J|A|C|S
ARTICLES
Published on Web 05/11/2006

A Designed Branched Three-Helix Bundle Protein Dimer

Gunnar T. Dolphin‡

Contribution from the Department of Chemistry—IFM, Linköping University, 58183 Linköping, Sweden

Received January 23, 2006; E-mail: gunnar.dolphin@ujf-grenoble.fr

Abstract: The ultimate goals of de novo protein design are the construction of novel tertiary structures and functions. Here is presented the design and synthesis of a uniquely branched three-helix bundle that folds into a well-folded dimeric protein. The branching of this protein was performed by the method of native chemical ligation, which provides a chemoselective and stable amide bond between the unprotected fragments. This ligation strategy was possible by the presented facile preparation of a peptide (43 amino acids) with a specific side chain thioester, which is synthesized by general Fmoc solid phase peptide synthesis. From the presented structural analysis, it is seen that the folded protein is present as a stable and highly helical dimer, thus forming a six-helix bundle. This unique tertiary structure, composed of a dimer of three individual α-helices branched together, offers different possibilities for protein engineering, such as metal and cofactor binding sites, as well as for the construction of novel functions.

Introduction

The construction of novel tertiary structures and functions is now the ultimate goal of de novo protein design.[1] Previously, de novo design was used to test our understanding of protein folding and structure. Although the folding pathways of proteins are still not fully understood, we now have enough knowledge of the forces that contribute to the thermodynamic stability and conformational specificity of proteins to be able to form uniquely well-folded structures. To date, a number of proteins with well-defined tertiary structures have been successfully designed, and there is intense interest for designed protein structures as they provide novel scaffolds for the engineering of, for example, biosensors,[4] novel catalysts,[5] cofactor-linked proteins,[6] and the construction of novel biomaterials for nanobiotechnology.[7] From the progress of chemoselective ligation strategies,[8−10] large monomeric de novo designed structures[11] and template-assembled proteins have been produced.[10,12] From this laboratory, research is performed on the design and functionalization of four-helix bundles with linear sequences.[4,11] To broaden the repertoire of designed protein tertiary structures, here is now reported the research on a designed uniquely branched three-helix bundle, 2α-α, that folds into a well-folded globular protein (Figure 1). Also reported is the synthesis of the branched protein, which was performed by an extension of the native chemical ligation (NCL) methodology that provides a chemoselective and stable amide bond between the unprotected fragments. Usually, NCL is used for the formation of linear peptide sequences,[8] but here is presented a strategy to prepare branched peptides in aqueous solution. A key factor for this research is the presented facile preparation of the required peptide fragment (43 amino acids) with a side chain thioester, which was synthesized by solid phase peptide synthesis (SPPS) using general fluorenylmethoxycarbonyl (Fmoc) methodology. From the presented structural analysis, it is seen that the folded protein is present as a highly helical dimer, thus forming a six-helix bundle. This unique tertiary structure, composed of a dimer of three individual α-helices branched together, offers different possibilities for

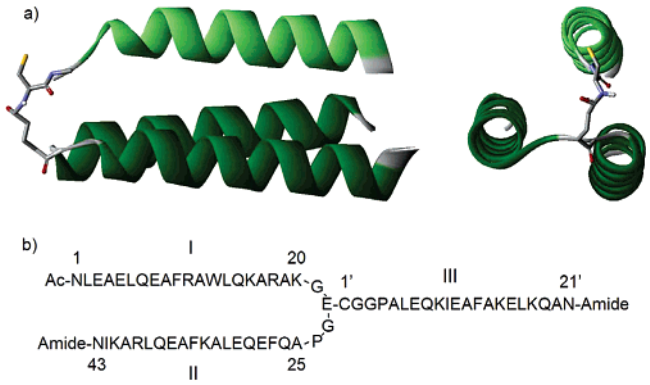


Figure 1. (a) Modeled representation of monomeric 2α-α, side and top view. The helix−loop−helix fragment is presented in dark green, and the branching helix is in green. (b) Amino acid sequence of the branched protein. Amino acid numbering is from the unligated fragments and indicates the individual helices' (I, II, and III) direction.

‡ Current address: LEDSS 5, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France.

(1) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779.
(2) Baltzer, L.; Nilsson, H.; Nilsson, J. *Chem. Rev.* **2001**, *101*, 3153.
(3) Kraemer-Pecore, C. M.; Wollacott, A. M.; Desjarlais, J. R. *Curr. Opin. Chem. Biol.* **2001**, *5*, 690.
(4) Enander, K.; Dolphin, G. T.; Liedberg, B.; Lundström, I.; Baltzer, L. *Chem.—Eur. J.* **2004**, *10*, 2375.
(5) Johnsson, K.; Allemann, R. K.; Widmer, H.; Benner, S. A. *Nature* **1993**, *365*, 530.
(6) Cochran, F. V.; Wu, S. P.; Wang, W.; Nanda, V.; Saven, J. G.; Therien, M. J.; DeGrado, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 1346.
(7) Yeates, T. O.; Padilla, J. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 464.
(8) Dawson, P. E.; Muir, T. W.; Clarklewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776.
(9) Liu, C. F.; Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6584.
(10) Rose, K. *J. Am. Chem. Soc.* **1994**, *116*, 30.
(11) Dolphin, G. T. *Chem.—Eur. J.* **2006**, *12*, 1436.
(12) Mutter, M.; Tuchscherer, G. G.; Miller, C.; Altmann, K. H.; Carey, R.; Wyss, D.; Labhardt, A.; Rivier, J. E. *J. Am. Chem. Soc.* **1992**, *114*, 1463.

protein engineering, such as metal and cofactor binding sites, as well as for the construction of novel functions.

Experimental Section

Peptide Synthesis: The fragments 2α-SBzl and Cys-α were synthesized on a Pioneer automated peptide synthesizer (Applied Biosystems) using general fluorenylmethoxycarbonyl (Fmoc) procedures and standard Fmoc-PAL-PEG-PS resin (Applied Biosystems) with a substitution level of 0.17 mmol. Amino acid coupling times of 60 min were used, except for Asn, Arg, and Gln (120 min).

For the synthesis of 2α-SBzl, the γ-carboxyl orthogonally protected Fmoc-Glu-γ-allyl (Applied Biosystems) was introduced into the sequence at the position for branching by the general procedures. The allyl group was selectively removed by treatment with tetrakis-(triphenylphosphine)palladium(0) ($Pd(PPh^3)^4$) (3 equiv) in a mixture of trichloromethane, acetic acid, and morpholine (17:2:1 v/v; 15 mL/g of polymer), for 3 h at room temperature. The resin was washed with 0.5% diethyldithiocarbamic acid in DMF, 0.5% DIPEA in DMF, 50 mM HOBt in DMF, and finally DMF washes. Thioesterification of the free γ-carboxyl group on Glu22 was performed by coupling benzyl mercaptan (10 equiv) with *N*,*N*′-diisopropylcarbodiimide (10 equiv) in DCM (6 mL/g of polymer). The reaction mixture was left to stand overnight with swirling at room temperature.

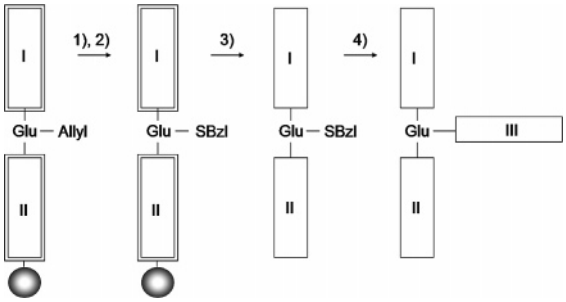
The deprotection and cleavage from the resin of 2α-SBzl was accomplished with a mixture of TFA:TIS:H_2O (95:2.5:2.5 v/v), 15 mL/g of resin, with swirling for 2 h. After filtration and reducing the TFA by N_2 bubbling, the peptide was precipitated and washed three times with cold diethyl ether. Purification was performed by reversed-phase HPLC on a semipreparative Kromasil C-8 with a gradient from 34 to 48% acetonitrile during 40 min, with 0.1% TFA and a flow rate of 10 mL min^{-1}. The retention time of 2α-SBzl is 33 min. The identity was verified by MALDI-MS and recorded on a Voyager System 4212 (Applied Biosystems). The calculated weight for 2α-SBzl is 5076.6 Da and found 5075.8 Da $[M + H]^+$.

Cys-α was deprotected and cleaved from the resin with a mixture of TFA:H_2O:1,2-ethanedithiol:TIS (94:2.5:2.5:1 v/v). Workup, purification, and identification were as for 2α-SBzl. For HPLC purification, Cys-α has a retention time of 13.5 min, with a gradient from 26 to 42% acetonitrile, 0.1% TFA, during 30 min, with a flow rate of 10 mL min^{-1}. The calculated weight for Cys-α is 2244.6 Da, found 2243.8 Da $[M + H]^+$.

Chemoselective Ligation. The ligation of the two unprotected synthetic peptide fragments was performed using standard methodology.[8] Briefly, the thioester, 2α-SBzl (15 mg, 2.96 μmol), and the C-terminal segment Cys-α (10 mg, 4.46 μmol) were dissolved in 6 M GuHCl with 100 mM sodium phosphate (pH 7.5) to give a final concentration of 2 and 3 mM, respectively. Thiophenol (100 μL) was added, and the mixture was stirred for 16 h. The reaction was followed by analytical HPLC with a linear gradient of 0.4% min^{-1} from 30 to 46% acetonitrile + 0.1% TFA. After ligation, the sulfhydryl group on the Cys was capped with *N*-ethylmaleimide (NEM). To do this, the reaction was acidified with TFA and the excess thiophenol was extracted away six times with diethyl ether. After lyophilization, the peptide was dissolved in 1.5 mL of H_2O, and 11 mg of TCEP (10 equiv) was added. NEM (11 mg, 10 equiv) was added after the pH was adjusted to 6.8, and the reaction was complete within 10 min. The reaction mixture was acidified to pH 3, and the peptide was purified by semipreparative reversed-phase C-8 HPLC, by using a 34−50% acetonitrile + 0.1% TFA linear gradient at 0.4% min^{-1} to afford 12 mg (55% yield based on limiting 2α-SBzl) of the full-length protein 2α-α. The product was characterized by MALDI-MS: observed, 7322.8 Da; calculated, 7322.2 Da $[M + H]^+$.

NMR Spectroscopy. One-dimensional ^1H NMR spectra were recorded on a Varian Inova Unity 600 NMR spectrometer equipped with a matrix shim system from Resonance Research Inc. using a 90° pulse of 4.15 μs. Solvent suppression was accomplished by weak preirradiation of the water resonance for 2 s. NMR spectra were the average of 16 transients. Samples were 0.5 mM in peptide in H_2O:D_2O (90:10). The temperature dependence of the ^1H NMR spectrum of 2α-α was determined at pH 3.3, and the pH dependence was performed at 15 °C.

Scheme 1. Strategy for the Preparation of Amide Bonded Branched Peptides by Fmoc−SPPS and Chemoselective Ligation[a]



[a] Conditions: (1) Selective removal of allyl by $Pd(PPh_3)_4$ (3 equiv) in $CHCl_3$:AcOH:morpholine (17:2:1 v/v). (2) Side chain thioester formation with benzyl mercaptan (10 equiv) and DIPCDI (10 equiv). (3) Cleavage from the resin and side chain deprotection with TFA:TIS:H_2O (95:2.5:2.5 v/v). (4) Chemoselective ligation. 2α-SBzl (1 equiv), Cys-α (1.5 equiv) in 6 M GuHCl, 100 mM phosphate, pH 7.5, and 2% thiophenol.

Circular Dichroism Spectroscopy. Circular dichroism (CD) spectra were acquired with signal averaging on a CD6 spectrodichrograph (Jobin-Yvon Instruments SA), and a baseline was recorded separately and subtracted. The instrument was routinely calibrated using d-10-(+)-camphorsulfonic acid. The spectra were processed and presented using the IGOR software from WaveMetrics Inc. Far-UV spectra were recorded from 185 to 260 nm, with quartz cells and 2 s response times. Far-UV spectra are the averages of three scans, and ellipticities are reported as mean residue ellipticities. Stock peptide concentrations were determined spectrophotometrically in 6.0 M GuHCl using an extinction coefficient at 280 nm of 5690 cm^{-1} M^{-1}.

A solution of 20 μM 2α-α was denatured by the addition of 0−6 M GuHCl at 21 °C. The samples were prepared from a 198 μM stock solution of peptide by the addition of aliquots from a 7.68 M stock solution of GuHCl followed by dilution with water to the desired volume, the pH was set to 3.3 in all solutions. The samples were equilibrated for 1 h before analysis. The concentration of the stock solution of GuHCl was determined from its refractive index. Spectra were recorded from 210 to 260 nm, using 0.05 cm quartz cell and 2 s response time.

The CD spectrum of 2α-α was recorded as a function of concentration at pH 3.3 and 21 °C. Aqueous concentrations studied and respective cell lengths are 1.0 μM and 1.0 cm, 2.0 μM and 1.0 cm, 20 μM and 0.05 cm, and 200 μM and 0.01 cm.

Results and Discussion

For the design of the branched three-helix bundle, a synthetic approach was first determined. To ligate two fragments by the methodology of NCL, a thioester and a N-terminal Cys are required.[8] Recently, the author has prepared a 45-residue peptide with a C-terminal thioester by Fmoc SPPS for use in NCL.[11] Using this same strategy, a specific glutamic acid side chain thioester could be formed. To do this, an allyl γ-carboxyl orthogonally protected glutamic acid (Glu-γ-allyl) was incorporated into the sequence at the desired position during SPPS. After selective allyl deprotection, the free γ-carboxyl was available for thioester functionalization before cleavage from the resin. Thereafter, chemoselective amide bond formation between the Glu-γ-thioester and an N-terminal cysteine would form a branched peptide. This site-selective approach was used to ligate the helix, Cys-α, to the loop region of the helix−loop−helix, 2α-SBzl (Scheme 1), where Cys indicates the N-terminal

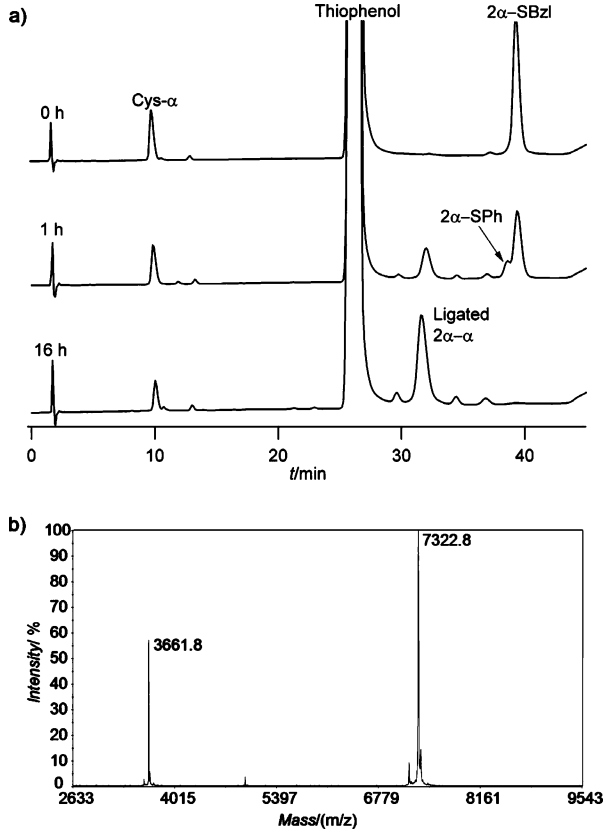


Figure 2. (a) Chemoselective ligation between 2α-SBzl and Cys-α was monitored by analytical HPLC (chromatograms at 230 nm are shown). After 1 h, the active thiophenyl fragment 2α-SPh can be seen as well as the ligation product 2α-α. After 16 h, the formation of 2α-α was complete. The gradient is 30−46% acetonitrile during 40 min and 0.1% TFA as counterion. (b) The MALDI mass spectrum of 2α-α shows single and doubly charged species. The calculated M_w is 7322.2 Da, found M_w is 7322.8 Da $[M + H]^+$.

cysteine and SBzl indicates the side chain benzyl thioester on Glu22.

The fragment 2α-SBzl was based on the de novo designed sequence of GTD-43, which forms a well-folded helix−loop−helix dimer in aqueous solution.[13,14] To create complementary hydrophobic surfaces and charged interactions to the branching helix Cys-α, 12 mutations were made on the parent 43-residue sequence. The de novo design procedure for Cys-α was performed as previously described for GTD-43. Briefly, this helix was designed to be amphiphilic with a complementary hydrophobic surface and a hydrophilic surface with designed salt bridges to the other two helices. From the point of branching, the loop of Cys-α consists of C−G−G−P−, where cysteine is at the N-terminal for thiol ligation, the glycines function as a flexible spacer, and the proline induces the helical structure of the branching peptide. Negative design was incorporated by destabilizing salt bridges for the folded conformation where the branching helix is on the opposite side of the helix−loop−helix as compared to the side-view structure seen in Figure 1a. An aromatic ensemble was incorporated in the hydrophobic core to restrict the dynamics of the folded structure, as found in GTD-43.[13] This new ensemble consists of Phe10 and Trp13 in helix I, Phe27 and Phe34 in helix II, and Phe13′ in helix III. The sequence of the branched peptide 2α-α is seen in Figure 1b.

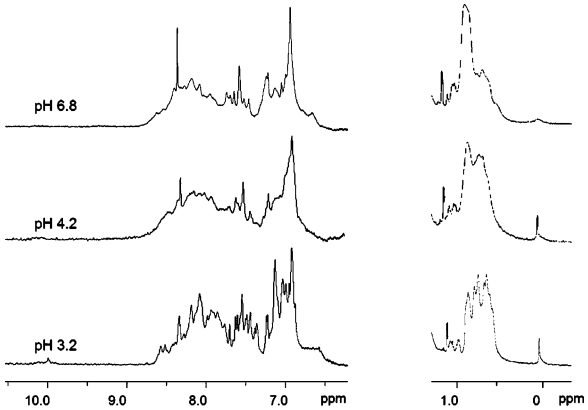


Figure 3. Dependence of pH on the structure of 2α-α (0.5 mM), studied by the ^{1}H NMR spectrum at 15 °C in H_2O:D_2O (90:10). The aromatic amide and methyl regions are shown.

The helix−loop−helix fragment, 2α-SBzl, contains 43 residues with a Glu-γ-benzyl thioester at position 22; this fragment and the Cys-α fragment were synthesized using standard TBTU/DIPEA activation on a peptide amide resin. Chemoselective ligation of the two synthetic peptide fragments was performed using standard NCL methodology (Scheme 1).[8] Ligation was followed by analytical RP-HPLC and was complete within 16 h (Figure 2a). After completion of the ligation, the sulfhydryl group on Cys1′ was capped with *N*-ethylmaleimide, and 2α-α was readily purified with good yield and identified by MALDI-MS (Figure 2b).

The structure and stability of 2α-α was studied by a number of biophysical techniques. With ^{1}H NMR spectroscopy, the pH and temperature dependence of the folded structure were determined. In Figure 3 is seen the pH dependence, which shows that the protein has a more well-defined structure at pH 3.2 where the spectrum is well-dispersed and has narrow line widths when compared to the spectra at higher pH. Well-dispersed spectra with narrow line widths are criteria that distinguish between a native-like structure and the molten globular state. The temperature dependence of the ^{1}H NMR shows, with increasing temperature, broader line widths and decreased chemical shift dispersion; this shows that 2α-α is in slow exchange on the NMR time scale (see Supporting Information).[15]

The far-UV circular dichroism (CD) spectra of 2α-α as a function of concentration are seen in Figure 4, measured range is 1−200 μM, and no precipitation was observed. The spectra are characteristic of a highly α-helical protein, with a maximum at 195 nm and minima at 208 and 222 nm. As concentration dependence is seen, this indicates that 2α-α is aggregating. Analyzing a plot of the CD data at 222 nm and [2α-α] (log scale), it is obvious that a transition is occurring and that the 200 μM sample is near the transition end (Figure 4, inset). From the experimental data, the low concentration transition end is not seen; however, as the monomeric three-helix bundle is designed to be helical, it is expected to have a high helical content. From these data, it is assumed that the observed concentration dependence is from the self-association of the helical monomer to a helical dimer and that dimerization induces a higher helical content. By fitting the data with an equation that describes a monomer−dimer equilibrium, it was possible to give a preliminary estimate of the dissociation constant, K_{diss}

(13) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 11297.
(14) Brive, L.; Dolphin, G.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 8598.
(15) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

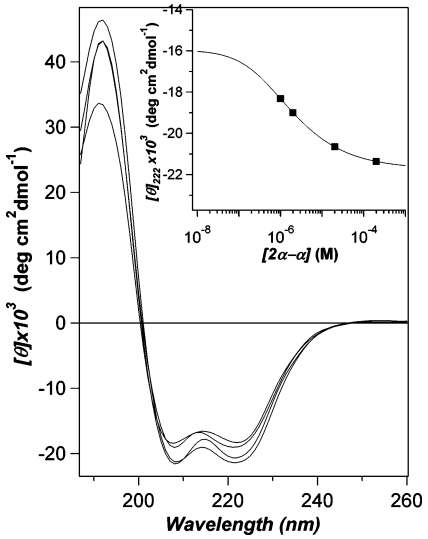


Figure 4. Concentration dependence of the far-UV CD spectra of 2α-α at pH 3.3. Studied concentrations are 1 (top trace at 222 nm), 2, 20, and 200 μM (bottom trace). Inset: The mean residue ellipticity at 222 nm as a function of concentration. The solid line is the nonlinear least-squares fit to the $[\theta]_{222}$ data with an equation that describes a monomer−dimer equilibrium.

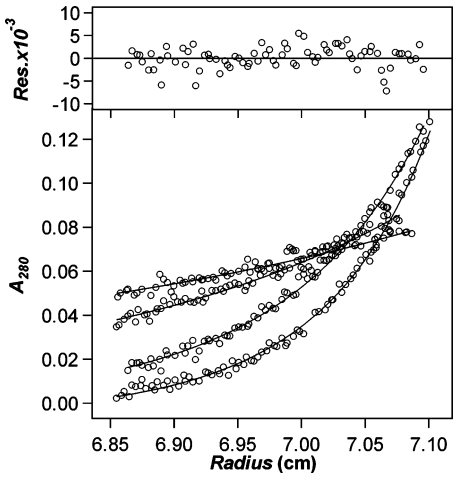


Figure 5. Equilibrium analytical ultracentrifugation of 2α-α (15 μM) at pH 3.3. Data are from 18 000, 23 000, 35 000, and 40 000 rpm and globally fitted to monomer−dimer equilibrium. The solid lines represent the calculated fit. Above are the residuals of the fit to the 35 000 rpm data.

$= 2$ μM, as well as an estimate of the mean residue ellipticity for the monomer at 222 nm, $[\theta]_{222,\mathrm{mono}} = -16\,000$ deg cm^2 dmol^{-1}.

To determine the aggregation behavior of 2α-α more accurately, the protein was further analyzed by sedimentation equilibrium ultracentrifugation. The peptide was analyzed at three concentrations, 15, 25 and 45 μM, to probe for concentration dependence. In Figure 5 is shown the 15 μM data globally fitted to a monomer−dimer equilibrium with $M_{\mathrm{W,calc}}$ fixed to 7322 Da; from this, the dissociation constant for the monomer−dimer equilibrium was determined as $K_{\mathrm{diss}} = 5$ μM, indicating that 2α-α exists as a dimer at the concentrations used for structural studies. The residuals of the 35 000 rpm data fit are shown in Figure 5 and are representative of the goodness of the fit as the residuals are random and centered on zero, fitting with higher order aggregates gave less good fits.

(16) Hill, R. B.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 1138.

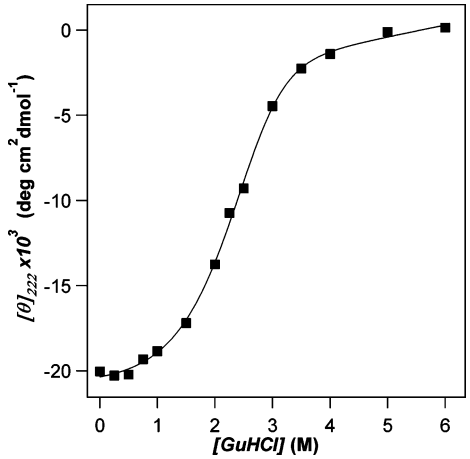


Figure 6. GuHCl-induced unfolding of 2α-α (20 μM) monitored by the mean residue ellipticity at 222 nm, pH 3.3, and 21 °C. The solid line shows the nonlinear least-squares best fits to the experimental results.

The stability of 2α-α was determined by chemical denaturation monitored by $[\theta]_{222}$ as a function of GuHCl concentration (Figure 6). The denaturation forms a sigmoid curve with a midpoint of the transition at 2.5 M GuHCl. Nonlinear least-squares fitting of a two-state model, including only native dimer and unfolded random coil monomer, the Gibbs free energy of unfolding in aqueous solution $\Delta G°(H_2O)$ is determined to 10.3 $\pm$ 0.3 kcal mol^{-1}. The measure of the cooperativity of the denaturation (m_G) is determined to 1.8 kcal mol^{-1} M^{-1}.

Conclusions

In conclusion, 2α-α forms a stable and highly helical branched three-helix bundle dimer. The branched peptide has been successfully synthesized by NCL, which forms a stable amide bond between two peptide fragments that were prepared by general Fmoc chemistry. So far, the exact folding arrangement of the individual helices in the dimeric tertiary structure is currently not determined; a stable tertiary configuration would be a bisecting-w, in the same manner as the bisecting-u found for the helix−loop−helix dimer, α2D.[16] A bisecting-w would lock the helices and form a well-shielded hydrophobic core. Future detailed structural analysis will determine the precise nature of the folded dimer. The presented design and synthesis is of general applicability for the simple construction of site-selective branched peptides and proteins and may be extended for use in template and MAP synthetic strategies, as well as for specific bioconjugation of affinity and functional groups to synthetic proteins. The presented six-helix tertiary structure, composed of a dimer of three individual α-helices branched together, expands the repertoire with its unique architecture, future possibilities for the construction of novel functions, as well as for the engineering of binding sites for cofactors and metals.

Acknowledgment. I thank Prof. L. Baltzer for fruitful discussions and support, and Dr. M. Lundqvist for assistance with the equilibrium sedimentation ultracentrifugation.

Supporting Information Available: NMR temperature study, equilibrium sedimentation ultracentrifugation, and data analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

JA060524K