Secondary Structure of a Pair of Fibronectin Type 1 Modules by Two-Dimensional Nuclear Magnetic Resonance†

Michael J. Williams, Isabelle Phan, Martin Baron,§ Paul C. Driscoll, and Iain D. Campbell*

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Received March 29, 1993

ABSTRACT: The fourth and fifth type 1 module pair, corresponding to residues 151–244 from the amino terminus of human fibronectin, has been produced as a recombinant protein using a yeast expression system and studied by two-dimensional homonuclear ^{1}H nuclear magnetic resonance (NMR) spectroscopy. The sequence-specific resonance assignment of the ^{1}H NMR spectrum has been completed using a combination of 2D ^{1}H nuclear Overhauser effect (NOE) spectroscopy, homonuclear Hartmann–Hahn, and correlated spectroscopy spectra recorded under a variety of pH and temperature conditions. Slow exchanging amide protons have been identified and estimates of many backbone $^3J_{NH-C^{\alpha}H}$ coupling constants were obtained by line shape fitting. The secondary structures of each module conform closely to the "consensus" fibronectin type 1 module structure determined previously for two other single type 1 modules. In the module pair described here, the two modules are linked by a short five-residue linker which appears to form a turn. The intermodule interface is defined by NOEs observed between a hydrophobic three-residue sequence from the fourth type 1 module and residues in the first double-stranded β-sheet of the fifth type 1 module. The interaction is dominated by a tryptophan residue (unconserved in other type 1 sequences) within the fourth module, which causes large upfield ring current shifts for several proton resonances from the β-sheet of the fifth module. The NMR data indicate that there is little or no relative reorientation of the two modules about the linker region but rather that the two modules combine with a fixed and intimate hydrophobic contact.

Fibronectin is a large extracellular glycoprotein which is involved in numerous protein binding interactions and plays important roles in many diverse biological activities such as wound healing, cell adhesion, and cell migration (Mosher et al., 1989; Hynes, 1990; Schwarzbauer, 1991). The protein is found as a soluble dimer in the plasma or laid down as part of an insoluble matrix, closely associated with cellular surfaces. The amino acid sequence of the monomer is almost entirely made up of multiple copies of three types of protein module, called types 1, 2, and 3 (Petersen et al., 1983; Ruoslahti, 1988). The type 1 module is repeated 12 times in the fibronectin monomer. Six consecutive type 1 modules are located at the N-terminus, followed by two type 2 modules and a further three type 1 modules. A chain of 15–17 type 3 modules then leads into a final three type 1 modules at the C-terminus. So far only two other type 1 modules have been identified, one in tissue-type plasminogen activator (t-PA)[1] and the other in factor XII. Members of the type 1 family consist of about 45 residues and are identified by a conserved consensus sequence encoded on a single exon and bordered by phase 1 introns (Patel et al., 1987). The consensus sequence is characterized by a number of conserved hydrophobic residues, glycines, and four cysteine residues which form two intramodule disulfide bonds, connecting cysteines 1 to 3 and 2 to 4 in all the modules studied so far (Skorstengaard et al., 1982). The twelfth type 1 module from fibronectin is an exception and has two extra cysteines. A second anomaly of this particular module is that it is encoded on two exons rather than one (Patel et al., 1987).

The solution structures of the seventh type 1 module from fibronectin (Baron et al., 1991) and the unique type 1 module from t-PA (Downing et al., 1992) have previously been determined by ^{1}H NMR spectroscopy. The conserved "consensus" structure derived from these studies is dominated by two regions of antiparallel β-sheet. The amino terminus leads into a short double-stranded sheet which folds onto a larger triple-stranded sheet to enclose a hydrophobic core. The structure is constrained by the two conserved disulfide bonds. One disulfide links the first strand of the double-stranded β-sheet to the second strand of the triple-stranded β-sheet, and the other joins the second and third strands of the triple-stranded β-sheet.

It is now important to confirm that this consensus structure is conserved in larger fragments of human fibronectin and to gain an insight into how neighboring modules interact. The fibronectin polypeptide contains nine pairs of type 1 modules (denoted here in the form ^{n}F1-$^{n+1}$F1). The intermodule polypeptide linker regions, defined as the sequence found between the last conserved cysteine of one module and the first conserved cysteine of the following module, are of various lengths. For the module pairs ^{2}F1-^{3}F1, ^{4}F1-^{5}F1, ^{8}F1-^{9}F1, ^{10}F1-^{11}F1, and ^{11}F1-^{12}F1 the linker region comprises five amino acid residues. Six residues separate ^{3}F1-^{4}F1 and nine residues separate ^{1}F1-^{2}F1 and ^{7}F1-^{8}F1. The ^{5}F1 and ^{6}F1 modules are linked via a proteolytically sensitive stretch of 37 residues and

† This paper is a contribution from the Oxford Centre for Molecular Sciences, which is funded by the U.K. Medical Research Council and Science and Engineering Research Council. P.C.D. is a Royal Society University Research Fellow.

* Author to whom correspondence should be addressed.

§ Present Address: Boyer Center for Molecular Medicine, Yale University School of Medicine, 295 Congress Ave., New Haven, CT 06536.

[1] Abbreviations: ^{4}F1-^{5}F1, fourth and fifth type 1 module pair from human fibronectin; ^{4}F1, fourth type 1 module; ^{5}F1, fifth type 1 module; t-PA, tissue-type plasminogen activator; PCR, polymerase chain reaction; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlated spectroscopy; DQF-COSY, double-quantum-filtered correlated spectroscopy; P-COSY, primitive correlated spectroscopy; HOHAHA, homonuclear Hartmann–Hahn; ppm, parts per million; HPLC, high-performance liquid chromatography.

0006-2960/93/0432-7388$04.00/0

so probably do not interact closely as a pair. In the present study a recombinant protein, corresponding to the fourth and fifth type 1 modules (^{4}F1·^{5}F1) from the amino terminus of fibronectin, has been studied by two-dimensional homonuclear ^{1}H NMR spectroscopy. This module pair was selected due to the relatively short length of its linker and the considerable amount of data on biological activity concerning the five N-terminal type 1 modules of fibronectin. Here we report the ^{1}H resonance assignments and the secondary structure of the protein. The conformation of the linker and the location of the intermodule interface is also discussed with reference to the observed NOE connectivities and the unusual chemical shifts of protons found at the intermodule interface.

EXPERIMENTAL PROCEDURES

Expression and Purification. The expression and purification of ^{4}F1·^{5}F1 has been described in detail elsewhere (Williams et al., 1993). In brief, the 93-residue protein was produced using an expression system adapted from the yeast α-factor secretion pathway (Baron et al., 1990; Blake et al., 1984). A DNA fragment was amplified by the polymerase chain reaction (PCR) from a cloned fragment of human fibronectin cDNA, using a sense-strand oligonucleotide primer with the sequence 5′-GCT GAG AAG TGT TTT GAT CAT G and an antisense-strand oligonucleotide primer with the sequence 5′-T AAT GGA TCC TTA AGA GGT GTG CCT CTC ACA CT. The former corresponded to the N-terminal region of ^{4}F1·^{5}F1, and the latter to the C-terminus of ^{4}F1·^{5}F1 followed by a stop codon, a *Bam*HI site, and a four-residue tail, present to aid efficient restriction digestion. The PCR fragment was designed to encode from the 5′-intron, located between the G and C nucleotides of the first codon of the sense-strand primer (Patel et al., 1987) as far as the estimated end of the C-terminal β-strand, based on the type 1 module consensus structure. The PCR reaction was performed over 25 cycles, each one consisting of a 1-min DNA melting step at 94 °C, a 1-min annealing step at 56 °C, and a 3-minute extension step at 72 °C. After *Bam*HI digestion, the PCR fragment was ligated between unique *Stu*I/*Bam*HI restriction sites in the vector pMB50, downstream and in-phase with the α-factor leader sequence. The α-factor–^{4}F1·^{5}F1 combination was then excised as a *Bgl*II/*Bam*HI fragment and ligated into the single *Bgl*II site of the pMA91 yeast expression plasmid (Mellor et al., 1983). Competent yeast cells (*Saccharomyces cerevisae* MD50/a,α/*leu2,3*) were transformed with pMA91-^{4}F1·^{5}F1 and selected by their ability to grow on leucine-minus medium. Cultures (0.5 L) were grown in baffled flasks for 48–60 h at 30 °C in a rotary shaking incubator, using yeast nitrogen base medium without amino acids (Difco), supplemented with 0.5% (w/v) ammonium sulfate and 2% (w/v) glucose. Recombinant expression of the α-factor–^{4}F1·^{5}F1 fusion protein was directed by the phosphoglycerate kinase promoter and the α-factor leader sequence was cleaved off by the native yeast KEX2 cathepsin B-like peptidase during secretion (Kuryan & Herskiwitz, 1982) to leave the mature folded recombinant protein in the supernatant. This was isolated by binding to C-18 silica beads present during fermentation and, after elution, the recombinant protein was purified by reverse-phase HPLC. Its primary structure and purity were confirmed by electrospray mass spectrometry, performed on a VG Biotech BioQ atmospheric pressure quadrupole spectrometer equipped with an electrospray interface. Samples of ca. 20 pmol/1 μL were run from an acetonitrile/water (1:1) solution, containing 1% formic acid.

NMR Analysis. Protein samples of 2–3 mM were dissolved in 0.5 mL of D_2O or 90%H_2O/10%D_2O and the sample pH was titrated to either 4.5 or 5.5 (uncorrected meter reading) with NaOD and DCl. NMR spectra were acquired at 37 and 30 °C on Bruker AM500 (500.1 MHz) and AM600 (600.1 MHz) spectrometers. Data processing was performed using the FELIX 1.1 software package (Hare Research Inc.). All experiments were recorded in a phase-sensitive manner with time-proportional phase incrementation (TPPI) for quadrature detection in t_1. Two-dimensional nuclear Overhauser effect spectroscopy (NOESY) (Kumar et al., 1980) and homonuclear Hartmann–Hahn (HOHAHA) spectroscopy (Braunschweiler & Ernst, 1983; Davies & Bax, 1985) experiments were recorded with 2K real t_2 data points and 800–1024 t_1 increments. The receiver phase and preacquisition delay were optimized to reduce base-line distortions (Marion & Bax, 1988a). A WALTZ-17 sequence was used for HOHAHA experiments (Bax et al., 1987), with mixing times ranging from 40 to 50 ms. NOESY spectra were acquired with mixing times of 125 and 200 ms. A "jump–return" read pulse sequence was used for solvent suppression in NOESY and HOHAHA experiments collected in 90% H_2O solution, and the resulting free induction decays were deconvoluted in the time domain with a Gaussian function to remove the residual water signal (Marion et al., 1989). For all D_2O experiments, the residual solvent signal was suppressed by selective irradiation of the H_2O resonance. Double-quantum-filtered correlated (DQF-COSY) spectra (Rance et al., 1983) and primitive COSY (P-COSY) spectra (Marion & Bax, 1988b) were acquired in D_2O and 90% H_2O/10% D_2O solution, respectively, with 4K real t_2 data points and 512–1024 t_1 increments.

Data matrices from NOESY and HOHAHA experiments were processed in t_2 with a Lorentz–Gaussian multiplication (typically LB = –15 Hz and GB = 0.15) and a linear baseline correction. A 50–70° phase-shifted squared sine bell window was applied for apodization in t_1. DQF-COSY and P-COSY data sets were processed with an unshifted sine bell applied in both t_2 and t_1 time domains.

The coupling constants between amide and C^α protons ($^3J_{HN-C^\alpha H}$) were obtained from ω_2 cross sections of a P-COSY spectrum, zero-filled to a digital resolution of 1.72 Hz/point. Two cross sections were taken from the ω_2 dimension through each well-resolved NH–C^αH cross peak. These were simultaneously best-fitted to simulated line shapes that incorporated the effects of the experimentally applied zero-filling and apodization functions (Redfield & Dobson, 1990).

Amide NH protons which exchanged slowly with the solvent were identified by reconstituting a sample of ^{4}F1·^{5}F1, lyophilized from H_2O solution, in D_2O and recording a short (ca. 7 h) 2D HOHAHA spectrum.

RESULTS

Expression and Purification. Recombinant ^{4}F1·^{5}F1 was the major protein constituent of the elutant from the C18 silica beads and was purified by two reverse-phase HPLC separations on a C8 column. The derived molecular weight from mass spectrometry was sufficiently close to the calculated average value to confirm accurate N-terminal processing and formation of disulfide bonds. Typical yields of purified ^{4}F1·^{5}F1 were of the order of 1 mg/L of culture medium.

Assignment of NMR Spectra. For a 93-residue protein, the ^{4}F1·^{5}F1 peptide gave extremely well-resolved spectra with very little resonance overlap (Figures 1 and 2). Hence the assignment of amino acid spin systems and sequential connectivities, using well-established procedures (Wüthrich, 1986), was relatively straightforward. At a fairly early stage in the assignment process, it became clear that the pattern of

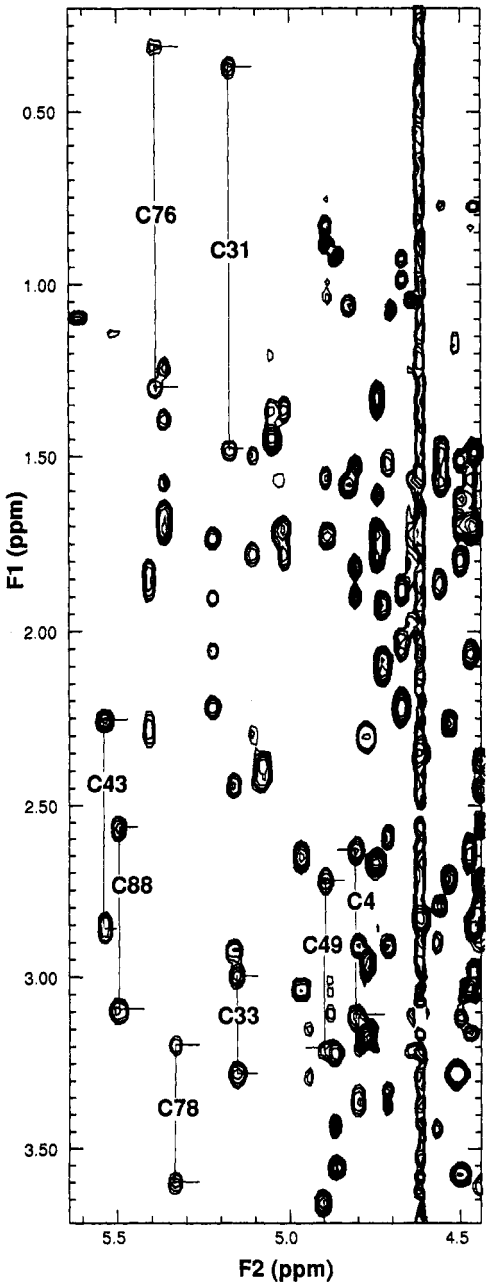

FIGURE 1: Region of a HOHAHA spectrum of ^{4}F1-^{5}F1, recorded in D_2O solution at 37 °C and pH 5.5. Resonances corresponding to $C^\alpha H$ protons lie parallel to the F2 axis and the $C^\alpha H$–$C^\beta H_2$ spin system assignments are shown for all eight cysteine residues. Note that protons corresponding to equivalent cysteines in the two modules have very similar chemical shifts.

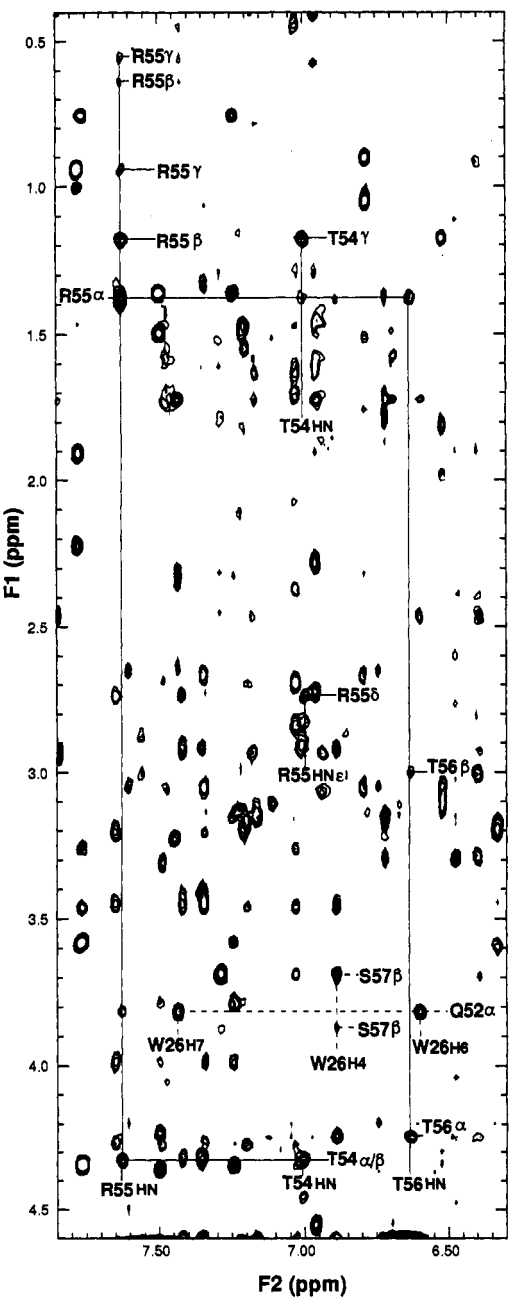

FIGURE 2: Portion of a NOESY spectrum of ^{4}F1-^{5}F1, recorded in 90% H_2O/10% D_2O at 37 °C and pH 4.5, with a mixing time of 200 ms. The sequential assignment of Arg55 is traced from the intraresidue NH–$C^\alpha H$ cross peak of T54 to the intraresidue NH–$C^\alpha H$ cross peak of T56. Dashed lines indicate a selection of four intermodule NOE connectivities involving connections of various ring protons of Trp26 to both the C^α proton of Gln52 and to the C^β protons of Ser57.

long-range NOEs was often repeated in both modules and closely mimicked the results obtained for the two previously studied single fibronectin type 1 modules. In particular, strong $C^\alpha H_i$–$C^\alpha H_j$ NOEs were observed between equivalent residues of all four modules. Thus, as the conservation of the type 1 secondary structure emerged for both ^{4}F1 and ^{5}F1, it became possible to confirm the sequential assignment of several residues by the presence of similar patterns of NOEs across equivalent β-sheets in both modules. Complete assignment of backbone C^α and amide protons was achieved, except for the exchange-labile N-terminal amino group of Ala 1. In addition, almost all side-chain proton resonances were assigned unambiguously. A list of chemical shift assignments is included in Table I.

Experimental Data. All NOE, slow-exchanging amide, and coupling constant data were obtained from spectra recorded at 310 K and pH 4.5. Interproton NOE distance restraints were collated from NOESY spectra with mixing times of 125 ms for samples in D_2O solution and 200 ms for samples in H_2O solution. These were classified as either strong, medium, or weak depending on the relative NOE cross-peak intensities.

In addition, 46 $^3J_{HN-C^\alpha H}$ coupling constants were determined and 12 backbone amide protons were found to exchange slowly with the bulk solvent. Slow solvent exchange is characteristic of amide NH protons involved in hydrogen-bonding interactions. Polypeptide backbone hydrogen-bond donor/acceptor pairs could be tentatively assigned by comparison to the secondary structure determined from the pattern of NOEs.

Table I: Assignment of Proton Resonances at 310 K and pH 4.5[a]

residue	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	others
Ala 1	–	4.15	1.49		
Glu 2	8.71	4.37	2.01*	2.27, 2.35	
Lys 3	8.02	5.07	1.48*	1.22*	δH 1.38*, εH 2.81*
Cys 4	9.03	4.85	3.14, 2.67		
Phe 5	9.12	4.77	2.70*		2,6H 7.02, 3,5H 7.18, 4H 7.19
Asp 6	8.25	4.80	3.16, 2.33		
His 7	8.76	4.28	3.33*		4H 7.44, 2H 8.54
Ala 8	8.34	4.24	1.51		
Ala 9	7.67	4.38	1.37		
Gly 10	8.09	3.99, 3.79			
Thr 11	7.26	4.35	3.58	0.76	
Ser 12	7.76	4.90	3.46, 3.26		
Tyr 13	8.31	4.58	2.82, 1.89		2,6H 6.23, 3,5H 5.78
Val 14	8.63	4.69	2.23	1.01, 0.95	
Val 15	7.76	3.60	1.92	0.95*	
Gly 16	9.13	4.44, 3.63			
Glu 17	8.41	4.47	2.47, 2.40	2.76, 2.61	
Thr 18	8.29	5.63	4.04	1.12	
Trp 19	8.61	4.98	3.30, 3.17		2H 6.71, 4H 6.45, 5H 6.39, 6H 6.17, 7H 6.72, NH 9.44
Glu 20	8.42	5.42	1.82, 1.78	2.34, 2.31	
Lys 21	8.17	5.07	1.72*	1.78*	δH 1.31*, εH 2.93*
Pro 22	–	5.32	2.23, 1.76	1.98, 1.93	δH 3.81, 3.77
Tyr 23	8.46	4.56	2.74, 2.30		2,6H 6.95, 3,5H 6.68
Gln 24	8.18	3.61	1.48, 1.44	1.59, 1.92	NH –
Gly 25	8.45	4.32, 3.41			
Trp 26	7.34	4.60	3.47, 2.92		2H 7.41, 4H 6.88, 5H 6.38, 6H 6.59, 7H 7.42, NH 10.24
Met 27	7.02	3.69	1.63, 1.28	0.42, 0.44	ϵCH_3 1.74
Met 28	7.29	5.13	1.59, 1.45	2.40, 2.24	ϵCH_3 1.79
Val 29	8.52	4.27	1.59	0.58, 0.42	
Asp 30	8.04	5.11	2.43, 2.42		
Cys 31	9.08	5.20	1.51, 0.40		
Thr 32	8.43	4.88	3.59	0.94	
Cys 33	8.03	5.18	3.30, 3.02		
Leu 34	9.17	4.58	1.61, 1.54	1.51	δCH_3 0.79*
Gly 35	8.51	3.98, 3.47			
Glu 36	8.29	4.09	2.08*	2.29, 2.17	
Gly 37	9.09	4.27, 3.75			
Ser 38	8.03	4.70	4.09, 3.87		
Gly 39	8.97	4.21, 3.91			
Arg 40	8.76	4.29	1.74, 1.61	1.65, 1.42	δH 3.16, 3.08, εNH 7.17
Ile 41	8.64	4.91	1.75	1.06, 1.02	γCH_3 0.92, δCH_3 2.02
Thr 42	8.28	4.75	3.90	1.09	
Cys 43	8.53	5.56	2.87, 2.28		
Thr 44	8.88	4.70	4.14	1.08	
Ser 45	8.74	4.73	4.00, 3.79		
Arg 46	8.40	4.06	1.65, 1.62	1.55, 1.52	δH 3.14*, εNH 7.11
Asn 47	8.65	4.99	3.05, 2.67		NH 7.38, 6.81
Arg 48	7.49	5.04	1.51*	1.59, 1.79	δH –, εH –
Cys 49	8.72	4.93	3.22, 2.75		
Asn 50	9.23	5.18	2.95, 2.47		NH 7.85, 7.18
Asp 51	7.96	4.74	2.94, 1.57		
Gln 52	8.85	3.84	2.36, 2.31	2.64, 2.62	NH –
Asp 53	8.79	4.47	2.91, 2.83		
Thr 54	7.01	4.34	4.34	1.18	
Arg 55	7.63	1.38	0.64, 1.19	0.58, 0.94	δH 2.73*, εNH 6.98
Thr 56	6.62	4.25	3.01	.0.20	
Ser 57	8.24	4.93	3.88, 3.68		
Tyr 58	8.59	4.38	3.18, 2.22		2,6H 6.21, 3,5H 5.75
Arg 59	9.16	4.67	1.89, 1.75	1.70, 1.59	δH 3.24*, εNH 7.44
Ile 60	8.05	3.44	1.70	0.98*	δCH_3 0.91, γCH_3 0.91
Gly 61	9.17	4.52, 3.59			
Asp 62	8.48	4.83	3.38, 2.94		
Thr 63	8.13	5.53	4.13	1.16	
Trp 64	8.28	4.91	3.13, 3.05		2H 6.51, 4H 6.10, 5H 6.32, 6H 6.15, 7H 5.90, NH 9.64
Ser 65	8.36	5.48	3.78*		
Lys 66	8.43	4.67	2.00, 1.27	1.08*	δH 1.82*, εH 3.07, 2.82
Lys 67	8.40	5.38	1.73, 1.70	1.60, 1.42	δH 1.28*, εH 2.93*
Asp 68	8.50	4.74	3.37, 2.62		
Asn 69	8.42	4.49	3.02, 2.88		NH 7.57, 6.87
Arg 70	8.21	4.49	2.09, 1.71	1.64, 1.59	δH 3.18*
Gly 71	8.00	4.27, 3.55			
Asn 72	8.79	4.64	2.86, 2.37		NH 8.74, 7.04
Leu 73	8.55	4.50	1.72*	1.51	δH 0.77, 0.72
Leu 74	9.21	4.78	1.53, 1.01	1.53	δH 0.77, 0.68
Gln 75	8.68	4.42	1.87, 1.69	2.05*	NH –
Cys 76	8.93	5.42	1.35, 0.35		
Ile 77	8.48	4.92	1.58	0.85, 0.78	γCH_3 0.91, δCH_3 1.43

Table I (Continued)

residue	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	others
Cys 78	8.41	5.35	3.62, 3.22		
Thr 79	8.78	4.54	4.20	1.19	
Gly 80	9.35	4.53, 3.31			
Asn 81	8.61	4.50	3.04, 2.66		NH 7.66, 6.74
Gly 82	9.18	4.15, 3.65			
Arg 83	7.87	4.82	1.84, 1.55	1.92*	δH 3.16*
Gly 84	8.65	4.04, 3.86			
Glu 85	8.53	4.28	1.83*	2.28, 2.10	
Trp 86	7.98	5.62	4.01, 3.46		2H 7.33, 4H 7.65, 5H 7.30, 6H 7.50, 7H 7.74, NH 10.16
Lys 87	8.44	4.79	1.82, 1.74	1.37*	δH 1.63*, εH 2.93*
Cys 88	8.87	5.54	3.11, 2.59		
Glu 89	8.83	4.71	2.06*	2.27, 1.91	
Arg 90	8.77	4.53	1.82, 1.72	1.63, 1.54	δH 3.14*, εNH 7.22
His 91	8.84	4.80	3.20, 3.00		2H 8.57, 4H 7.21
Thr 92	8.30	4.39	4.24	1.18	
Ser 93	7.97	4.26	3.85*		

[a] Chemical shifts are with respect to dioxane (3.74); asterisks indicate protons with degenerate chemical shifts and (–) indicates unassigned resonances.

DISCUSSION

Secondary Structure. The secondary structure of ^{4}F1·^{5}F1 has been defined by the pattern of short-, medium- and long-range backbone–backbone NOEs, spin–spin $^3J_{HN-C^{\alpha}H}$ coupling constants, and slow-exchanging amide NH protons (Wüthrich, 1986). These data (Figures 3 and 4) confirm that both type 1 modules consist of five β-strands connected by β-turns as expected from the consensus module structure.

Figure 4 shows a summary of the experimentally observed backbone–backbone NOEs and slowly exchanging amide NH protons, with the likely hydrogen-bonding pattern for the regions of regular β-sheet secondary structure. The highly similar secondary structure pattern observed for both modules highlights the level of structural conservation within the type 1 module family. The five extended strands of each module form an N-terminal short double-stranded β-sheet, followed by a larger triple-stranded β-sheet, as expected from the consensus type 1 module structure. In addition, the lengths of the equivalent regions of β-sheet shown in figure 4 are very similar and, in both modules, the respective strands are connected by turns which involve the same number of residues. This degree of homology reflects the fact that there are no insertions or deletions in the amino acid sequence of ^{4}F1 relative to ^{5}F1.

It is of interest to consider how the modules fit together, and one significant feature of the secondary structure of the pair is found in the linker region of the polypeptide. The β-strand conformation is only maintained for one or two residues following Cys43 at the end of ^{4}F1. It would appear that the carbonyl of Thr44 may be involved in forming a hydrogen bond with the amide proton of Asp30, suggesting that the β-sheet extends one residue into the linker. A $C^{\alpha}H_i$–$C^{\alpha}H_j$ NOE of medium strength is also observed between Ser45 and Val29, which confirms the close proximity of these two protons as expected in an antiparallel β-sheet. Figure 3 indicates that the linker region between residues Thr44 and Arg48 is characterized by sequential $C^{\alpha}H_i$–NH_{i+1} NOEs of smaller intensity than regions of β-sheet structure, sequential NH_i–NH_{i+1} connectivities, and several other medium-range backbone–backbone NOEs. In particular, a medium NH_i–NH_{i+1} NOE was observed between Arg46 and Asn47, a strong NH_i–NH_{i+1} NOE between Asn47 and Arg48, and a weak NH_i–NH_{i+2} NOE connectivity between Arg46 and Arg48. Hence it is most probable that the linker forms a four-residue turn, involving residues Ser46–Arg48. The end of this proposed turn would then lead directly into the first β-strand of ^{5}F1, with a medium-strength $C^{\alpha}H_i$–$C^{\alpha}H_j$ NOE connectivity observed between the terminal residue of the linker Arg^{48} and Arg^{59} (Figure 4) in the β-sheet.

Chemical Shifts. A notable feature of the spectra of ^{4}F1·^{5}F1 is the degree of chemical shift conservation for many conserved residues of the two modules, which provides a further indication of the level of similarity between the tertiary folds of the two type 1 modules. This observation is highlighted by the chemical shift index plot (Wishart et al., 1992) shown in Figure 5. The plot provides an independent assessment of secondary structure based on $C^{\alpha}H$ chemical shifts. An index value of +1 indicates a downfield shift of the $C^{\alpha}H$ resonance by more than 0.1 ppm compared to the random coil shift; an index value of –1 is for an upfield shift greater than 0.1 ppm. Stretches of three residues with an index value of +1 generally indicate regions of β-structure. The chemical shift index plot for ^{4}F1·^{5}F1 is mostly consistent with the deduced secondary structure shown in Figure 3. One discrepancy is found for strands B and B′ (see Figure 3 for strand notation) for which the conserved residues Y13 and Y58 display upfield chemical shifts, which may be due to an intraresidue ring current effect.

Figure 2 shows a region of a D_2O HOHAHA spectrum with the assignments of all eight cysteine $C^{\alpha}H$–$C^{\beta}H_2$ spin systems. It is clear that pairs of homologous cysteines, occupying equivalent positions in the consensus sequence of ^{4}F1 and ^{5}F1, have highly conserved patterns of proton chemical shifts. The assignment of the secondary conserved cysteine residue of both modules (Cys31 and Cys76) is especially interesting as these spin systems were not observed in spectra obtained for either the seventh fibronectin type 1 module or t-PA single type 1 module (Baron, 1990; Downing et al., 1992). Downing et al. speculated that the $C^{\beta}H_2$ resonances of this cysteine may be broadened due to slow internal motions of the molecule. This type of resonance broadening may be due to "flipping" or "wobbling" of proximal aromatic rings, and this cysteine is indeed found to be close to a conserved tryptophan residue in the type I consensus structure (Trp19 and Trp64; see Figure 6). The fact that both of the pairs of cysteine $C^{\beta}H$ resonances from ^{4}F1·^{5}F1 are shifted upfield from the random coil shifts by 1.5–2.0 ppm suggests that they do experience a large ring current effect from this proximal tryptophan. Thus the observation of these resonances with normal line widths in the NMR spectra of the ^{4}F1·^{5}F1 pair suggests that the internal motions present in the single modules are suppressed in the double module. It is possible that this observation reflects an increase in stability of the individual module folds within the module pair due to close interactions between the modules.

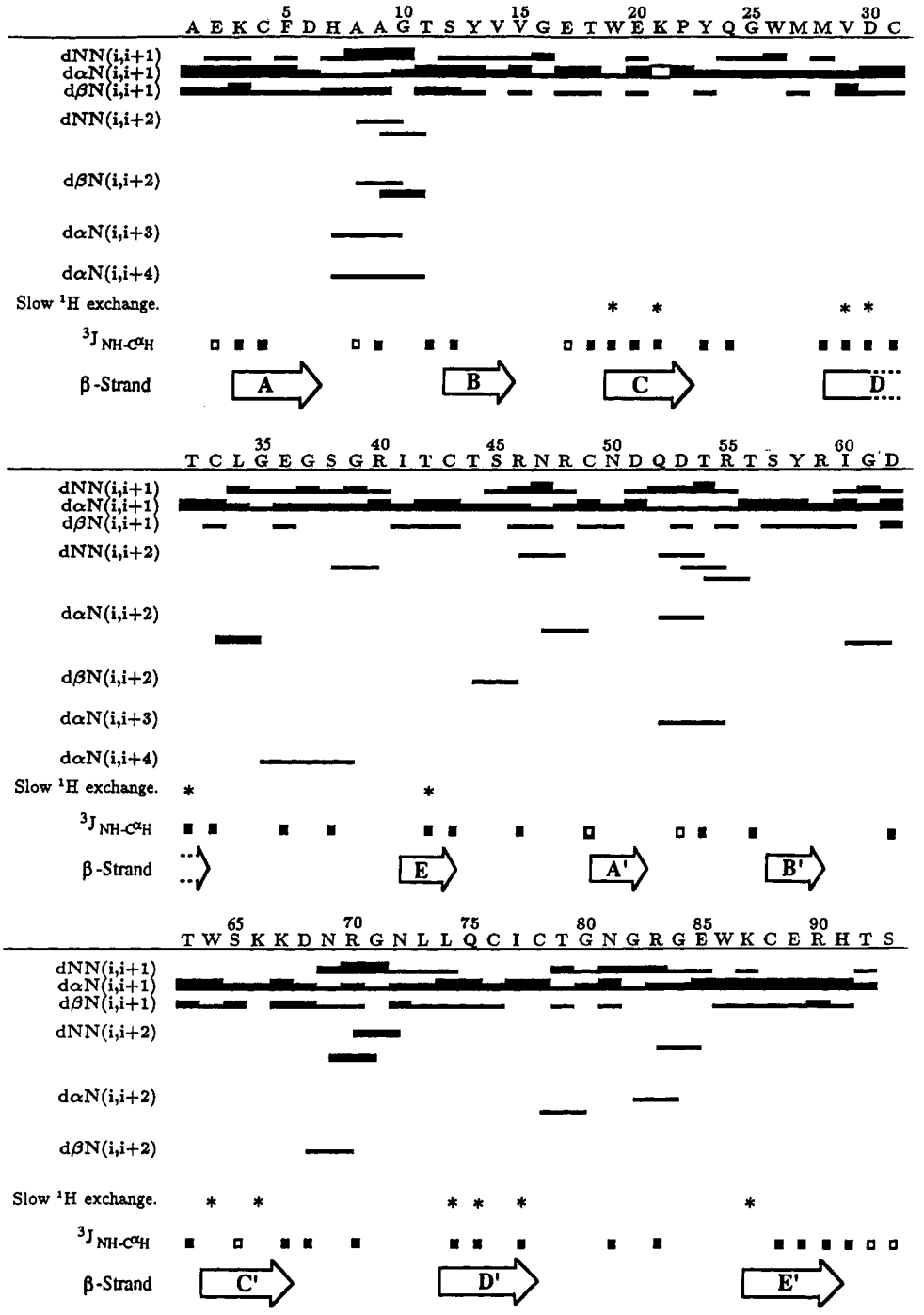

FIGURE 3: Summary of $^3J_{NH-C^\alpha H}$ coupling constants, slow-exchanging amides, and sequential and short-range NOEs for ^{4}F1·^{5}F1. Bars stretch from the residue *i* to residue *i*+*n* as indicated and the width of each bar corresponds to either a strong, medium, or weak NOE, categorized by the relative cross-peak intensity. A white bar is used for the NOE between C$^\alpha$H of Lys21 and the C$^\delta$H protons of Pro22. $^3J_{NH-C^\alpha H}$ coupling constants of 6–8 Hz are shown as open squares, and values >8 Hz are denoted as filled squares. Arrows correspond to approximate regions of β-strand deduced from the data, labeled A–E for the ^{4}F1 module and A′–E′ for the ^{5}F1 module.

Insight into the Intermodule Interface. Numerous intermodule NOEs have been observed between specific patches of side chains on each module in the ^{4}F1·^{5}F1 pair. These long-range NOE connectivities are represented schematically in Figure 6. The major interface is formed by the short hydrophobic Trp26–Met27–Met28 sequence, at the end of the third turn from ^{4}F1, which has many observable long-range NOE connectivities to residues Asn50–Gln52 and Arg55–Ser57 from the first β-sheet of ^{5}F1. A total of 58 NOEs have been observed between the two modules at this interface. The majority of these NOEs (four of which have been assigned and labeled in the spectrum shown in Figure 2) involve the nonconserved Trp26 residue. The NOEs observed between the indole protons of Trp26 and the aliphatic side chain of Arg55 are of particular interest as the C$^\alpha$H resonance of Arg55 is shifted upfield by approximately 2.6 ppm from the random coil shift position (Figure 2). To achieve this shift, it is necessary for the proton to be located above or below, and extremely close to the center of, the indole ring so that it can experience the maximum ring current effect (Perkins, 1982). Any small fluctuations in this relative orientation, on the NMR time scale, could cause significant changes in the degree of shielding and the C$^\alpha$H resonance would become broadened. The fact that all the protons of the Arg55 aliphatic chain form sharp, discrete resonances and

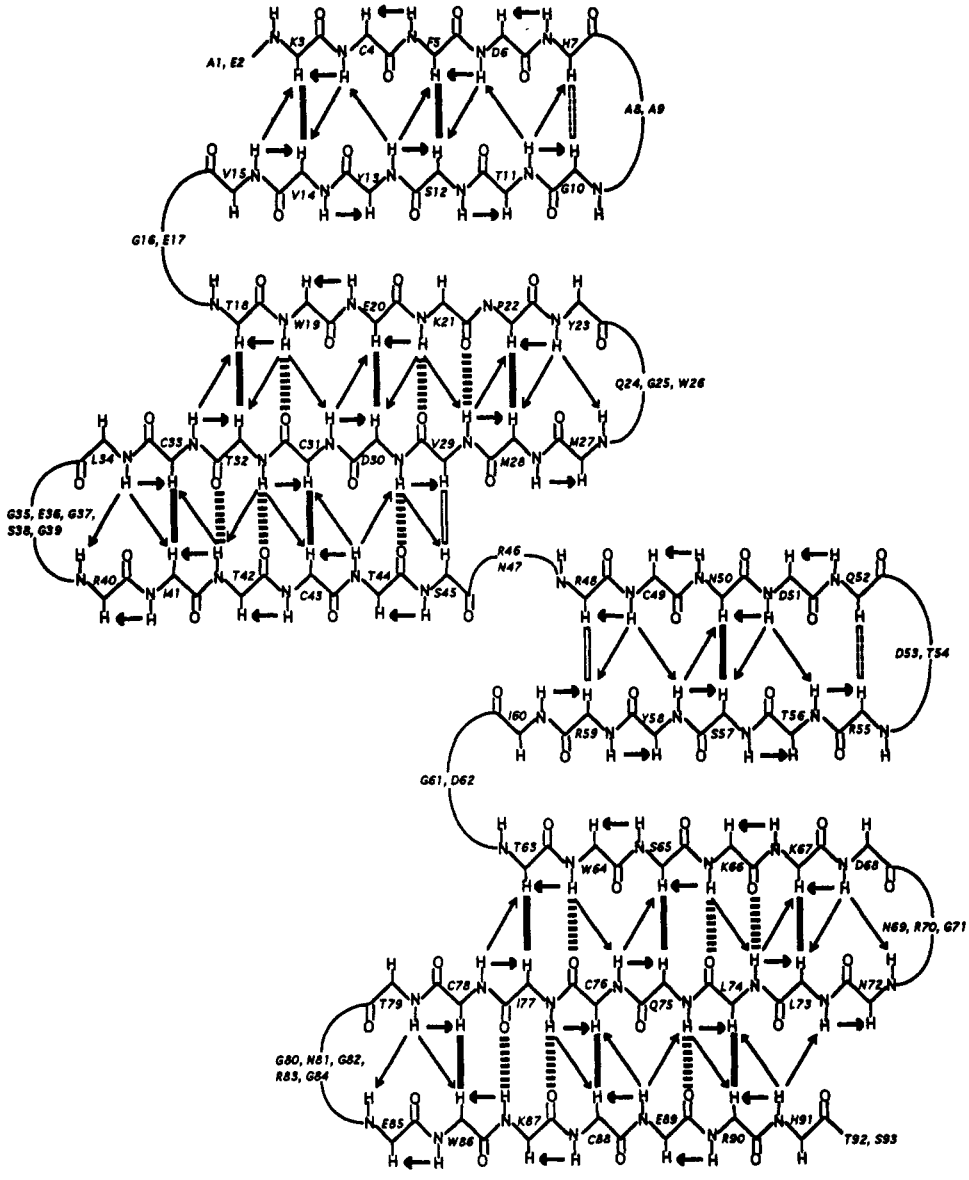

FIGURE 4: Secondary structure of ^{4}F1·^{5}F1 deduced from backbone–backbone NOEs and slow-exchanging proton data. Only the backbone of β-strand elements is traced out in detail. Sequential and interstrand NOEs are denoted by arrows or vertical bars. Vertical bars correspond to interstrand $C^{\alpha}H_i$–$C^{\alpha}H_j$ NOEs; closed bars are strong, open bars are medium, and dashed bars are weak NOEs. Slow-exchanging amide protons are connected to suitable hydrogen-bond acceptors via striped vertical lines.

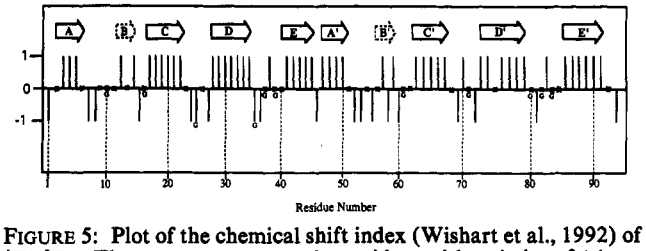

FIGURE 5: Plot of the chemical shift index (Wishart et al., 1992) of ^{4}F1·^{5}F1. Three or more consecutive residues with an index of +1 are indicative of a β-strand. These are shown as arrows, labeled A–E for the ^{4}F1 module and A′–E′ for the ^{5}F1 module. Strands B and B′ are shown as dotted arrows as they do not conform to the standard criteria. Crosses signify chemical shift index values of zero, and G's locate the positions of all glycine residues.

that the $C^{\alpha}H$ resonance display such a large upfield chemical shift suggests there is a fixed relative orientation of Arg55 and Trp26. As these two residues are from different modules in the ^{4}F1·^{5}F1 pair, this observation strongly implies that the two modules themselves adopt a fixed relative orientation, with little or no reorientation of the intermodule linker.

The conformation of the module linker region will play an important role in defining the relative orientation of modules within the ^{4}F1·^{5}F1 pair. The presence of a turn rather than an extended β-strand between the modules may well allow them to pack closer together. Several NOEs have been observed between the linker and the final turn of ^{5}F1 (separating strands D′ and E′ in Figure 3). These include two NOEs between Arg46 and Gly82, six NOEs between Arg46 and Arg83, and two NOEs between Asn47 and Gly80. The presence of a medium-strength NOE connectivity between the $C^{\alpha}H$ protons of Arg46 and Arg83 indicates that the polypeptide backbones come close together in this region. When these connectivities are taken into account with the backbone–backbone NOEs observed from residues at the two ends of the linker back to their parent modules, it would suggest that the linker is in a constrained conformation, reducing the degree of flexibility possible between the two modules.

It is clear that the secondary structures of both modules are well conserved and that the modules interact closely via a well-defined hydrophobic interface. Calculations of the three-dimensional solution structure of ^{4}F1·^{5}F1 is presently underway

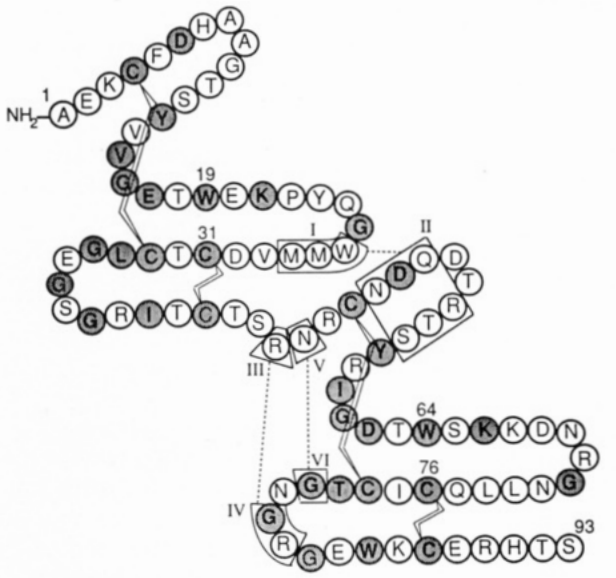

FIGURE 6: Schematic representation of the structure of ^{4}F1·^{5}F1 based on the consensus type 1 structure and the data presented here. Heavily shaded residues correspond to conserved type 1 consensus residues. Arrows link boxed residues that are involved in intermodule or important linker NOE connectivities. Numerous NOEs were observed between the hydrophobic tripeptide Trp26–Met27–Met28 (box I) and the six residues in box II, consisting of Asn51–Gln52 and Arg55–Ser57. Linker NOEs were observed between Arg46 (box III) and Gly82–Arg83 (box IV) and between Asn47 (box V) and Gly82 (box VI). The C$^{\beta}$H protons of Cys31 and Cys76 experience large ring current shifts due to the proximity of Trp19 and Trp64, respectively.

and will provide a more detailed insight into how type 1 modules link up together to form the larger groups of consecutive type 1 modules found in the fibronectin polypeptide. The structure will also give valuable information concerning the surface-accessible residues, which could be available to form part of potential binding sites for the wide range of protein ligands known to bind to the amino-terminal region of fibronectin.

ACKNOWLEDGMENT

We thank Robin Aplin for the mass spectrometry analysis of ^{4}F1·^{5}F1 and Helen Mardon for providing the fibronectin cDNA.

REFERENCES

Baron, M. (1990) D. Phil. Thesis, Oxford University, Oxford, England.

Baron, M., Kingsman, A. J., Kingsman, S. M., & Campbell, I. D. (1990) *Protein Production in Biotechnology* (Harris, T. J. R., Ed.), pp 49–60, Elsevier, London.

Baron, M., Norman, D. G., Willis, A. & Campbell, I. D. (1991) *Nature 345*, 642–646.

Bax, A., Sklenar, V., Clore, G. M., & Gronenborn, A. M. (1987) *J. Am. Chem. Soc. 109*, 6511–6513.

Blake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F., R., Mullenbach, G. T., Urdea, M. S., Valenzuela, P., & Barr, P. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4642–4646.

Braunschweiler, L. R., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Davies, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.

Downing, A. K., Driscoll, P. C., Harvey, T. S., Dudgeon, T. J., Smith, B. O., Baron, M., & Campbell, I. D. (1992) *J. Mol. Biol. 225*, 821–833.

Hynes, R. O. (1990) *Fibronectins*, Springer Verlag, New York.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.

Kuryan, J., & Herskiwitz, I. (1982) *Cell 30*, 933–943.

Marion, D., & Bax, A. (1988a) *J. Magn. Reson. 79*, 352–356.

Marion, D., & Bax, A. (1988b) *J. Magn. Reson. 80*, 528–533.

Marion, D., Ikura, M., & Bax, A. (1989) *J. Magn. Reson. 84*, 425–430.

Mellor, J., Dobson, M. J., Roberts, N. A., Tuite, M. F., Emtage, J. S., White, S., Lowe, P. A., Patel, T., Kingsman, A. J., & Kingsman, S. M. (1983) *Gene 24*, 1–14.

Mosher, D. F. (1989) *Fibronectin*, Academic Press, New York.

Patel, R. S., Odermatt, E., Schwarzbauer, J. E., & Hynes, R. O. (1987) *EMBO J. 6*, 2565–2572.

Perkins, S. J. (1982) *Biol. Magn. Reson. 4*, 193–255.

Petersen, T. E., Thogersen, H. C., Skorstengaard, K., Vibe-Petersen, K., Sahl, P., Sottrup-Jensen, L., & Magnusson, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 137–141.

Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 177*, 479–485.

Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201–7214.

Ruoslahti, E. A. (1988) *Annu. Rev. Biochem. 57*, 375–413.

Schwarzbauer, J. E. (1991) *Curr. Opin. Cell Biol. 3*, 786–791.

Skorstengaard, K., Thogersen, H. C., Vibe-Petersen, K., Petersen, T. E., & Magnusson, S. (1982) *Eur. J. Biochem. 161*, 441–453.

Williams, M. J., Phan, I., Baron, M., Aplin, R., & Campbell, I. D. (1993) *Techniques in Protein Chemistry IV*, Academic Press, San Diego (in press).

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647–1651.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.