Solution Structure of the N-Terminal F1 Module Pair from Human Fibronectin[†,‡]

Jennifer R. Potts, Jeremy R. Bright, David Bolton, Andrew R. Pickford, and Iain D. Campbell*

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

Received January 27, 1999; Revised Manuscript Received May 4, 1999

ABSTRACT: Multiple sites within the N-terminal domain ($^{1-5}$F1) of fibronectin have been implicated previously in fibronectin matrix assembly, heparin binding, and binding to cell surface proteins of pathogenic bacteria. The solution structure of ^{1}F1^{2}F1, the N-terminal F1 module pair from human fibronectin, has been determined using NMR spectroscopy. Both modules in the pair conform to the F1 consensus fold. In ^{4}F1^{5}F1, the only other F1 module pair structure available, there is a well-defined intermodule interface; in ^{1}F1^{2}F1, however, there is no detectable interface between the modules. Comparison of the backbone ^{15}N-{^{1}H} NOE values for both module pairs confirms that the longer intermodule sequence in ^{1}F1^{2}F1 is flexible and that the stabilization of the ^{4}F1 C−D loop observed in ^{4}F1^{5}F1, as a result of the intermodule interface, is not observed in ^{1}F1^{2}F1.

Fibronectin is a large extracellular glycoprotein that exists in the plasma as a soluble disulfide-linked dimer and in an insoluble multimeric form as a major component of the extracellular matrix. Fibronectin is a mosaic protein with each monomer being composed almost entirely of three types of protein module. Fibronectin type 1, 2, and 3 (F1, F2, and F3) modules were first identified in fibronectin but have since been found in a number of other proteins. Structural studies of recombinant modules have identified a well-defined consensus three-dimensional fold for each of the three module types (*1*−*3*).

Consistent with fibronectin's involvement in many vital physiological processes (for review see ref *4*), binding sites for a wide range of molecules have been identified along the length of the fibronectin monomer. Fibronectin modules are organized into a number of proteolytically resistant domains that retain affinity for the various ligands, and in many cases, attempts to localize binding sites more precisely within these domains have shown that module pairs or larger fragments are required for optimal binding activity (*5*).

The 29 kDa N-terminal domain, a frequently studied proteolytic fragment of fibronectin, contains a short sequence (residues 1−18), containing a covalent binding site for fibrin, which is followed by a string of five F1 modules. This domain is involved in a number of important physiological processes; in particular, it is vital for incorporation of fibronectin into the extracellular matrix (*6*). Binding sites within $^{1-5}$F1, for ^{1}F3 (*7*) and $^{12-14}$F3 (*8*) of fibronectin, and for an as yet unidentified cell surface receptor (*7*) have been proposed to play important roles during fibronectin matrix assembly. $^{1-5}$F1 also contains binding sites for fibrin, heparin, and cell surface proteins of pathogenic bacteria such as *Staphylococcus aureus* and *Streptococcus dysgalactiae*. Bacterial binding via these cell surface proteins (or MSCRAMMs) to extracellular matrix proteins such as fibronectin is a putative mechanism for host invasion by the microorganism (for a review see ref *9*). The fibronectin-binding MSCRAMMs share a similar structural organization, with the ligand binding region consisting of repeated units (3−5 times) of 40−50 residues.

Much of the fibrin (*10*, *11*) and ^{1}F3 (*7*) binding activity of $^{1-5}$F1 has been further localized to the ^{4}F1^{5}F1 module pair, and these interactions seem to require the module pair, rather than the isolated ^{4}F1 or ^{5}F1 module. Studies using synthetic peptides from fibronectin-binding MSCRAMMs have also located binding sites for ^{4}F1^{5}F1 in each of three repeats of an *S. aureus* protein (*12*) and for ^{1}F1^{2}F1 and ^{2}F1^{3}F1 in a single repeat from *S. dysgalactiae* (*13*). However, the intact $^{1-5}$F1 domain, as well as being required for matrix assembly, also appears to be necessary for optimal binding to *S. aureus* (*14*). Sites in ^{1}F1, ^{2}F1, and ^{3}F1 have also been implicated in heparin binding (*15*). The relative orientations of F1 modules in intact $^{1-5}$F1 may play a role in all these processes.

In ^{4}F1^{5}F1, the only F1 module pair structure previously available, the relative orientation of the two modules is determined by a well-defined intermodule interface (*11*). In the present work high-resolution NMR spectroscopy is used to study the ^{1}F1^{2}F1 module pair and to compare the backbone dynamics of the ^{1}F1^{2}F1 and ^{4}F1^{5}F1 module pairs.

MATERIALS AND METHODS

Expression and Purification. Uniformly ^{15}N-labeled ^{1}F1^{2}F1 (mature human fibronectin residues 17−109) and ^{4}F1^{5}F1 (residues 152−244) were expressed in *Pichia pastoris*. Construction of vectors, transformation of *P. pastoris*, and expression were carried out using procedures similar to those described previously (*3*). For ^{1}F1^{2}F1, ^{15}N-labeling was achieved by using 0.2% (w/v) $(^{15}NH_4)_2SO_4$ as the sole

[†] This research was supported by the Wellcome Trust, the Biotechnology and Biological Sciences Research Council, the Medical Research Council, and the Engineering and Physical Sciences Research Council. This is a contribution from the Oxford Centre for Molecular Sciences.

[‡] The coordinates for the family of NMR structures have been deposited with the Protein Data Bank and have the ID code 1qgb.

* Corresponding author. Telephone: 01865 275 346. Fax: 01865 275 253. Email: idc@bioch.ox.ac.uk.

10.1021/bi990202b CCC: $18.00
Published on Web 06/11/1999

nitrogen source in the growth and induction phase. Expression of ^{4}F1^{5}F1 was carried out in a 1 L fermentor from Electrolab Ltd. (Tewkesbury, U.K.), using 1.0% (w/v) $(^{15}NH_4)_2SO_4$ as described elsewhere (*16*).

^{1}F1^{2}F1 and ^{4}F1^{5}F1 were purified using procedures similar to those described previously (*3*). However, after harvesting the culture supernatant, the initial chromatography step utilized cation exchange (20 mM sodium acetate, pH 4.8) on SP-Sepharose Fast Flow (Pharmacia), rather than anion exchange. Prior to cation exchange, the supernatant containing ^{1}F1^{2}F1 was filtered (0.2 μm), concentrated, and diafiltrated into the cation-exchange buffer. The supernatant containing ^{4}F1^{5}F1 was adjusted to pH 3 and filtered (0.2 μm), prior to cation exchange. Elution was achieved with a 0−1 M NaCl or 0−0.5 M NaCl gradient for ^{1}F1^{2}F1 and ^{4}F1^{5}F1, respectively. After further purification by reverse-phase HPLC, the identity and purity (>95%) of ^{1}F1^{2}F1 and ^{4}F1^{5}F1 were confirmed by electrospray mass spectrometry and N-terminal sequence analysis.

NMR Spectroscopy. A 2.2 mM solution of ^{15}N-labeled ^{1}F1^{2}F1 in 90% H_2O/10% D_2O was adjusted to pH 4.0. NMR spectra were acquired at ^{1}H frequencies of 500.1, 600.1, and 750.1 MHz on spectrometers built in-house at the Oxford Centre for Molecular Sciences and incorporating Oxford Instruments Magnets and GE/Omega computers. Spectra were acquired at 37 °C unless stated otherwise. For the assignment and structure determination the following spectra were recorded in H_2O: a 2D ^{1}H-^{1}H NOESY (100 ms mixing time) (*17*) and a 2D ^{1}H-^{1}H TOCSY (42 ms mixing time) (*18*, *19*) incorporating a DIPSI3 mixing sequence (*20*, *21*) and both incorporating a 1-1 read pulse for solvent suppression (*22*), a 2D ^{1}H-^{1}H COSY (*23*), and a 3D ^{15}N-filtered gradient-enhanced NOESY−HSQC (100 ms mixing time) (*24*, *25*). The sample was then exchanged into D_2O, and the following 2D homonuclear spectra were acquired: a DQF-COSY (*26*), a TOCSY incorporating a DIPSI2 mixing sequence (38 ms mixing time), and four NOESY spectra (50, 100, 150, and 200 ms mixing times) at 37 °C and a TOCSY (38 ms mixing time) and a NOESY (100 ms mixing time) at 25 °C. ^{15}N decoupling during acquisition was carried out using a GARP1 decoupling sequence (*27*). $^3J_{NH-C\alpha H}$ coupling constants were measured by line-shape fitting of ω_1 traces from a ^{15}N-^{1}H HMQC-J spectrum (*28*). Slowly exchanging backbone amide protons were identified by lyophilizing ^{1}F1^{2}F1 from H_2O, redissolving in D_2O, and recording two TOCSY spectra (38 ms mixing time) at 25 °C.

To compare the backbone dynamics of ^{1}F1^{2}F1 and ^{4}F1^{5}F1, the steady-state ^{15}N-{^{1}H} NOE was measured for all residues with well-resolved ^{1}H-^{15}N cross-peaks. A 1.6 mM solution of ^{15}N-labeled ^{4}F1^{5}F1 in 95% H_2O/5% D_2O was adjusted to pH 4.6. The ^{15}N-labeled ^{1}F1^{2}F1 sample was as described above. The pulse sequence was based on sequences previously described (*29*) and incorporated gradient enhancement (*25*). Experiments were performed at a ^{15}N frequency of 76.0 MHz. A relaxation delay of 5 s was used to ensure complete solvent recovery. For the experiment with NOE, ^{1}H saturation was applied during the relaxation delay. A WALTZ16 decoupling sequence (*30*) was applied for ^{15}N decoupling during acquisition. For the assignment of Ser45 of ^{15}N-labeled ^{4}F1^{5}F1, 2D TOCSY and NOESY spectra were acquired at 37 °C.

Data Processing and Analysis. Data processing was performed using the FELIX 2.3 software package (Biosym Technologies Inc.) on Sun workstations. Homonuclear NOESY, TOCSY, COSY, and DQF-COSY experiments were processed as described previously (*3*). The 3D NOESY−HSQC was processed using linear prediction in t_1 from 96 to 128 points prior to apodization using a Lorentzian−Gaussian function in t_2 and a Kaiser function in t_1 and t_3. ^{15}N-{^{1}H} NOE spectra were processed using methods similar to those described previously (*31*). The HMQC-J spectrum was processed by multiplication with a Lorentzian−Gaussian function in t_2 and using linear prediction in t_1 from 256 to 384 points.

For ^{1}F1^{2}F1, spectral assignment and the extraction of NOE intensities for the structure calculations were performed with the aid of the program NMRView (Merck and Co. Inc) on an SGI Indigo2 workstation. Spectra of ^{1}F1^{2}F1 were assigned using established techniques for 2D ^{1}H-^{1}H spectra (*32*). Additional assignments were obtained using a 3D ^{1}H-^{15}N HSQC−NOESY spectrum.

Assignments for the ^{4}F1^{5}F1 ^{15}N-{^{1}H} NOE spectra were from Phan et al. (*31*), with the exception of Ser45 which was assigned using 2D NOESY and TOCSY spectra of ^{15}N-labeled ^{4}F1^{5}F1 and previously reported homonuclear chemical shifts (*33*).

Structure Calculations. Interproton distances were estimated from the relative intensities of cross-peaks (NOEs) in NOESY spectra acquired with a 100 ms mixing time. This mixing time was chosen on the basis of an NOE build-up curve (data not shown) drawn using the intensities of NOEs in NOESY spectra acquired using mixing times of 50, 100, 150, and 200 ms. NOE intensities were calibrated using known interproton distances in regions of secondary structure (*34*) and classified as strong (≤2.8 Å), medium (≤3.5 Å), or weak (≤5.0 Å). Intraresidue NOEs between protons on adjacent carbon atoms were not included as restraints. Hydrogen bond restraints were included in the later stages of the structure calculations and only when three criteria were met: slow exchange of the corresponding amide proton and an NH···O distance of ≤2.3 Å and an N−H···O angle of ≥120° in >70% unrestrained structures (*35*). For each hydrogen bond, two restraints were added: 1.58 Å < d_{HN-O} < 2.3 Å and 1.58 Å < d_{N-O} < 3.2 Å. Estimates of ϕ backbone dihedral angles were obtained from $^3J_{NH-C\alpha H}$ using the Karplus equation (*36*) and using initial structures to eliminate ambiguity. ϕ angle restraints were added with errors of ±40°.

^{1}F1^{2}F1 structures were calculated from the experimental restraints with simulated annealing using X-PLOR v3.851 (*37*) with v4.02 and v4.01 parameter and topology files, respectively (kindly provided by Michael Nilges). A total of 100 structures were calculated using the *ab initio* simulated annealing protocol with 60 ps of high-temperature (2000 K) restrained molecular dynamics, followed by two cooling stages of 40 and 20 ps to a final temperature of 100 K. The nonbonded energy was calculated using a purely repulsive function with a final value of the van der Waals radii scaled by a REPEL value of 0.75. An iterative procedure was used; NOEs which could not be assigned unambiguously were included as ambiguous restraints during initial structure calculations (*38*) followed, where possible, by resolution of the ambiguity by inspection of preliminary structures. The

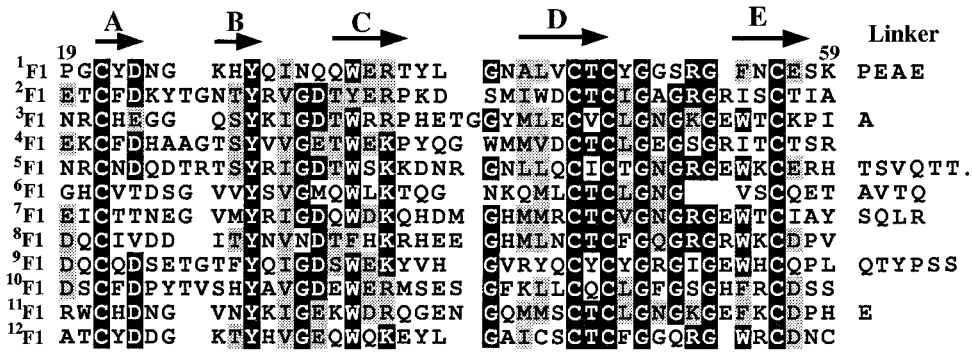

FIGURE 1: Sequence alignment of the 12 F1 modules from fibronectin. The definition of module and linker residues is as described in the Results and Discussion section. The sequence alignment was performed with the assistance of the programs PILEUP (Genetics Computer Group Inc.) and BOXSHADE. Residues conserved in >50% of the sequences are highlighted with a black box, and conservative substitutions are highlighted with a gray box. The secondary structure of ^{1}F1 is represented above the sequence alignment. For clarity, residue numbering for ^{1}F1 is indicated.

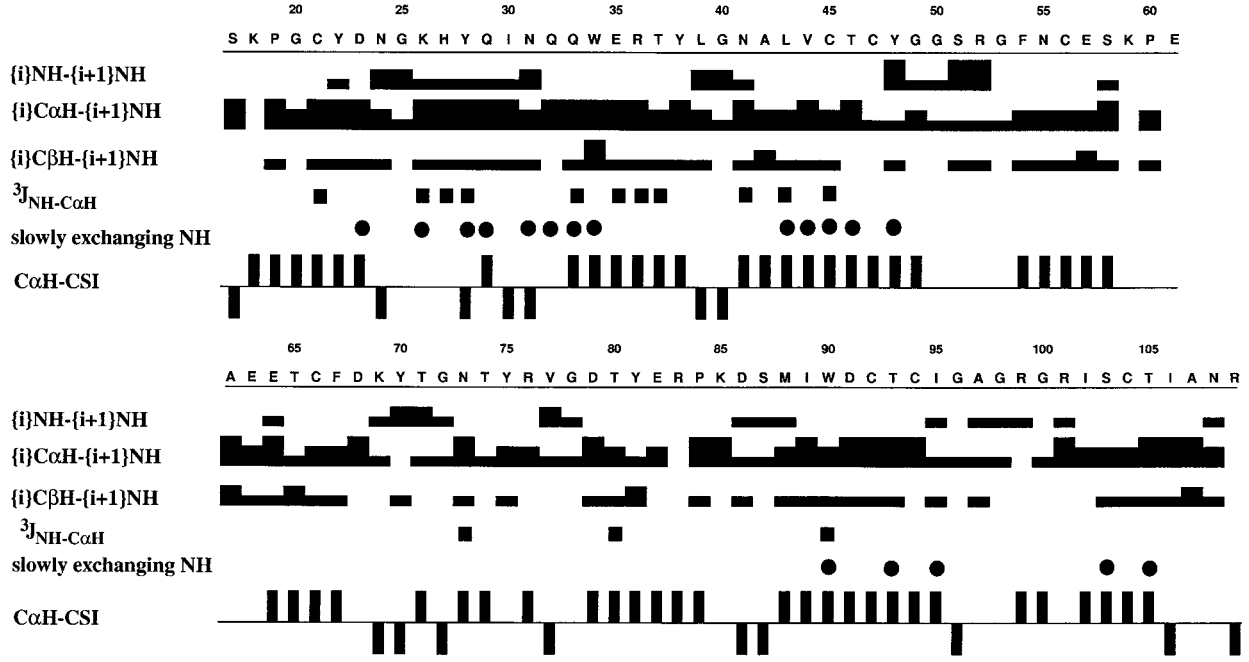

FIGURE 2: Summary of NH(i)−NH(i+1), CαH(i)−NH(i+1), and CβH(i)−NH(i+1) NOEs, $^3J_{NH-C\alpha H}$, slowly exchanging amide protons, and chemical shift index (CSI) (*41*) observed in spectra of ^{1}F1^{2}F1. The relative strengths of the NOEs (strong, medium, or weak) are indicated by the width of the horizontal bars. Values of $^3J_{NH-C\alpha H}$ >8 Hz are indicated by a filled square and slowly exchanging amide protons by a filled circle. Positive or negative values of the CSI were calculated using CαH chemical shifts from Wishart et al. (*47*) and are indicated by a solid vertical bar above or below the line, respectively.

protocol utilized "sum averaging" and "floating stereospecific assignment" of valine and leucine methyl carbons and methylene protons (*39*). The structures were refined using an additional cycle of simulated annealing. A total of 24 refined structures were chosen on the basis of their low overall energy and agreement with the experimental restraints. The average structures and rmsd for ^{1}F1 and ^{2}F1 were calculated separately by superimposing over the backbone heavy atoms (N, Cα, C) of secondary structure elements of ^{1}F1 and ^{2}F1, respectively. Geometric strain was removed from the average structures by restrained energy minimization in X-PLOR.

Calculation of ^{15}N-{^1H} NOE. The steady-state ^{15}N-{^{1}H} NOE was measured for the backbone ^{15}N nuclei of ^{1}F1^{2}F1 and ^{4}F1^{5}F1. The NOE was calculated as the ratio of resonance intensities in spectra collected with and without ^{1}H saturation. A number of residues were not included in the analysis due to spectral overlap. Thus, the heteronuclear NOE was measured for 78 out of 90 non-proline residues in ^{1}F1^{2}F1 and for 80 out of 92 non-proline residues in ^{4}F1^{5}F1. Errors on ^{15}N-{^{1}H} NOE values were estimated from a Monte Carlo simulation, which used one standard deviation of baseline noise.

RESULTS AND DISCUSSION

Resonance Assignments. Complete assignment of the backbone ^{1}H and ^{15}N and most of the side chain ^{1}H of ^{1}F1^{2}F1 was achieved. In some cases overlap of peaks precluded specific assignment of side chain resonances. The assignments are provided in Supporting Information.

The structure of ^{1}F1 as a single module (fibronectin residues 1−61, referred to below as ^{1}F1s) was determined previously in this laboratory (*40*). The ^{1}F1s construct included residues 1−18, which are N-terminal to the ^{1}F1 module in native fibronectin. For residues 18−61, the chemical shifts of the assigned carbon-attached protons (backbone and side

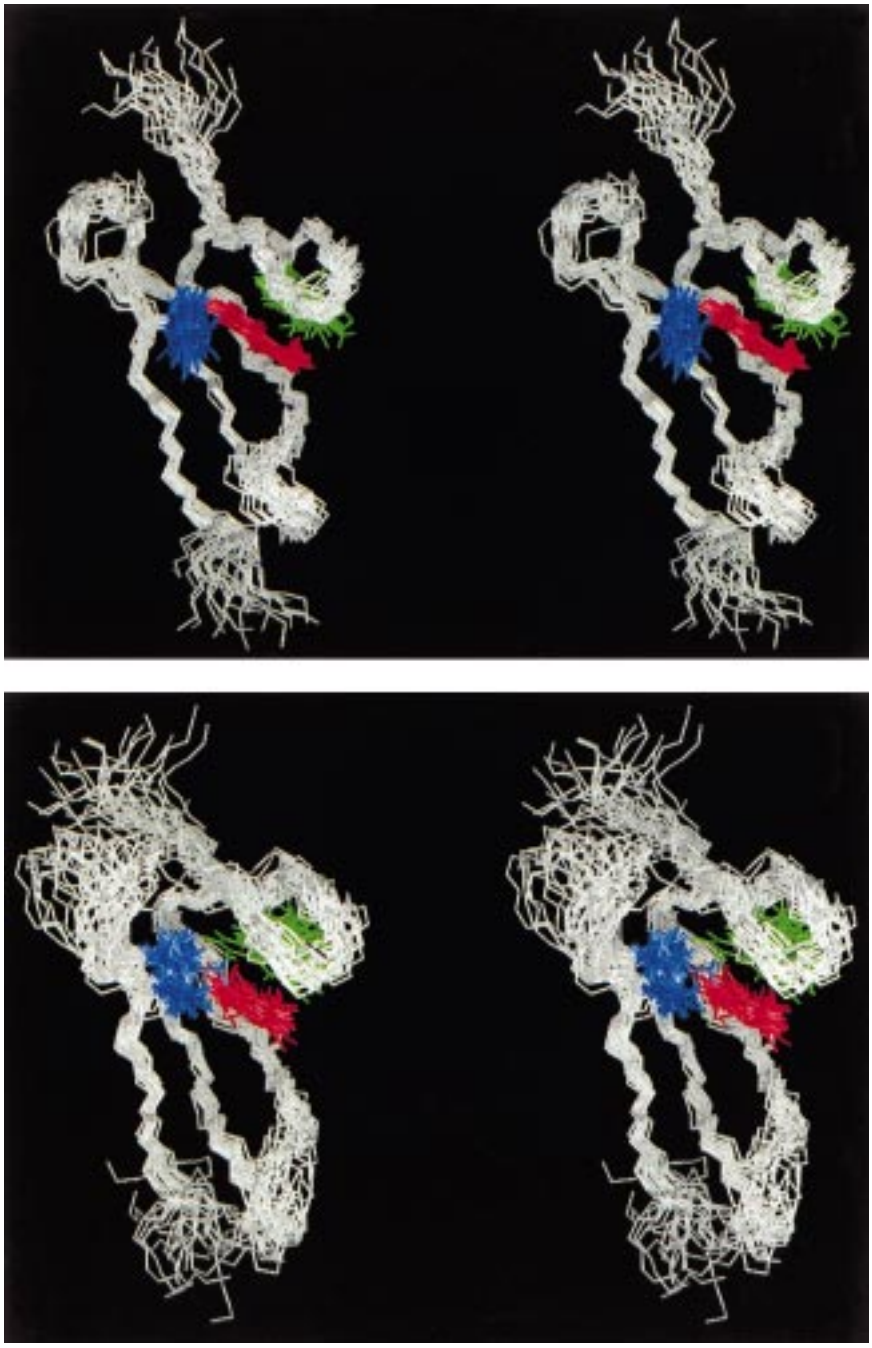

FIGURE 3: Solution structure with side chains of selected core residues of ^{1}F1 and ^{2}F1. (A, top) Stereoview of the 24 final structures superimposed over the backbone heavy atoms (N, Cα, C) of the secondary structure of ^{1}F1 (only residues 17−61 are shown). Side chains of the core residues Tyr28, Trp34, and Phe54 are shown in green, red, and blue, respectively. (B, bottom). Stereoview of the 24 final structures superimposed over the backbone heavy atoms (N, Cα, C) of the secondary structure of ^{2}F1 (only residues 62−109 are shown). Side chains of the core residues Tyr75, Tyr81, and Ile102 are shown in green, red, and blue, respectively. The figure was prepared using the program INSIGHT II (Biosym Technologies, San Diego).

chain) differ by ≤0.07 ppm between ^{1}F1s and ^{1}F1^{2}F1, and CαH chemical shifts of residues 19−58 differ by ≤0.04 ppm.

Figure 1 shows a sequence alignment of the 12 F1 modules from fibronectin. With the structures of several F1 modules now available, it is possible to define the extent of the modules on the basis of structure rather than amino acid sequence. For example, the observed long-range intramodule NOEs and the rmsd of backbone residues in the calculated structures of ^{1}F1, ^{2}F1 (present work), ^{4}F1, and ^{5}F1 (*11*) suggest that two residues N-terminal to the first cysteine and three residues after the last cysteine of each F1 module can be considered part of the module fold. Using this definition there are no "linker" residues in the ^{4}F1^{5}F1 pair, and in the ^{1}F1^{2}F1 pair, the intermodule linker is formed by residues 60−63.

Secondary Structure. Potential β-strands of ^{1}F1^{2}F1 were identified on the basis of medium−strong CαH(*i*)−NH(*i*+1) and weak NH(*i*)−NH(*i*+1) NOEs, $^{3}J_{HN-C\alpha H}$ coupling constants >8 Hz, nonrandom coil CαH chemical shifts (*41*), slowly exchanging backbone amide protons (Figure 2), and characteristic long-range interstrand NOEs, particularly CαH(*i*)−CαH(*j*) NOEs. The ^{1}F1 module has the consensus F1 fold

Table 1: Structural Statistics for the Family of 24 Structures

rmsd (±sd) from experimental restraints[a]		
all NOE restraints (Å)		0.0083 ± 0.0007
intraresidue (Å)		0.0116 ± 0.0013
sequential (Å)		0.0060 ± 0.0013
short range (Å)		0.0046 ± 0.0016
long range (Å)		0.0041 ± 0.0013
ambiguous (Å)		0.0029 ± 0.0019
hydrogen bonds (Å)		0.0006 ± 0.0014
dihedral angle (deg)		0.0306 ± 0.0421
rmsd (±sd) from idealized covalent geometry		
bond lengths (Å)		0.0009 ± 0.0001
angles (deg)		0.2654 ± 0.0029
impropers (deg)		0.1134 ± 0.0064
rmsd (±sd) of Cartesian coordinates (Å)[b]		
atoms	backbone N, Cα, C	all heavy
residues 21−58	0.63 ± 0.17	1.07 ± 0.18
residues 21−23, 26−28, 33−37, 42−47, 54−58	0.46 ± 0.11	0.91 ± 0.14
residues 66−106	1.12 ± 0.21	1.60 ± 0.22
residues 66−68, 73−75, 80−84, 89−94, 102−106	0.69 ± 0.15	1.14 ± 0.20

[a] None of the 24 structures had NOE violations >0.2 Å or dihedral angle violations >2°. Analysis of the ϕ and ψ angles of the 24 structures using PROCHECK_NMR (*48*) shows that 93% of the (non-glycine and non-proline) residues fall within the most favored or additionally allowed regions of the Ramachandran plot. All residues with ϕ and ψ angle order parameters ≥0.9 (*49*) had average ϕ and ψ angles in the most favored or additionally allowed regions of the Ramachandran plot. [b] The average structure was calculated by superimposing N, Cα, and C atoms of the selected residues. Rmsd was calculated by finding the mean and standard deviation between the average structure and each member of the family for the selected residues.

with a short N-terminal double-stranded antiparallel β-sheet (strands A and B: residues 21−23 and 26−28) followed by a triple-stranded antiparallel β-sheet (strands C, D, and E: residues 33−37, 42−47, and 54−58). However, not all the expected slowly exchanging amides were observed (Figure 2) or predicted (based on interatom distances and angles) in the 24 final structures. In ^{2}F1, residues 66−68 and 73−75 (strands A′ and B′) have an extended antiparallel arrangement. However, no slowly exchanging amide protons and not all the expected interstrand NOEs were observed for these residues. Residues 80−84, 89−94, and 102−106 (strands C′, D′, and E′) form three extended strands in an antiparallel triple-stranded arrangement; again, however, this β-sheet is not as well-defined as in ^{1}F1, and not all the expected slowly exchanging amides were observed, or predicted.

The triple-stranded β-sheet residues reported for ^{1}F1s (*40*) differ slightly from those above for ^{1}F1^{2}F1; this primarily arises from an improvement in the data analysis in the present work. By including hydrogen bond restraints more conservatively, it was clear that in ^{1}F1^{2}F1 the hydrogen bond acceptor for the slowly exchanging Glu32 HN (Figure 2) was more likely to be Glu29 O (in the BC turn) than Cys47 O (as assigned for ^{1}F1s). Additionally, the hydrogen bond acceptor for the slowly exchanging Tyr48 HN was not reliably predicted in initial structures; therefore, Tyr48 was no longer classified as a strand residue. Ile95, the ^{2}F1 residue in a homologous position to Tyr48, also has a slowly exchanging amide proton (Figure 2), and the most likely acceptor is Thr93 O with the formation of an intrastrand hydrogen bond.

Tertiary Structure of the F1 Modules. The ^{1}F1^{2}F1 structure was calculated using a total of 1216 restraints. For residues 18−61, these included 258 intraresidue, 185 sequential, 56 short-range ($1 < i - j < 5$), 146 long-range ($i - j \geq 5$), and 19 ambiguous interproton distance restraints, 11 ϕ backbone torsion angle restraints, and 8 hydrogen bond restraints. For residues 62−109, 206 intraresidue, 131 sequential, 25 short-range, 123 long-range, and 40 ambiguous interproton distance restraints, 2 ϕ backbone torsion angle restraints, and 2 hydrogen bond restraints were used. In addition, 3 sequential restraints between Glu61 and Ala62 and 1 short-range restraint between Pro60 and Ala62 were included.

A total of 24 refined structures were chosen on the basis of their low overall energy and agreement with the experimental restraints. Figure 3 shows an overlay of the 24 final ^{1}F1^{2}F1 structures, and Table 1 shows the statistics for the structure calculations. Both ^{1}F1 (Figure 3A) and ^{2}F1 (Figure 3B) adopt the previously determined F1 consensus fold, but (as expected from the larger number of distance restraints) the ^{1}F1 module is better defined with an rmsd of 0.46 Å, compared with 0.69 Å for ^{2}F1 when the 24 final structures are superimposed over the backbone N, Cα, and C atoms of the β-strands. When the minimized average structure of ^{2}F1 is superimposed over the backbone C, Cα, and N atoms of the common secondary structure of the minimized average structures of ^{1}F1, ^{4}F1, and ^{5}F1 (*11*) and ^{7}F1 (*1*), the rmsds are 1.2, 1.2, 1.3, and 1.1 Å, respectively.

In both ^{1}F1 and ^{2}F1 a short double-stranded β-sheet folds over the triple-stranded β-sheet to enclose a hydrophobic core (Figure 3). The four conserved cysteine residues (Figure 1) form disulfide bonds in a 1−3, 2−4 configuration, connecting the two β-sheets and two strands in the second β-sheet, respectively. Figure 4 shows that in both ^{1}F1 and ^{2}F1 these disulfide bonds are almost completely buried in the core of the module. As expected, in general, residues with fewer distance restraints also have a high rmsd, and in both modules the turn between the two β-sheets is well-defined (low rmsd; Figure 4). The clockwise twist (when viewed from the N-terminus) of the turn between the first two strands of the triple-stranded β-sheet is a feature of all F1 module structures determined to date.

There is a significant difference between ^{1}F1 and ^{2}F1 in the number of slowly exchanging amide protons observed; fewer slowly exchanging amides are observed for ^{2}F1 (Figure 2). In addition, preliminary measurements of the T_2 relaxation time constant (data not shown) for backbone ^{15}N nuclei of

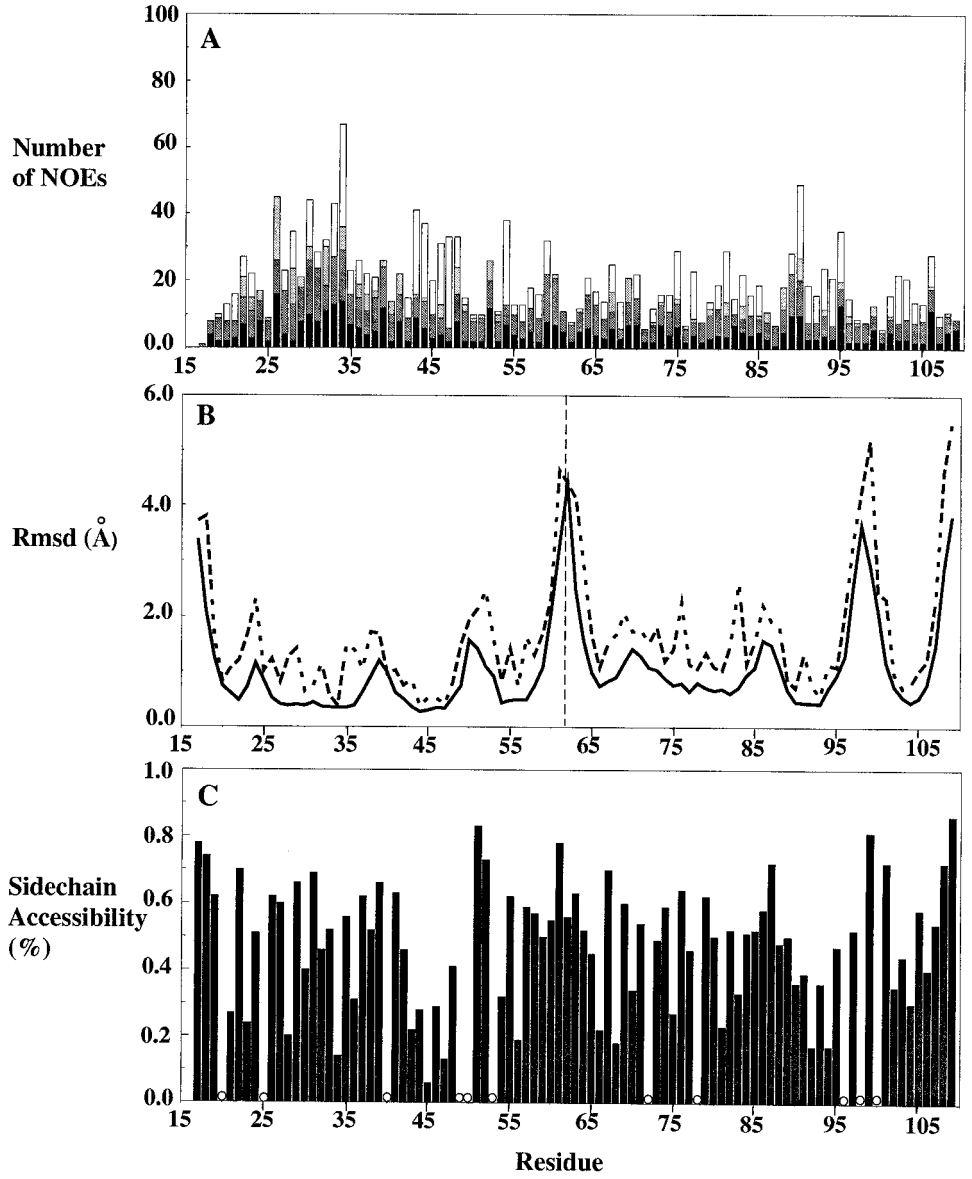

FIGURE 4: Correlation between the number of restraints, backbone and side chain atom rmsd, and side chain accessibility in ^{1}F1^{2}F1. (A) The number of NOESY-derived restraints as a function of residue number. The stacked bars from bottom to top (with progressively lighter shading) show the number of intraresidue, sequential, short-range, and long-range restraints, respectively. (B) Average atomic rmsd between the 24 lowest energy structures calculated for the backbone heavy atoms (solid line) and the side chain heavy atoms (dashed line). The vertical dashed line indicates that the rmsd was calculated for residues in ^{1}F1 and ^{2}F1 separately, by superimposing over residues in ^{1}F1 and ^{2}F1, respectively. (C) Mean side chain solvent accessibility (calculated over the 24 final structures) as a function of residue number. The accessibility was calculated in X-PLOR using a probe radius of 1.4 Å and is expressed as a percentage of the accessibility for the corresponding residue in an extended chain Gly-Gly-X-Gly-Gly (*3*). Circles show the position of glycine residues in the sequence.

^{1}F1^{2}F1 indicate significant chemical exchange contributions for a larger number of residues in ^{2}F1 than ^{1}F1. The increased line width (shorter T_2s) observed for these residues is likely to contribute to the observation of fewer NOEs for ^{2}F1. Thus the higher rmsd for ^{2}F1 compared with ^{1}F1, which is expected from the difference in the number of restraints used in the structure calculations, may also reflect a real difference between the modules.

The particular importance of three conserved residues, a tyrosine in strand B, a tryptophan in strand C, and a hydrophobic residue in strand E (Figure 1), for the stability of the F1 module fold has been suggested previously (*11*). This idea is supported by the conserved side chain orientation of the tryptophan and tyrosine residues in all F1 module structures to date. The tryptophan is conserved in all but two F1 modules, ^{2}F1 and ^{8}F1 (Figure 1), where the equivalent residue is a tyrosine and a phenylalanine, respectively. The side chain orientation of residues Tyr28, Trp34, and Phe54 in ^{1}F1, and Tyr75, Tyr81, and Ile102 in ^{2}F1 is shown in panels A and B of Figure 3, respectively. The substitution of tyrosine in place of tryptophan, resulting in a smaller buried hydrophobic surface area, may reduce the thermodynamic stability of ^{2}F1, resulting in fewer slowly exchanging amides.

A putative salt bridge was noted previously between a highly conserved basic residue in strand C and an acidic residue in strand A of the F1 fold (*42*) (Figure 1). In ^{1}F1, the side chains of Asp23 and Arg36 are in close proximity and have low solvent accessibility (Figure 4). In ^{2}F1, the side chains of the homologous residues, Asp68 and Arg83,

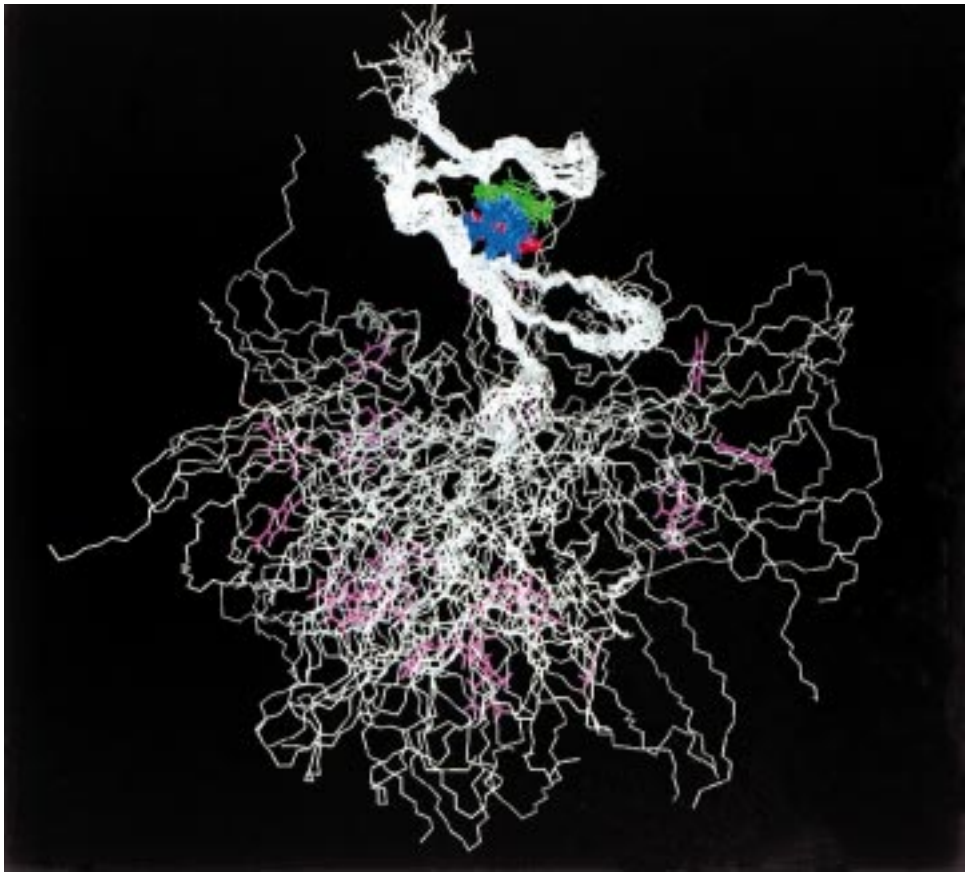

FIGURE 5: Intermodule orientation of ^{1}F1^{2}F1. A total of 24 final structures of ^{1}F1^{2}F1 are shown superimposed over the secondary structure of ^{1}F1. The side chains of residues Tyr28, Trp34, and Phe54 (in ^{1}F1) are colored as in Figure 3A. The side chain of Tyr81 (in ^{2}F1) is shown in magenta. The figure was prepared using the program INSIGHT II (Biosym Technologies, San Diego).

are less well-defined but also have low solvent accessibility compared with other charged residues. The chemical shifts of NϵH of Arg36 and Arg83 are 7.93 and 7.82 ppm, respectively, compared with values of 6.94, 7.23, 7.24, and 7.17 ppm for Arg52, Arg99, Arg101, and Arg109, respectively, for which there is no evidence for a salt bridge. Although the effects of aromatic ring current shifts from core residues cannot be ruled out, shifts to high frequency of NϵH of arginines involved in putative salt bridges have been observed previously (*3*, *43*).

Residues Arg52 and Arg99 have been implicated in binding of the N-terminal domain of fibronectin to heparin (*15*). Mutation of a single positively charged residue in homologous positions in either ^{1}F1 (Arg52), ^{2}F1(Arg99), or ^{3}F1 (Lys143) completely eliminated the heparin binding activity of $^{1-5}$F1; thus multiple sites appear to be required for heparin binding. Figure 4 shows that in ^{1}F1^{2}F1 the side chains of Arg52 and Arg99 are highly solvent exposed and therefore would be available for interaction with the ligand.

Intermodule Orientation. Figure 5 shows the 24 final ^{1}F1^{2}F1 structures superimposed over the secondary structure of ^{1}F1. For clarity, the side chains of Tyr28, Trp34, and Phe54 in ^{1}F1 and of Tyr81 in ^{2}F1 are shown. The intermodule orientation is very poorly defined as would be expected from the complete lack of intermodule NOEs observed in the NOESY spectra of ^{1}F1^{2}F1 and is consistent with the similarity in chemical shift between ^{1}F1s and ^{1}F1^{2}F1. This lack of an intermodule interface is in contrast to the well-defined interface observed in the only other F1 module pair structure available, ^{4}F1^{5}F1 (*11*).

In ^{4}F1^{5}F1, the major intermodule interface is formed between the N-terminal end of strand D in ^{4}F1 and the N-terminal β-sheet of ^{5}F1 (*11*). A tryptophan residue in ^{4}F1, which appears to play an important role in this interface, is not conserved in ^{1}F1. In addition, in ^{4}F1^{5}F1, the C-terminal residue of ^{4}F1 and the N-terminal residue of ^{5}F1 are both constrained by NOEs to the DE loop of ^{5}F1. In ^{1}F1^{2}F1, however, these C- and N-terminal residues are separated by the four-residue intermodule linker whose orientation is ill-defined with respect to both modules. The NMR-observed difference in the intermodule interface of these two module pairs is consistent with previous spectroscopic and calorimetric studies which detected an intermodule-stabilizing interaction in ^{4}F1^{5}F1 but not in ^{1}F1^{2}F1 (*10*). In intact $^{1-5}$F1, it has been suggested that ^{2}F1 is stabilized by interactions with ^{3}F1 (*10*).

^{15}N-{^{1}H} NOE. One advantage of NMR spectroscopy for studying protein structure is that the relaxation properties of ^{15}N nuclei can provide direct information about molecular dynamics in solution (*44*). To obtain further evidence about the interface in ^{1}F1^{2}F1, the backbone dynamics of ^{1}F1^{2}F1 and ^{4}F1^{5}F1 were compared. ^{15}N-{^{1}H} NOE values for ^{4}F1^{5}F1 are available in the literature (*31*); however, the present studies were carried out on a sample of higher protein concentration. This resulted in lower experimental errors on the NOE values which was important for comparison of the

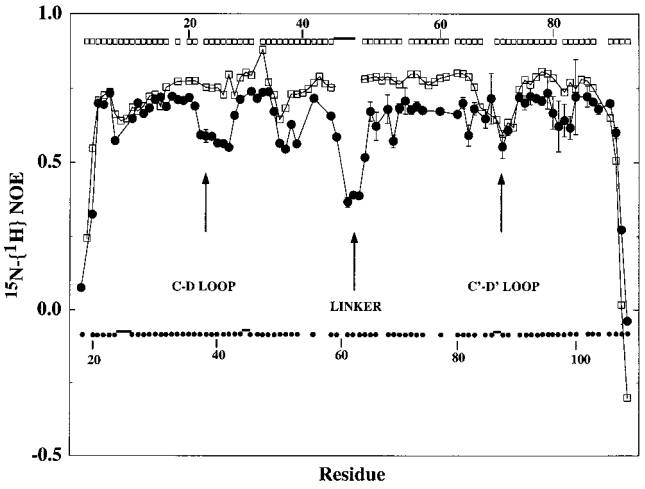

FIGURE 6: ^{15}N-{^{1}H} NOE for ^{1}F1^{2}F1 (residues 18−109) and ^{4}F1^{5}F1 (residues 2−93). The data for ^{1}F1^{2}F1 and ^{4}F1^{5}F1 are indicated by filled circles and open squares, respectively. The larger symbols show the NOE value, and the smaller symbol indicates the residue number. A horizontal bar above the sequence indicates where gaps in the sequence occur as a result of the sequence alignment. Otherwise, a gap in the sequence indicates a lack of data for this residue. The positions of the CD loops and the intermodule linker, which are mentioned in the text, are indicated. In HSQC spectra of ^{1}F1^{2}F1, peaks for a number of residues in ^{2}F1 are weak, giving rise to larger errors in the NOE for these residues. Where no error bar is visible, the error falls within the dimensions of the symbol. A sequence alignment of the two module pairs was performed using the program PILEUP (Genetics Computer Group Inc.).

NOE for homologous residues in the two module pairs. In addition, in the previous study the HSQC peak of Ser45 (near the ^{4}F1 C-terminus) was not assigned, and thus an NOE value was not available for this residue.

The results of the NOE experiments are consistent with the calculated structures of the ^{1}F1^{2}F1 and ^{4}F1^{5}F1 module pairs. As mentioned above, in ^{4}F1^{5}F1 (*11*) Arg46, the C-terminal residue of ^{4}F1, as well as being constrained by NOEs to the adjacent β-strand of ^{4}F1, is close to the DE turn of ^{5}F1, forming part of the intermodule interface. Asn47 is the N-terminal residue of ^{5}F1 and is restrained (as in other F1 modules) by NOEs to the BC turn of ^{5}F1 and is also close to the DE turn (*11*). In the ^{15}N-{^{1}H} heteronuclear NOE experiment, there is no evidence of increased flexibility (lower heteronuclear NOE value) for residues Arg46 and Asn47 when compared with β-sheet residues (Figure 6). In contrast, in ^{1}F1^{2}F1, there are no short- or long-range NOEs between the linker (residues 60, 61, 62, and 63) and residues in either ^{1}F1 or ^{2}F1. From the low value of the ^{15}N-{^{1}H} heteronuclear NOE for residues 61, 62, and 63 (residue 60 is a proline; Figure 6) it seems likely that the high rmsd observed for these residues (Figure 4) results from the flexibility of the intermodule linker. While there is a significant difference in the heteronuclear NOE value for residues in the C−D loop of ^{4}F1 compared with the same loop in ^{5}F1, residues in both C−D loops of ^{1}F1^{2}F1 show increased flexibility compared with β-sheet residues. Again, this is consistent with the ^{4}F1^{5}F1 structure in which residues at the N-terminal end of strand D of ^{4}F1 form a major part of the intermodule interface with residues in the N-terminal β-sheet of ^{5}F1. There is no evidence for such a stabilizing interaction in ^{1}F1^{2}F1. As expected, the N- and C-termini of both ^{1}F1^{2}F1 and ^{4}F1^{5}F1 have low NOE values, indicating significant backbone flexibility for these residues.

CONCLUDING REMARKS

The results presented here show that although both modules of the ^{1}F1^{2}F1 module pair adopt the consensus F1 fold, the F1−F1 intermodule interface identified previously in the ^{4}F1^{5}F1 pair is not conserved. No intermodule NOEs were observed in spectra of the ^{1}F1^{2}F1 pair, and significant backbone flexibility was detected in the intermodule linker. In addition, the ^{2}F1 module is less well-defined and has fewer slowly exchanging amide protons than ^{1}F1, suggesting that the ^{2}F1 module is less stable and may be undergoing conformational exchange.

There is evidence in the literature for the role of intact module pairs or larger fragments of the N-terminal domain in fibronectin function. The intermodule interface identified previously in ^{4}F1^{5}F1 may well be functionally important as neither ^{4}F1 nor ^{5}F1 binds fibrin as single modules and yet the ^{4}F1^{5}F1 pair contains most of the fibrin binding activity of this region of fibronectin (*10*, *11*). In addition, the ^{4}F1^{5}F1 module pair binds to a synthetic repeat from the fibronectin binding protein from *S. aureus*, while the individual modules lack activity (*12*). Although a synthetic peptide from an MSCRAMM of *S. dysgalactiae* binds ^{1}F1^{2}F1 (*13*), it is not known whether both modules are required for binding.

There is also evidence for the importance of multiple binding sites in $^{1-5}$F1 in matrix assembly (*7*), bacterial binding (*13*), and heparin binding (*15*). Studies using circular dichroism and fluorescence anisotropy indicated that heparin binding causes changes in conformation and flexibility of the N-terminal domain (*45*, *46*). Although the nature of this conformational change has yet to be identified, a flexible ^{1}F1^{2}F1 linker may allow the sites in ^{1}F1 (Arg52) and ^{2}F1 (Arg99) to bind to heparin simultaneously. Similarly, the other interactions involving multiple binding sites in the N-terminal domain may rely on a combination of well-defined intermodule interfaces such as in ^{4}F1^{5}F1 and intermodule flexibility as observed here in ^{1}F1^{2}F1, which would allow a new intermodule orientation to be adopted on ligand binding. Such variation in intermodule interactions may be another evolutionary variable that allows the formation of a wide range of binding sites in fibronectin and other mosaic proteins, from relatively few basic module folds.

ACKNOWLEDGMENT

The authors gratefully acknowledge the assistance of Dr. Moshe Werber, BioTechnology General (Israel) Ltd., Rehovot, Israel, in the early stages of the project. We also thank Dr. Graeme Shaw for providing pulse sequences, Dr. Robin Aplin for mass spectrometry, and Tony Willis for N-terminal sequencing.

SUPPORTING INFORMATION AVAILABLE

One table containing ^{1}H and backbone ^{15}N chemical shifts for ^{1}F1^{2}F1. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Baron, M., Norman, D. G., Willis, A., and Campbell, I. D. (1990) *Nature 345*, 642−646.

2. Main, A. L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. D. (1992) *Cell 71*, 671−678.
3. Pickford, A. R., Potts, J. R., Bright, J. R., Phan, I., and Campbell, I. D. (1997) *Structure 5*, 359−370.
4. Hynes, R. O (1990) *Fibronectins*, Springer-Verlag, New York.
5. Potts, J. R., and Campbell, I. D. (1994) *Curr. Opin. Cell Biol. 6*, 648−655.
6. Quade, B. J., and McDonald, J. A. (1988) *J. Biol. Chem. 263*, 19602−19609.
7. Sottile, J., and Mosher, D. F. (1997) *Biochem. J. 323*, 51−60.
8. Bultmann, H., Santas, A. J., and Peters, D. M. P. (1998) *J. Biol. Chem. 273*, 2601−2609.
9. Patti, J. M., Allen, B. L., McGavin, M. J., and Höök, M. (1994) *Annu. Rev. Microbiol. 48*, 585−617.
10. Matsuka, Y. V., Medved, L. V., Brew, S. A., and Ingham, K. C. (1994) *J. Biol. Chem. 269*, 9539−9546.
11. Williams, M. J., Phan, I., Harvey, T. S., Rostagno, A., Gold, L. I., and Campbell, I. D. (1994) *J. Mol. Biol. 235*, 1302−1311.
12. Huff, S., Matsuka, Y. V., McGavin, M. J., and Ingham, K. (1994) *J. Biol. Chem. 269*, 15563−15570.
13. Joh, D., Speziale, P., Gurusiddappa, S., Manor, J., and Höök, M. (1998) *Eur. J. Biochem. 258*, 897−905.
14. Sottile, J., Schwarzbauer, J., Selegue, J., and Mosher, D. F. (1991) *J. Biol. Chem. 266*, 12840−12843.
15. Kishore, R., Samuel, M., Khan, M. Y., Hand, J., Frenz, D. A., and Newman, S. A. (1997) *J. Biol. Chem. 272*, 17078−17085.
16. Bright, J. R., Pickford, A. R., Potts, J. R., and Campbell, I. D. (1999) in *Methods in Molecular Biology*, Humana Press Inc., Totowa, NJ (in press).
17. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
18. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
19. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
20. Shaka, A. J., Lee, C. J., and Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
21. Rucker, S. P., and Shaka, A. J. (1989) *Mol. Phys. 68*, 509−517.
22. Plateau, P., and Guéron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
23. Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229−2246.
24. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 28*, 6150−6156.
25. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
26. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
27. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
28. Kay, L. E., and Bax, A. (1990) *J. Magn. Reson. 86*, 110−126.
29. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003
30. Shaka, A. J., Keeler, J., Frenkiel, T., and Freeman, R. (1983) *J. Magn. Reson. 52*, 335−338.
31. Phan, I. Q. H., Boyd, J., and Campbell, I. D. (1996) *J. Biomol. NMR 8*, 369−378.
32. Wüthrich, K. (1986) *NMR of Proteins and Amino Acids*, John Wiley & Sons, New York.
33. Williams, M. J., Phan, I., Baron, M., Driscoll, P. C., and Campbell, I. D. (1993) *Biochemistry 32*, 7388−7395.
34. Barsukov, I. L., and Lian, L.-Y. (1993) in *NMR of Macromolecules* (Roberts, G. C. K., Ed.) pp 315−357, Oxford University Press, Oxford.
35. Sticht, H., Pickford, A. R., Potts, J. R., and Campbell, I. D. (1998) *J. Mol. Biol. 276*, 177−187.
36. Pardi, A., Billeter, M., and Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741−751.
37. Brünger, A. T. (1992) *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT.
38. Nilges, M. (1995) *J. Mol. Biol. 245*, 645−660.
39. Folmer, R. H. A., Hilbers, C. W., Konings, R. N. H., and Nilges, M. (1997) *J. Biomol. NMR 9*, 245−258.
40. Potts, J. R., Phan, I., Williams, M. J., and Campbell, I. D. (1995) *Nat. Struct. Biol. 2*, 946−950.
41. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
42. Downing, A. K., Driscoll, P. C., Harvey, T. S., Dudgeon, T. J., Smith, B. O., Baron, M., and Campbell, I. D. (1992) *J. Mol. Biol. 225*, 821−833.
43. Aumelas, A., Chiche, L., Kubo, S., Chino, N., Tamaoki, H., and Kobayashi, Y. (1995) *Biochemistry 34*, 4546−4561.
44. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972−8979.
45. Khan, M. Y., Jaikaria, N. S., Frenz, D. A., Villanueva, G., and Newman, S. A. (1988) *J. Biol. Chem. 263*, 11314−11318.
46. Khan, M. Y., Medow, M. S., and Newman, S. A. (1990) *Biochem. J. 270*, 33−38.
47. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
48. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
49. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) *Protein Sci. 1*, 736−751.

BI990202B