Published in final edited form as:

J Mol Biol. 2009 October 23; 393(2): 397–408. doi:10.1016/j.jmb.2009.07.086.

Crystal Structure of Three Tandem FF Domains of Transcription Elongation Regulator CA150

Ming Lu^{1,2}, Jun Yang^{1,2}, Zhiyong Ren¹, Subir Sabui³, Alexandra Espejo⁴, Mark T. Bedford^{2,4}, Raymond H. Jacobson^{1,2}, David Jeruzalmi⁵, John S. McMurray^{2,3}, and Xiaomin Chen^{1,2,*}

¹Department of Biochemistry and Molecular Biology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030

²Graduate School of Biomedical Sciences, University of Texas Health Science Center, Houston, TX 77030

³Department of Experimental Therapeutics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030

⁴Department of Carcinogenesis, University of Texas M.D. Anderson Cancer Center Science Park, Smithville, TX 78957

⁵Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

Summary

FF domains are small protein-protein interaction modules that have two flanking conserved phenylalanine residues. They are present in proteins involved in transcription, RNA splicing, and signal transduction, and often exist in tandem arrays. Although several individual FF domain structures have been determined by NMR, the tandem nature of most FF domains has not been revealed. Here we report the 2.7 Å resolution crystal structure of the first three FF domains of human transcription elongation factor CA150. Each FF domain is composed of three α -helices and a 3_{10} helix between α -helices 2 and 3. The most striking feature of the structure is that an FF domain is connected to the next by an α -helix that continues from helix 3 to helix 1 of the next. The consequent elongated arrangement allows exposure of many charged residues within the region that can be engaged in interaction with other molecules. Binding studies using a peptide ligand suggests that a specific conformation of the FF domains might be required to achieve higher affinity binding. Additionally, we explore potential DNA binding of the FF construct used in this study. Overall we provide the first crystal structure of a FF domain and insights into the tandem nature of the FF domains and suggest that in addition to protein binding, FF domains might be involved in DNA binding.

Keywords

FF domain; CA150; X-ray crystallography; protein-ligand interaction; transcription

© 2009 Elsevier Ltd. All rights reserved.

*Corresponding author, Tel: (713) 834-6279, Fax: (713) 834-6273, xichen@mdacc.tmc.edu.

Accession Number Coordinates and structure factors have been deposited in the PDB with accession number 3HFH.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Since the discovery of the Src homology 2 (SH2) domain in the late 1980s^{1, 2}, many protein-protein interaction modules have been shown to play important roles in numerous cellular processes. Compared to the better-characterized domains such as SH2, SH3, PDZ, and WW, structure-function relationships of the recently discovered FF domain are still not very clear. These domains get their name from the two conserved phenylalanine residues near the N- and C-termini (we shall refer to them as F_N and F_C, respectively) (Figure 1). FF domains are typically of ~60 residues and are found in a limited set of eukaryotic proteins, including transcription elongation factor CA150 (also known as TCERG1), pre-mRNA splicing factors PRPF40A (Prp40, FBP11, FBNP3, and HYPA) and PRPF40B (HYPC), and RhoGAP proteins P190A (GRLF1) and P190B (RHG05)³. One of the intriguing features of the FF domains is that they often occur as multiple copies in tandem. As noted by Ester and Uetz⁴, all of the metazoan FF domain-containing proteins in the SMART database⁵ have two or more FF domains, whereas more than 83% of them have four to six. In addition, FF-containing proteins often have WW domains (Figure 1). Another noteworthy feature is the presence of a FXXLL motif in the FF domain. Figure 1B shows the sequence alignment of the human FF-containing proteins. The pair-wise sequence homology is in the range of 15-30%, with most of the conserved residues in the hydrophobic core.

CA150 (co-activator of 150 kDa), the best-studied FF domain-containing protein thus far, is a regulator of RNA polymerase II (RNAP II) transcription elongation^{6, 7} and mRNA processing⁸ and serves as a link between the two processes^{9, 10}. As shown in Figure 1A, human CA150 is a large molecule of 1098 residues⁶. Its N-terminal half is highly compositionally biased, along with three WW domains, whereas most of the C-terminal half is occupied by six FF domains. The interaction regions between CA150 and RNAP II have been mapped to the FF domains of CA150 and the phosphorylated C-terminal domain (pCTD) of the polymerase¹¹. The N-terminal PGM motif of CA150 is arginine methylated by the transcriptional coactivator, CARM1, and this methylation event generates a docking site for the tudor domain of the splicing factor SMN¹². Consistent with its role in coupling transcription and mRNA processing, CA150 has a potential nuclear localization signal (NLS) upstream of the first FF domain⁶ and is localized in nuclear speckles, rich in mRNA processing factors¹³. Dysregulation of CA150 expression has been implicated in the onset of Huntington's disease¹⁴.

Like CA150, FF-containing splicing factors PRPF40A and PRPF40B are also involved in mRNA processing¹⁵ and bind to the pCTD of RNAP II¹⁶. They both have two WW domains and respectively four and two FF domains. In contrast, each of the FF-containing RhoGAP proteins P190A and P190B has four FF domains but no WW domains. Instead, they have an N-terminal GTPase region and a C-terminal RhoGAP domain. Recently P190 FF domains have been shown to mediate a growth factor-stimulated, phosphorylation-dependent, signaling pathway, leading to transcriptional activation of some serum-inducible genes¹⁷.

To date, structures of several single FF domains from various proteins have been determined by nuclear magnetic resonance spectroscopy (NMR). These include human Prp40 FF domains 1 (PDB ID: 1UZC)¹⁸ and 5 [PDB ID: 2CQN; Riken Structural Genomics/Proteomics Initiative (RSGI)], yeast Prp40 FF domain 1 (PDB ID: 2B7E)¹⁹, the (only) FF domain of yeast pre-mRNA splicing factor URN1 (PDB ID: 2JUC)²⁰, and the first four individual FF domains of CA150 (PDB ID's: 2DOD, 2E71, 2DOE, and 2DOF, respectively; all from RSGI). As shown in Figure 1C, the overall architecture of these FF domains is very similar: three α helices forming a bundle with a 3_{10} helix (η) in the linker loop connecting the second and third helices. One variation to this so-called "FF fold" is observed in the first

FF domain of P190A21. This recent NMR structure revealed that the 3_{10} helix is replaced by an α helix that is a full-turn longer.

Nevertheless, none of these structures is able to shed light on the tandem nature of the FF domains, thus there is no structural information on their orientation with respect to each other. We sought to study this using X-ray crystallography starting with crystallization of several tandem FF constructs and crystallized a construct of the first three FF domain of human CA150. Here we present this first crystal structure of an FF domain laying bare the spatial arrangement of these domains.

FF domains from different proteins exhibit a variety of ligand preferences⁴. Through our binding studies we suggest that a favorable conformation may be necessary in achieving higher binding affinity to its protein partner. Moreover, we provide biochemical evidence to propose that CA150 FF2, resembling a homeodomain, may be involved in binding to DNA.

Results and Discussion

Protein Modification and Structure Determination

The crystal structure reported here is that of human CA150 FF1-3 (amino acid residues 661-845). During the early stage of this study we were able to crystallize the native protein under various conditions. However, none of the crystal forms gave rise to sufficient diffraction quality for us to solve the structure. Inspired by a report by Walter et al.²², we carried out reductive lysine methylation of the protein. The modification procedure is relatively simple and has been used as a routine rescue strategy in protein crystallization²². CA150 FF1-3 was modified essentially according to the published protocol and further purified by gel filtration. The completion of the reaction was confirmed by mass spectrometry. As shown in Figure 2, all 30 lysine side chains as well as the N-terminal NH_2 group of the protein were quantitatively di-methylated after the treatment, leading to an increase in the molecular weight of 867.8 Da, almost identical to the calculated MW increase of 868.0 Da (31×28.0 Da). Therefore we used exclusively dimethylated lysine residues in the model building.

Purified, methylated FF1-3 protein was crystallized at 4°C under several conditions, most of which overlapped with those used for the native protein crystallization. One novel condition was discovered, which gave rise to a new morphology (hexagonal rods) of the crystals. Preliminary diffraction analysis showed a significantly improved quality of these crystals: the best crystals of FF1-3 without lysine methylation diffracted X-ray to ~ 4 Å Bragg spacing at a synchrotron source, in spite of their much bigger size (typically $500 \times 500 \times 500 \mu\text{m}^3$), whereas those with methylation diffracted to beyond 2.4 Å.

Structure determination was carried out by single-wavelength anomalous dispersion (SAD) on a selenomethionine-substituted (SeMet) crystal. The structure was refined using data to 2.70 Å resolution, with the data collection and refinement statistics reported in Table 1. There are two molecules (A and B) in each asymmetric unit, with molecule B being more ordered than A. The final model has all of the residues for two molecules except residues 737-757 in molecule A. In addition, five extra N-terminal residues (GPLGS) introduced from the cloning vector are also modeled in each molecule. The conventional and free R factors of the final model are 23.0% and 28.3%, respectively. Of note, we were unable to solve the structure by molecular replacement using the available NMR structures as search models.

Overall Architecture

As shown in Figure 3, the three FF domains of CA150 (in red; boundaries as defined in Figure 1B) are well separated in space and each one is connected to the next by a long helix (4.5 turns) (in green) that extends contiguously from the last helix (α_3) of an FF domain to the first helix (α_1) of the next. Given this architecture, the orientation of the first FF domain dictates that of the second and so on. The first connector helix (between FF1 and FF2) has a slight curvature along the axis of the helix, whilst the second one is rather straight. The two connector helices have a cross angle of $\sim 110^\circ$ (Figure 3B). Overall the molecule has a wide-open V shape and is very elongated, with the dimensions 100 by 45 by 25 Å (Figure 3). The structure has a very large and extensive exposed surface area for interacting with other proteins. There are 25 residues between the F_C residue of FF1 and F_N of FF2 and between F_C of FF2 and F_N of FF3. Consequently, the two arms of the V shape are of the same length. However, the connecting regions between FF3 and FF4, FF4 and FF5, and FF5 and FF6 have 62, 14, and 14 residues, respectively. The variation in the connector length is common among the FF domains in different proteins.

Out of the 185 CA150 residues in this construct, there are 15 arginines, 30 lysines, 13 aspartates, and 22 glutamates. In other words, more than 43% of the side chains are either positively or negatively charged, creating a molecular surface that can be engaged in binding to other molecules through interdigitation of complementary charges. Figures 4A highlights the locations of these charged residues (in blue and red) and 4B shows the electrostatic potential as calculated by APBS23. Overall, this region is rich in basic residues with a calculated isoelectric point of 10.

Comparison of the CA150 NMR and Crystal Structures

Figure 3D shows the superimpositions of the individual CA150 FF domain NMR structures (PDB ID's: 2DOD, 2E71, and 2DOE) onto this crystal structure. Overall they all can be superimposed very well, with the RMSD values of 0.77, 1.74, and 1.25 Å (values from PyMol <www.pymol.org>), respectively. Indeed, these NMR structures provided casual guidance during our model building of the crystal structure. They have a three-residue overlap between each adjacent pair. However, it was not clear how an FF domain would connect to the next because the very N- and C-termini in each individual structure are very flexible and have lost their helical conformation. Moreover, it is interesting to see that in the NMR structures of FF1 and FF3, residues 659-660 and 846-848 (not included in our construct) are α -helical, suggesting that the helices may extend beyond the construct in this study. It is tempting to propose that this FF-containing region of the protein forms a big open structure, assuming all of the connecting regions are also α -helical. Based upon secondary structure prediction this assumption is likely to be true.

In a more recent NMR structure reported in the accompanying article²⁴, Murphy et al. used a CA150 construct containing the CA150 FF1 and some extra residues at the C-terminus and showed that the extension is α -helical in solution, suggesting that the helical nature of the connectors in our structure was not caused by crystallographic artifacts.

Peptide binding studies

The FF domain was originally suggested to be a protein-protein interaction module³ and this concept received further experimental support from different groups. Common to this theme is that FF domains from CA150 and Prp40 are clearly shown to bind to RNA polymerase II via the enzyme's phosphorylated heptad repeats YpSPTpSPS in the CTD (pS: phosphoserine)⁹; 11¹ 16. On the other hand, FF domains are also known to bind to other sequences. For example, FF1 of yeast Prp40 can bind to the crn-TRP1 repeat with a

sequence of GAMGSTNIDILDLEELREYORRKRTEYEGYLKRNRLD19. Nevertheless, there is no detailed structural information available on any FF domain-ligand interactions.

More directly relevant to our study is the report by Smith et al., in which a spot peptide array scanning a CA150 binding protein, HIV Tat-specific factor 1 (Tat-SF1), was screened against CA150 FF1-3 as a C-terminal fusion to GST25. The authors identified a consensus sequence of (D/E)₂₋₅-F/W/Y-(D/E)₂₋₅ as a FF ligand²⁵. Using fluorescence polarization (FP) they demonstrated that the peptide with the highest binding affinity (dissociation constant $K_d = 48 \mu\text{M}$) to GST-FF1-3 has the sequence of WFHVEED. To gain structural insights into an FF binding to its ligand, we synthesized the same peptide and used it either as a pre-mixture with the protein in the crystallization trials or to soak the peptide into preformed crystals, with a peptide:protein molar ratio ranging from 3:1 to 10:1. However, in the electron density maps generated from the crystals grown in the presence of, or soaked with, the peptide, we could not obtain any evidence of the peptide's existence.

To further examine the peptide binding properties of the FF domains, we synthesized the fluorescent-labeled peptide, (FAM)-AWFHVEED (FAM: 5-carboxyfluorescein), and used it in FP binding studies. Three protein constructs used were GST-FF1-3, FF1-3 alone, and GST alone, the latter two being generated by a single PreScission protease cleavage between GST and FF1-3. As shown in Figure 5 (top curve), binding between the labeled peptide and GST-FF1-3 reached saturation at 1 mM protein concentration. Data analysis using one-site specific binding mode yielded a K_d of $\sim 32 \mu\text{M}$, similar to the reported $48 \mu\text{M}$ by Smith et al.²⁵. However, significantly weaker binding ($K_d > 400 \mu\text{M}$) was detected for FF1-3 alone (middle curve). At even 1 mM protein concentration, binding still did not reach saturation. When GST alone was used (bottom curve), even weaker binding was observed ($K_d > 740 \mu\text{M}$), suggesting that the observed binding between GST-FF1-3 was unlikely to be caused by the presence of GST. Since GST is well known to dimerize²⁶, perhaps the FF domains in the fusion context adopt a more favorable conformation for peptide binding in the context of a GST-mediated dimer. The exact cause of this apparent increased binding of GST-FF1-3 remains to be elucidated. Alternatively, the linker between the GST and FF1-3 in the fusion context is somehow involved in the binding when it was not cleaved. Nonetheless, the apparently weak binding between FF1-3 and the potential ligand could perhaps also explain the failure to observe the peptide in the electron density maps. Further investigations are necessary to determine the nature of a FF ligand and its association with an FF domain. Of note, we have not been able to crystallize the GST-FF1-3 fusion protein.

One possible explanation to the increased binding affinity of GST-FF1-3 could be that FF1-3 could dimerize, either in isolation or in the context of the full-length CA150 protein. We used multi-angle light scattering (MALS) to characterize the oligomerization states of purified FF1-3 and GST-FF1-3 in solution. As shown in Figure 6, the measured molecular weight of FF1-3 was 23.2 kDa (red curve for LS signal and green for MW), compared to the calculated 22.6 kDa. This clearly showed that FF1-3 was monomeric in solution. In comparison, the majority of GST-FF1-3 fusion protein had a measured MW of 102.4 kDa (second peak in the blue curve), suggesting a dimer of GST-FF1-3 (calculated MW of 49.7 kDa). There were also some heterogeneous and higher molecular-weight aggregates of more than 230 kDa in MW (first peak in the blue curve). Although the reason for the formation of these aggregates is not clear, our data suggest that it is unlikely to be caused by the presence of FF1-3 in the fusion. Furthermore, the hydrodynamic radii (R_{hyd}) of FF1-3 and GST-FF1-3 are 29 and 56 Å, respectively, as measured by dynamic light scattering. Using the monomeric structure of FF1-3, we calculated R_{hyd} and radius of gyration (R_g). R_{hyd} obtained from HYDROPRO²⁷ was 30.4 Å, in good agreement with the DLS-measured 29 Å. This supports the monomeric state of FF1-3 in solution. R_g values obtained from MOLEMAN²²⁸ and HYDROPRO were respectively 33.0 and 32.8 Å. These values are

greater than R_{hyd} for FF1-3, consistent with the non-globular shape of the FF1-3 molecule. In summary, we ruled out the possibility of FF1-3 forming a dimer in solution. Whether the full-length CA150 could form a dimer/oligomer remains to be defined.

FF domains from different proteins, albeit with the same fold, are quite divergent in sequence (Figure 1B). Consequently, they most likely have different molecular surface properties. As summarized in Table 1 of Ester and Uetz⁴, FF domains can bind to a variety of protein motifs. Thus, it is clear that not all FF domains are created equal. For example, CA150 FF domains 1, 3, and 4 do not bind to the pCTD (as judged by Far Western), whereas FF 5 binds strongly and FF2 has weak binding¹¹. Not surprisingly, the binding surface on Prp40 for the Clf1 TPR1 motif¹⁹ is not the same as that on Prp40 for the RNAP II CTD phosphopeptide¹⁸. This can probably be rationalized, in part, by the different pKa values of the different FF domains. In the case of CA150, FF1 and FF2 have similar basic pKa's (9.2 and 9.1, respectively) and FF3 and FF4 have similar neutral values (6.9 and 6.1, respectively)²⁰. Considering all these, it seems more plausible that the tandem yet versatile FF domains act as a scaffold to bind to a diverse repertoire of molecules, allowing the subsequent functional events to occur in close proximity in a concerted manner.

Potential DNA binding of CA150 FF1-3

Structural analysis using the ProFunc server^{29, 30} suggested that the CA150 FF2 domain could bind to DNA: this domain is structurally similar to the yeast DNA binding homeodomain protein MAT $\alpha 2$ bearing three alanine mutations. The similarity is limited to the last helix of FF2 along with its C-terminal connector of CA150 and the DNA recognition helix ($\alpha 3$) of MAT $\alpha 2$ (Figures 7A and 7B). The crystal structure of the mutant MAT $\alpha 2$ was determined as a ternary complex with its binding partner MAT $\alpha 1$ and DNA³¹. In this structure, the wild-type MAT $\alpha 2$ DNA-interacting residues (S50, N51, and R54) were simultaneously mutated to alanines thus affording higher sequence homology with the FF construct used in this study (Figure 7C).

Computer modeling was used to explore the hypothesis that our structure can bind to DNA. Since there has been no report on FF domains binding to DNA we used the DNA duplex sequence reported by Ke et al.³¹ as a template in our trials. Results of energy minimization using steepest descent and conjugate gradient methods supported this hypothesis showing that with some local adjustment, the relevant helix can indeed interact with the major groove of the DNA (results not shown).

To experimentally test potential DNA binding of CA150 FF1-3, we used a fluorescein-labeled oligonucleotide duplex in a FP binding assay. We used the same oligonucleotide duplex employed by Ke et al.³¹ in their structural studies. Several protein samples were used, including wild-type FF1-3, GST-FF1-3, methylated FF1-3, and a mutant (E775A) in FF1-3. As shown in Figure 7D and Table 2, wild-type FF1-3 binds to DNA with a K_d of 3.87 μ M, close to that of MAT $\alpha 2$ -DNA binding, 4.4 μ M³¹. GST-FF1-3 displayed higher binding affinity with a K_d of 0.52 μ M, whereas the methylated FF1-3 showed a marginally lower affinity with a K_d of 5.65 μ M. Interestingly, the E775A mutant had a higher binding affinity (K_d of 0.81 μ M) than the wild-type. In MAT $\alpha 2$, E775 corresponds to N47, which is involved in interacting with several bases via either direct or water-mediated interactions. Without a structure of FF1-3 bound to DNA, it is difficult to rationalize this increased binding. One possible explanation is that in the modeled structure, E775 is along the side of the helix that is rather hydrophobic, other than E775 itself. By changing it to an Ala, this side of the helix now is more hydrophobic, which would favor the van der Waals interactions mediated by the Ala residues 50 and 51 in MAT $\alpha 2$. Also, the water molecules trapped between the protein and DNA, when E775 was not mutated to Ala, would be released, which is entropically favorable. Overall, consistent with the MAT $\alpha 2$ -3A binding

with DNA, the flexibility of this protein-DNA interface seems tolerate other mutations in the protein sequence.

It is not surprising that the overall DNA binding is modest since the optimal sequence is yet to be identified. Yet, FF1-3 has higher binding affinity for DNA than for the peptide by more than two orders of magnitude (3.87 vs. >400 μM). Also, it has been shown that the mutant MAT α 2 alone binds to DNA with a rather weak affinity³¹. In the presence of its partner MAT α 1, the binding affinity increased by 20-fold. Compared to the classic homeodomains, even the wild-type MAT α 2 binds to DNA with modest affinity. Therefore, it is conceivable that full-length CA150 alone or in complex with other associating factors may exhibit higher affinity of DNA binding.

MAT α 1 and α 2 belong to the large homeodomain transcription factor superfamily. DALI searches using individual FF domain structures indeed identified quite a few homeodomain proteins as structurally similar to FF domains. Significantly, a search using the structure of CA150 FF2 yielded a list on which seven out of the eight FF domain NMR structures currently in the PDB were the top seven hits. The next three hits were the yeast MAT α 2 structures (chain A of 1K6132, chain C of 1APL33, and chain D of 1K6132). These three have the same Z score (4.4) and respective RMSD of 2.9, 2.8, and 2.8 Å. Unexpectedly, the FF domain of yeast URN1 (2JUC) had a Z score of 4.1 and RMSD of 3.0 Å and ranks even lower on the list. In comparison, searches using CA150 FF1 and FF3 structures did not give rise to as highly ranked homeodomain hits as that using FF2. When FF1 was used the top homeodomain hit (#17 on the list) was 2HI3, with a Z score of 4.4 and RMSD of 5.7 Å. In the case of FF3, the top hit (#59 on the list) was 1YRN, with a Z score of 3.9 and RMSD of 3.2 Å. Overall, we demonstrated that FF2 could bind to DNA *in vitro*. Its resemblance to a DNA-binding homeodomain, the high positive-charge content in this region of the protein, and the nuclear location of CA150 are consistent with its potential of DNA binding and its functional role in transcription.

Concluding Remarks

In this study we present the first crystal structure of an FF domain and provide insights into the tandem nature of the first three FF domains in CA150. We reason that a specific conformation, as mediated by GST in the construct, might be required for the FF domains to bind to their peptide ligand WFHVEED. We also explored the potential of DNA binding of FF proteins. Although FF domains are typically involved in protein-protein interactions, our results suggest the activities of an FF domain may also include DNA binding. Future investigations are necessary to identify the DNA binding sites as well as relevant protein partners *in vivo*. In summary, our results are consistent with a model in which the versatile CA150, as an assistant to RNAP II, functions as a scaffold through its multiple protein modules to recruit other transcription and mRNA processing factors and couples the two processes near the polymerase.

Materials and Methods

Protein expression and purification

The DNA sequence encoding residues 661-845 of human CA150 (GenBank accession number AAB80727) was amplified by PCR and inserted into a pGEX-6P vector (GE Healthcare). Site-directed mutagenesis was performed using QuikChange (Stratagene) and the mutation was confirmed by DNA sequencing. *E. coli* strain Rosetta (DE3) (Novagen) was used as the host and the cells were grown at 37°C until OD₆₀₀ of the culture reached 0.7 to 0.8. Expression of the fusion protein was induced using 0.1 mM IPTG and the cells were allowed to grow at 18°C overnight. All of the subsequent steps were performed at 4°C. Cells

were resuspended in Buffer GA (50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT, 1 mM EDTA, 0.1 mM PMSF, and 0.7 mg/ml pepstatin A) and lysed by sonication. After centrifugation to remove the unlysed cells and cell debris, the supernatant fraction of the cell lysate was applied to a Glutathione S-Sepharose Fast Flow column (GE Healthcare) pre equilibrated with buffer GA. After extensive washing of the column, the fusion protein was eluted with buffer GB (GA with 10 mM reduced glutathione). The fusion protein was digested by PreScission protease overnight. The digestion products were applied onto an SP sepharose column (GE Healthcare) pre equilibrated with buffer SA (20 mM Tris-HCl, pH 7.5, 1 mM DTT, 1 mM EDTA, 0.1 mM PMSF, and 0.7 mg/ml pepstatin A), and separated with a linear gradient of 0-500 mM KCl in Buffer SA. The fractions containing FF1-3 were pooled, concentrated to a minimal volume, and applied onto a Superdex 75 column (GE Healthcare) pre equilibrated with GFB buffer (20 mM Hepes, pH 7.2, 100 mM KCl, 1 mM DTT, and 1 mM EDTA). The peak fractions were pooled, concentrated to more than 20 mg/ml, aliquoted, flash frozen in liquid N₂ and stored at -80°C.

Reductive lysine methylation

Reductive methylation reaction was performed essentially as published²². Briefly, purified FF1-3 protein was diluted with the methylation buffer (50 mM Hepes, pH 7.5 and 250 mM NaCl) to 1 mg/ml final concentration. To each milliliter of the diluted protein solution, 20 µl of freshly prepared 1 M dimethylamine-borane complex (Fluka) and 40 µl of 1 M formaldehyde (Fluka) were added. The reaction was carried out on ice for 2 h. Then 20 µl of 1 M dimethylamine-borane complex and 40 µl of 1 M formaldehyde were added for each milliliter of the reaction, which was allowed to continue for another 2 h. Then another 20 µl of 1M dimethylamine-borane complex per ml of solution was added and the reaction solution was kept on ice overnight. Finally, the reaction mixture was subject to another round of gel filtration using a Superdex 75 column pre-equilibrated with the GFB buffer.

Mass spectrometry

One microgram of native or chemically methylated FF1-3 protein was injected onto a 1 × 50 mm C18 Vydac 218MS5105 reversed phase HPLC column and eluted with a linear gradient in a 0.02% trifluoroacetic acid/acetonitrile buffer system. The eluent from the column was directly infused into an Agilent LC-MSD TrapSL ion-trap mass spectrometer for electrospray ionization and mass analysis. Masses were deconvoluted from the observed charge state envelopes using Bruker LC-MS data analysis software.

Protein Crystallization

Lysine-methylated CA150 FF1-3 was subject to crystallization trials at 4°C using the “hanging-drop” vapor diffusion method. Crystals were obtained by mixing 1 µl of 10-15 mg/ml protein, 1 µl of reservoir solution [50 mM Bis-Tris, pH 6.5, 50 mM ammonium sulfate, 30% pentaerythritol ethoxylate (Hampton Research)] and 0.5 µl of 5 mM ATP (without pH adjustment). The best crystals were obtained in about a week with a typical size of 400 × 400 × 150 µm³. SeMet-substituted FF1-3 was also purified and crystallized under similar conditions.

Data collection, Processing, and Structure Determination

FF1-3 crystals were treated with a series of cryo-protection solutions with 5-20% PEG400 (in 5% increments) in the reservoir solution and flash frozen in liquid N₂. Diffraction data were collected at Beamlines A1 and F1 of Cornell High Energy Synchrotron Source (CHESS), Ithaca, NY, and Beamline 8.3.1 of the Advanced Light Source (ALS), Berkeley, CA, using ADSC CCD Quantum detectors. Data were processed and reduced using DENZO and SCALEPACK programs³⁴. The crystals are in space group P6₂2 with cell dimensions

of $a=b=141.30$ Å, and $c=155.42$ Å, $\alpha=\beta=90^\circ$, and $\gamma=120^\circ$. There are two molecules in each asymmetric unit. Heavy atom positions were located and refined against a Se-SAD dataset collected at the selenium edge (0.9796 Å) using the Phaser module³⁵ of Phenix^{36, 37}. Solvent modification was carried out using the Resolve module³⁸ of Phenix. The resultant electron density map was of excellent quality, enabling us to build all of the residues of one molecule and most of the other, with casual guidance from the individual FF domain NMR structures. Manual model building was performed using programs O39 and Coot⁴⁰ and refinement using Phenix. The final model contains all of the construct residues except 737-757 in the first molecule. In addition, the five residues GLPSP in each molecule (introduced from cloning) are included. The model quality was assessed using PROCHECK⁴¹ in the CCP4 suite⁴². Most of the residues are in the most favored (92.0%) or additionally allowed (7.7%) regions and none in the disallowed region.

Peptide Synthesis

Two peptides (WFHVEED and AWFHVEED) were synthesized using manual solid phase Fmoc methodology. The former was used in crystallization experiments, whereas the latter was labeled at the N-terminus with 5-carboxy-fluorescein as an amide for fluorescence polarization binding studies. These peptides were purified and composition was confirmed by mass spectrometry.

Fluorescence Polarization

FP experiments were performed as previously described⁴³ to measure the binding between the peptide and the protein. Briefly, 50 µl of 20 nM fluorescent-labeled peptide in 50 mM NaCl, 10 mM Hepes, pH 7.5, 1 mM EDTA, 2 mM DTT, and 1% NP-40 were placed in the wells of a 96-well microtiter plate. To each well it was added 50 µl of a protein solution of up to 2 mM concentration in the same buffer. Fluorescence polarization was then read in a Tecan Polarian plate reader. With use of Prism (GraphPad Software), the millip values were plotted against the protein concentrations and affinity values were obtained from non-linear regression analysis in the “one site specific -binding” mode, as defined by Prism. For FP studies on DNA binding, the same procedure was followed with the following modifications: 10 µl of 20 nM fluorescein (FLC)-labeled oligonucleotide duplex was used to mix with 10 µl of a protein sample up to 0.66 mM in the wells of a 384-well plate. The sequences of the two strands of the duplex are respectively 5'-(FLC) ACATGTAAAAATTTACATCA-3' and 5'- TTGATGTAAATTTTACATG-3'.

Multi-angle Light Scattering and Dynamic Light Scattering

Purified FF1-3 and GST-FF1-3 were applied to a Superdex 200 gel filtration column (GE Healthcare), pre-equilibrated with 20 mM Tris, pH 7.5 and 250 mM NaCl. Light scattering and refractive index signals were measured using a Wyatt Optilab and Dawn EOS system. Scattering curves were processed using the provided Astra software package^{44, 45}.

Purified FF1-3 and GST-FF1-3 proteins at 10 mg/ml in the GFB buffer were centrifuged at 4°C for 20 min at a speed of $13,200 \times g$. Fifteen microliters of a protein solution were transferred to a quartz cuvette and allowed to equilibrate to 4°C before taking an average of twenty readings in a DynaPro instrument (Protein Solutions). Data were analyzed using Dynamics software suite (Protein Solutions).

Computer Modeling

Structural model of FF1-3 was superimposed onto that of the mutant MAT α 2 as in the MAT α 1- α 2-DNA complex (PDB ID: 1LE8)³¹, based on the sequence alignment as shown in Figure 7C. After setting the potentials using the Amber99 force field in InsightII

(Accelrys, Inc), the DNA duplex and CA150 residues 718-800 (FF2 and its C-terminal extension) were combined as an assembly and subject to 5000 steps each of steepest descent and conjugate gradient minimization using a distance dependent dielectric. In a separate experiment, the starting FF-DNA assembly was soaked with a layer of water molecules 6 Å thick. The same minimization protocol was used except that no distance dependence was employed.

Acknowledgments

We thank Ke Zhang and staff members at CHESS Beamlines A1 and F1 and ALS Beamline 8.3.1 for help on diffraction data collection. We are also grateful to the members of the Bedford and Chen groups for stimulating discussions and insightful suggestions and to Drs. James Murphy and Tony Pawson for communication of their results prior to publication. This work is supported in part by grants from the Robert Welch Foundation to M.T.B. (G-1495) and X.C. (G-1499) and the NIH to J.S.M. (CA096652), R.H.J. (GM069769), and X.C. (GM068556). CHESS is supported by the NSF (DMR-0225180) and the MacCHESS resource is supported by NIH/NCRR (RR-01646).

References

1. Sadowski I, Stone JC, Pawson T. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol Cell Biol.* 1986; 6:4396–408. [PubMed: 3025655]
2. Pawson T. Non-catalytic domains of cytoplasmic protein-tyrosine kinases: regulatory elements in signal transduction. *Oncogene.* 1988; 3:491–5. [PubMed: 3078956]
3. Bedford MT, Leder P. The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem Sci.* 1999; 24:264–5. [PubMed: 10390614]
4. Ester C, Uetz P. The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. *BMC Biochem.* 2008; 9:29. [PubMed: 19014439]
5. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* 2006; 34:D257–60. [PubMed: 16381859]
6. Sune C, Hayashi T, Liu Y, Lane WS, Young RA, Garcia-Blanco MA. 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.* 1997; 17:6029–39. [PubMed: 9315662]
7. Sune C, Garcia-Blanco MA. Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner. *Mol Cell Biol.* 1999; 19:4719–28. [PubMed: 10373521]
8. Pearson JL, Robinson TJ, Munoz MJ, Kornblihtt AR, Garcia-Blanco MA. Identification of the cellular targets of the transcription factor TCERG1 reveals a prevalent role in mRNA processing. *J Biol Chem.* 2008; 283:7949–61. [PubMed: 18187414]
9. Goldstrohm AC, Albrecht TR, Sune C, Bedford MT, Garcia-Blanco MA. The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol Cell Biol.* 2001; 21:7617–28. [PubMed: 11604498]
10. Carty SM, Greenleaf AL. Hyperphosphorylated C-terminal repeat domain-associating proteins in the nuclear proteome link transcription to DNA/chromatin modification and RNA processing. *Mol Cell Proteomics.* 2002; 1:598–610. [PubMed: 12376575]
11. Carty SM, Goldstrohm AC, Sune C, Garcia-Blanco MA, Greenleaf AL. Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc Natl Acad Sci U S A.* 2000; 97:9015–20. [PubMed: 10908677]
12. Cheng D, Cote J, Shaaban S, Bedford MT. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol Cell.* 2007; 25:71–83. [PubMed: 17218272]
13. Sanchez-Alvarez M, Goldstrohm AC, Garcia-Blanco MA, Sune C. Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol Cell Biol.* 2006; 26:4998–5014. [PubMed: 16782886]

14. Holbert S, Denghien I, Kiechle T, Rosenblatt A, Wellington C, Hayden MR, Margolis RL, Ross CA, Dausset J, Ferrante RJ, Neri C. The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis. *Proc Natl Acad Sci U S A*. 2001; 98:1811–6. [PubMed: 11172033]
15. Kao HY, Siliciano PG. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol Cell Biol*. 1996; 16:960–7. [PubMed: 8622699]
16. Morris DP, Greenleaf AL. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*. 2000; 275:39935–43. [PubMed: 10978320]
17. Jiang W, Sordella R, Chen GC, Hakre S, Roy AL, Settleman J. An FF domain-dependent protein interaction mediates a signaling pathway for growth factor-induced gene expression. *Mol Cell*. 2005; 17:23–35. [PubMed: 15629714]
18. Allen M, Friedler A, Schon O, Bycroft M. The structure of an FF domain from human HYPA/ FBP11. *J Mol Biol*. 2002; 323:411–6. [PubMed: 12381297]
19. Gasch A, Wiesner S, Martin-Malpartida P, Ramirez-Espain X, Ruiz L, Macias MJ. 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*. 2006; 281:356–64. [PubMed: 16253993]
20. Bonet R, Ramirez-Espain X, Macias MJ. 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*. 2008; 73:1001–9. [PubMed: 18536009]
21. Bonet R, Ruiz L, Aragon E, Martin-Malpartida P, Macias MJ. NMR structural studies on human p190-A RhoGAP FF1 revealed that domain phosphorylation by the PDGF-receptor alpha requires its previous unfolding. *J Mol Biol*. 2009 Apr 22. Epub ahead of print.
22. Walter TS, Meier C, Assenberg R, Au KF, Ren J, Verma A, Nettleship JE, Owens RJ, Stuart DI, Grimes JM. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*. 2006; 14:1617–22. [PubMed: 17098187]
23. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*. 2001; 98:10037–41. [PubMed: 11517324]
24. Murphy JM, Hansen DF, Wiesner S, Muhandiram DR, Borg M, Smith MJ, Sicheri F, Kay LE, Forman-Kay JD, Pawson T. Structural studies of FF domains of the transcription factor, CA150, provide insights into the organization of FF domain tandem arrays.
25. Smith MJ, Kulkarni S, Pawson T. FF domains of CA150 bind transcription and splicing factors through multiple weak interactions. *Mol Cell Biol*. 2004; 24:9274–85. [PubMed: 15485897]
26. Riley LG, Ralston GB, Weiss AS. Multimer formation as a consequence of separate homodimerization domains: the human c-Jun leucine zipper is a transplantable dimerization module. *Protein Eng*. 1996; 9:223–30. [PubMed: 9005444]
27. Garcia De La Torre J, Huertas ML, Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J*. 2000; 78:719–30. [PubMed: 10653785]
28. Kleywegt GJ. Validation of protein models from C α coordinates alone. *J Mol Biol*. 1997; 273:371–6. [PubMed: 9344745]
29. Laskowski RA, Watson JD, Thornton JM. Protein function prediction using local 3D templates. *J Mol Biol*. 2005; 351:614–26. [PubMed: 16019027]
30. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res*. 2005; 33:W89–93. [PubMed: 15980588]
31. Ke A, Mathias JR, Vershon AK, Wolberger C. Structural and thermodynamic characterization of the DNA binding properties of a triple alanine mutant of MAT α 2. *Structure*. 2002; 10:961–71. [PubMed: 12121651]
32. Aishima J, Gitti RK, Noah JE, Gan HH, Schlick T, Wolberger C. A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res*. 2002; 30:5244–52. [PubMed: 12466549]
33. Wolberger C, Vershon AK, Liu B, Johnson AD, Pabo CO. Crystal structure of a MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*. 1991; 67:517–28. [PubMed: 1682054]

34. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997; 276:307–326.
35. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst.* 2007; 40:658–674. [PubMed: 19461840]
36. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1948–54. [PubMed: 12393927]
37. Zwart PH, Afonine PV, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, McKee E, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Storoni LC, Terwilliger TC, Adams PD. Automated structure solution with the PHENIX suite. *Methods Mol Biol.* 2008; 426:419–35. [PubMed: 18542881]
38. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr.* 2000; 56:965–72. [PubMed: 10944333]
39. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A.* 1991; 47(Pt 2):110–9. [PubMed: 2025413]
40. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–32. [PubMed: 15572765]
41. Laskowski RA, MacArthur MW, Moss DS, Thornton J. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst.* 1993; 26:283–91.
42. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 1994; 50:760–3. [PubMed: 15299374]
43. Coleman, DRt; Ren, Z.; Mandal, PK.; Cameron, AG.; Dyer, GA.; Muranjan, S.; Campbell, M.; Chen, X.; McMurray, JS. Investigation of the binding determinants of phosphopeptides targeted to the SRC homology 2 domain of the signal transducer and activator of transcription 3. Development of a high-affinity peptide inhibitor. *J Med Chem.* 2005; 48:6661–70. [PubMed: 16220982]
44. Foltá-Stogniew E, Williams K. Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J Biomol Tech.* 1999; 10:51–63. [PubMed: 19499008]
45. Wyatt PJ. Combined differential light scattering with various liquid chromatography separation techniques. *Biochem Soc Trans.* 1991; 19:485. [PubMed: 1889645]
46. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–8. [PubMed: 17846036]
47. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics.* 1999; 15:305–8. [PubMed: 10320398]

Figure 1.

Structures of human FF domain-containing proteins. (A) Domain structure of CA150. The N-terminal half of the molecule is strongly compositionally biased with two proline-rich, one alanine/glutamine-rich, one threonine-rich, and one glutamate-rich regions (all in grey). There are also three scattered WW domains (red) in the N-terminal region. The C-terminal half is mostly occupied by six FF domains (green). Three potential coiled-coil regions (blue) and one NLS (magenta) have also been predicted within the molecule. (B) Sequence alignment of the FF domains in human proteins: CA150 (SWISS-Prot ID: O14776), PRPF40A (O75400), PRPF40B (Q6NWX9), P190A (Q9NRY4), and P190B (Q13017). The FF domain sequences were aligned using ClustalX (2.0.10)46 with manual adjustments and alignment outcome was rendered using web-based Esript47. (C) Superimposition of the individual NMR structures of the FF domains currently in the PDB. This diagram and all of the structure diagrams throughout the manuscript were prepared using PyMOL. The structural models used are chain A of each of the following NMR structures: human FBP11/FNBP3 FF1 (PDB ID: 1UZC) and FF5 (PDB ID: 2CQN); yeast Prp40 FF1 (PDB ID: 2B7E); human CA150 FF1, FF2, FF3, and FF4 (PDB ID's: 2DOD, 2E71, 2DOE, and 2DOF, respectively); and the yeast URN1 FF domain (PDB ID: 2JUC).

Figure 2.

Compound mass spectrum deconvolution confirming the identity of the purified protein (top panel; measured [M+H]⁺ molecular weight $22,646.0 \pm 0.69$ Da, as compared to the calculated MW of 22,643.1 Da) and the quantitative di-methylation of the 30 lysine side chains as well as the N-terminal NH₂ group (bottom panel; measured [M+H]⁺ molecular weight $23,513.8 \pm 0.83$ Da). The difference between the two MW's is 867.8 Da, as Compared to 868 Da for 100% dimethylation of the 31 NH₂ groups.

Figure 3. Overall structure of CA150 FF1-3. (A, B, and C) “Cartoon” diagrams of the structure related by $\pm 90^\circ$ rotations along the x-axis. FF domains (boundaries defined in Figure 1B) are colored in red and the connectors are in green. The dimensions of the structure and the cross angle of the two connector helices are indicated. (D) Superimposition of the NMR structures of CA150 FF1 (cyan), FF2 (magenta), and FF3 (yellow), onto the crystal structure of FF1-3 (green; same orientation as in Fig. 3B). The termini of the NMR structures are labeled by their corresponding residues.

Figure 4. Electrostatic properties of FF1-3. (A) "Cartoon" diagram showing the location of the charged residues on the protein. Positively (KR) and negatively (DE) charged residues are colored in blue and red, respectively, whereas the other residues are in lighter colors. (B) Surface diagram showing the electrostatic potential of the protein, as calculated using APBS23.

Figure 5.

FP studies showing the binding of peptide FAM-WFHVEED to protein samples. Top curve (closed square): GST-FF1-3 fusion, middle curve (closed circle): FF1-3, bottom curve (closed diamond): GST. The average and standard deviation values reported here are from six independent experiments.

Figure 6.

MALS characterization of FF1-3 and GST-FF1-3. Purified FF1-3 and GST-FF1-3 sample was subjected to size exclusion chromatography on a Superdex 200 column (20 mM Tris-HCl, pH 7.5 and 250 mM NaCl) with a flow rate of 0.5 ml/min. The filtrate was analyzed online by a multi-angle laser light scattering device (Wyatt). The light scattering signal at 90° (left axis, linear scale) and the measured molecular weight in Da (right axis, log scale) were plotted against the elution volume in ml. FF1-3 (red curve) appeared monodisperse with an apparent molecular mass of 23.2 kDa (green curve), very close to the calculated MW of 22.6 kDa. The majority of GST-FF1-3 (blue curve) appeared also monodisperse and dimeric (purple curve), whereas a small portion of the fusion protein appeared heterogeneous, forming tetramer and/or higher molecular weight aggregates (brown curve).

Table 1

Summary of crystallographic analysis (a single SeMet crystal was used for SAD data collection).

Data Collection Statistics	
	SeMet-SAD
Space group	P622
Unit cell	
a, b, c (Å)	141.30, 141.30, 155.42
α, β, γ (°)	90, 90, 120
Energy (wavelength)	12,656.7 eV (0.97959 Å)
Resolution range (Å)	50.0-2.70 (2.80-2.70)
Completeness (%)	99.9 (100.0)
Redundancy	11.7 (11.2)
I/ σ I	34.7 (3.2)
R _{merge} (%)	10.9 (56.3)
Refinement statistics	
Numbers of Reflections	20.0-2.70 Å
Working set	39877
Test set	4331
Number of atoms	3184
Rmsd Bonds (Å)	0.007
Rmsd Angles (°)	1.153
R _{conventional} ^a (%)	23.0
R _{free} ^b (%)	28.3
Ramachandran plot	
Most favored	311 (92.0%)
Additionally allowed	26 (7.7%)
Generously allowed	1 (0.3%)
Disallowed	0 (0.0%)

^aR_{conventional} = $\Sigma |F_{obs} - F_{calc}| / \Sigma F_{obs}$.

^bR_{free} was calculated as R_{conventional} by using 10% of the data not included in refinement.

Table 2

DNA binding properties as determined by fluorescence polarization. Data presented in Figure 7D were fitted with “one-site specific binding” mode [$Y=B_{\max} * X / (K_d + X)$], as defined in Prism (GraphPad Software). The resultant binding dissociation constant K_d and goodness of fit R^2 values are reported here.

Protein	K_d (μ M)	R^2
FF1-3 wild-type	3.87 $\pm$ 0.31	0.9856
FF1-3 E775A mutant	0.81 $\pm$ 0.06	0.9866
GST-FF1-3	0.52 $\pm$ 0.06	0.9546
Methylated FF1-3	5.65 $\pm$ 0.53	0.9819