called a "protection factor".[40e] The larger the protection factor, the slower the rate of exchange of the amide proton with the solvent and hence the proton is said to be "protected" from exchange. Typical protection factors for native proteins are in the range of $10^5$–$10^8$, where as molten globule-like structures exhibit protection factors in the range of $10^1$–$10^3$.[40d,40e] A representative stack plot for the N–H/D exchange for LG3 is shown in Fig. 18. The first-order exchange rates and protections factors for the LG3/LG2 and LG3/LG2C hetero-TASP families are tabulated in Tables 7 and 8, respectively. No protection factors were calculated for the LG3/AG3 family, excluding LG3, as all of the amide protons exchanged with solvent before the first spectrum could be acquired.

LG3 and LG2 were found to have the largest protection factors determined from the N–H/D exchange experiments, suggesting that these amide protons were uniquely protected from the solvent. The amide protons in the remaining hetero-TASPs exchanged with the solvent within 1.5 hours.

**ANS binding.** 1-Anilinonaphthalene-8-sulfonate (ANS) is a hydrophobic, nonpolar dye, which binds preferentially to exposed hydrophobic surfaces of a protein, and can be effectively detected using fluorescence spectroscopy.[41] Molten globule-like structures typically expose more hydrophobic surfaces through their mobility, and therefore typically bind nonpolar molecules more strongly than native-like proteins.[42] Furthermore, ANS seldom binds to unfolded proteins. ANS binding was studied by fluorescence spectroscopy, and under standard experimental conditions[43] negligible binding was observed for any of the caviteins (the emission spectra for the caviteins can be found in the ESI†).

## Conclusions

*De novo* protein design entails the study of protein structure and the forces involved in protein folding and stability, as well as the design and synthesis of novel compounds, including the creation of functional proteins. Designing *de novo* proteins with concrete native-like structures remains as an important but largely unachieved goal.

The present work reports the development of methods for the efficient synthesis of cavitand-based hetero-TASPs, which will hopefully expedite the exploration of hetero-TASPs, as few examples have been reported.[4g,18d] Of the caviteins studied, the LG3/LG2 family resulted in tertiary structures with similar properties of other native-like *de novo* proteins.[18c,43b] 2LG3·2LG2_ab and 2LG3·2LG2_ac were the first cavitand-based hetero-TASPs to

**Table 7** Tabulated data from the mmide H/D exchange experiments of the LG3/LG2 substituted caviteins in a 50 mM acetate buffer, pD 5.02 at 20 °C

| Cavitein name | Amide proton chemical shift[a]/ppm | First-order rate constant/$h^{-1}$ | Half-life/h | Protection factor[b] |
|---|---|---|---|---|
| LG3 | 8.5 | $2.97 \times 10^{-2}$ | 23 | $(7.3 \pm 0.5) \times 10^3$ |
| LG2 | 8.5 | $3.37 \times 10^{-2}$ | 20 | $(6.3 \pm 0.5) \times 10^3$ |
| 2LG3·2LG2_ab | 8.5 | $4.62 \times 10^{-1}$ | 1.0 | $(3.1 \pm 0.3) \times 10^2$ |
| 2LG3·2LG2_ac | 8.5 | $6.93 \times 10^{-1}$ | 1.5 | $(4.7 \pm 0.3) \times 10^2$ |

[a] Only the data on the most protected proton is included. [b] These values are based on the half-life of an unprotected proton at 20 °C at pD 5.02 to be $3.18 \times 10^{-3}$ h.

**Table 8** Tabulated data from the amide H/D exchange experiments of the LG3/LG2C substituted caviteins in a 50 mM acetate buffer, pD 5.02 at 20 °C

| Cavitein name | Amide proton chemical shift[a]/ppm | First-order rate constant/$h^{-1}$ | Half-life/h | Protection factor[b] |
|---|---|---|---|---|
| LG3 | 8.5 | $2.97 \times 10^{-2}$ | 23 | $(7.3 \pm 0.5) \times 10^3$ |
| LG2C | 8.5 | 1.39 | 0.50 | $(1.6 \pm 0.2) \times 10^2$ |
| 2LG3·2LG2C_ab | 8.5 | 2.77 | 0.25 | $(7.9 \pm 0.2) \times 10^1$ |
| 2LG3·2LG2C_ac | 8.5 | 2.10 | 0.33 | $(1.0 \pm 0.3) \times 10^2$ |

[a] Only the data on the most protected proton is included. [b] These values are based on the half-life of an unprotected proton at 20 °C at pD 5.02 to be $3.18 \times 10^{-3}$ h.

show promising native-like properties, and therefore serve as useful precursors for future designs. The hetero-TASPs of the remaining two families displayed well-defined secondary structures and were highly stable towards chemical denaturation, although they did not exhibit such highly native-like structures (less well-resolved signals in the $^1$H NMR spectra) as those caviteins in the LG3/LG2 family.

These three series of hetero-TASP prototypes represent the establishment of an enhanced versatility that is now available for *de novo* design. An ideal linker between the peptide and cavitand has been shown to be key to the native-like properties of homo-caviteins.[14a,14d] The next step for hetero-TASPs is to refine the linkers and peptide sequences in order to develop native-like structure that surpasses that of homo-TASPs.

With a better understanding of the elements influencing the secondary and tertiary structures of our systems, it would be interesting to explore caviteins with specific applications. For example, one could model a specific native four-helix bundle, typically made up of non-identical helices, using the hetero-TASP approach. Other research groups have introduced functionality into their protein models by creating *de novo* protein ion-channels,[44] proteins binding cofactors,[45] peptide receptors,[46] and catalysts to name a few.[47] These examples of *de novo* proteins designed with specific functions, combined with our ability to now create hetero-TASPs, may lead us to new caviteins that manifest well-defined functions.

## Experimental

### Cavitand Synthesis

The synthesis of the arylthiol cavitand was synthesized following literature procedures.[11a]

### General

All reagents for the peptide and cavitein syntheses were reagent grade. The peptides and caviteins were purified by preparative reversed-phase HPLC using a Perkin–Elmer Biocompatible Pump 250 with a PE LC90 BIO spectrophotometric UV detector and a KIPP and ZONEN chart recorder. A Phenomenex Selectosil $C_{18}$ reversed-phase HPLC column (preparative: 250 mm × 10 mm, 10 μM particle size, 300 Å pore size) was used. The wavelength for the UV detection was set at 229 nm for recognition of the amide chromophore. The samples were filtered through a 0.45 μM Nylon™ syringe filter (Phenomenex) prior to injection and run at a flow rate of 10 mL $min^{-1}$ using helium-sparged filtered water (0.1% TFA)–HPLC-grade acetonitrile (0.05% TFA) gradient. The purity of the peptides and caviteins were analyzed by analytical reversed-phase HPLC and were filtered prior to injection onto a Varian 9010 pump with a Varian 9050 UV detector and a Varian 4290 integrator. A Phenomenex Selectosil $C_{18}$ reversed-phase HPLC column (analytical: 250 mm × 4.5 mm, 5 μM particle size, 100 Å pore size) was used. Analytical samples were run at a flow rate of 1 mL $min^{-1}$ using the same solvents as for the preparative purification. The purified samples were evaporated *in vacuo* and lyophilized. The mass spectra were run on a MALDI-MS Bruker Biflex IV in reflection mode using 50 μM cinnamic acid, in 1 : 1 $H_2O$–MeCN, as a matrix. Cavitein concentrations were determined using a Bradford assay[48] measured on a CARY UV-visible spectrophotometer. The pH values of the buffers were determined using a Fisher Scientific Accumet pH-meter 915 calibrated with two purchased buffer standards (pH = 4.0 and 10.0).

**Peptide synthesis.** The peptides were synthesized following to a large extent literature procedures, with peptides **1** and **2** having previously been synthesized.[14a] The synthesis of peptide **1** (lg3) is described below, and similar procedures were followed for the synthesis of peptides **2** (lg2), and **3** (ag3). The peptide synthesis involved using standard Fmoc techniques on an automated Applied Biosystems peptide synthesizer attached to an Apple IIsi MacIntosh computer. All Fmoc protected amino acids, solvents and coupling reagents were purchased from Advanced Chemtech (Louisville, KY, USA). The peptides were synthesized on a 0.25 mmol scale using the FastMoc™ protocols. Side-chain-protected amino acids were used for

**Table 9** % yields and MALDI-MS characterization of the "activated" peptides (note, calculated masses agreed within 1 Da).

| Peptide | % yield | Mass/Da |
|---|---|---|
| lg3 | 82 | 1916 |
| lg2 | 85 | 1860 |
| ag3 | 87 | 1706 |
| lg2c | 19 | 2037 |

chemoselective synthesis of the peptide, which was in turn bound by its C-terminus to a resin developed by Rink[49] to afford a C-terminal amide upon cleavage. A single amino acid coupling cycle included a: (i) 13 min Fmoc deprotection using piperidine, (ii) 6 min wash step using *N*-methylpyrrolidone (NMP), (iii) 30 min coupling step to 1.0 mmol of the next Fmoc amino acid using 2-(1*H*-benzotriazole)1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and 1-hydroxybenzotriazole (HOBt) as coupling reagent (note activation of the amino acid started with DIPEA), and lastly (iv) 6 min wash with NMP. NMP was the solvent throughout the synthesis with each cycle having an approximate time of 55 minutes.

Thereafter, chloroacetylation of the free N-terminus was achieved through manual treatment of the resin (600 mg peptide resin, ~300 mg peptide **1**, ~0.160 mmol) with chloroacetyl chloride (75 μL, 0.96 mmol, 6 equiv.) and DIPEA (165 μL, 0.95 mmol, 6 equiv.) in DMF for 1 hour at room temperature under nitrogen. The last step included cleavage of the peptides from the resin in addition to removing the side-chain protecting groups simultaneously using a 2 hour treatment with 95% TFA–$H_2O$. An ice bath was used for the first 10 minutes of the reaction. After completion, the resin was removed by suction filtration through a medium frit filter with a $CH_2Cl_2$ wash. The TFA–$CH_2Cl_2$ filtrate was evaporated to a few mL *in vacuo* and the crude peptide was precipitated using ice-cold diethyl ether. The peptide was recovered by suction filtration using a fine frit filter. The peptide was then dissolved in distilled water, filtered, and purified by reversed-phase HPLC. The peptide was lyophilized until a fluffy white solid peptide was obtained (105 mg, 23%). MALDI-MS and yields of the peptides are given in Table 9.

Slight modifications to the above procedure were used for the preparation peptide **4** (lg2c) and the details for this synthesis can be found in the supporting information†.

### Cavitein synthesis

Caviteins **5–8** were synthesized by following literature procedures.[11a,14a] The synthesis of cavitein **5** is described below, and similar procedures were followed for the synthesis of caviteins **6**, **7** and **8**. A solution of the arylthiol cavitand (1.1 mg, 1.4 μmol, 1 equiv.) and peptide **1** (21 mg, 11.3 μmol, 8 equiv.) were stirring in degassed DMF under $N_2$. DIPEA (2.5 μL, 15 μmol, 10 equiv.) was added in excess until the solution turned cloudy. The reaction was monitored (appearance of the cavitein peak) by analytical reversed-phase HPLC and was complete after 4 hours. The crude reaction mixture was evaporated *in vacuo*, dissolved in water, filtered, and purified by reversed-phase HPLC to yield cavitein **5** as a fluffy white solid (5.5 mg, 49%) after lyophilization. The additional unwanted tris-cavitein byproduct was separated and removed during purification.

The synthesis of hetero-TASP **9** entailed mixing a solution of the cavitand (5 mg, 6.9 μmol, 1 equiv.) with lg3 peptide (33.3 mg, 17.4 μmol, 2.5 equiv.) in degassed DMF (5 mL) under $N_2$. DIPEA (100 μL) was then added in excess and the reaction was left to stir for 4 hours. The crude reaction mixture was then evaporated *in vacuo* and purified by reversed-phase HPLC to isolate the mono-, di- (a,b and a,c), tri-, and tetrasubstituted caviteins (see ESI† for identification). The lg3 a,b disubstituted cavitein intermediate (2 mg, 0.4 μmol, 1 equiv.) and excess peptide lg2 (15 mg, 8.0 μmol, 20 equiv.) were dissolved in degassed DMF and stirring under $N_2$. DIPEA (50 μmol) was added in excess. The reaction was left to stir for 4 h. The crude reaction mixture was evaporated *in vacuo* and dissolved into distilled $H_2O$. It was filtered using a 45 μm nylon filter, and purified by reversed-phase HPLC to yield hetero-TASP **9** (3 mg, 83%). The mass was confirmed by MALDI-MS and determined to be >95% pure by analytical reversed-phase HPLC. Similar procedures were followed for the syntheses of the other hetero-TASPs. The caviteins and hetero-TASPs were characterized using MALDI-MS and the masses are outlined in Table 10.

### Circular dichroism (CD) spectroscopy

All CD spectra were recorded on a JASCO J-710 spectropolarimeter. The J-710 had a circulating water bath set to 25 °C, a 400 W xenon lamp, and an IBM-compatible PC computer for data acquisition. Some of the parameter settings include: 0.1 nm step resolution, 2 nm bandwidth, and 50 nm $min^{-1}$ scanning speed. The J-710 spectropolarimeter was calibrated routinely using $d_{10}$-(+)-camphorsulfonic acid.[50] Each spectrum was an average of three scans subtracted from a reference background scan. Individual samples were run three different times to ensure reproducibility. The caviteins were monitored at 4 μM and 40 μM in 50 mM sodium phosphate buffer (pH = 7.02) to check for concentration effects in a 1 cm and a 1 mm quartz cuvette, respectively.

The raw spectra were normalized to a mean residue ellipticity [$\theta$] at 222 nm using the following equation:

$$[\theta]_{222} = \theta_{obs}/(10lcn),$$

where $\theta_{obs}$ is the observed ellipticity measured in millidegrees, $l$ is the path length in cm, $c$ is the cavitein concentration in mol $L^{-1}$, and $n$ is the number of residues in the cavitein. Errors were on average ± 5%.

**Table 10** % yields and MALDI-MS characterization of the caviteins made from "activated" peptides (note, calculated masses agreed within 1 Da).

| Cavitein | % yield | MALDI mass/Da |
|---|---|---|
| LG3 | 34 | 8240 |
| LG2 | 29 | 8016 |
| AG3 | 45 | 7400 |
| LG2C | 32 | 8424 |
| 2LG3·2LG2_ab | 83 | 8128 |
| 2LG3·2LG2_ac | 81 | 8128 |
| 3LG3·1AG3 | 78 | 8030 |
| 2LG3·2AG3_ab | 75 | 7820 |
| 2LG3·2AG3_ac | 77 | 7820 |
| 1LG3·3AG3 | 83 | 7610 |
| 2LG3·2LG2C_ab | 65 | 8332 |
| 2LG3·2LG2C_ac | 57 | 8332 |

Guanidine hydrochloride (GuHCl) denaturation experiments were performed between 0 and 8.0 M GuHCl in a 50 mM sodium phosphate buffered (pH = 7.0) cavitein solution. Data points were collected at 1 molar units between 0 and 8.0 M to generate a rough unfolding curve. 0.25 interval denaturation studies were then completed to achieve accuracy in the unfolding region, and repeated three times to ensure reproducibility. Likewise, cavitein samples were monitored for unfolding at 4 μM and 40 μM to study concentration effects in a 1 mm and a 1 cm quartz cuvette, respectively. Samples were prepared immediately before data acquisition and equilibrated for 10 min (previously determined that any effect of GuHCl is immediate). The mean residue ellipticity was again monitored at $[\theta] = 222$ nm.

Protein unfolding was analyzed using the linear extrapolation method of Santoro and Bolen.[33] According to this method, unfolding is a reversible, two-state process and that the free energy of folding is a linear function of the GuHCl concentration. The GuHCl denaturation data were fit using a nonlinear least-squares analysis to fit the pre-transitional baseline using the following equation:

$$\theta_{obs} = \theta_N(f_N)(1\text{-}a[\text{GuHCl}]) + \theta_U(1 - f_N),$$

where $\theta_{obs}$ is the mean residue ellipticity at 222 nm at a certain concentration of GuHCl, $\theta_N$ is the mean residue ellipticity of the folded state in the absence of GuHCl, $\theta_U$ is the mean residue ellipticity of the unfolded state, $a$ is a constant and $f_N$ is the fraction of the protein in the folded state. $f_N$ is related to the free energy of unfolding, $\Delta G^0_{H_2O}$, by the following equation:

$$f_N = e^{((\Delta G^0_{H_2O} - m[GuHCl])/RT)}/[1 + e^{((\Delta G^0_{H_2O} - m[GuHCl])/RT)}],$$

where $\Delta G^0_{H_2O}$ is the free energy of unfolding in the absence of GuHCl, $m$ is the free energy change with respect to the concentration of GuHCl, $R$ is the universal gas constant, and $T$ is the temperature. A MacIntosh compatible computer program, KaleidaGraph V. 3.08d was used to calculate the values for $\Delta G^0_{H_2O}$ by a nonlinear least-squares regression analysis. The value of $\theta_N$ was normalized to one. The software analysis program calculated the reported errors.

### Analytical ultracentrifugation (AUC)

Sedimentation equilibrium studies were carried out on a temperature-controlled Beckman Coulter Optima™ XL-I analytical ultracentrifuge. Sedimentation equilibrium experiments were run using either an An60 Ti rotor, or an An50 Ti rotor (4 sample holders and 8 sample holders, respectively) and a UV photoelectric scanner. A six-sector cell, equipped with a 12 mm Epon centerpiece and quartz windows, was loaded with 3 × 120 μL of sample at 3 different concentrations made up in 50 mM sodium phosphate buffer at pH = 7.0, and 3 × 130 μL of reference solvent. Data were collected at 20 °C and at rotor speeds of 27 000, 35 000, and 40 000 rpm until equilibrium was established. Samples were equilibrated for 40 hours and single scans 3 hours apart were overlaid to determine that equilibrium had been reached. Scanning parameters included: radial step size of 0.001 cm, step mode, 10 replicate scans, radial scan range between 5.8 cm and 7.3 cm, and UV detection at 270 nm. The solution density of the samples in sodium phosphate buffer was taken to be 1.000 g mL$^{-1}$. The partial specific volumes of the caviteins were calculated based on their amino acid compositions.[51]

The sedimentation equilibrium data was analyzed on a PC compatible software program called NONLIN.[37] This program uses a nonlinear least-squares analysis in order to generate a reduced molecular weight, $\sigma$, from which the actual experimental molecular weight, $M_w$, can be calculated. Nine sets of data (3 different concentrations at the 3 different rotor speeds) per cavitein were analyzed at a time. The data were initially fit to a single non-associating ideal species model using the Lamm[52] equation below:

$$A_r = \exp[\ln(A_o) + M_w\omega^2(1 - \bar{\nu}\rho/RT)\,(r^2 - r_o^2)] + E,$$

where $A_r$ is the absorbance at radius $r$, $A_o$ is the absorbance at a reference radius $r_o$ (the meniscus), $M_w$ is the molecular weight in g mol$^{-1}$, $\omega$ is the angular velocity of the rotor in rad s$^{-1}$, $\bar{\nu}$ is the partial specific volume of the peptide, $\rho$ is the density of the solvent in g mL$^{-1}$, $R$ is the universal gas constant, $T$ is the temperature in K, and $E$ is the baseline correction factor or baseline offset. For the NONLIN fitting details see the ESI†.

### $^1$H NMR spectroscopy

The 1D $^1$H NMR spectra were run at 20 °C on a 500 MHz Varian Unity instrument, and the samples were dissolved in 45 mM sodium phosphate buffer at pH = 7.0 (90 : 10, $H_2O$–$D_2O$) to a final concentration of approximately 1.5 mM. Spectra were processed using a PC "Windows XP" compatible NMR processing program, MestRe-C 2.3.

The 1D $^1$H NMR N–H/D exchange spectra were run at 20 °C, and the samples were prepared as follows: ~1.5 mM cavitein solutions in a 50 mM acetic acid–acetate buffer at pH = 4.62 were lyophilized to a white solid. $D_2O$ was then added to the lyophilized samples in the NMR room to the previous volume before lyophilization of 0.5 mL. The resulting sample in a deuterated acetic acid–acetate buffer at pD = 5.02 was transferred quickly to an NMR tube. The pH was re-checked after the exchange experiments were completed to ensure a correct reading of pD = 5.02, since pH has a dramatic effect on exchange rates. The pD was corrected for isotope effects using the equation:[53]

$$\text{pD} = \text{pH}_{read} + 0.4,$$

where $\text{pH}_{read}$ is the reading of the pH electrode. The first scan was acquired 5 minutes after the addition of $D_2O$ and subsequent scans were collected at various time intervals until all of the amide protons had completed exchanged with deuterium. The spectra were analyzed using the same processing program mentioned above. The peak heights were integrated and normalized with the non-exchangeable cavitand proton ($H_{out}$) at ~6.1 ppm. The first-order rate constants were calculated using the first-order rate equation:

$$\ln([H_0]/[H_t]) = k_{obs}t,$$

where $k_{obs}$ is the first-order rate constant, $t$ is the time at which the scan was taken, $[H_0]$ is the integration of the proton at time zero, and $[H_t]$ is the integration of the same proton at time $t$. The half-lives, $t_{1/2}$, of the amide protons were then calculated using the equation:

$$t_{1/2} = \ln(2/k_{obs}).$$

Protection factors were then calculated using the equation:

$$P = k_{int}/k_{obs}.$$

Where $P$ is the protection factor, $k_{obs}$ is the experimental first-order rate constant, and $k_{int}$ is the first-order rate constant for an "unprotected" amide proton at pH = 4.62 at 20 °C. $k_{int}$ can be calculated from the intrinsic half-life, $t_{1/2\text{-int}}$, which is determined using the following equation:[41a]

$$t_{1/2\text{-int}} = 200/[10^{(pH-3)} + 10^{(3-pH)}][10^{0.05T}].$$

Where $t_{1/2\text{-int}}$ is the intrinsic half-life for an unprotected proton, and $T$ is the temperature in °C. Errors represent one standard deviation from three rate constant estimates.

### ANS binding

ANS fluorescence measurements were made on a Varian CARY Eclipse fluorescence spectrophotometer equipped with a Xenon Arc lamp. Samples were run at 20 °C using a 1 cm path length with concentrations of 50 μM and 100 μM, and contained 2 μM ANS, respectively, in 50 mM sodium phosphate buffer at pH = 7.02. Reference emission spectra were collected for 95% ethanol and 100% HPLC-grade methanol with 2 μM ANS. Excitation was at 370 nm and emission was recorded between 385 and 600 nm.

## Acknowledgements

We would like to thank the Natural Sciences and Engineering Research Council of Canada for their financial support. We also like to thank M. Okon for acquiring the $^1$H NMR spectra and sharing his expertise in protein NMR. Additional acknowledgements are given to The Center for Biothermodynamics and support from the Michael Smith Foundation for Health Research for data obtained on the AUC, and to F. Rosell and G. Mauk for the use of their Varian CARY Eclipse fluorescence spectrophotometer in the Life Sciences Center.

## References

1 C. B. Anfinsen, *Science*, 1973, **181**, 223–230.
2 (*a*) W. F. Degrado, *Adv. Protein Chem.*, 1988, **39**, 51–124; (*b*) M. H. Hecht, J. S. Richardson and W. F. DeGrado, *Biochemistry*, 1997, **36**, 2450–2458; (*c*) P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber and P. S. Kim, *Science*, 1998, **282**, 1462–1467; (*d*) L. Baltzer, H. Nilsson and J. Nilsson, *Chem. Rev.*, 2001, **101**, 3153–3163.
3 S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.*, 1987, **109**, 6751–6758.
4 (*a*) J. Kaplan and W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11566–11570; (*b*) J. M. Mason and K. M. Arndt, *ChemBioChem*, 2004, **5**, 170–176; (*c*) Pinto, G. R. Dieckmann, C. S. Ghandi, C. G. Papworth, J. Braman, M. A. Shaughnessy, J. D. Lear, R. A. Lamb and W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 11301–11306; (*d*) J. D. Lear, H. Gratkowski and W. F. DeGrado, *Biochem. Soc. Trans.*, 2001, **29**, 559–564; (*e*) P. Rossi, L. Tecilla and P. Baltzer, *Chem.–Eur. J.*, 2004, **10**, 4163–4170; (*f*) W. F. DeGrado, *J. Am. Chem. Soc.*, 1997, **119**, 3212–3217; (*g*) H. K. Rau, N. DeJonge and W. Haehnel, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 11526–11531.
5 (*a*) M. Mutter and S. Vuilleumier, *Angew. Chem., Int. Ed. Engl.*, 1989, **28**, 535–554; (*b*) M. Mutter, *Trends Biochem. Sci.*, 1988, **13**, 260–265.
6 (*a*) K. J. Jensen and G. J. Barany, *J. Pept. Res.*, 2000, **56**, 3–11; (*b*) K. J. Jensen and Brask, *Cell. Mol. Life Sci.*, 2002, **59**, 859–869.
7 (*a*) T. Sasaki and E. T. Kaiser, *J. Am. Chem. Soc.*, 1989, **111**, 380–381; (*b*) T. Sasaki and E. T. Kaiser, *Biopolymers*, 1990, **29**, 79–88; (*c*) K. S. Akerfeldt, R. S. Kim, D. Camac, J. T. Groves, J. D. Lear and W. F. DeGrado, *J. Am. Chem. Soc.*, 1992, **114**, 9656–9657; (*d*) H. Mihara, N. Nishino, R. Hasegawa and T. Fujimoto, *Chem. Lett.*, 1992, **9**, 1805–1808.
8 (*a*) R. Hirschmann, P. A. Sprengeler, T. Kawasaki, J. W. Leahy, W. C. Shakespeare and A. B. Smith, III, *J. Am. Chem. Soc.*, 1992, **114**, 9699–9701; (*b*) H. Li and L. X. Wang, *Org. Biomol. Chem.*, 2003, **1**, 3507–3513.
9 (*a*) Y. Hamuro, M. C. Calama, H. S. Park and A. D. Hamilton, *Angew. Chem.*, 1997, **109**, 2797–2800; (*b*) K. A. Wong, M. P. Jacobsen, D. J. Winzor and D. P. Fairlie, *J. Am. Chem. Soc.*, 1998, **120**, 3836–3841.
10 (*a*) M. R. Ghadiri, C. Soares and C. Choi, *J. Am. Chem. Soc.*, 1992, **114**, 4000–4002; (*b*) M. W. Mutz, M. A. Case, J. F. Wishart, M. R. Ghadiri and G. L. McLendon, *J. Am. Chem. Soc.*, 1999, **121**, 858–859.
11 (*a*) B. C. Gibb, A. R. Mezo, A. S. Causton, J. R. Fraser, F. C. S. Tsai and J. C. Sherman, *Tetrahedron*, 1995, **51**, 8719–8732; (*b*) A. R. Mezo and J. C. Sherman, *J. Am. Chem. Soc.*, 1999, **121**, 8983–8994; (*c*) A. S. Causton and J. C. Sherman, *J. Pept. Sci.*, 2002, **6**, 275–282.
12 (*a*) V. Nanda, M. M. Rosenbkatt, A. Osyczka, H. Kono, Z. Getahun, P. L. Dutton, J. G. Saven and W. F. DeGrado, *J. Am. Chem. Soc.*, 2005, **127**, 5804–5805; (*b*) J. Brask, J. M. Dideriksen, J. Nielsen and K. Jensen, *Org. Biomol. Chem.*, 2003, **1**, 2247–2252; (*c*) W. F. DeGrado, Z. R. Wasserman and J. D. Lear, *Science*, 1989, **243**, 622–628; (*d*) M. H. Hecht, J. S. Richardson, D. C. Richardson and R. C. Ogden, *Science*, 1990, **249**, 884–891; (*e*) O. Nyanguile, M. Mutter and G. Tuchscherer, *Lett. Pept. Sci.*, 1994, **1**, 9–16.
13 (*a*) D. P. Raleigh, S. F. Betz and W. F. DeGrado, *J. Am. Chem. Soc.*, 1995, **117**, 7558–7559; (*b*) R. B. Hill, J. K. Hong and W. F. DeGrado, *J. Am. Chem. Soc.*, 2000, **122**, 746–747; (*c*) C. P. Hill and W. F. DeGrado, *Structure*, 2000, **8**, 471–479; (*d*) M. H. Hecht, A. Das, A. Go, L. H. Bradley and Y. Wei, *Protein Sci.*, 2004, **13**, 1711–1723.
14 (*a*) A. R. Mezo and J. C. Sherman, *J. Am. Chem. Soc.*, 1999, **121**, 8983–8994; (*b*) A. S. Causton and J. C. Sherman, *Bioorg. Med. Chem.*, 1999, **7**, 23–27; (*c*) A. Causton and J. C. Sherman, *J. Pept. Sci.*, 2002, **8**, 275–282; (*d*) E. S. Seo, W. R. P. Scott, S. Straus and J. C. Sherman, *Chem.–Eur. J.*, 2007, **13**, 3596–3605.
15 W. F. DeGrado and J. D. Lear, *J. Am. Chem. Soc.*, 1985, **107**, 7684–7689.
16 F. H. C. Crick, *Acta Crystallogr.*, 1953, **6**, 689–697.
17 A. S. Causton, Ph.D. Thesis, University of British Columbia, Vancouver, Canada, 2001.
18 (*a*) S. F. Betz and W. F. DeGrado, *Biochemistry*, 1996, **35**, 6955–6962; (*b*) S. F. Betz, P. A. Liebman and W. F. DeGrado, *Biochemistry*, 1997, **36**, 2450–2458; (*c*) R. B. Hill and W. F. DeGrado, *J. Am. Chem. Soc.*, 1998, **120**, 1138–1145; (*d*) H. K. Rau, H. Snigula, A. Struck, B. Robert, H. Scheer and W. Haehnel, *Eur. J. Biochem.*, 2001, **268**, 3284–3295.
19 (*a*) S. Cusack, C. Berthet-Colominas, M. Härtlein, N. Nassar and R. Leberman, *Nature*, 1990, **347**, 249–255; (*b*) S. M. Soisson, B. MacDougall-Shackleton, R. Schleif and C. Wolberger, *Science*, 1997, **276**, 421–425; (*c*) V. Biou, A. Yaremchuk, M. Tukalo and S. Cusack, *Science*, 1994, **263**, 1404–1410.
20 (*a*) R. Aurora, R. Srinivasan and G. D. Rose, *Science*, 1994, **264**, 1126–1130; (*b*) R. Aurora and G. D. Rose, *Protein Sci.*, 1998, **7**, 21–38.
21 (*a*) R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149–2154; (*b*) B. C. Gibb, A. R. Mezo and J. C. Sherman, *Tetrahedron Lett.*, 1995, **36**, 7587–7590.
22 (*a*) J. R. Moran, S. Karbach and D. J. Cram, *J. Am. Chem. Soc.*, 1982, **104**, 5826–5828; (*b*) D. J. Cram, J. M. Cram, *Container Molecules and Their Guests*, Royal Society of Chemistry, Cambridge, UK, 1994.
23 G. Tuchscherer and M. J. Mutter, *J. Biotechnol.*, 1995, **41**, 197–210.
24 According to a crystal structure, the distance between adjacent bromines when Br atoms replace SH groups of the arylthiol cavitand are 6.8 Å and the distance between opposing bromines is 9.6 Å. See: D. J. Cram, S. Karbach, H.-E. Kim, C. B. Knobler, E. F. Maverick, J. L. Ericson and R. C. Helgeson, *J. Am. Chem. Soc.*, 1988, **110**, 2229–2237.
25 B. V. Reddy and T. L. Blundell, *J. Mol. Biol.*, 1993, **233**, 464–479.
26 (*a*) W. C. Johnson, Jr., *Proteins: Struct., Funct., Genet.*, 1990, **7**, 205–214; (*b*) R. W. Woody, *Methods Enzymol.*, 1995, **246**, 34–70; (*c*) G. D. Fasman, *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York, 1996.
27 S. C. Kwok and R. S. Hodges, *J. Biol. Chem.*, 2003, **278**, 35248–35254.
28 F. Offredi, F. Dubail, P. Kischel, K. Sarinski, A. S. Stern, C. Van de Weerdt, J. C. Hoch, C. Prosperi, J. M. Francois, S. L. Mayo and J. A. Martial, *J. Mol. Biol.*, 2003, **325**, 163–174.
29 K. Kuwijima, *Proteins: Struct., Funct., Genet.*, 1989, **6**, 87–103.
30 P. C. Kahn, *Methods Enzymol.*, 1979, **61**, 339–377.
31 T. E. Creighton, *Proteins Structure and Molecular Properties*, W. H. Freeman, New York, 2nd edn, 1993.

32 (*a*) C. N. Pace, *Methods Enzymol.*, 1986, **131**, 266–280; (*b*) C. N. Pace, *Trends Biochem. Sci.*, 1990, **15**, 14–17.
33 (*a*) M. M. Santoro and D. W. Bolen, *Biochemistry*, 1988, **27**, 8063–8068; (*b*) M. M. Santoro and D. W. Bolen, *Biochemistry*, 1992, **31**, 4901–4907.
34 J. K. Myers, C. N. Pace and J. M. Sholtz, *Protein Sci.*, 1995, **4**, 2138–2148.
35 F. Ahmad and C. C. Bigelow, *Biopolymers*, 1996, **25**, 1623–1633.
36 (*a*) S. E. Harding, A. J. Rowe, J. C. Horton, *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, Royal Society of Chemistry, Oxford, 1992; (*b*) J. C. Hansen, J. Lebowitz and B. Demeler, *Biochemistry*, 1994, **33**, 13155–13163.
37 M. L. Johnson, J. J. Correia, D. A. Yphantis and H. R. Halvorson, *Biophys. J.*, 1981, **36**, 575–588.
38 (*a*) J. Cavanagh, W. J. Fairbrother, A. G. Palmer, III, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego, 1996; (*b*) S. Roy, K. J. Helmer and M. H. Hecht, *Folding Des.*, 1997, **2**, 89–92.
39 O. B. Ptitsyn, *Adv. Protein Chem.*, 1995, **47**, 83–229.
40 (*a*) S. W. Englander, N. W. Downer and H. Teitelbaum, *Annu. Rev. Biochem.*, 1972, **41**, 903–905; (*b*) T. M. Raschke and S. Marqusee, *Curr. Opin. Biotechnol.*, 1998, **9**, 80–86; (*c*) S. W. Englander, L. Mayne, Y. Bai and T. R. Sosnick, *Protein Sci.*, 1997, **6**, 1101–1109; (*d*) M.-F. Jeng, S. W. Englander, G. A. Elove, A. J. Wand and H. Roder, *Biochemistry*, 1990, **29**, 10433–10437; (*e*) F. M. Hughson, P. E. Wright and R. L. Baldwin, *Science*, 1990, **249**, 1544–1548.
41 K. Kuwajima, *Proteins: Struct., Funct., Genet.*, 1989, **6**, 87–103.
42 G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas and R. I. Gilmanshin, *Biopolymers*, 1991, **31**, 119–128.
43 (*a*) S. Roy, G. Ratnaswamy, J. A. Boice, R. Fairman, G. McLendon and M. H. Hecht, *J. Am. Chem. Soc.*, 1997, **119**, 5302–5306; (*b*) D. P. Raleigh, S. F. Betz and W. F. DeGrado, *J. Am. Chem. Soc.*, 1995, **117**, 7558–7559.
44 (*a*) K. S. Akerfeldt, R. M. Kim, D. Camac, J. T. Groves, J. D. Lear and W. F. DeGrado, *J. Am. Chem. Soc.*, 1992, **114**, 9656–9657; (*b*) W. F. DeGrado, *J. Am. Chem. Soc.*, 1997, **119**, 3212–3217; (*c*) A. Senes, D. E. Engel and W. F. DeGrado, *Curr. Opin. Struct. Biol.*, 2004, **14**, 465–479.
45 (*a*) W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 6298–6305; (*b*) F. V. Cochran, S. P. Wu, W. Wang, V. Nanda, J. G. Saven, M. J. Therien and W. F. DeGrado, *J. Am. Chem. Soc.*, 2005, **127**, 1346–1347.
46 A. Lombardi, J. W. Bryson and W. F. DeGrado, *J. Am. Chem. Soc.*, 1997, **119**, 12378–12379.
47 (*a*) K. Severin, H. Lee, A. J. Kennan and M. R. Ghadiri, *Nature*, 1997, **389**, 706–709; (*b*) B. R. Gibney, F. Rabanal, K. S. Reddy and P. L. Dutton, *Biochemistry*, 1998, **37**, 4635–4643.
48 M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248–254.
49 H. Rink, *Tetrahedron Lett.*, 1987, **28**, 3787–3790.
50 G. C. Chen and J. T. Yang, *Anal. Lett.*, 1977, **10**, 1195–1207.
51 E. J. Cohn, J. T. Edsall, *Amino, Acids, Peptides and Proteins as Ions and Dipolar Ions*, Reinhold Publishing Corporation, New York, 1943.
52 O. Lamm, "Die Differentialgleichung der Ultrazentrifugierung", *Ark. Mat. Astron. Fys.*, 1929, **21B**, 1–4.
53 P. K. Glasoe and F. A. Long, *J. Phys. Chem.*, 1960, **64**, 188–190.