doi:10.1016/j.jmb.2009.08.049 J. Mol. Biol. (2009) 393, 409–424

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structural Studies of FF Domains of the Transcription Factor CA150 Provide Insights into the Organization of FF Domain Tandem Arrays

James M. Murphy[1]*, D. Flemming Hansen[2,3,4], Silke Wiesner[5], D. Ranjith Muhandiram[2,3,4], Mikael Borg[2,3,5], Matthew J. Smith[1,2], Frank Sicheri[1,2], Lewis E. Kay[2,3,4], Julie D. Forman-Kay[3,5]* and Tony Pawson[1,2]*

[1]*Program in Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5*

[2]*Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8*

[3]*Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8*

[4]*Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8*

[5]*Department of Molecular Structure and Function, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8*

Received 12 May 2009; received in revised form 6 August 2009; accepted 20 August 2009
Available online 26 August 2009

Edited by K. Morikawa

FF domains are poorly understood protein interaction modules that are present within eukaryotic transcription factors, such as CA150 (TCERG-1). The CA150 FF domains have been shown to mediate interactions with the phosphorylated C-terminal domain of RNA polymerase II (phosphoCTD) and a multitude of transcription factors and RNA processing proteins, and may therefore have a central role in organizing transcription. FF domains occur in tandem arrays of up to six domains, although it is not known whether they adopt higher-order structures. We have used the CA150 FF1 + FF2 domains as a model system to examine whether tandem FF domains form higher-order structures in solution using NMR spectroscopy. In the solution structure of FF1 fused to the linker that joins FF1 to FF2, we observed that the highly conserved linker peptide is ordered and forms a helical extension of helix α3, suggesting that the interdomain linker might have a role in orientating FF1 relative to FF2. However, examination of the FF1 + FF2 domains using relaxation NMR experiments revealed that although these domains are not rigidly orientated relative to one another, they do not tumble independently. Thus, the FF1 + FF2 structure conforms to a dumbbell-shape in solution, where the helical interdomain linker maintains distance between the two dynamic FF domains without cementing their relative orientations. This model for FF domain organization within tandem arrays suggests a general mechanism by which individual FF domains can manoeuvre to achieve optimal recognition of flexible binding partners, such as the intrinsically-disordered phosphoCTD.
© 2009 Elsevier Ltd. All rights reserved.

Keywords: relaxation; diffusion tensor; dynamics; ModelFree; phosphopeptide

**Corresponding authors.* J. M. Murphy is to be contacted at Division of Molecular Medicine, Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, VIC 3052, Australia. J. D. Forman-Kay, Department of Molecular Structure and Function, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8. T. Pawson, Program in Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5. E-mail addresses: jamesm@wehi.edu.au; forman@sickkids.ca; pawson@lunenfeld.ca.

Abbreviations used: TCERG1, transcription elongation regulator 1; Prp40, pre-mRNA processing factor 40; FBP11, formin-binding protein 11; LTR, long terminal repeat; CTD, C-terminal domain; phosphoCTD, phosphorylated CTD; NOE, nuclear Overhauser effect; ITC, isothermal titration calorimetry.

0022-2836/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.

Introduction

FF domains are poorly understood modular protein domains that are present in the eukaryotic proteome within two distinct families: (i) a subset of proteins that contain WW domains at the N-terminus and (ii) the p190 family of RhoGAPs. The WW domain-containing family of FF proteins, comprising CA150 (also termed transcription elongation factor 1 or TCERG1), the yeast pre-mRNA processing factor 40 (Prp40) and the mammalian Prp40 orthologs known as formin-binding protein 11 (FBP11), serve as nuclear transcription and splicing factors.[1-5] In contrast, the p190 RhoGAPs are cytoplasmic proteins, and in the case of p190A, the component FF domains serve to sequester the transcription factor TFII-I in the cytoplasm.[6] Although the FF domains within both protein families mediate distinct sets of protein–protein interactions,[6,7] both protein families are pivotally involved in the regulation of gene transcription processes.

Interestingly, within both of these protein families, FF domains occur in tandem arrays of four to six domains separated by linker sequences of variable length.[4] The organization of FF domains in tandem arrays is likely to be important for their biological function. For example, several studies indicate that FF domains bind protein interactors *via* multiple weak interactions, and the tandem arrangement of FF domains serves to increase avidity for binding partners.[7-9] To date, it is unclear whether the FF domains within these tandem arrays show any 3D organization or whether the FF domains are merely tethered to one another *via* sequences of flexible random coil. Although the possibility of the tandem arrays adopting higher-order structures has been suggested,[6] this hypothesis has not been tested.

One of the WW domain-containing FF domain proteins, CA150, is an evolutionarily conserved transcription factor for which orthologs are known from nematodes to humans. CA150 was initially identified as a cellular repressor of transcriptional elongation from the HIV-1 long terminal repeats (LTR) locus.[1,2] Full-length CA150 is composed of three WW domains at the N-terminus of the protein (WW1–3), followed by three FF domains (FF1–3), a segment of 48 amino acids, and then three C-terminal FF domains (FF4–6) (Fig. 1a). Several studies have established that both the CA150 WW domains[5,10-12] and the FF domains[7] make extensive protein–protein interactions within the nucleus. While the WW domains were found to interact primarily with components of the spliceosome,[5,10,12] the FF domains were found to interact with a plethora of transcription factors and splicing factors.[7] However, to date, FF domains of the WW domain-containing protein family have been primarily described as phosphopeptide interaction modules and the FF domains of CA150, FBP11/HYPA and Prp40 have all been implicated in binding the phosphorylated C-terminal domain (CTD) of RNA polymerase II.[7-9,11,13,14] The CTD is an intrinsically disordered polypeptide composed of tandem repeats of the heptad consensus motif Y1-S2-P3-T4-S5-P6-S7. The number of repeats of this heptad consensus varies between organisms, from 17 in *Plasmodium falciparum* to 52 in the mammalian CTD.[15,16] Each heptad repeat within the CTD is subject to phosphorylation at Ser2 and/or Ser5 *in vivo*[17] and this phosphorylation pattern governs which transcription and pre-mRNA processing factors are recruited to the transcription complex. In the case of CA150, both the WW domains and FF domains were found to bind the phosphorylated CTD (phosphoCTD) in Far Western assays.[8] The predominant phosphoCTD binding activity was attributed to the FF domains, and of the six individual FF domains, only FF2 and FF5 displayed detectable phosphoCTD binding.[8]

In the present work, we sought to establish whether the FF domains within tandem arrays are organized into higher-order structures by using the N-terminal two FF domains (FF1 and FF2) of CA150 as a model system for structural examination by NMR. Here, we present the solution structure of the FF1 domain of CA150 and the interdomain linker that joins FF1 to FF2. The core FF domain structure resembles the FF domain structures described to date[13,18,19] but, unexpectedly, the interdomain linker is structured and forms contacts with FF1, which led us to ask whether the interdomain linker has a role in positioning FF2 relative to FF1. NMR relaxation experiments performed on an FF1 + FF2 fragment of CA150 suggest that FF1 and FF2 are not orientated rigidly with respect to one another, although the high degree of anisotropy in tumbling indicates that these domains do not tumble independently. These data indicate that the interdomain linker probably serves to maintain the domains at a specific distance without cementing their relative orientations. By performing chemical shift perturbation NMR studies, we probed whether CA150 FF1, FF2 and FF3 each possess an intrinsic phosphoCTD binding ability. Surprisingly, we were unable to detect interactions between any of these domains and tandem phosphoCTD peptides, indicating that CTD binding by CA150 FF domains is dependent on the multiplicity of FF domains and CTD heptad motifs and thus the avidity arising from multiple weak interactions. Our data suggest a model for the organization of FF domains within tandem arrays and provide a rationale for the existence of both FF domains and CTD heptad repeats within tandem arrays in nature.

Results

FF1–3 represent an evolutionarily conserved cassette

The alignment of amino acid sequences revealed remarkably high levels of sequence conservation

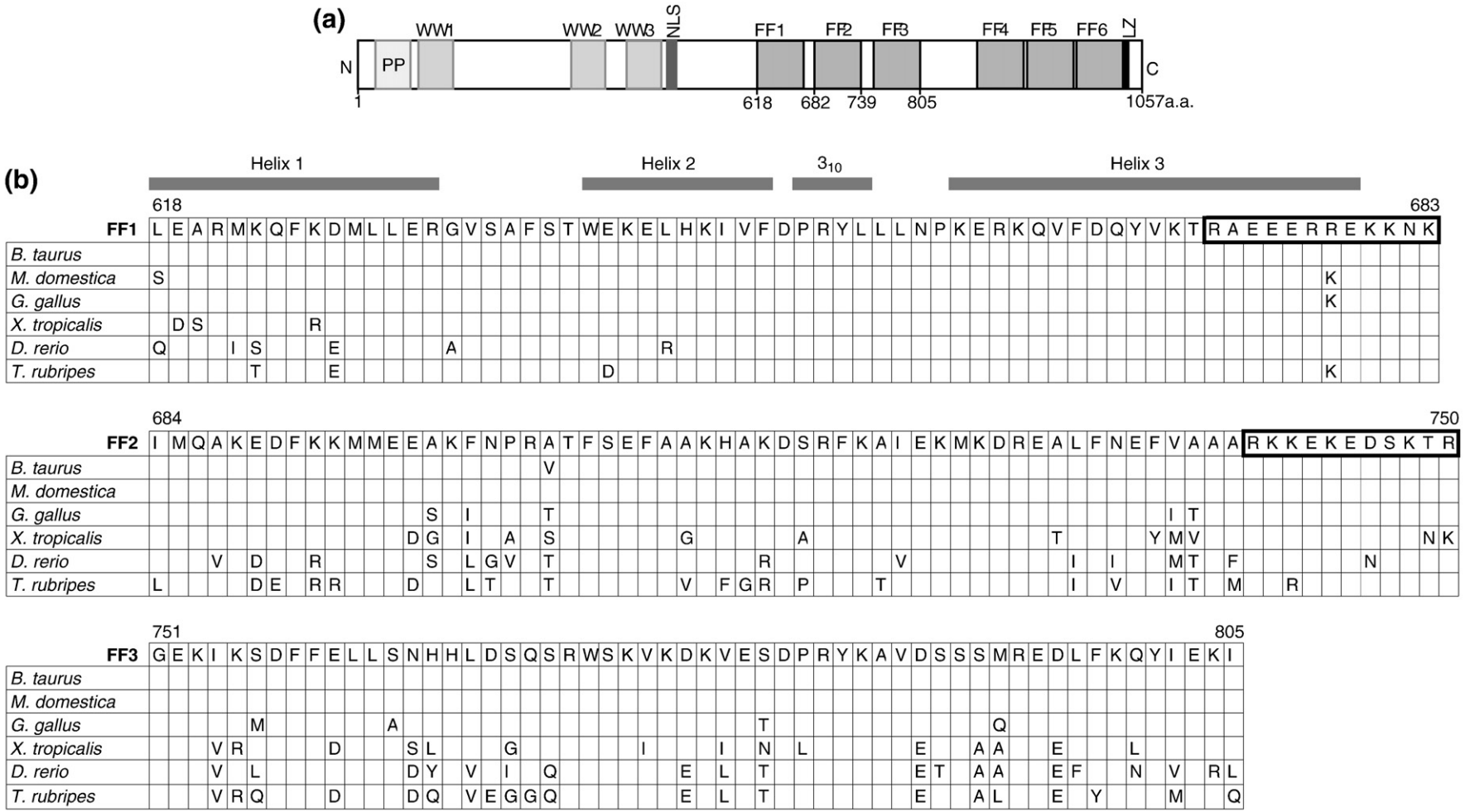

Fig. 1. (a) Domain architecture of CA150. The Pro-rich region (PP), WW domains (WW1–3), nuclear localization signal peptide (NLS), FF domains (FF1–6) and putative leucine zipper (LZ) are depicted schematically. Numbers below the figure correspond to the murine CA150 amino acid sequence. (b) Sequence alignment of vertebrate CA150 FF1–3 amino acid sequences. The uppermost sequence is absolutely conserved in *H. sapiens* (human), *M. musculus* (mouse), *R. norvegicus* (rat), *P. troglodytes* (chimpanzee), *M. mullata* (rhesus monkey) and *C. familiaris* (dog). Amino acid differences in other vertebrate sequences are written below the relevant residue. Secondary structure elements within the FF1 + linker structure are drawn above the sequence. Boldly boxed sequences represent the linker sequences between FF domains. *B. taurus* (bovine), *M. domestica* (opposum), *G. gallus* (chicken), *X. tropicalis* (frog), *D. rerio* (zebrafish), *T. rubripes* (pufferfish).

between the FF domains of CA150 orthologs (Fig. 1b). We focused on the N-terminal three FF domains (denoted FF1–3) because these FF domains were predicted to be joined by short "linker" sequences of 11 or 12 residues. Like FF1–3, the FF4–6 domains are joined by short linkers (two amino acids), whereas the sequence connecting FF3 to FF4 is markedly longer (48 residues). The long sequence dividing FF3 and FF4 led us to consider FF1–3 and FF4–6 as two distinct modules. It is noteworthy that the high level of interspecies conservation extends to the linker sequences that connect the FF1-FF2 and FF2-FF3 domains (boxed in Fig. 1b), suggesting an essential role for the linker sequences in CA150 structure and/or function.

The FF1–FF2 interdomain linker interacts with the FF1 domain

In order to probe the organization of the CA150 FF domains, we examined the FF1 and FF2 domains using NMR. Comparison of 2D ^{15}N-^{1}H HSQC spectra of ^{15}N-labelled CA150 fragments corresponding to (a) FF1 + FF2 (residues 611–739 in full-length murine CA150), (b) FF1 alone (residues 618–671) and (c) FF2 alone (residues 682–739) indicated that, on the whole, FF2 resonances transposed directly onto a subset of the FF1 + FF2 peaks (Fig. 2b), while numerous chemical shift changes were observed in the positions of FF1 peaks in the FF1 + FF2 spectrum (Fig. 2a). In order to examine whether the chemical shift perturbations observed for FF1 resonances in the FF1 + FF2 construct resulted from interactions between the FF1 and FF2 domains, ^{15}N-^{1}H HSQC spectra were recorded on a ^{15}N-labelled FF1 sample alone and in the presence of unlabelled FF2 up to a 10-fold molar excess. No chemical shift perturbations were observed, indicating that there is no direct interaction between FF1 and FF2 in solution (data not shown).

Subsequently, we examined whether FF1 interacts with the linker sequence that joins FF1 and FF2 in native CA150 by comparing ^{15}N-^{1}H HSQC spectra of FF1 (618–671) and a longer FF1 construct incor-

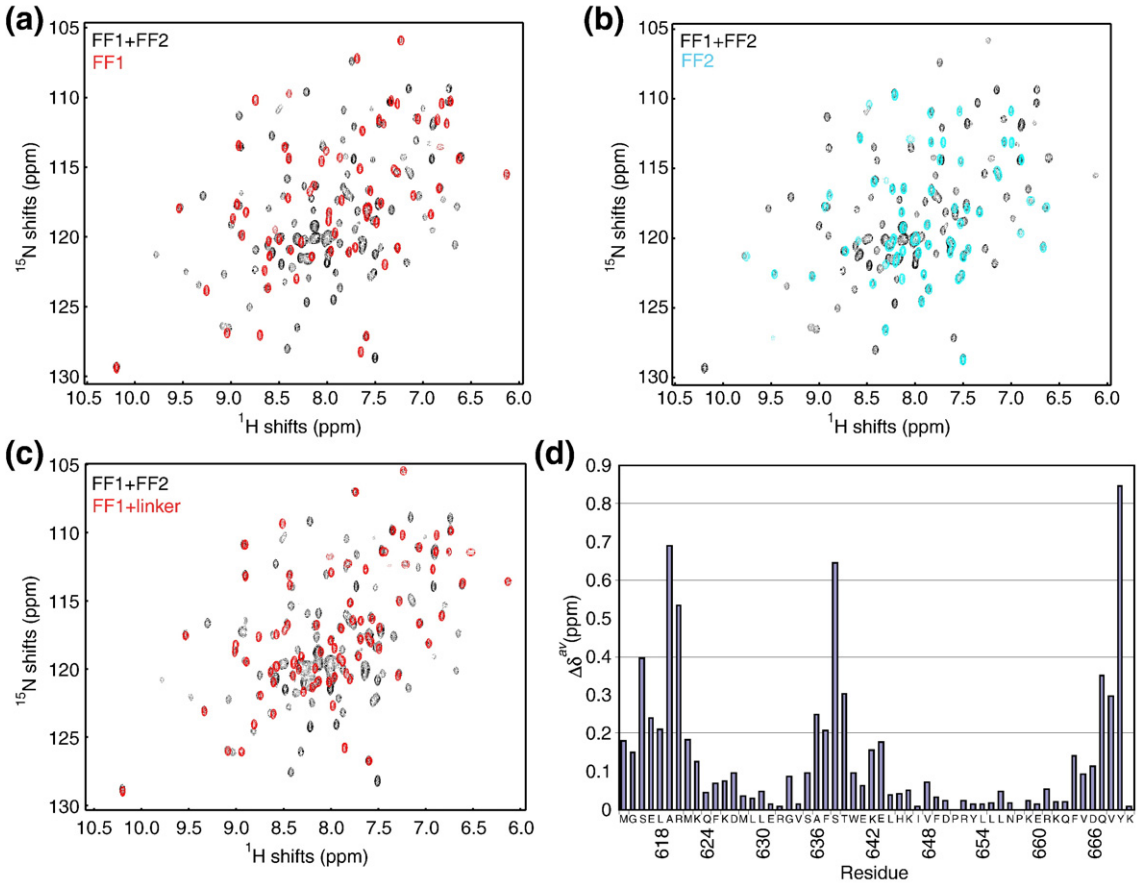

Fig. 2. The linker that connects CA150 FF1 and FF2 interacts with FF1. Overlays of FF1 + FF2 (residues 611–739; black peaks) and: (a) FF1(residues 618–671; red peaks); (b) FF2 (residues 682–739; cyan peaks); and (c) FF1 + linker (residues 618–683; red peaks) ^{1}H-^{15}N HSQC spectra. Most FF2 peaks correspond with peaks within the FF1 + FF2 spectrum (b) but extensive chemical shift changes are observed for FF1 peaks within the FF1 + FF2 construct (a). (c) The chemical shift changes in the FF1 spectrum — relative to the FF1 + FF2 spectrum — arise from connection of FF1 to the FF1–FF2 interdomain linker. (d) Average chemical shift perturbations for peaks in ^{1}H-^{15}N HSQC spectra of FF1 + linker (618–683) peaks *versus* FF1 (618–671). The inclusion of the linker that connects FF1 to FF2 in the FF1 + linker construct induces chemical shift perturbations relative to cognate peaks within the spectrum of FF1 (618–671).

porating the linker between FF1 and FF2 (618–683). This comparison showed that the presence of the linker sequence induced the changes in FF1 chemical shifts, implicating the linker in contacts with FF1. Comparison of the ^{15}N-^{1}H HSQC spectra obtained for the FF1 + linker and FF1 + FF2 proteins show excellent correspondence of backbone peaks (Fig. 2c), indicating that the FF1 + linker protein is representative of the FF1 structure in the context of the FF1 + FF2 construct. After assigning the backbone amide resonances of FF1(618–671) and FF1 + linker(618–683), we were able to map backbone amide chemical shift changes induced by the presence of the interdomain linker to the N-terminal sequence of FF1, residues 636–639 and residues in the C-terminus (Fig. 2d).

Solution structure of CA150 FF1 + linker

In order to better understand the nature of the interaction between FF1 and the linker that joins FF1 and FF2, we determined the structure of FF1 + linker (residues 618–683) using 3D NMR methods (Fig. 3a, c and d). Analysis of 3D NMR spectra enabled nearly complete assignment of the ^{1}H, ^{15}N and ^{13}C resonances. A total of 3226 non-redundant nuclear Overhauser effects (NOE) restraints that were applied with 29 hydrogen bond restraints and 40 dihedral angle restraints derived from secondary chemical shifts and NOE patterns (data not shown) were used for structure calculation. The statistics for the 20 lowest energy structures of the FF1 + linker are shown in Table 1.

The CA150 FF1 domain adopts a three α helix fold consisting of: α1 (residues 618–632), α2 (residues 640–649), α3 (residues 659–679) and a 3_{10} helix in the α2-α3 loop (residues 651–654) that begins with the proline of the conserved DPRY motif (Fig. 3a). The three-helical bundle fold is characteristic of the FF domain structures described to date (Fig. 3b).[13,18,19] Consistent with the previously determined FF domain structures, the FF1 + linker structure contains a network of semi-conserved aromatic (Phe625, Trp640, Tyr653, Phe665, Tyr668) and aliphatic (Leu629, Leu644, Ile647, Leu656) sidechains that form the core of the domain. However, in contrast to the previously described FF domains, the FF1 + linker structure shows an extended α3 helix (Fig. 3a and c). This helical extension, which was unexpected, arises from the inclusion of the highly conserved linker sequence between FF1 and FF2 — a sequence of previously unknown structure. Notably, this C-terminal helical extension is in proximity to the α1-α2 loop residues, Ser635, Ala636 and Phe637, in the FF1 + linker structure — accounting for the chemical shift changes of residues in this loop when compared to the resonances for the corresponding residues in the shorter FF1(618–671) (Fig. 2d). The C-terminal helical extension also induces chemical shift changes in the N- and C-terminal residues (Fig. 2d). The chemical shift changes in the C-terminal residues are a consequence of the presence of additional residues at the C-terminus, while the chemical shift changes in the N-terminal residues arise from the interaction of the neighboring α1-α2 loop with the α3 helix stabilizing the packing of the α1 helix onto the domain structure and preventing rigid body movements of the helix. The α1-α2 loop residues, Phe637, Ser638 and Thr639, show an extended β-strand conformation within the structure (Fig. 3a and c). This extended conformation is consistent with the large couplings observed for these residues in a $^{3}J(H^{N},H^{\alpha})$ experiment (data not shown). The solution structure of the core CA150 FF1 domain was determined by the RIKEN Structural Genomics/Proteomics Initiative during the course of this work (PDB code 2dod). A structural superimposition (Supplementary Data Fig. 1) illustrates that the core FF1 domain shows excellent agreement between the structure described here and the PDB structure 2dod (RMSD 1.45 Å). In the present work, however, the inclusion of the C-terminal interdomain linker has enabled the identification of the previously unrecognized helical extension from the FF1 α3 helix and led us to characterize the role of this interdomain helix in orienting FF1 relative to the neighboring FF2 domain (described below).

Table 1. Structural Statistics For The 20 Lowest Energy Structures Of CA150 FF1+Linker(618-683)

Number of structural restraints	
All	3226
Sequential ($\|i-j\| = 1$)	771
Medium range ($2 \le \|i-j\| \le 4$)	549
Long range ($\|i-j\| > 4$)	482
Intraresidual	1352
Ambiguous	0
Hydrogen bonds	29
r.m.s.d. from experimental restraints[a]	
All distance restraints (Å)	0.018±0.003
Hydrogen bonds (Å)	0.045±0.003
Dihedral angles (°)	0.11±0.10
r.m.s.d. from idealized covalent geometry	
Bond lengths (Å)	0.0027±0.0002
Bond angles (°)	0.45±0.007
Improper dihedral angles (°)	0.34±0.02
Average atomic r.m.s.d. from the mean structure (Å)	
Residues in secondary structure elements (N, C^{α}, C′)	0.22±0.03
Residues in secondary structure elements (all heavy atoms)	0.85±0.09
All residues (N, C^{α}, C′)	1.44±0.33
All residues (all heavy atoms)	1.75±0.23
Ramachandran plot statistics	
Residues in most favored regions[b] (%)	77.6
Residues in additionally allowed regions (%)	20.7
Residues in generously allowed regions (%)	1.7

[a] No distance restraint was violated by >0.3 Å and no dihedral angle restraint was violated by >5°.
[b] Excluding glycine and proline residues.

Characterization of conformational dynamics within the CA150 FF1 + linker

We used ^{15}N-relaxation experiments to examine the internal motions and overall tumbling of the FF1 + linker fragment of CA150 (Fig. 4a and b). The $\{^{1}H\}$-^{15}N heteronuclear NOEs are almost uniform, with values between 0.68 and 0.83 for residues 618–

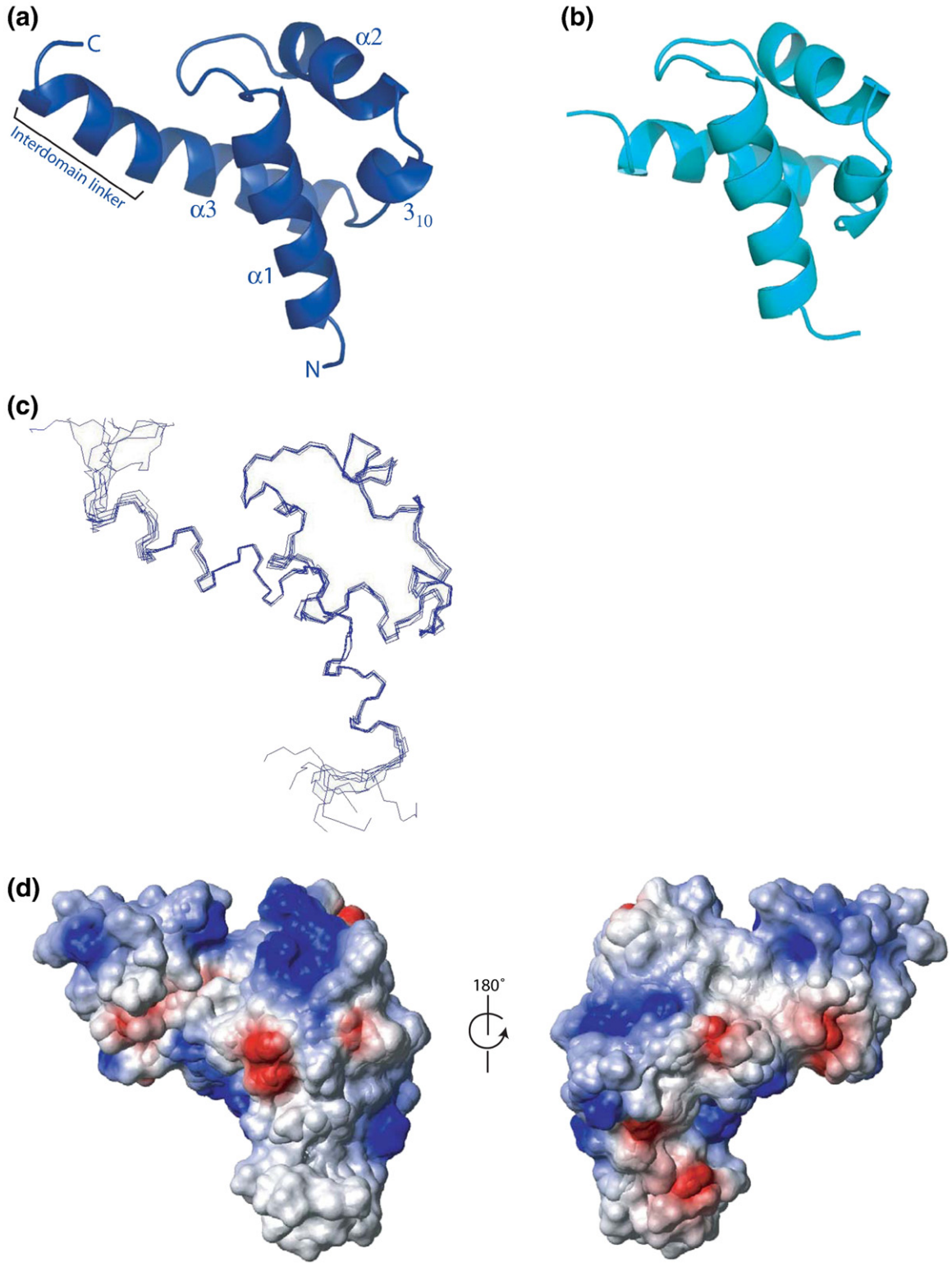

Fig. 3. Solution structure of CA150 FF1 + linker(618–683) reveals that the interdomain linker has stable helical structure. (a) Cartoon representation of the solution structure of CA150 FF1 + linker. Helices and the C-terminal helical extension formed by the interdomain linker are labeled. This figure was prepared using PyMOL (www.pymol.org). (b) Cartoon representation of the archetypal FF domain structure: HYPA/FBP11 FF1 (PDB accession number 1UZC) prepared using PyMOL. (c) Backbone superimposition of the seven lowest energy CA150 FF1 + linker structures (of the 20 calculated). This figure was prepared using MOLMOL.[20] (d) Electrostatic surface potential representation of CA150 FF1 + linker illustrating the extensive basic surface patches. This figure was prepared using MOLMOL.[20] Left: Surface potential drawn in the same orientation as FF1 + linker in a and c. Right: 180° rotation about the *y*-axis.

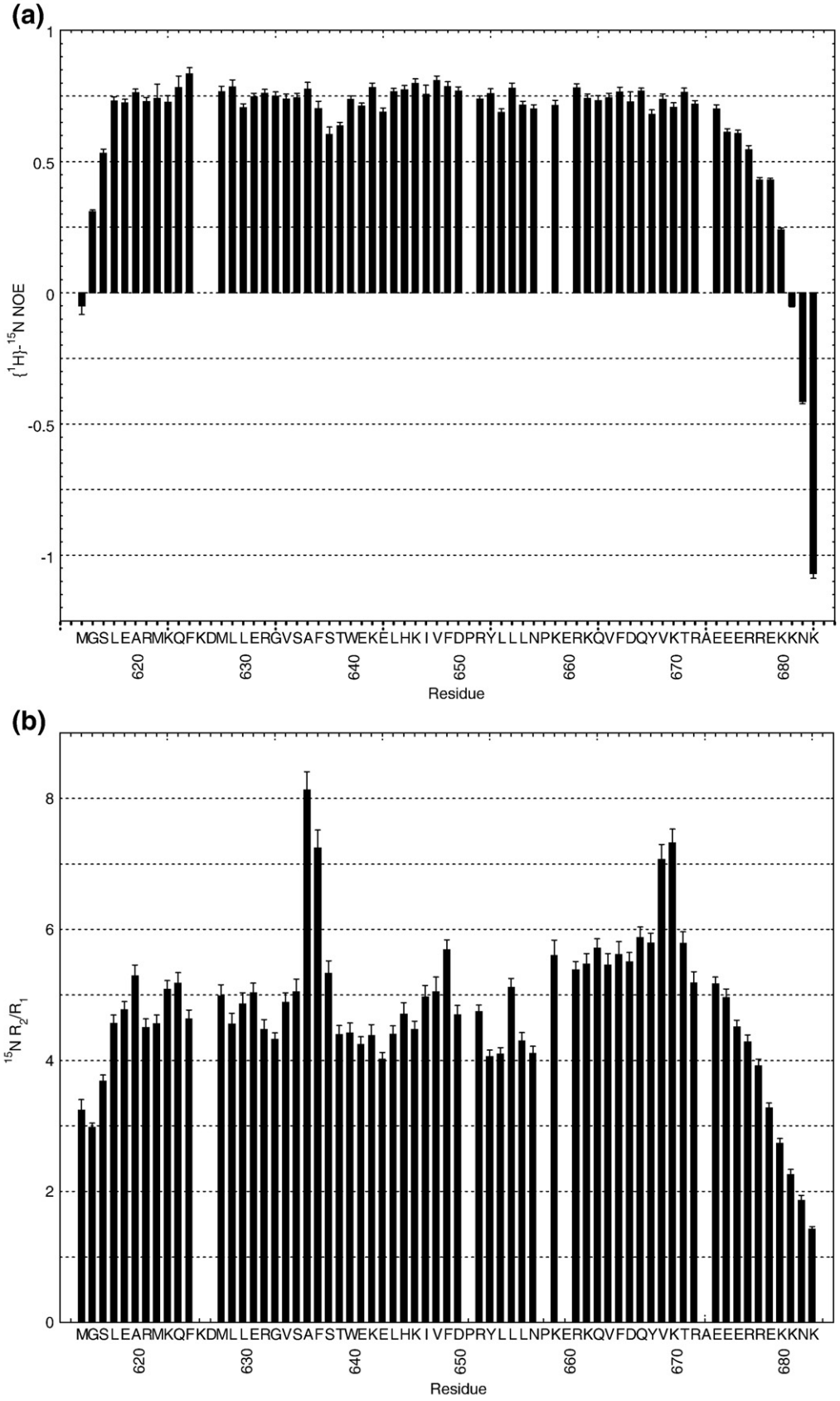

Fig. 4. ^{15}N-Relaxation studies of FF1+linker. (a) Heteronuclear {^{1}H}-^{15}N NOEs; (b) R_2/R_1 ratios. Data were not obtainable for proline residues due to absence from the ^{1}H-^{15}N HSQC spectra and some overlapped peaks. Heteronuclear NOEs confirm that the interdomain linker is ordered and, with the exception of terminal residues, relatively rigid on the nano- to picosecond timescale.

674 (i.e. excluding the highly flexible residues at the N- and C-termini). This indicates that the FF1 + linker structure tumbles as a single, folded unit, is relatively rigid and devoid of extensive backbone motions on the nano- to picosecond timescale. Notably, the R_2 values in the α1-α2 loop residues, Ala636 and Phe637, and the spatially adjacent α3 helix residues, Val669 and Lys670, are elevated relative to the remainder of the FF1 + linker residues (Supplementary Data Fig. 2A), providing strong evidence for these regions undergoing chemical exchange on the milli- to microsecond timescale.

Interdomain motions of the FF domains within the FF1 + FF2 fragment

In order to examine whether the helical interdomain linker has a role in orientating FF1 with respect to the neighboring FF2 domain, we turned our attention to the FF1 + FF2(611–739) fragment of CA150. Initially, we prepared a ^{13}C-^{15}N-labelled FF1 + FF2(611–739) sample and employed 3D NMR methods to assign the backbone ^{1}H-^{15}N resonances. Analysis of spectra enabled assignment of 88% of backbone resonances, as peaks corresponding to residues 677–680, Glu689, 693–696, Phe719 and Met725 were either unable to be assigned due to spectral overlap or were absent from the spectra, presumably due to exchange broadening. This is in contrast to the lack of broadening the linker region found in the case of the FF1 + linker. Subsequently, we performed ^{15}N relaxation NMR experiments in order to determine whether the tumbling of the two component FF domains shows a degree of autonomy, or whether the two FF domains tumble uniformly within a rigid structure held in place by the interdomain helix. NMR relaxation provides information about the overall tumbling of the molecule as well as information of local motions. As a result, relaxation analysis – using the orientational dependence of nuclear spin relaxation – can be used to determine the relative orientations of the two component FF domains within the FF1 + FF2 tandem domain construct without residual dipolar couplings. The ratio of the transverse and longitudinal ^{15}N relaxation rates, R_2/R_1 was calculated (Fig. 5) and used to determine the diffusion tensors of each domain within the FF1 + FF2 construct. The ratio R_2/R_1 is dominated by contributions from the overall diffusion of the protein; the contributions of local picosecond dynamics to R_2/R_1 can, to good approximation, be neglected.[21] The magnitudes (τ_{iso}) and the anisotropies ($D_{||}/D_{\perp}$) of the diffusion tensors of FF1 and FF2 within the FF1 + FF2 fragment were determined independently, assuming an axially symmetric diffusion model. The magnitudes and anisotropies of the diffusion tensors amount to values:

$$\tau_{1,iso} = 0.5/\left(D_{1,\parallel} + 2D_{1,\perp}\right) = 18.5 \pm 0.6 \text{ ns and}$$
$$D_{1,\parallel}/D_{1,\perp} = 1.93 \pm 0.18 \text{ for FF1; and}$$

$$\tau_{2,iso} = 15.5 \pm 0.4 \text{ ns and } D_{2,\parallel}/D_{2,\perp} = 2.21 \pm 0.20 \text{ for FF2}$$

with errors determined from a jack-knife analysis.[22] *F*-statistics comparing different diffusion models show that both diffusion tensors are axially symmetric rather than isotropic; the probability that the improved fit of the axially symmetric model over the isotropic model as the result of chance is less than 2×10^{-5} for FF1 and less than 2×10^{-8} for FF2. No further improvement to the fits was obtained when a completely anisotropic model of diffusion was invoked. The results of these calculations demonstrate that the two domains do not tumble as a rigid entity, with the magnitude of the diffusion tensors (τ_{iso}) of FF1 and FF2 being significantly different (15.5±0.4 ns for FF2 *versus* 18.5±0.6 ns for FF1). Both the magnitude and the anisotropy ($D_{||}/D_{\perp}$) of the two diffusion tensors would be identical if the two domains in the FF1 + FF2 fragment tumbled as a single rigid unit. Critically, this analysis of ^{15}N relaxation NMR data demonstrates that the two component FF domains within the FF1 + FF2 construct are distanced by the helical interdomain linker, as helix α3(FF1) and α1(FF2) are approximately colinear when the structures of FF1 and FF2 are orientated in the principal coordinate system of their individual diffusion tensors (Fig. 6). Owing to the absence of several amide chemical shift assignments in the FF1-FF2 interdomain linker, presumably due to exchange-broadening, we cannot decisively determine the secondary structure of the linker that joins the two FF domains. However, it is noteworthy that the ^{15}N-^{1}H heteronuclear NOEs that can be measured in the FF1-FF2 linker helix (i.e. those between residues 658 and 695) are similar to those obtained for residues within each of the component FF domains (Fig. 5a). These data demonstrate that the picosecond timescale dynamics of the interdomain linker residues are akin to residues known to reside within folded regions: an observation consistent with the interdomain linker existing as a helix, as observed in the crystal structure of FF1–3 reported in the accompanying paper,[23] where the CA150 inter-FF domain linker sequences were found to adopt extended α helices. Moreover, the tumbling of the two fragments is not independent. Relaxation experiments performed on the FF1 + linker alone establish that it tumbles with τ_{iso} = 7.5±0.1 ns and $D_{||}/D_{\perp}$ = 1.35±0.04, values that are substantially different from those obtained for FF1 in the FF1+FF2 construct ($\tau_{1,iso}$ = 18.5±0.6 ns, $D_{1,||}/D_{1,\perp}$ = 1.93±0.18). Yet, the FF domains are not completely rigid relative to each other. For example, the χ^2 value of a least-squares fit of experimental to calculated ^{15}N R_2/R_1 ratios when FF1 and FF2 domains are analyzed simultaneously using the model structure of the FF1 + FF2 complex (Fig. 6), is significantly higher (265) than the sum of χ^2 values when each of the domains is considered independently χ^2 = 50.1 (FF1) and χ^2 = 28.2 (FF2) (with corresponding *p*-value of 6×10^{-10}). Finally, hydrodynamic calculations of $D_{||}/D_{\perp}$ performed using the *diffc* program within the DASHA package†,[24,25] based on the

† http://www.nmr.ru/dasha.html

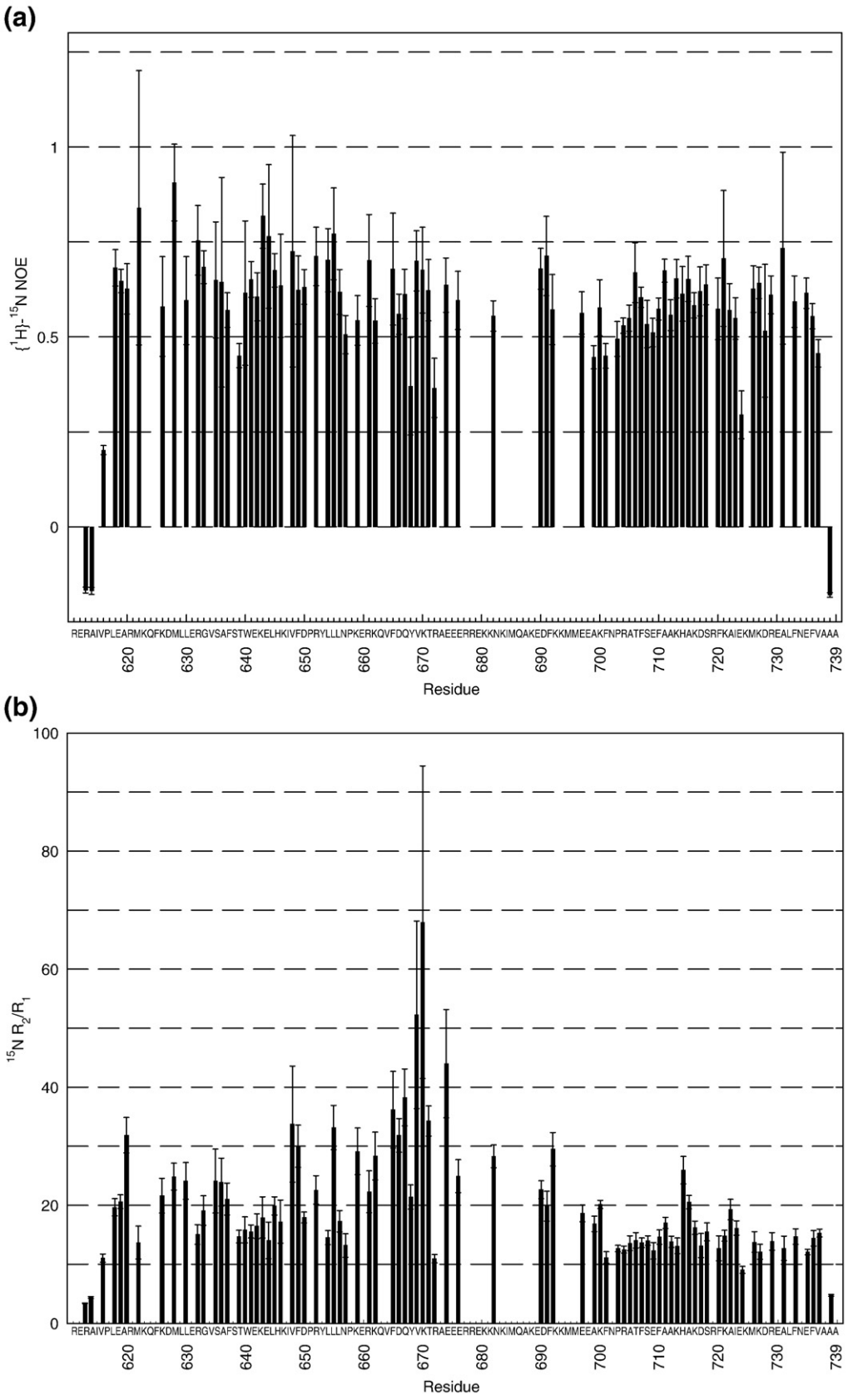

Fig. 5. ^{15}N-Relaxation studies of FF1 + FF2(611–739) provide insights into the dynamics of FF domains within tandem arrays. (a) Heteronuclear {^{1}H}-^{15}N NOEs; (b) R_2/R_1 ratios. Data were not obtainable for all residues owing to some peaks being absent from spectra or overlapped and consequently unable to be dissected by the program FuDA.

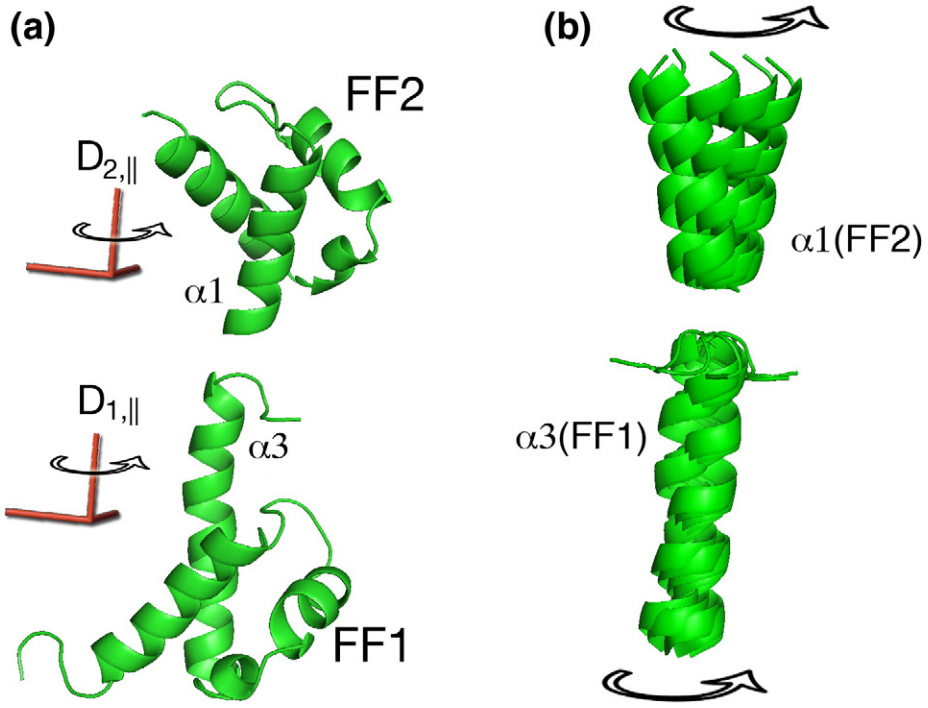

Fig. 6. The relative orientation of FF1 and FF2. The FF1-FF2 moiety tumbles as a dynamic dumbbell structure in solution with the FF1 and FF2 domains connected by a long dynamic linker helix consisting of α3 of FF1 and α1 of FF2. The diffusion tensors of FF1 and FF2, $\mathbf{D}_1$ and $\mathbf{D}_2$, were determined independently using an axially symmetric diffusion model (see the text). (a) The model of the FF1 + FF2 structure was generated by orienting the structures of FF1 and FF2 so that the unique diffusion tensor axes ($D_{1,||}$ and $D_{2,||}$) are colinear, and so that the C-terminus of FF1 is spatially close to the N-terminus of FF2. The unique axis of the diffusion tensor of FF1 is colinear with helix α3(FF1) (8°) and the unique diffusion tensor axis of FF2 is nearly colinear with helix α1(FF2) (15°). Thus, the NMR relaxation data shows that helix α3 (FF1) extends into α1(FF2) to form a long (dynamic) FF1 + FF2 linker helix. Only the orientation of the unique diffusion axis can be determined from the NMR relaxation experiments, whereas the rotation about this unique axis cannot be determined from the data. All the possible orientations of α3(FF1) and α1(FF2) are approximately colinear, as shown in (b) where the possible orientations are generated by rotating the FF1 and FF2 structures about their individual $D_{||}$ axis. The angle between the helix axes of α3(FF1) and α1(FF2) shown in b vary from 6° to 23°.

model of the FF1 + FF2 complex generated by diffusion anisotropy (Fig. 6), predict a value of 3.1, significantly larger than 1.9 or 2.2 determined experimentally that suggests at least some degree of relative flexibility between the domains. Taken together, these results are consistent with the notion that the interdomain helix and N-terminal portion of the FF2 domain are conformationally-flexible, and therefore permissive of some relative motion of the two FF domains within the FF1 + FF2 fragment.

It is noteworthy that there is greater variation between relaxation values for FF1 + linker residues within the FF1 + FF2 tandem domain construct (Fig. 5; Supplementary Data Fig. 2B) than corresponding residues within the isolated FF1 + linker (Fig. 4 and Supplementary Data Fig. 2A). This variation can be attributed to the strong anisotropic tumbling of the FF1 + FF2 tandem domains ($D_{||}/D_{\perp} \sim 2$), because the axially symmetric diffusion tensors give rise to variation in $\tau_{c,eff}$ (and thus R_2/R_1 values) for FF1 + FF2 resonances (Supplementary Data Fig. 3). In contrast, the isolated FF1 + linker undergoes more isotropic tumbling ($D_{||}/D_{\perp} \sim 1.3$) and consequently exhibits a more homogeneous distribution of R_2/R_1 values.

CA150 FF1, FF2 and FF3 do not detectably bind phosphorylated CTD peptides

To date, phosphoCTD recognition by individual FF domains is poorly characterized, probably as a consequence of the weak affinity of individual FF domains for phosphoCTD peptides. The HYPA/FBP11 FF1 domain was reported to bind a phosphoCTD peptide (SYpSPTpSPSYpSPTpSPSY) with an affinity of ~50 μM and bind *via* an FF domain interface composed of residues at the N-terminal end of α1 and N-terminal end of α3.[13] In contrast, the Prp40 FF1 domain did not detectably bind phosphoCTD peptides in a chemical shift perturbation NMR experiment,[18] although Prp40 FF1–6 were shown biochemically to have phosphoCTD binding activity.[9,14]

In order to probe phosphoCTD binding by the CA150 FF1, FF2 and FF3 domains, we exploited the capacity of NMR for detecting weak intermolecular interactions. ^{1}H-^{15}N HSQC spectra were recorded for ^{15}N-labelled FF1(618–671), FF2(682–739), FF3 (750–805) or FF1 + FF2(611–739) alone and in the presence of several unlabelled peptides corresponding to the phosphoCTD (Supplementary

Data Fig. 4a–d). Surprisingly, we did not observe chemical shift perturbations for residues within any of the three individual FF domains or the dual FF1 + FF2 construct upon addition of phosphoCTD peptides. Titrations extending to 10-fold molar excesses of peptides representing one repeat of the CTD heptad consensus motif in a Ser2-monophosphorylated state (PSYpSPTSPS), Ser5-monophosphorylated state (PSYSPTpSPS) or a Ser2,Ser5-diphosphorylated state (PSYpSPTpSPS or SPSYpSPTpSPS) did not induce chemical shift perturbations. We repeated these titrations using a longer peptide representing a tandem repeat of Ser2,Ser5-diphosphorylated CTD heptad SPS(YpSPTpSPS)$_2$, as used in other studies.[13,18] Addition of this tandem peptide to FF1, FF2, FF3 or FF1 + FF2 induced only slight chemical shift perturbations for histidine residues and residues in proximity of the histidine residues (Supplementary Data Fig. 4a–d). However, we attributed these shifts to a slight change of pH that resulted from the addition of the acidic phosphopeptide, rather than the FF domains binding the phosphopeptide. We observed that pH reductions induced marked chemical shift perturbations in histidine and neighboring residues in control spectra, probably owing to protonation of the histidine imidazole nitrogen (pK_a~6.0) (data not shown). One would anticipate interactions between an FF domain and phosphoCTD peptide could be detected in ^{1}H-^{15}N HSQC NMR spectra, since amide ^{15}N and ^{1}H chemical shifts are remarkably sensitive to small perturbations in backbone structure, electrostatic fields (charges), and ring current effects (from aromatic sidechains) that would arise from a ligand interaction. However, the absence of detectable chemical shift perturbations within FF domain HSQC spectra in the presence of phosphoCTD is not surprising, because a recent comparison of structural and ligand-binding properties of FF domains led Bonet *et al.*[19] to conclude that the overall pK_a and the presence of charged patches on an FF domain were not predictive of the nature of the ligand or of a ligand-binding site.

As an independent measure of the CA150 FF domain's capacity to bind the phosphoCTD peptides, we performed isothermal titration calorimetry (ITC) experiments to determine the heat release upon serial injection of the SPS(YpSPTpSPS)$_2$ peptide into a solution containing either FF1 or FF1–3. Consistent with the findings of our NMR titration experiments, we were unable to detect an interaction between CA150 FF1 and the SPS(YpSPTpSPS)$_2$ peptide using ITC at near-physiological pH (pH 7.25; Supplementary Data Fig. 4E, left-hand panel). Interestingly, we observed a very subtle, barely detectable interaction of FF1–3 with the SPS (YpSPTpSPS)$_2$ peptide by ITC (Supplementary Data Fig. 4E, right-hand panel), in which some, albeit small, dependence on the molar ratio of peptide to FF1–3 was observed: suggesting that FF1–3 might interact with SPS(YpSPTpSPS)$_2$ peptide extremely weakly. However, the signal arising from these titrations was too weak to allow reliable extraction of thermodynamic parameters. These data reinforce our findings in NMR titration experiments where the individual CA150 FF domains do not detectably bind the phosphoCTD peptides, while the potential ultra-weak interaction observed between FF1–3 and the tandem phosphoCTD peptide suggests a rationale for the occurrence of both CA150 FF domains and the phosphoCTD heptad motif within tandem arrays in nature.

Discussion

A defining characteristic of FF domains is their linkage in nature in arrays of between four and six domains. In the present work, we set out to probe how FF domains are organized within a tandem array using CA150 FF1 + FF2 as a model system for structural studies by NMR. Initially, we observed that CA150 FF1 interacted with the linker sequence that joins FF1 to FF2, but not FF2, enabling us to simplify the system for structural determination. In order to illuminate the nature of the interaction of the FF1-FF2 linker with FF1, we determined the solution structure of this fragment of CA150 (Fig. 3). Within the CA150 FF1 + linker structure, the topology of the core FF1 domain was comparable to the three-helix bundles described for the FBP11/HYPA FF1, Prp40 FF1 and URN1 FF domain.[13,18,19] However, the CA150 FF1 + linker structure differs from the previously described FF domains due to the inclusion of the C-terminal interdomain linker, which is ordered and forms a helical extension to α3 at the C-terminus (Fig. 3). The ordered nature of this linker was unexpected, but in light of the very high conservation between the linker sequences across species (Fig. 1b), this can be rationalized. If FF domains were merely tethered by sequences of random coil, there would be no selective pressure for the linker sequences to be evolutionarily conserved and thus divergence would be anticipated. But within the CA150 FF1–3 module there is a high level of conservation within linker sequences, consistent with these regions being integral to CA150 function. Indeed, the CA150 inter-FF domain linker sequences were found to adopt extended α helices in the crystal structure of FF1–3.[23]

We further characterized CA150 FF1 + linker using ^{15}N-relaxation NMR experiments to provide insights into dynamics within the domain and the relationship between the C-terminal helical extension and the canonical FF1 domain. Heteronuclear NOE data demonstrated that the core CA150 FF1 domain and the helical extension formed by the linker behave as a single, folded domain that tumbles as a rigid unit (Fig. 4a). Interestingly, when we extended our study to the CA150 FF1 + FF2 peptide, we observed that amide resonances corresponding to residues within the interdomain helix and the FF2 N-terminus were subject to line-broadening, indicating that the interdomain region was undergoing conformational exchange on the milli- to microsecond timescale. Due to the absence of amide resonances corresponding to interdomain linker

residues from NMR spectra, we were unable to directly examine the secondary structure of the interdomain region to determine the extent to which the extended α3/interdomain helix observed in the FF1+linker solution structure (Fig. 3a) was maintained in the linked FF1+FF2 domains and whether the relative dynamics of the two FF domains arises from a folded to unfolded transition within the interdomain region. Notwithstanding this, the line-broadening of resonances corresponding to interdomain residues suggested flexibility in the interdomain region, which was borne out in the ^{15}N-relaxation NMR data where ModelFree analysis demonstrated that FF1 and FF2 do not tumble as a single unit, with distinct differences in the anisotropies of their diffusion tensors. Notably, both FF domains within the FF1+FF2 peptide exhibited pronounced axial anisotropy, indicating that the protein conforms to a dumbbell shape, with the interdomain linker acting as a tether to maintain the relative proximity of the two FF domains without locking their relative orientations (Fig. 6). The differences in dynamics between the two FF domains reinforce the fact that there must be some flexibility afforded by the interdomain linker – a feature that distinguishes our solution structure from the CA150 FF1–3 crystal structure in which a helical conformer of the interdomain linker has been crystallized.[23]

To date, FF domains have been predominantly characterized as interactors of the phosphorylated C-terminal domain (phosphoCTD) of RNAPII. However, evidence has accumulated that FF domains are multifunctional units with the capacity to act as scaffolds that can recruit transcription and splicing factors to the transcription complex,[7,18,26,27] and can even bind RNA.[3] The pre-mRNA processing factor Prp40 was initially identified in a yeast genetic suppressor screen as an interactor of the U1 small nuclear RNA.[3] Prp40 was rendered unable to bind the U1 small nuclear RNA by a point mutation of Ser240 to Phe. Ser240 is located within the FF2 domain of Prp40, just C-terminal to the DPXY motif (in this case, EPIY) and thus the conserved 3_{10} helix.[3] Intriguingly, Prp40 FF2 is highly basic, with a predicted pK_a of 9.9, which raises the prospect that other highly basic FF domains, such as CA150 FF1–3, FF5 and FF6, may be RNA interactors. In fact, we observed that purification of recombinant CA150 FF1, FF2 and FF3 from *Escherichia coli* resulted in the copurification of bacterial RNA oligomers (data not shown). This interaction was reproducibly observed in 0.2 M NaCl, but was disrupted when purification was performed in 0.5 M NaCl (data not shown), inferring that the interaction is ionic in nature. Consequently, we consider it a distinct possibility that nucleic acid binding might be a more widespread property of basic FF domains. It will be of interest to investigate whether FF domains, such as those from CA150, have biologically relevant nucleic acid binding partners, and to undertake a more detailed examination of the molecular basis of nucleic acid recognition by FF domains.

In addition to RNA binding, FF domains have been shown to mediate many diverse protein–protein interactions. The sole existence of FF domains within tandem arrays in nature raises the possibility that these proteins serve as scaffolds to recruit numerous factors that regulate transcriptional processes. Consistent with this idea, the yeast Prp40 FF1 domain was shown to directly bind the tetratricopeptide protein recognition domain of the spliceosome scaffold protein, Crooked neck protein-like factor (Clf1)[18,26] and the U1 snRNP protein Luc7[27], whereas the neighboring Prp40 FF2+FF3 domains bind the U1 snRNP protein Snu71.[3,27] Although it is clear that FF domains mediate many different protein–protein interactions, they have predominantly been described as phosphoCTD recognition modules. The FF domains of Prp40 have been implicated in phosphoCTD binding;[9,14] however, Gasch *et al*. did not observe chemical shift perturbations when the isolated Prp40 FF1 domain was subjected to NMR titrations with phosphoCTD peptides.[18] This FF domain is highly acidic (pK_a 4.7); in contrast, the FBP11/HYPA FF1 domain shown by NMR to bind phosphoCTD peptide is highly basic (pK_a 9.6). Consequently, Gasch *et al*. postulated that FF domain basicity might be a requirement for phosphoCTD peptide binding.[18] As we describe above, we were unable to detect any interaction between the highly basic CA150 FF1, FF2, FF3 or FF1+FF2 domains and phosphoCTD peptides, even utilizing the sensitivity of NMR spectroscopy for detecting weak protein–protein interactions. This finding is surprising and indicates that the individual CA150 FF1, FF2 and FF3 domains are unable to detectably bind phosphoCTD peptides despite their high pI values (FF1, 9.5; FF2, 9.4; FF3, 8.2), and that FF domain basicity does not solely govern phosphoCTD interactions. However, these studies do not preclude the possibility that the six CA150 FF domains act in concert to bind the highly repetitive phosphoCTD heptad motif of RNAPII *via* multiple weak interactions: a model which suggests a rationale for the existence of both FF domains and the phosphoCTD heptad motif in tandem arrays in nature. This model is, in part, supported by our ITC experiments (Supplementary Data Fig. 4E) in which isotherms measured for titration of the tandem phosphoCTD peptide with FF1–3, but not FF1 alone, show some dependence on molar ratio, suggesting that the peptide may indeed interact with FF1–3, albeit with ultra-weak affinity.

Complementary to their proposed roles as phosphoCTD interactors, the CA150 FF domains may serve a scaffolding function for recruiting transcription and splicing factors to the elongating RNAPII. A recent proteomic screen for CA150 FF1–6 binding partners illustrated that the FF domains interact with a network of diverse nuclear proteins that regulate transcription and RNA processing.[7] Further investigations will shed light on the roles of the individual FF domains in recruiting transcription and splicing factors, and will clarify whether the six CA150 FF domains have individual specificities for their binding partners or whether they act

in concert to recruit interactors. The possibility of each FF domain within a tandem array mediating a distinct repertoire of interactions is suggested by a recent study in which CA150 FF5 was implicated as the critical FF domain for CA150 localization to splicing-rich nuclear speckles.[28]

A recent examination of CA150 splicing variants in the trematode *Schistosoma mansoni* provides further validation for the biological relevance of performing structural studies on a two FF domain construct.[29] DeMarco *et al*.[29] identified two isoforms of CA150 in *S. mansoni* that are truncated after the linker sequence that would join FF2 to FF3 in the full-length protein. Though no such alternative splicing isoforms have been identified to date in mammals, a CA150 homolog, TCERG1-like, is encoded by an alternative locus within the mammalian genome. The human homolog of TCERG1-like is composed of two N-terminal WW domains and two C-terminal FF domains (Gene ID, 256536); like CA150 FF1–3 (Fig. 1b), the FF domains and intervening linker of TCERG1-like are highly conserved amongst mammals. The human TCERG1-like FF1 + FF2 domains have an overall identity with human CA150 FF1 + FF2 of 64% at the amino acid level, with the FF1 + linker portion exhibiting much greater identity (77%). These similarities suggest that the TCERG1-like locus arose from the duplication of the CA150 gene within the mammalian genome, and suggest subsequent sequence divergence, especially in FF2, and truncation of the C-terminal four FF domains has occurred. No function has been attributed to TCERG1-like in mammals, although the *S. mansoni* splicing variants of similar domain composition were postulated to be involved in sex-differentiation processes due to a gender bias in the representation of the different isoforms.[29]

The repetition of modular protein domains within proteins is not uncommon, and the example of tandem WW domains provides a useful point of comparison for the tandem FF domains due to their presence within many FF domain-containing proteins. In particular, the helical nature of the linker between CA150 FF1 and FF2 is reminiscent of the 12-residue helix that connects the yeast Prp40 WW domain pair.[30] These WW domains were found to adopt a fixed orientation by virtue of conserved leucine residues at either end of the interdomain helix contacting the core of their adjacent WW domain to restrict domain mobility.[30] Superficially, the organization of the Prp40 WW domain pair shows some similarity to that of CA150 FF1 + FF2, as our studies of the FF1 + linker construct indicate that a helix is formed by the interdomain linker, and this helix interacts with residues within FF1 to restrict its mobility. However, in contrast to the WW domain pair, the FF1 and FF2 within CA150 FF1 + FF2 exhibit relative motion and neither domain adopts a fixed orientation. The relative flexibility of FF1 and FF2 within CA150 suggests that these domains are able to manoeuvre to engage different interactors while the interdomain helix maintains their spatial proximity. This concept of FF domain organization yields a general model for how FF domains are organized within tandem arrays. In particular, this model of FF domain organization is pertinent to recognition of the intrinsically disordered phosphoCTD, because mobility of FF domains within a tandem array would be necessary for optimal recognition of a highly flexible binding partner. These insights into FF domain organization provide a starting point for understanding the structural basis of the multitude of potential protein–protein and protein–nucleic acid interactions mediated by FF domains.

Materials and Methods

Preparation of protein samples

DNA sequences encoding FF1(618-671), FF1 + linker (618-683), FF2(682-739), FF1 + FF2(611-739) and FF3(750-805) were amplified by PCR from the murine I.M.A.G.E. clone, 3982736 (accession BC039185; Open Biosystems, Huntsville, AL), and cloned in-frame into the pProEX Htb expression vector (Invitrogen) in order to incorporate an N-terminal hexa-histidine tag that can be cleaved using tobacco etch virus (TEV) protease. These constructs were expressed in *E. coli* BL21(DE3) cells grown at 37 °C to an A_{600} of 0.6–0.8 before a 16 h induction with isopropyl 1-thio-β-D-galactopyranoside at 15 °C. ^{15}N and ^{13}C labelling was achieved by expressing proteins in M9 minimal medium containing 2 g/L of D-[^{13}C]glucose and/or 1 g/L of [^{15}N]NH_4Cl as the sole carbon and nitrogen source, respectively. Cells were lysed by sonication in 0.5 M NaCl, 5 mM imidazole, 20 mM Hepes (pH 7.5), 1 mM PMSF and the cleared supernatant loaded onto a HiTrap Chelating HP column (GE Healthcare) charged with $NiSO_4$ according to the manufacturers' instructions. The His_6-tagged protein was eluted from the Ni column using a 250 mM–500 mM imidazole (pH 7.5) gradient and the His_6-tag cleaved by incubation with TEV protease for 2 h at 25 °C. Dialysis was used to eliminate imidazole from the cleavage mixture and the dialysate was passed over a Ni column to remove the His_6-tagged proteins. The flow-through was concentrated and applied to a Superdex-75 size-exclusion column (GE Healthcare) and eluted in either PBS (137 mM NaCl, 10.1 mM Na_2HPO_4, 2.7 mM KCl, 1.8 mM KH_2PO_4, pH 6.5) or a buffer composed of 50 mM NaCl, 10 mM Tris–HCl, pH 7.0. All purified proteins contain an N-terminal sequence of GAMGS or GAMGSGI as a result of cloning artifacts.

NMR spectroscopy

Experiments were performed on a Varian INOVA 500 MHz spectrometer equipped with a triple resonance probe and pulsed field gradient unit. All data were processed with the NMRPipe/NMRDraw package,[31] with subsequent analysis performed using XEASY.[32] All 3D and ^{15}N-relaxation experiments were performed on 1–1.7 mM proteins in PBS containing 7.5% ^2H_2O at 15 °C. Backbone resonances were assigned for FF1(618–671), FF1 + linker(618–683), FF2(682–739), FF1 + FF2(611–739) and FF3(750–805) using standard triple-resonance experiments.[33,34] FF1 + linker sidechain resonances were assigned using standard experiments,[33,34] and analysis

of the ^{13}C-edited NOESY spectrum was sufficient to enable assignment of aromatic sidechain resonances.

The ^{15}N-relaxation experiments were performed with ^{15}N-labelled samples of FF1+linker (1.7 mM) and FF1+FF2 (1 mM) in PBS containing 7.5% $^{2}H_2O$ at 15 °C. Heteronuclear {^{1}H}-^{15}N NOE experiments were performed using standard 2D methods,[35] with the reference and proton saturated spectra collected in an interleaved fashion. The {1H}-15N NOEs were calculated from the peak intensity ratio between the reference and proton-saturated experiments using the software FuDA (developed by DFH and Dr Soren Kristensen, University of Copenhagen). Longitudinal (R_1) and transverse (R_2) relaxation rates were calculated from T_1 (Ref. 35) and $T_1\rho$,[36] respectively, which were measured using standard 2D methods. The following relaxation delays were applied in an interleaved fashion: 10.1, 80.6, 161.3, 252.0, 362.9, 483.9, 625.0, 806.5 ms for the T_1 experiment; and 2, 7, 15, 24, 33, 43, 55, 68, 83, 100 ms for the $T_1\rho$ experiment. Relaxation rates and associated errors were calculated using the software FuDA.

Structure calculation

For structure calculation, a NOESY experiment with simultaneous ^{15}N and ^{13}C chemical shift evolution was recorded with a mixing time of 125 ms.[37] Interproton distance restraints were derived from assigned manually peaks within ^{13}C- and ^{15}N-edited NOESY spectra that were analyzed and integrated in the XEASY software. Hydrogen bond restraints were applied according to $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts and NOE patterns. Dihedral angle restraints were derived from the program TALOS.[38] and were applied where consistent with experimental ϕ angles determined from a quantitative ^{3}J (H^{N},H^{α}) correlation experiment.[39] Structures were calculated in eight iterations using a standard simulated annealing protocol with the programs, CNS[40] and ARIA 1.2.[41,42] The final 20 lowest energy structures were assessed with PROCHECK-NMR.[43]

Chemical shift perturbation experiments

^{1}H-^{15}N Heteronuclear single quantum correlation (^{1}H-^{15}N HSQC) spectra were recorded for ^{15}N-labelled CA150 FF1, FF2 or FF3 domains in the presence of increasing amounts of unlabelled phosphoCTD peptides on Varian INOVA 500 MHz spectrometers at 25 °C in two distinct buffers: PBS (pH 6.5) or 50 mM NaCl, 10 mM Tris–HCl (pH 7.0). The peptides SPSYpSPTpSPS and SPS $(YpSPTpSPS)_2$ representing a single and tandem repeat of the Ser2,Ser5-diphosphorylated CTD heptad motif, respectively, were synthesized and purified to >95% homogeneity by Dalton Chemical Laboratories (Toronto, Ontario, Canada). The peptides PSYpSPTpSPS, PSYpSPTSPS and PSTSPTpSPS were synthesized in-house on an AbiMed 431 synthesizer with the standard Fastmoc protocol and purified to >90% by preparative HPLC. Peptides were added to samples of 0.4 mM ^{15}N-labelled FF domains to a molar peptide/protein ratio of ~10:1. The pH was maintained at 6.5 in PBS or at 7.0 in the Tris buffer over the course of the titration.

Average chemical shift perturbations were calculated using the relationship:

$$\Delta\delta^{av} = \left((\Delta\delta_{1H})^2 + (\Delta\delta_{15N}/5)^2\right)^{1/2}$$

where $\Delta\delta_{1H}$ and $\Delta\delta_{15N}$ are the chemical shift differences in ^{1}H and ^{15}N, respectively. Chemical shift perturbations were deemed significant if >0.2 ppm.

Rotational diffusion analyses

The program *r2r1_tm* was used to calculate an effective correlation time $\tau_{c,eff}$ for each residue using the R_2/R_1 ^{15}N ratio. Subsequently, the diffusion tensors and the principal coordinate systems of FF1+linker and FF1+FF2 were calculated using the *quadric_diffusion* program with $\tau_{c,eff}$ as input. The programs *r2r1_tm* and *quadric_diffusion* are both components of the *ModelFree4.16* package (A.G. Palmer, ModelFree 4.1). Only residues from the structured helix regions were included, since omitting relaxation data from flexible loop regions ensures that internal dynamics do not obscure the determination of the diffusion tensor.[44] Isotropic and axially symmetric diffusion models were tested and evaluated based on *F*-statistics. The rotational diffusion tensor, **D**, of FF1+linker was determined from ^{15}N R_1 and R_2 relaxation rates of residues 619–630, 641–645, and 661–675. The rotational diffusion tensors of FF1 and FF2 within the FF1+FF2 construct were determined independently for the following reasons. (1) The relative orientation of the FF1 and FF2 domain within FF1+FF2 complex is not known, which precludes a simultaneous analysis. (2) Possible inter-domain motions (see below) will lead to different effective diffusion tensors for each of the two domains. Relaxation rates of residues 620–630, 641–645 and 661–671 were included for the FF1 domain and residues 690–691, 706–715, and 727–737 for the FF2 domain were included in the determination of the diffusion tensors of the individual domains within the FF1+FF2 moiety. All residues in the α helices were included in diffusion tensor calculations, because: (a) the bond vector orientation of residues in α helices is significantly more accurate than those of residues in loop regions; (b) residues that experience chemical exchange could not simply be excluded from analysis — because the diffusion tensor of the FF1-FF2 tandem is highly anisotropic ($D_{||}/D_{\perp}$~2) — instead, we measured the ^{15}N T_2 values with a spin-lock experiment (2 kHz spin-lock field) where contributions from chemical exchange processes slower than ~0.5 ms are quenched; and (c) strict selection criteria (e.g. het-NOE<0.65) would exclude too many residues from the analysis and thus preclude determination of the orientation and magnitude of the diffusion tensors.

Isothermal titration calorimetry (ITC)

All ITC measurements were performed using a VP-ITC MicroCalorimeter (MicroCal, Inc., Northampton, MA). Titration experiments were performed in 3.3 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 at pH 7.25. Both protein and peptide were dialyzed against the same buffer. FF1 protein (25.7 μM) in the calorimetric cell was titrated by a series of 10 μL volume injections of the peptide (386 μM) with an interval of 150 s at 13 °C. FF1–3 protein (100 μM) in the calorimetric cell was titrated by a series of 10 μL volume injections of the peptide (1.54 mM) with an interval of 150 s at 25 °C. Control titrations were performed to measure the heat of dilution arising from titration of the peptides into buffer alone, which were in turn subtracted from peptide titrations with FF domains.

Protein Data Bank accession numbers

Structural coordinates have been deposited in the Protein Data Bank with accession number 2kis; chemical shift assignments have been deposited in the BioMagResBank with accession number 16293.

Acknowledgements

We thank Dr Soren Kristensen (University of Copenhagen) for his role in developing the relaxation data analysis program FuDA. Dr Remco Sprangers is thanked for helpful discussions and computational wizardry. We thank Dr Xiaomin Chen for sharing data before publication. Support was gratefully received by J.M.M. (C.J. Martin Fellowship of the National Health and Medical Research Council of Australia, Grant 305546), D.F.H. (post-doctoral fellowship from the Canadian Institutes of Health Research (CIHR)), M.B. (post-doctoral fellowships from the CIHR Training Program in Protein Folding and Swedish-America Foundation), and M.J.S. (postgraduate scholarship from Natural Sciences and Engineering Research Council (NSERC), Canada). This work was supported by grants from the National Cancer Institute of Canada (NCIC) and CIHR to F.S. (CIHR grant #MOP-36399), J.D.F.-K. and T.P. (CIHR grant #MOP-6849) and Genome Canada to T.P.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.08.049

References

1. Sune, C., Hayashi, T., Liu, Y., Lane, W. S., Young, R. A. & Garcia-Blanco, M. A. (1997). CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* **17**, 6029–6039.
2. Sune, C. & Garcia-Blanco, M. A. (1999). Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner. *Mol. Cell. Biol.* **19**, 4719–4728.
3. Kao, H.-Y. & Siliciano, P. G. (1996). Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* **16**, 960–967.
4. Bedford, M. T. & Leder, P. (1999). The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem. Sci.* **24**, 264–265.
5. Lin, K.-T., Lu, R.-M. & Tarn, W.-Y. (2004). The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing in vivo. *Mol. Cell. Biol.* **24**, 9176–9185.
6. Jiang, W., Sordella, R., Chen, G. C., Hakre, S., Roy, A. L. & Settleman, J. (2005). An FF domain-dependent protein interaction mediates a signalling pathway for growth factor-induced gene expression. *Mol. Cell*, **17**, 23–35.
7. Smith, M. J., Kulkarni, S. & Pawson, T. (2004). FF domains of CA150 bind transcription and splicing factors through multiple weak interactions. *Mol. Cell. Biol.* **24**, 9274–9285.
8. Carty, S. M., Goldstrohm, A. C., Sune, C., Garcia-Blanco, M. A. & Greenleaf, A. L. (2000). Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc. Natl Acad. Sci. USA*, **97**, 9015–9020.
9. Morris, D. P. & Greenleaf, A. L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **275**, 39935–39943.
10. Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M. & Ajuh, P. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genet.* **20**, 46–50.
11. Goldstrohm, A. C., Albrecht, T. R., Sune, C., Bedford, M. T. & Garcia-Blanco, M. A. (2001). The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* **21**, 7617–7628.
12. Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182–185.
13. Allen, M., Friedler, A., Schon, O. & Bycroft, M. (2002). The structure of an FF domain from human HYPA/FBP11. *J. Mol. Biol.* **323**, 411–416.
14. Phatnani, H. P., Jones, J. C. & Greenleaf, A. L. (2004). Expanding the functional repertoire of CTD Kinase I and RNA Polymerase II: Novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry*, **43**, 15702–15719.
15. Corden, J. L., Cadena, D. L., Ahearn, J. M. & Dahmus, M. E. (1985). A unique structure at the carboxyl terminus of the largest subunit of the eukaryotic RNA polymerase II. *Proc. Natl Acad. Sci. USA*, **82**, 7934–7938.
16. Li, W.-B., Bzik, B., Gu, H., Tanaka, M., Fox, B. A. & Inselburg, J. (1989). An enlarged largest subunit of Plasmodium falciparum RNA polymerase II defines conserved and variable RNA polymerase domains. *Nucleic Acids Res.* **17**, 9621–9636.
17. Zhang, J. & Corden, J. L. (1991). Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J. Biol. Chem.* **266**, 2290–2296.
18. Gasch, A., Wiesner, S., Martin-Malpartida, P., Ramirez-Espain, X., Ruiz, L. & Macias, M. J. (2006). The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains. *J. Biol. Chem.* **281**, 356–364.
19. Bonet, R., Ramirez-Espain, X. & Macias, M. J. (2008). Solution structure of the yeast URN1 splicing factor FF domain: Comparative analysis of charge distributions in FF domain structures – FFs and SURPs, two domains with a similar fold. *Proteins: Struct. Funct. Genet.* **73**, 1001–1009.
20. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 29–32.

21. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
22. Efron, B. & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* **37**, 36–48.
23. Lu, M., Yang, J., Ren, Z., Sabui, S., Espejo, A., Bedford, M. T. *et al.* (2009). Crystal structure of the three tandem FF domains of the transcription elongation regulator CA150. *J. Mol. Biol.* **393**, 397–408.
24. Korzhnev, D. M., Billeter, M., Arseniev, A. S. & Orekhov, V. Y. (2001). NMR studies of Brownian tumbling and internal motions in proteins. *Prog. NMR Spectrosc.* **38**, 197–266.
25. Orekhov, V. Y., Nolde, D. E., Golovanov, A. P., Korzhnev, D. M. & Arseniev, A. S. (1995). Processing of heteronuclear NMR relaxation data with the new software DASHA. *Appl. Magn. Reson.* **9**, 581–588.
26. Chung, S., McLean, M. R. & Rymond, B. C. (1999). Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA*, **5**, 1042–1054.
27. Ester, C. & Uetz, P. (2008). The FF domains of yeast U1 snURP protein Prp40 mediate interactions with Luc7 and Snu71. *BMC Biochem.* **9**, 29.
28. Sanchez-Alvarez, M., Goldstrohm, A. C., Garcia-Blanco, M. & Sune, C. (2006). Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxy regions. *Mol. Cell. Biol.* **26**, 4998–5014.
29. DeMarco, R., Oliveira, K. C., Venancio, T. M. & Verjovski-Almeida, S. (2006). Gender biased differential alternative splicing patterns of the transcriptional cofactor CA150 gene in Schistosoma mansoni. *Mol. Biochem. Parasitol.* **150**, 123–131.
30. Wiesner, S., Stier, G., Sattler, M. & Macias, M. J. (2002). Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J. Mol. Biol.* **324**, 807–822.
31. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
32. Bartels, C., Xia, T.-H., Billeter, M., Güntert, P. & Wüthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR*, **6**, 1–10.
33. Kay, L. E. (1995). Pulsed field gradient multidimensional NMR methods for the study of protein structure and dynamics in solution. *Prog. Biophys. Mol. Biol.* **63**, 277–299.
34. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34**, 93–158.
35. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M. & Gish, G. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
36. Korzhnev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. & Kay, L. E. (2002). An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J. Am. Chem. Soc.* **124**, 10743–10753.
37. Pascal, S., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. (1994). Simultaneous acquisition of ^{15}N and ^{13}C edited NOE spectra of proteins dissolved in H_2O. *J. Magn. Reson. B*, **103**, 197–201.
38. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
39. Kubinowa, H., Grzesiek, S., Delaglio, F. & Bax, A. (1994). Measurement of H^N-H^α J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR*, **4**, 871–878.
40. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P. & Grosse-Kunstleve, R. W. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
41. Folmer, R. H., Hilbers, C. W., Konings, R. N. & Nilges, M. (1997). Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data. *J. Biomol. NMR*, **9**, 245–258.
42. Nilges, M. & O'Donoghue, S. I. (1998). Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spectrosc.* **32**, 107–139.
43. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kapstein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
44. Mandel, A. M., Akke, M. & Palmer, A. G., 3rd (1995). Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144–163.