Total Chemical Synthesis of a Folded β-Sandwich Protein Domain: An Analog of the Tenth Fibronectin Type 3 Module

Michael J. Williams,* Tom W. Muir,*
Mark H. Ginsberg, and Stephen B. H. Kent

Contribution from The Scripps Research Institute
10666 North Torrey Pines Road, La Jolla, California 92037

Received July 7, 1994

The fibronectin type 3 module is one of the most common protein domains identified to date.[1] Members of this family have been identified in over 60 different proteins, in addition to 17 repeats found in the fibronectin monomer itself. The conserved type 3 module structure is characterized by a seven-stranded β-sandwich, consisting of a three-stranded sheet that folds onto a four-stranded sheet, to enclose a hydrophobic core.[2,3] This fold shows striking similarity to structures described for cytokine receptor and immunoglobulin domain superfamilies.[4] We report the total chemical synthesis of one of these folded β-sandwich protein domains by the chemoselective ligation of two mutually reactive unprotected synthetic peptides. This 94-residue domain corresponds to the tenth type 3 module (^{10}F3) from fibronectin[5] which contains the -Arg-Gly-Asp- (RGD) motif, a key determinant of the binding interaction between fibronectin and a number of heterodimeric integrin cell surface receptors, including $\alpha_{IIb}\beta_3$, $\alpha_3\beta_1$, $\alpha_5\beta_1$ and most of the α_v integrins.[6] The tertiary structure of recombinant ^{10}F3 (rec^{10}F3) has previously been determined by NMR,[2] and we demonstrate that the synthetic protein maintains the same global fold.

The protocol used for the synthesis of ^{10}F3 has been described previously.[7] In brief, the chemoselective reaction of unique thioacid and bromoacetyl moieties results in the efficient ligation of two synthetic peptides, corresponding to each half of the target protein domain.[8] The ligation leads to formation of a backbone thioester bond between the two peptides, that replaces a peptide bond in the native structure; hence the backbone NH is replaced with a sulphur atom. Backbone amide protons are important hydrogen bond donors. Moreover, the thioester bond lacks the resonance stabilization of a planar amide bond and therefore exhibits increased flexibility. Thus, placement of the thioester linkage has important implications for the structure of the synthetic target. Examination of the ^{10}F3 module structure revealed that a Gly-Gly pair (residues 40–41) was located in a loop, near the middle of the sequence. Furthermore, backbone amide ^{15}N–^{1}H heteronuclear NOE measurements have shown that the protein backbone of this loop is relatively flexible.[2] Therefore, we anticipated that a thioester bond between Gly^{40} and Gly^{41} would have minimal effect on the overall protein fold.

In preliminary experiments, the type 3 module analog [(COS)$^{40-41}$]^{10}F3 was successfully synthesized by this technique.[8] However, extensive NMR and electrospray mass spectrometry (ESMS) studies revealed that, in addition to the folded protein, considerable amounts of unfolded or misfolded ligated material were present that possessed the same molecular weight. This discrepancy was resolved by subsequent studies that identified Ser^{43} as the site of significant levels of an N → O acyl shift rearrangement in the synthetic peptide segment BrAc(42–94)^{10}F3, resulting in the generation of an O-linked synthetic isomer of the target protein module.[12]

To overcome this problem, we synthesized a second analog [Ala^{43},(COS)$^{41-42}$]^{10}F3, in which Ser^{43} was replaced with alanine.[8] The product was initially characterized by ESMS (Figure 1), and its folded state was examined by 1D ^{1}H NMR through comparison to rec^{10}F3 (Figure 2).[13] In addition to the overall conservation of chemical shifts and relative peak heights,

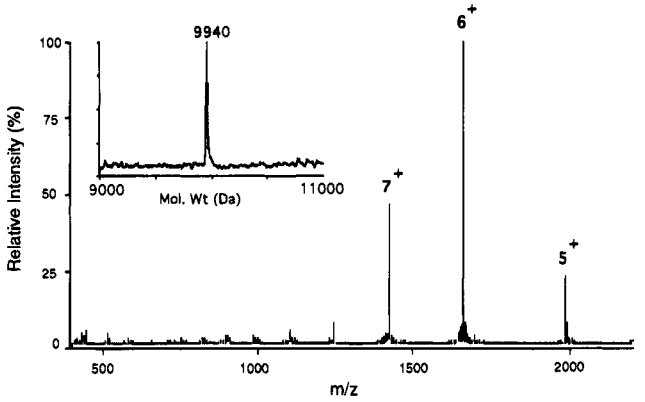

Figure 1. Characterization of purified [Ala43,(COS)$^{40-41}$]^{10}F3. Main panel: electrospray mass spectrum showing the distribution of charge states. Insert: hypermass reconstruction of the spectrum [Obsd MW 9939.84 ± 0.54; Calcd (average isotope composition) MW 9939.14].

* To whom correspondence should be addressed.

(1) Doolittle, R. F.; Bork, P. *Sci. Am.* **1993**, *268* (Oct), 50–56.

(2) Main, A. L.; Harvey, T. S.; Baron, M.; Boyd, J.; Campbell, I. D. *Cell* **1992**, *71*, 671–678. Baron, M.; Main, A. L.; Driscoll, P. C.; Mardon, H. J.; Boyd, J.; Campbell, I. D. *Biochemistry* **1992**, *31*, 2068–2073.

(3) Leahy, D. J.; Hendrickson, W. A.; Aukhil, I.; Erickson, H. P. *Science* **1992**, *258*, 987–991.

(4) Bazan, J. F. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6934–6938. Williams, A. F.; Barclay, A. N. A. *Annu. Rev. Biochem.* **1988**, *6*, 381–405.

(5) The amino acid sequence of ^{10}F3, VSDVPRDLEVVAATPTSLLISW-DAPAVTVRYYRITY GETGGNSPVQEFTVPGSKSTARISGLKPGVD-YTITVYAVTGRGDSPASSKPISINYRT, corresponds to residues 1416–1509 from human fibronectin, denoted as residues 1–94.[2]

(6) D'Souza, S. E.; Ginsberg, M. H.; Plow, E. F. *TIBS* **1991**, *16*, 46–250. Hynes, R. O. *Cell* **1992**, *69*, 11–25.

(7) Schnölzer, M.; Kent, S. B. *Science* **1992**, *256*, 221–225.

(8) Peptides were synthesized using a machine-assisted *in situ* neutralization/HBTU activation protocol for Boc solid-phase chemistry described previously.[9] The thioacid peptide ^{10}F3(1–40)$^{\alpha}$COSH was synthesized on a 4-[α-(Boc-Gly-S)-benzyl]phenoxyacetamidomethyl resin.[10] Boc-aminoacyl-OCH$_2$-Pam resins were used for all other peptides. The bromoacetyl group was coupled onto the peptide N$^{\alpha}$ as the preformed symmetrical anhydride. Peptides were deprotected and cleaved from the resin by liquid HF treatment with 4% *p*-anisole for 1 h at 0 °C. The crude peptide was precipitated and washed with diethyl ether, dissolved in aqueous acetic acid (50%), and lyophilized. To increase solubility, a few drops of hexafluoroacetone trihydride were first added to lyophilized ^{10}F3(1–40)$^{\alpha}$COSH. Peptides were purified by preparative reverse phase HPLC on a C18 column with linear gradients of CH_3CN in 0.1% TFA. Products were characterized both crude and after HPLC by ESMS using an API-III quadupole ion-spray instrument (Sciex, Toronto) as described.[11] In a typical ligation, 5 mL of cold reaction buffer (8 M urea, 50 mM NH_4OAc (pH 5.3)) was added to lyophilized ^{10}F3(1–40)$^{\alpha}$COSH (20.2 mg, 4.6 μmol, 1.0 equiv) and BrAc[Ala43](42–94)^{10}F3 (31 mg, 5.2 μmol, 1.2 equiv). Reactants were stirred for 80 h at 4 °C, and the reaction was monitored by reverse phase HPLC and ESMS. The target molecule [Ala^{43},(COS)$^{40-41}$]^{10}F3 was dialyzed against decreasing concentrations of urea (4 M → 2 M → 0) in the presence of NH_4OAc at pH 4.5. The analog was purified by preparative reverse phase HPLC using a linear gradient of 25–50% CH_3CN in 0.1% TFA over 60 min (yield: 30 mg, 3 μmol, 65%). Lyophilized [Ala^{43},(COS)$^{40-41}$]^{10}F3 adopted a folded conformation when redissolved in water up to concentrations of 2–3 mM at pH 4.0, without requiring further refolding. This was also the case for rec^{10}F3.[2]

(9) Schnölzer, M.; Alewood, P.; Jones, A.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180–193.

(10) Yamashiro, D.; Li, C. H. *Int. J. Pept. Protein Res.* **1988**, *31*, 22.

(11) Schnölzer, M.; Jones, A.; Alewood, P. F.; Kent, S. B. H. *Anal. Biochem.* **1992**, *204*, 335–343.

(12) Identification of this rearrangement and optimization of product yield are the subject of a second manuscript in preparation (Muir, T. W.; Williams, M. J.; Kent, S. B. H.).

(13) NMR spectra were acquired on a Bruker AMX500 series spectrometer, and FELIX 2.05 was used for data processing. Recombinant protein was repurified from the original NMR sample used to determine the ^{10}F3 structure.[2] Lyophilized samples were reconstituted in 0.5 mL of ultrapure D_2O, previously titrated to pH 4.0. Protein solutions of [Ala^{43},(COS)$^{40-41}$]^{10}F3 at 0.6 mM and rec^{10}F3 at 0.5 mM were titrated to pH 4.0 with NaOD (uncorrected meter readings), and NMR spectra were recorded at 39 °C using standard procedures.

0002-7863/94/1516-10797$04.50/0

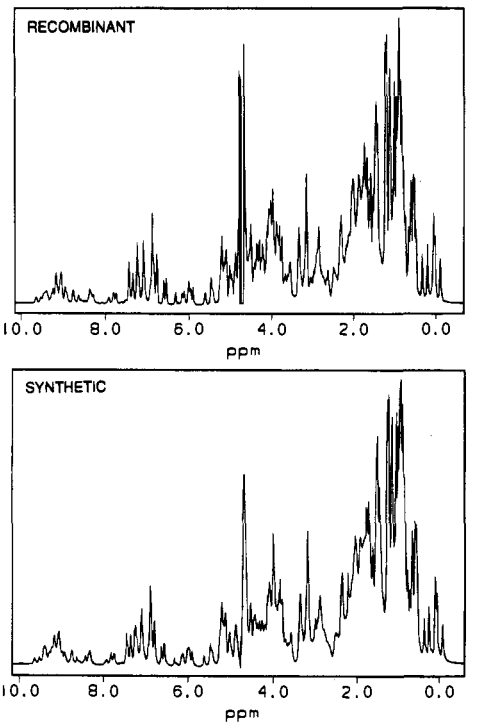

Figure 2. 1D ^{1}H NMR spectra of rec^{10}F3 (upper) and synthetic [Ala^{43},(COS)$^{40-41}$]^{10}F3 (lower). Broader line shapes in the spectrum of the synthetic protein can be attributed to small amounts of residual misfolded material, as line broadening was reduced through further HPLC purification. The thioester linkage was stable under NMR conditions.

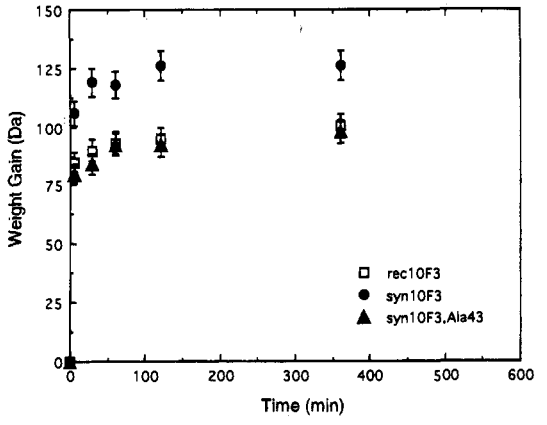

Figure 3. Determination of the folded state by proton exchange with deuterons in the bulk solvent for samples of [Ala^{43},(COS)$^{40-41}$]^{10}F3, rec^{10}F3, and [(COS)$^{40-41}$]^{10}F3 containing an O-linked chain at Ser^{43}. The plot shows the average of three experiments for each data point, with error bars corresponding to mean deviation over a data point set.

it is clear that both spectra have the same pattern of upfield-shifted methyl resonances above 0.5 ppm and numerous C$^{\alpha}$H resonances shifted downfield of the water resonance at 4.7 ppm. The methyl resonances are upfield shifted due to proximal aromatic residues within the protein core. While the C$^{\alpha}$H resonances are shifted downfield as a consequence of the β-strand backbone conformation. The fact that these regions of the spectra are conserved indicates that the proteins share the same global fold. Furthermore, both spectra exhibit similar patterns of peaks between approximately 7.7 and 9.7 ppm. These peaks correspond to backbone amide protons that exchange slowly with deuterons in the bulk solvent, due to their location within the core of the compact native structure where they are involved in hydrogen bonding with backbone carbonyl oxygens on proximal β-strands.[2]

NMR gives an indication of the average exchange rate for equivalent protons within all the protein molecules in a D_2O solution. However, it is unable to distinguish subpopulations of molecules in which equivalent protons possess different exchange rates, unless these give rise to unique resonances. ESMS, on the other hand, can identify differently folded subpopulations of molecules, when these gain mass at rates different from those of the natively folded protein in D_2O.[14] The rec^{10}F3 protein contains a maximum of 127 labile protons (85 backbone amide and 42 side chain protons) that could potentially exchange. Of these, Baron et al. identified 35 backbone amide protons by NMR that exchanged slowly (between ca. 1 h to several days) at 39 °C and pH 4.0,[2] all of which are present in the synthetic analogs. We prepared protein samples in D_2O under similar conditions and monitored the increase in protein mass, as deuterons exchanged with the labile protons (see Figure 3).[15] The weight gains for folded [Ala^{43},(COS)$^{40-41}$]^{10}F3 and rec^{10}F3 follow nearly identical profiles, confirming that they possess similar structural properties. [(COS)$^{40-41}$]^{10}F3 containing an O-linked chain at Ser^{43} was also isolated by reverse phase HPLC. Although this sample had the expected initial mass, the amide protons exchanged much more quickly (Figure 3). This indicates diminished proton protection due to the loss of the compact native fold. Significantly, only one species was identified in each of the three samples analyzed by this technique. Hence, rec^{10}F3 and [Ala^{43},(COS)$^{40-41}$]^{10}F3 are free of observable subpopulations of unfolded or misfolded material.

We now intend to modify the RGD site and analyze the folding characteristics of this synthetic protein module, through the synthesis of a range of chemical analogs. Moreover, this strategy provides an efficient method for the synthesis of other similar protein domains, opening up whole families of biologically important modular proteins to precise chemical investigation.

Acknowledgment. We thank Prof. Iain Campbell of Oxford University for providing rec^{10}F3. This work was supported by a grant from the Markey Foundation (S.B.H.K.) and NIH Grants HL28235 and HL31950 (M.H.G. and S.B.H.K). T.W.M. was supported by an Amgen Research Fellowship, and M.J.W. is the recipient of an Arthritis Foundation Fellowship.

(14) Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Science* **1993**, *262*, 896–900.

(15) Proteins were dissolved in D_2O (previously titrated to pH 4.0 with DCl) to a concentration of 1 mg/mL and left at 20 °C. Aliquots (20 μL) were removed periodically, and 10 μL of CH_3CN was added before immediate analysis by ESMS. Weight gain was calculated relative to the mass of each protein recorded in H_2O.