Supporting information

Materials and methods:

Protein expression and purification.

Protein expression was performed in *Escherichia coli* strains BL21 CodonPlus-RP (Stratagene) for the pGEX-6P-2 constructs and BL21(DE3) CodonPlus-RIPL (Stratagene) for the pET-16 constructs in Luria Broth or M9 minimal media containing ^{15}N-enriched NH_4Cl (Spectra Stable Isotopes) and unenriched or ^{13}C-enriched glucose (Spectra Stable Isotopes) as sole nitrogen and carbon sources, respectively. Expression was induced at 30 ^{o}C by adding isopropyl-β-D-thiogalactopyranoside (Melford) to a final concentration of 0.2 mM. Cells were harvested after 12 hours, resuspended in a buffer containing 150 mM NaCl, 20 mM NaPi pH 7.3, 10 mM MgCl, DNAse I and 1% Triton-X, and lysed by repeated freeze/thaw cycles. Cell lysates were centrifuged at 20,000 g for 20 min.

GST-fused protein constructs were purified as follows: the lysate supernatant was loaded in a glutathione-sepharose column (GE Biosciences) and eluted with a 10 mM reduced glutathione (Sigma), 50 mM Tris-Cl pH 7.8 buffer. EDTA was added to a final concentration of 9 mM and the protein fusion was digested overnight at 4 ^{o}C by 3C protease. After cleavage, the protein was buffer exchanged to 150 mM NaCl, 20 mM NaPi pH 7.3, passed again through the glutathione-sepharose column to retain the cleaved GST and buffer exchanged to a final buffer. In contrast, His-tag fused constructs were purified from lysate inclusion bodies by metal affinity chromatography (GE biosciences) under denaturing conditions (8 M Urea, 100 mM NaCl, 20 mM Tris-Cl pH 7.8). The proteins were subsequently refolded by slow dialysis against a 100 mM NaCl, 20 mM Tris-Cl pH 7.8 buffer and digested by 3C protease as described. Final purification was performed by anion exchange chromatography and buffer exchange against a final buffer. Wild type $^{1\text{-}2}$FNIII prepared by both methods yielded identical ^{15}N HSQC spectra (data not shown). Centrifugal filtering devices (Amicon, Millipore) were used for protein concentration to levels suitable

for NMR (~1-2 mM). Protein concentration was determined by absorbance at 280 nm, using a corrected extinction coefficient as described (Pace et al., 1995).

NMR experiments.

^{1}FNIII chemical shift assignments were reported previously (Gao et al., 2003). ^{2}FNIII sequential ^{13}C, ^{15}N and ^{1}H, backbone and sidechain chemical shift assignments and $^{1\text{-}2}$FNIII backbone chemical shift assignments were performed using standard triple resonance experiments. Stereospecific assignments, $^{3}J^{HNHA}$ coupling measurements, χ_1 and χ_2 dihedral angle determinations and proline isomerization state determinations were performed as described previously (Vakonakis et al., 2004a; Vakonakis et al., 2004b). Backbone dynamics measurements (^{15}N T_1, ^{15}N T_2 and $\{^{1}H\text{–}\}^{15}N$ NOE) were performed and analyzed as described previously (Vakonakis et al., 2004a). Residual dipolar coupling (RDC) restraints were obtained using an IPAP ^{1}H-^{15}N HSQC experiment (Ottiger et al., 1998) in 4% polyacrylamide gels radially compressed using an apparatus similar to that of Chou *et al*. (Chou et al., 2001), or a 4% $C_{12}E_5$ polyethylyne glycol / hexanol medium (Ruckert and Otting, 2000). RDC restraints were derived from single, un-overlapped peaks in the IPAP spectra and used only for structured residues ($\{^{1}H\text{–}\}^{15}N$ NOE > 0.6). The RDC rhombicity and anisotropy components in structure calculations were determined by grid-search using an initial protein structure and further refined in subsequent calculation iterations. Singular value decomposition fit of RDC data to protein structures was performed using the program PALES (Zweckstetter and Bax, 2000).

NMR structure calculations: linker

Hydrogen bond restraints were implemented using a specialized distance and geometry potential energy term derived from high resolution X-ray diffraction structures (Lipsitz et al.,

2002) and applied in accordance with hydrogen exchange protection data acquired by NMR, and the existence of regular secondary structure NOEs (Wüthrich, 1986). φ and ψ dihedral angle values were predicted using TALOS (Cornilescu et al., 1999). A total of 50 structures were calculated by a distance geometry simulated annealing protocol in Cartesian space using the XPLOR-NIH software package. Only NOE, hydrogen bond, dihedral angle, RDC and $^{3}J^{HNHA}$-coupling potential energy terms were used as restraints during simulated annealing. During structure refinement additional potential energy terms were used, including a radius of gyration restraint, with a calculated value of 11.29 Å applied to residues 735-608 (Kuszewski et al., 1999), a conformational database potential term (Kuszewski et al., 1996) and direct refinement against $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (Kuszewski et al., 1995). The final ensemble consists of the 25 lowest energy structures.

NMR structure calculations: linker

A number of short simulations on possible interdomain linker conformations were performed using XPLOR-NIH (Schwieters et al., 2003). Amino acid chains of the specific linker sequence were subjected to a simulated slow cooling process under full Van der Waals and chemical structure potentials. In addition, potential terms for direct refinement against the measured $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (Kuszewski et al., 1995) and conformational database refinement (Kuszewski et al., 1996) were included. This ensured that overall linker conformations remained those of random coil while simultaneously biasing the linker population against disallowed regions of the Ramachandran plot. A total of 500 linker conformations were calculated and used to obtain the average end-to-end distance.

NMR Structure calculations: $^{1\text{-}2}$FNIII

The $^{1\text{-}2}$FNIII structure was calculated using a rigid body docking protocol as described previously (Clore and Schwieters, 2003) modified to allow for interactions between two domains of the same polypeptide chain (^{1}FNIII, residues 609-701 and structured ^{2}FNIII, residues 734-809) connected through a flexible linker (residues 702-733). Distance restraints were implemented based on the consensus of the chemical shift perturbations observed, and

RDC restraints were derived from the strained polyacrylamide media experiments. The sidechains of all residues of the structured domains were fixed during the first few initial runs. Subsequently, the sidechains of residues at the interaction interface (E618, Q622, N624, S625, T667, K669, K672, S748, R751, E753, Y765, D767, I793, E795, D796, E798 and Q799) were allowed to move and a radius of gyration potential term (Kuszewski et al., 1999) with a calculated value of 13.25 Å was applied between ^{1}FNIII residues 617-627, 667-677, 697-701 and all the structured ^{2}FNIII residues, in order to improve the interdomain interface packing. 100 structures were generated and the 50 lowest energy structures were retained for further analysis. After site specific amino acid substitution and additional RDC experiments in PEG/hexanol media one final ensemble of 39 structures was selected.

Table I: ^{2}FNIII structure statistics and quality assessment

Experimental restraints		Structure quality				
NOE			25 Structure ensemble		Minimized Average Structure	
Intra-residue (*i-j=0*)	408	RMSDs from experimental restraints				
Sequential (*i-j=1*)	588	Distance restraints (Å)	0.0120±0.0009		0.0085	
Short range (*i-j<5*)	291	Dihedral angles (deg.)	0.25±0.04		0.21	
Long range (*i-j≥5*)	1063	$^{13}C^{\alpha}$ Chemical shifts (ppm)	0.98±0.07		0.94	
Ambiguous	27	$^{13}C^{\beta}$ Chemical shifts (ppm)	1.14±0.04		1.08	
Hydrogen bonds[a]	20	$^{3}J^{HNHA}$ couplings (Hz)	0.57±0.05		0.56	
Dihedral angles		$^{1}D_{NH}$ couplings (Hz)	0.42±0.02		0.39	
ϕ	50	$^{1}D_{NH}$ couplings R-factor[b] (%)	6.2±0.3		5.8	
ψ	50	Distance violations > 0.3 Å	0		0	
χ_1	48	Dihedral angle violations > 5°	0		0	
χ_2	11	RMSDs from idealized geometry				
$^{3}J^{HNHA}$ couplings	45	Bonds (Å)	0.0023±0.0002		0.0024	
$^{1}D_{NH}$ couplings	50	Angles (deg.)	0.51±0.01		0.55	
$^{13}C^{\alpha}$, $^{13}C^{\beta}$ chemical shifts	176	Impropers (deg.)	0.42±0.02		0.41	
Total number of restraints	2827	Ramachandran statistics	Core particle[c]	Ordered residues[d]	Core particle[c]	Ordered residues[d]
		Most favoured regions	88.4%	94.4%	86.4%	94.3%
		Additionally allowed regions	9.5%	5.4%	10.6%	5.7%
		Generously allowed regions	1.4%	0.1%	1.5%	0%
		Disallowed regions	0.7%	0.1%	1.5%	0%
		Structure precision[e]				
		Backbone atoms (Å)	0.28±0.04	0.20±0.04		
		All heavy atoms (Å)	0.58±0.08	0.43±0.09		

a) Restraints were applied between donor and acceptor groups using a specialized hydrogen bond distance/geometry potential term (Lipsitz et al., 2002).
b) The RDC R-factor was calculated as suggested by Clore and Garrett (Clore and Garrett, 1999).
c) Excluding the flexible N-terminus (residues -6-734).
d) Mobile residues ({^{1}H–}^{15}N NOE < 0.6) were excluded. Included are residues 736-741, 748-757, 765-807.
e) RMS deviations from the average structure.

Table II: ^{2}FNIII structural similarity search

Protein domain	PDB ID	Backbone RMSD	Aligned residues	Z score
H. sapiens FN ^{8}FNIII	1FNF	1.3 Å	72	10.9
G. gallus Tenascin ^{6}FNIII	1QR4-B	1.4 Å	72	10.6
H. sapiens FN ^{13}FNIII	1FNH	1.4 Å	73	10.6
R. norvegicus Tenascin ^{3}FNIII	1TDQ-A	1.6 Å	73	10.2
H. sapiens Tenascin ^{3}FNIII	1TEN	1.7 Å	73	9.9

Shown here are the five highest scoring superpositions of ^{2}FNIII achieved through a structural similarity search by Dali (Holm and Sander, 1998). For each superposition the matching RCSB PDB ID code, backbone RMS deviation, number of superimposed residues and Z score is shown. Z scores higher than 2 are considered significant.

Table III: $^{1-2}$FNIII structure analysis (population A of the ensemble).

Experimental restraints		Ensemble quality and precision	39 Structure ensemble	Lower E_{rgyr} subpopulation[a]	Higher E_{rgyr} subpopulation[a]
		Occupancy (%)		89.7%	10.3%
^{1}F3 interface residues	7	Ambiguous distance restraints RMSD (Å)	0.11 ± 0.06	0.10 ± 0.05	0.17 ± 0.05
^{2}F3 interface residues	7	$^{1}D_{NH}$ couplings R-factor[b] (%)	14.1 ± 1.2	14.2 ± 1.2	13.4 ± 0.7
^{1}F3 $^{1}D_{NH}$ couplings	30	E_{rgyr} (kcal/mol)	19.1 ± 7.1	17.5 ± 5.6	32.7 ± 0.5
^{2}F3 $^{1}D_{NH}$ couplings	42	Distance violations > 0.5 Å	4	2	2
		Overall ensemble precision (Å)[c]	0.52 ± 0.20	0.52 ± 0.20	0.39 ± 0.06
		^{1}F3 ensemble precision (Å)[d]	1.76 ± 0.93	1.72 ± 0.83	1.80 ± 0.33
		^{2}F3 ensemble precision (Å)[e]	1.37 ± 0.56	1.37 ± 0.59	0.90 ± 0.26

a) Population A was found to contain two subpopulations with different packing characteristics thus, different packing energy. Both subpopulations are included in the final ensemble deposited. An analysis performed by Clore and Schwieters (Clore and Schwieters, 2003) on their results found that the subpopulation with the best packing characteristics (Lower E_{rgyr} energy) had smaller deviations from the correct structure.

b) The RDC R-factor was calculated as suggested by Clore and Garrett (Clore and Garrett, 1999).

c) Atomic backbone RMSD between the average and individual structures calculated over both the ^{1}FNIII and ^{2}FNIII domains (residues 609-700 and 735-809).

d) Atomic backbone RMSD between the average structure, derived by superposition of the structured ^{2}FNIII domain (residues 735-809), and individual structures calculated over the ^{1}FNIII domain (residues 609-700).

e) Atomic backbone RMSD between the average structure derived by superposition of the ^{1}FNIII domain (residues 1-92) and individual structures, calculated over the structured ^{2}FNIII domain (residues 735-809).

Figure 1.

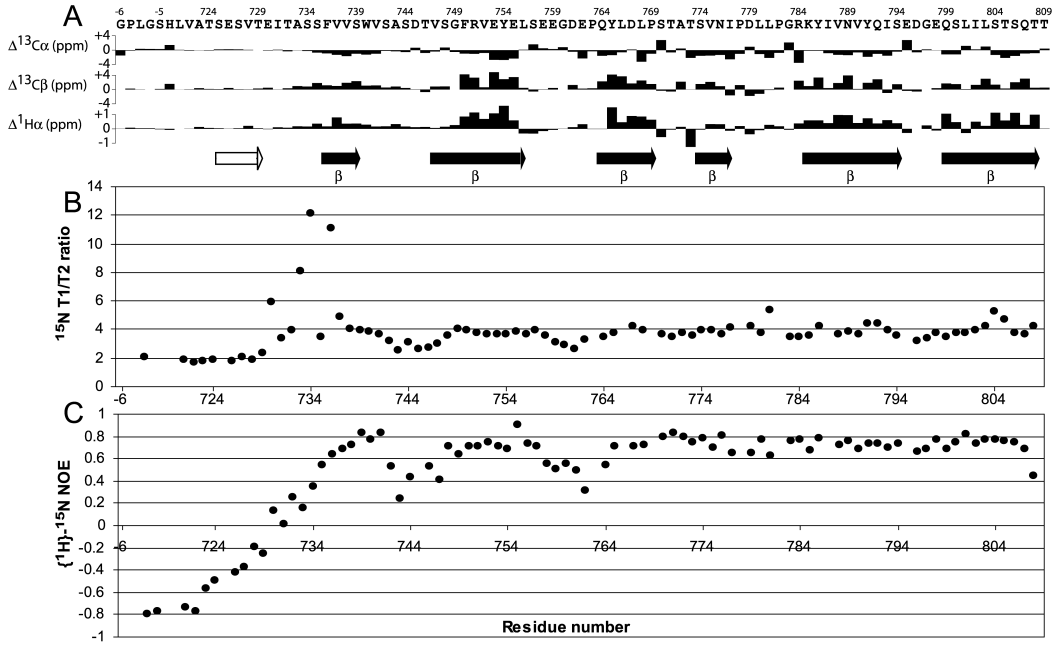

Figure 1. (A) $^{13}C_\alpha$, $^{13}C_\beta$ and $^{1}H_\alpha$ chemical shift differences compared to random coil chemical shifts (Wishart et al., 1995) versus residue number for ^{2}FNIII. The chemical shifts for ^{2}FNIII have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number 7127. The position of the secondary structure elements identified in the ^{2}FNIII are shown as filled arrows for reference. Also shown (open arrow) is the relative position of the first β-strand of typical FNIII domains, which is disordered in ^{2}FNIII. (B) ^{15}N T_1/T_2 ratios and (C) heteronuclear $\{^{1}H–\}^{15}N$ NOE versus residue number for ^{2}FNIII. Data were collected and analyzed as described previously (Vakonakis et al., 2004b) at 30° C and 14.1 T (600 MHz ^{1}H frequency). The estimated total correlation time (τ_c) under these conditions is ~5.2 ns which is consistent with a monomeric particle in solution.

Figure 2.

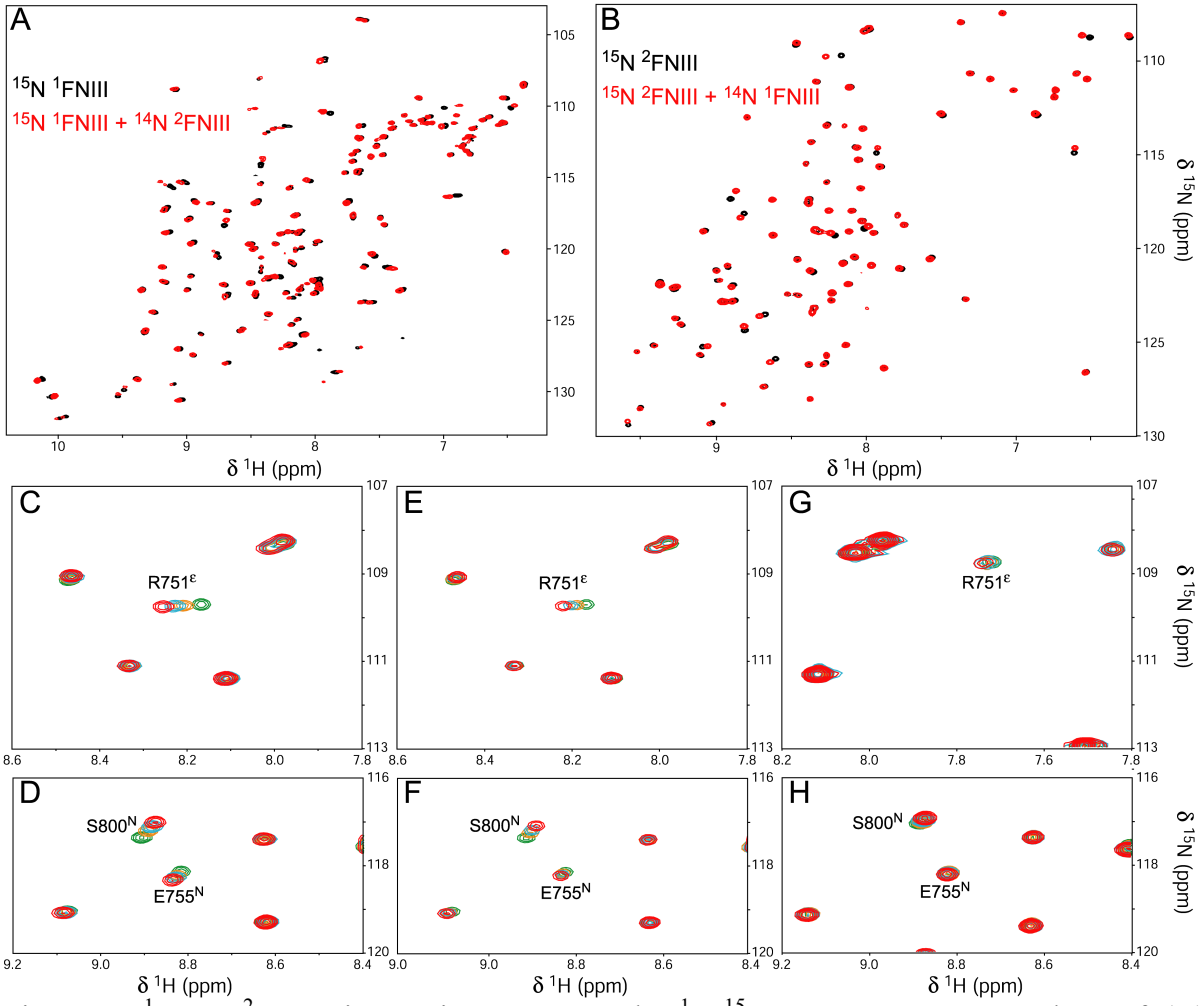

Figure 2. ^{1}FNIII–^{2}FNIII interaction. Top panels: ^{1}H-^{15}N HSQC spectra overlays of (A) isolated ^{15}N enriched ^{1}FNIII (black) or in the presence of unenriched ^{2}FNIII (red) and similar spectra of (B) ^{15}N enriched ^{2}FNIII and unenriched ^{1}FNIII. Lower panels: Details from ^{2}FNIII HSQC spectra overlays for different ^{1}FNIII/^{2}FNIII stoichiometric ratios of (C, D) wild-type proteins, (E, F) wild-type ^{2}FNIII with ^{1}FNIII K669A, and (G, H) ^{2}FNIII D767A with wild-type ^{1}FNIII. In all cases overlays of similar colour correspond to identical domain stoichiometric ratios and absolute protein concentrations. The larger effect observed upon D767 substitution can be attributed to additional electrostatic and hydrogen bonding contributions of this residue to ^{1}FNIII–^{2}FNIII binding and organization of the ^{2}FNIII ^{1}FNIII-binding interface.

Figure 3.

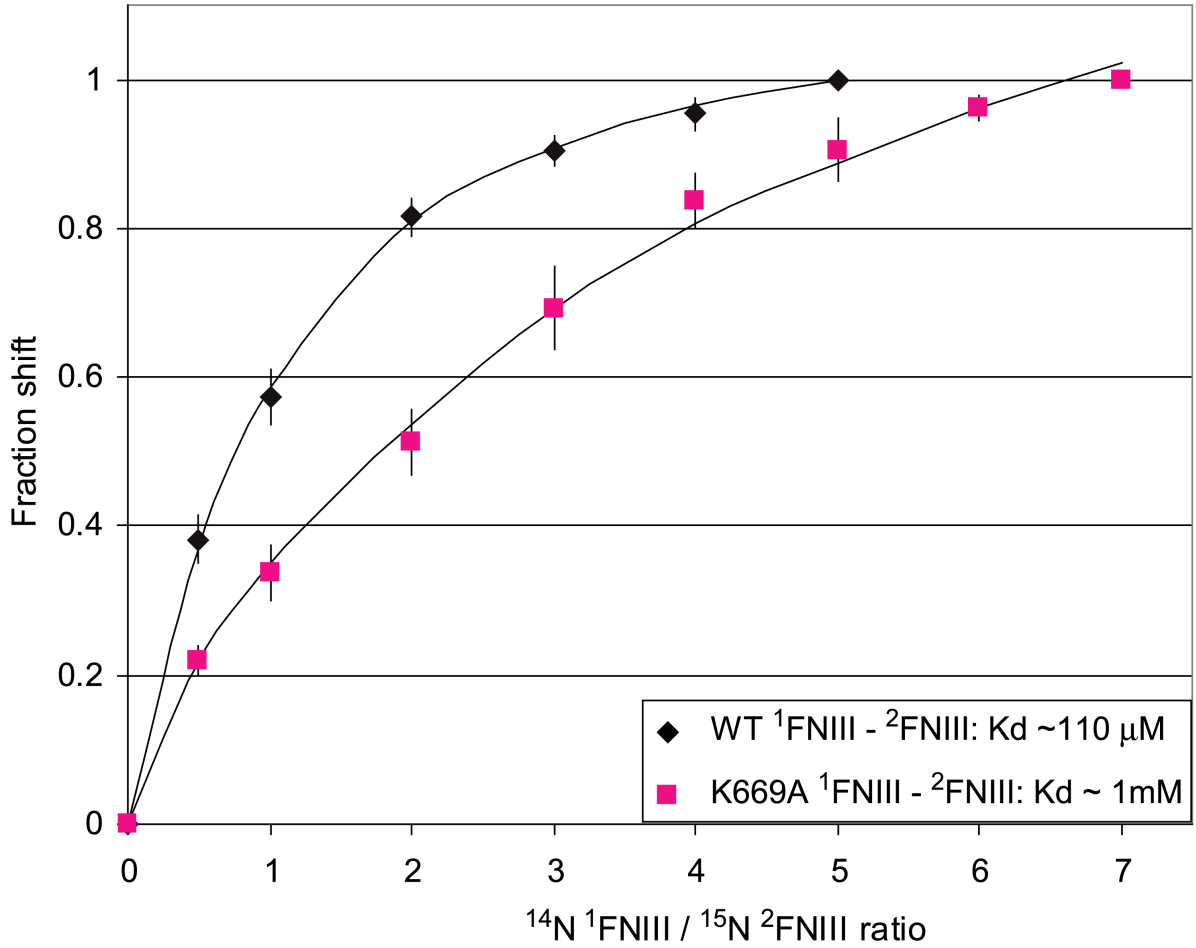

Figure 3. The chemical shifts of eight resonances (R751$^{\varepsilon}$, E755^{N}, E758^{N}, D767^{N}, V788^{N}, Q799$^{\varepsilon 2}$ and S800^{N}) of ^{15}N enriched ^{2}FNIII were monitored during titrations with unenriched wild-type ^{1}FNIII (black diamonds) or ^{1}FNIII K669A (purple squares) under the NMR conditions described. Averages of the fractional shifts $f_{\Delta\delta}=(|\Delta\delta^{H}|+|\Delta\delta^{N}/5|)/(|\Delta\delta^{H}{}_{max}|+|\Delta\delta^{N}{}_{max}/5|)$ of all eight resonances are plotted here against increasing stoichiometric ratios of the two domains. Error bars correspond to two standard deviations from the average. Data from all eight resonances were fit simultaneously using Origin assuming a single binding event but allowing for different chemical shift perturbation maxima. Solid lines correspond to calculated average chemical shift perturbations from the fit.

Figure 4.

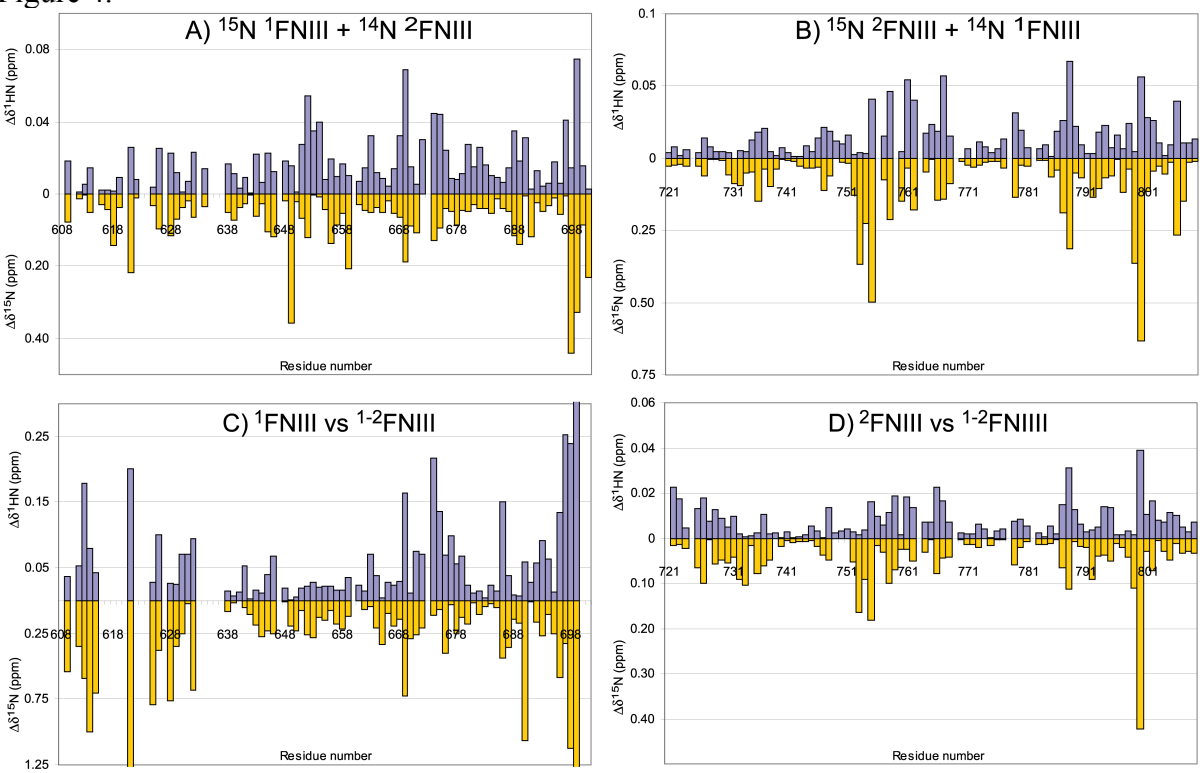

Figure 4. Shown here are ^{1}H and ^{15}N chemical shift perturbations versus residue number for (A,B) titrations of unenriched ^{2}FNIII in ^{15}N enriched ^{1}FNIII or unenriched ^{1}FNIII in ^{15}N enriched ^{2}FNIII, respectively and (C,D) ^{1}FNIII or ^{2}FNIII resonances, respectively, in the $^{1-2}$FNIII context compared to those of the independent domains.

Figure 5.

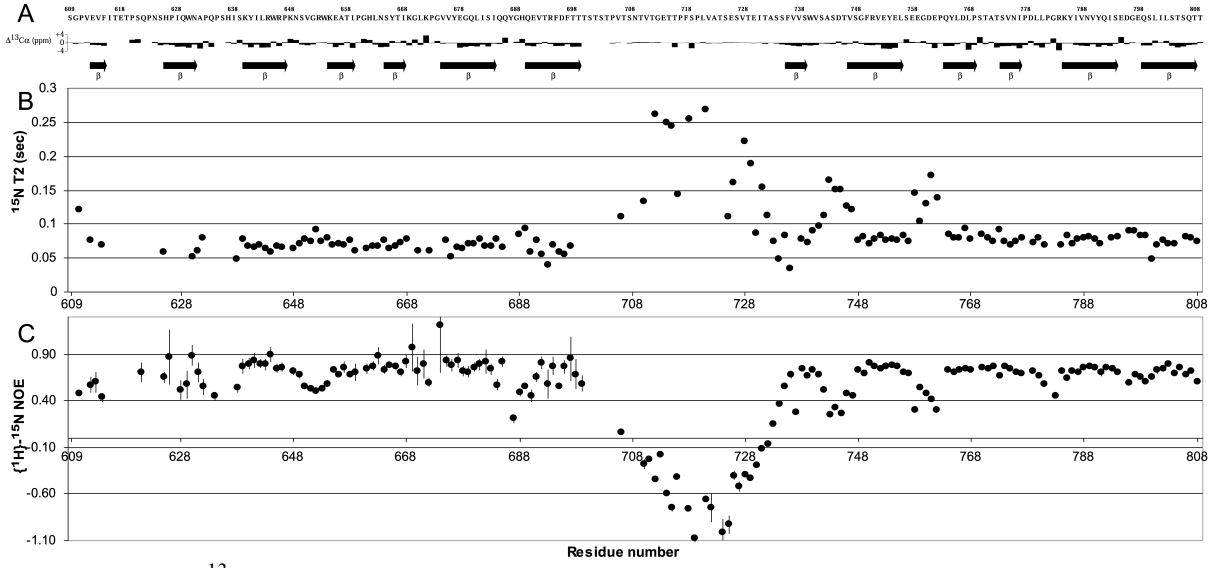

Figure 5. (A) $^{13}C_\alpha$ chemical shift differences compared to random coil chemical shifts (Wishart et al., 1995) versus residue number for $^{1\text{-}2}$FNIII. Backbone chemical shift assignments for $^{1\text{-}2}$FNIII have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number 7128. Secondary structure elements are indicated by arrows. (B) ^{15}N T_2 relaxation rates and (C) heteronuclear {^{1}H–}^{15}N NOE versus residue number for $^{1\text{-}2}$FNIII. Error bars were calculated from spectral noise and correspond to two standard deviations. Data were collected and analyzed as described previously (Vakonakis et al., 2004b) at 30^{o} C and 14.1 T (600 MHz ^{1}H frequency). The average ^{15}T2 relaxation rate observed for the structured domains is 73 msec, which is consistent with the two domains interacting in a closed conformation.

Figure 6.

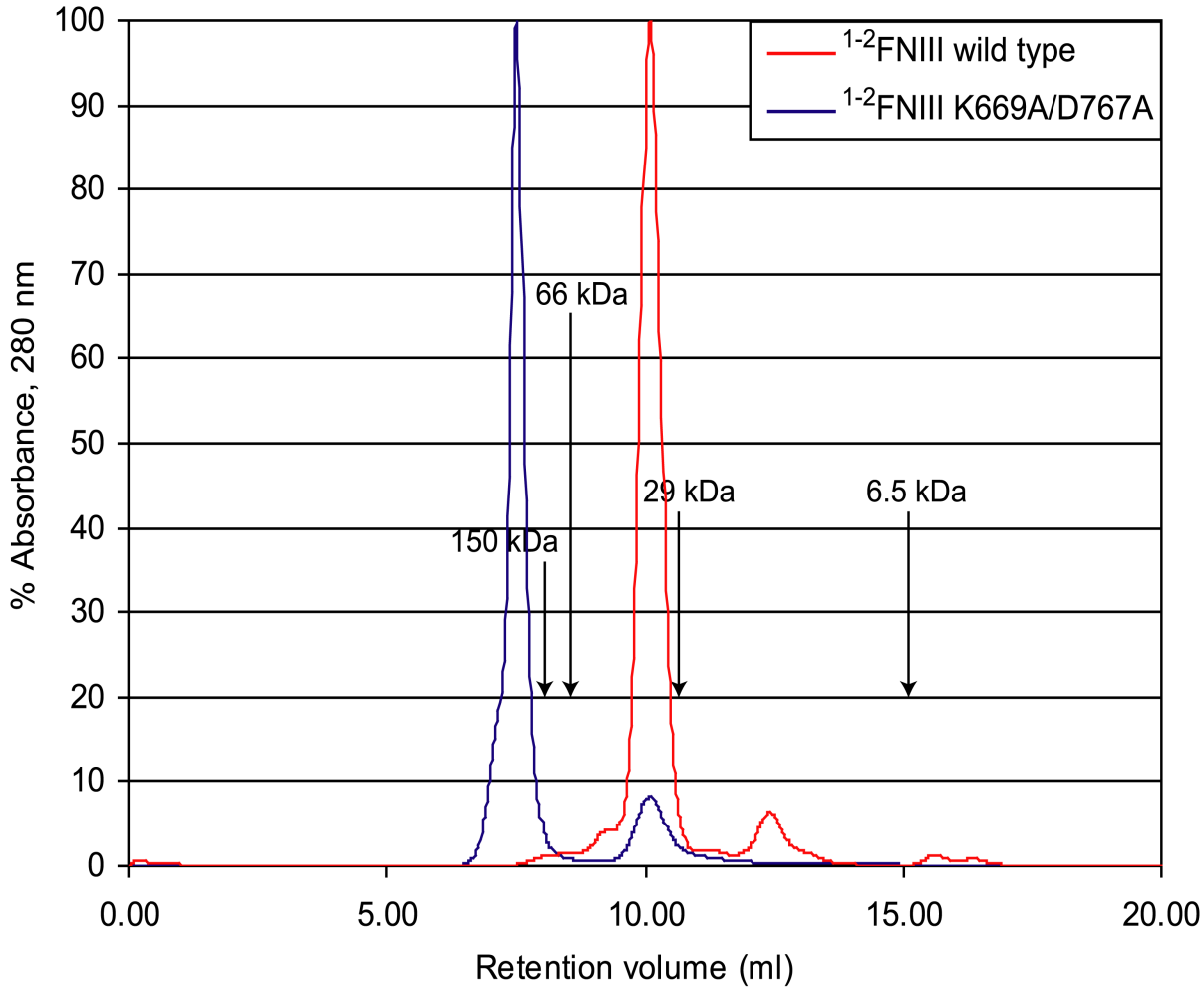

Figure 6. Analytical gel filtration elution profiles versus retention volume for $^{1\text{-}2}$FNIII wild-type (red) and K669A/D767A variant (blue). Retention volumes for the calibration markers, as well as their apparent molecular weights, are indicated. $^{1\text{-}2}$FNIII K669A/D767A is eluted at the column void volume indicating complete exclusion of this variant. This is consistent with the expected large average hydrodynamic radius of two non-interacting domains connected through a large flexible linker. The secondary, smaller peak observed for wild-type $^{1\text{-}2}$FNIII corresponds to a minor proteolytic fragment.

Figure 7.

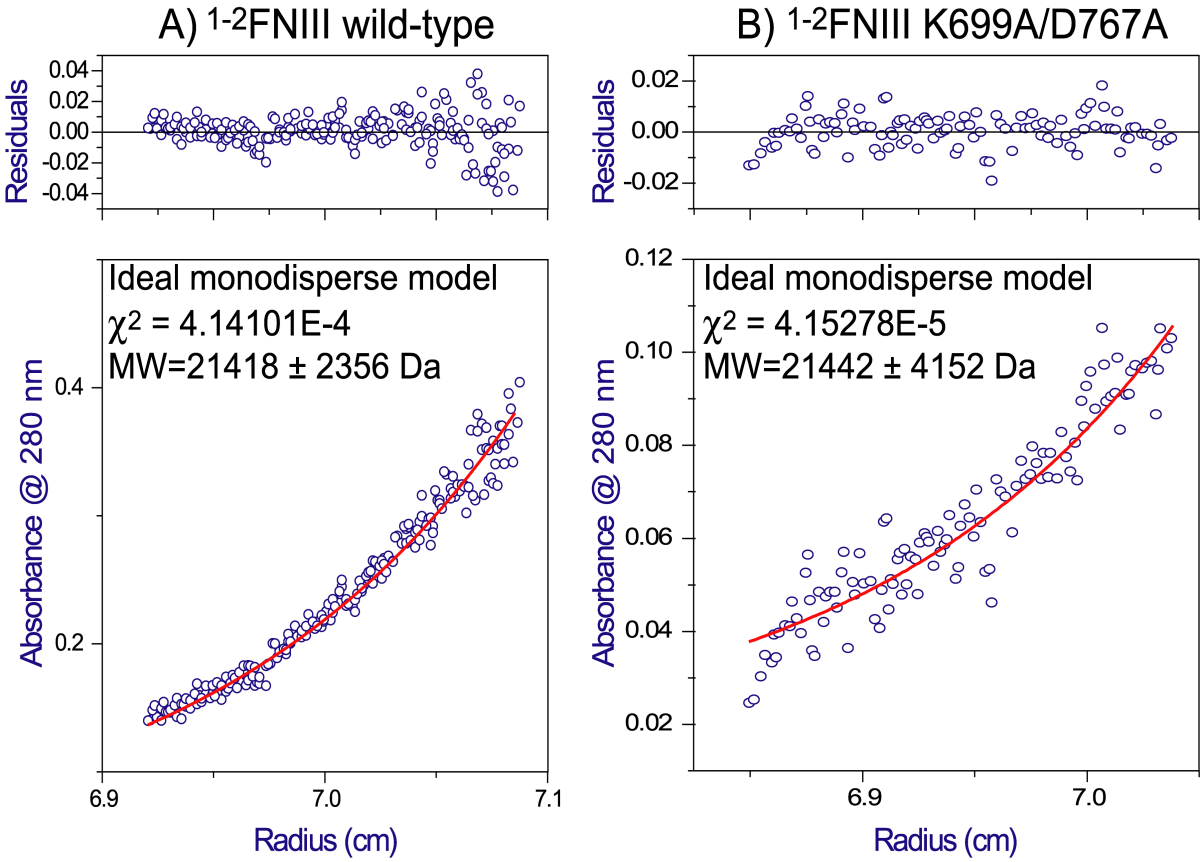

Figure 7. Analytical ultracentrifugation equilibrium absorbance versus rotor radius. Data were acquired on samples of (A) wild type or (B) K699A/D767A $^{1\text{-}2}$FNIII in PBS buffer and fit to an ideal monodisperse model shown as a solid line. Residuals of the fit are plotted against rotor radius at the top of the graphs. The molecular weights of both proteins from the fits are consistent with monomeric particles in solution (~22.5 kDa).

Figure 8.

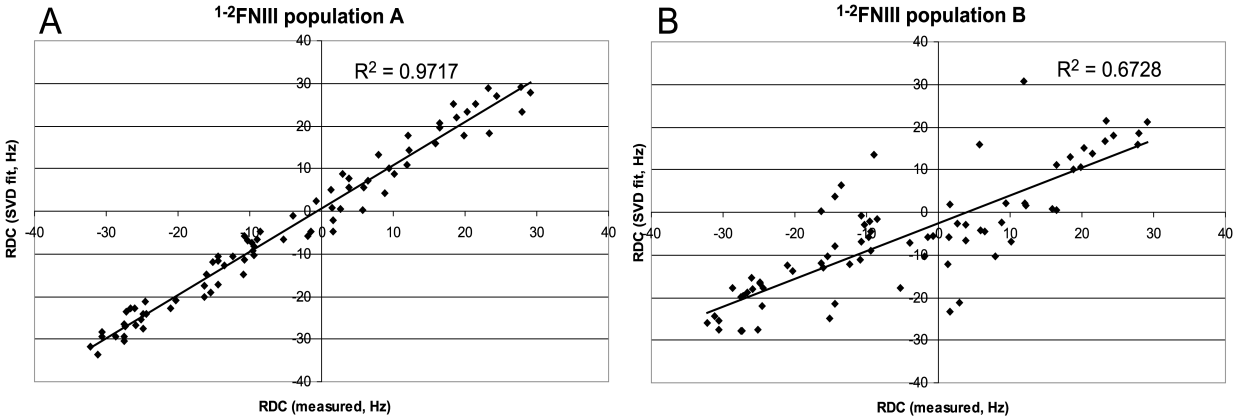

Figure 8. Singular value decomposition (SVD) fit of residual dipolar coupling (RDC) restraints to representative $^{1\text{-}2}$FNIII structural models of (A) population A of the calculated ensemble and (B) population B of the same ensemble. The RDC restraints used in this fit were derived from a $C_{12}E_5$ polyethylene glycol/hexanol medium (Ruckert and Otting, 2000). SVD fit was perform using the program PALES (Zweckstetter and Bax, 2000). The structural models used had the smallest backbone RMSD from the average of their respective population. The fit is significantly better for the population A model, indicating that this population better represents the correct $^{1\text{-}2}$FNIII structure.

Figure 9.

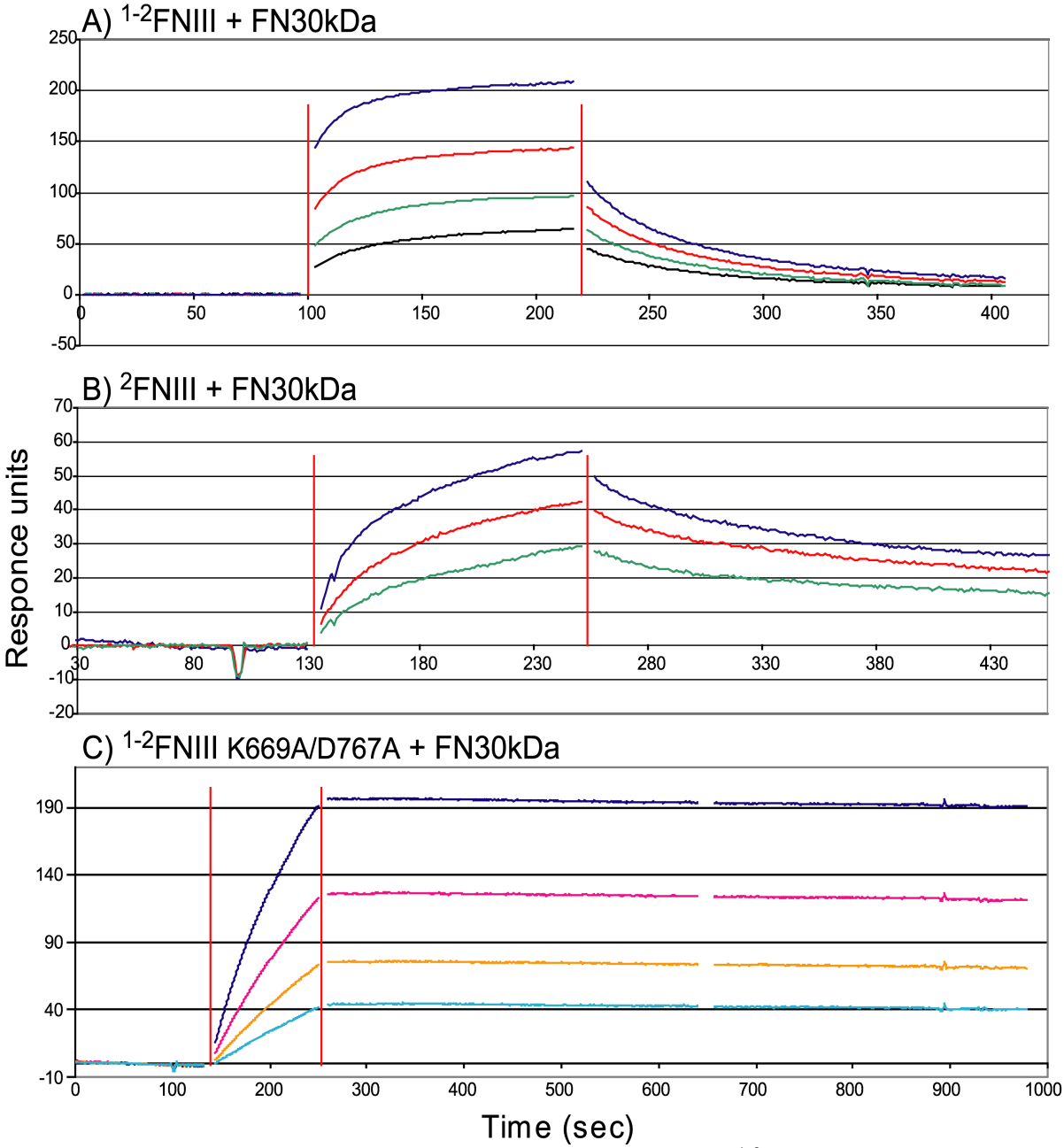

Figure 9. SPR sensogram traces of the interaction of wild type $^{1-2}$FNIII (17 to 270μM) (A), wild type ^{2}FNIII (20 to 80μM) (B) and $^{1-2}$FNIII K669A/D767A (31 to 250 nM) (C) with immobilized FN30kDa fragment. Data points close to injection start and end (denoted by red vertical lines) were removed for clarity. Sensogram traces of (C) featured a system spike at approximately 650 seconds, possibly as a result of system pump direction reversal. The corresponding data points were removed. Equilibrium analysis of (A) yielded a Kd value of approximately 85 μM for the interaction of wild type $^{1-2}$FNIII with FN30kDa. In contrast, kinetic analysis of (C) yielded a Kd value of approximately 1.4 nM for the equivalent interaction of $^{1-2}$FNIII K669A/D767A.

References

Chou, J.J., Gaemers, S., Howder, B., Louis, J.M. and Bax, A. (2001) A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR*, **21**, 377-382.

Clore, G.M. and Garrett, D.S. (1999) R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. *Journal of the American Chemical Society*, **121**, 9008-9012.

Clore, G.M. and Schwieters, C.D. (2003) Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1H/15N chemical shift mapping and backbone 15N-1H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. *J Am Chem Soc*, **125**, 2902-2912.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289-302.

Gao, M., Craig, D., Lequin, O., Campbell, I.D., Vogel, V. and Schulten, K. (2003) Structure and functional significance of mechanically unfolded fibronectin type III1 intermediates. *Proc Natl Acad Sci U S A*, **100**, 14784-14789.

Holm, L. and Sander, C. (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res*, **26**, 316-319.

Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1996) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci*, **5**, 1067-1080.

Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J Am Chem Soc*, **121**, 2337-2338.

Kuszewski, J., Qin, J., Gronenborn, A.M. and Clore, G.M. (1995) The impact of direct refinement against 13C alpha and 13C beta chemical shifts on protein structure determination by NMR. *J Magn Reson B*, **106**, 92-96.

Lipsitz, R.S., Sharma, Y., Brooks, B.R. and Tjandra, N. (2002) Hydrogen bonding in high-resolution protein structures: a new method to assess NMR protein geometry. *J Am Chem Soc*, **124**, 10621-10626.

Ottiger, M., Delaglio, F. and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson*, **131**, 373-378.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci*, **4**, 2411-2423.

Ruckert, M. and Otting, G. (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society*, **122**, 7793-7797.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Marius Clore, G. (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson*, **160**, 65-73.

Vakonakis, I., Klewer, D.A., Williams, S.B., Golden, S.S. and LiWang, A.C. (2004a) Structure of the N-terminal domain of the circadian clock-associated histidine kinase SasA. *J Mol Biol*, **342**, 9-17.

Vakonakis, I., Sun, J., Wu, T., Holzenburg, A., Golden, S.S. and LiWang, A.C. (2004b) NMR structure of the KaiC-interacting C-terminal domain of KaiA, a circadian clock protein: implications for the KaiA-KaiC interaction. *Proc. Natl. Acad. Sci. USA*, **101**, 1479-1484.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR*, **5**, 67-81.

Wüthrich, K. (1986) *NMR of proteins and nucleic acids*. Wiley, New York.

Zweckstetter, M. and Bax, A. (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *Journal of the American Chemical Society*, **122**, 3791-3792.