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

Alexander Gasch1#, Silke Wiesner1,2#, Pau Martin-Malpartida3, Ximena Ramirez-Espain3, Lidia Ruiz3 and Maria J. Macias3*

1Structural Biology Program, EMBL Heidelberg, Meyerhofstr. 1, 69117 Heidelberg, Germany, 2Present address: Structural Biology and Biochemistry, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada 1ICREA and the Biomedical Research Institute of Barcelona, Barcelona Science Park, Josep Samitier 1-5, E-08028 Barcelona, Spain

Running title: Structure and ligand binding of the Prp40 FF1 domain

Address correspondence to: Maria J. Macias, Josep Samitier 1-5, E-08028 Barcelona, Spain, Tel.: +34-93-403 71 89 / Fax: +34-93-403 71 14 e-mail: mmacias@pcb.ub.es

# Equal contribution

The yeast splicing factor Prp40 (pre-mRNA processing protein 40) consists of a pair of WW domains followed by several FF domains. The region comprising the FF domains has been shown to associate with the 5' end of U1 snRNA and to directly interact with two proteins, the Crooked neck-like factor 1 (Clf1) and the phosphorylated repeats of the carboxy-terminal domain of RNA polymerase II (CTD-RNAPII). In this work we report the solution structure of the first FF domain of Prp40 and the identification of a novel ligand binding site in FF domains. Using chemical shift assays we found a binding site for the N-terminal crooked neck tetratrico peptide repeat of Clf1 that is distinct and structurally separate from the previously identified CTD-RNAPII binding pocket of the Formin Binding Protein 11 (FBP11) FF1 domain. No interaction, however, was observed between the Prp40 FF1 domain and three different peptides derived from the CTD-RNAPII protein. Indeed, the equivalent CTD-RNAPII binding site in the Prp40 FF1 domain is predominantly negatively charged and, thus, unfavorable for an interaction with phosphorylated peptide sequences. Sequence alignments and phylogenetic tree reconstructions using the FF domains of three functionally related proteins, Prp40, FBP11, and CA150 reveal, that Prp40 and FBP11 are not orthologous proteins and support the different ligand specificities shown by their respective FF1 domains. Our results also reveal that not all FF domains in Prp40 are functionally equivalent. We propose that at least two different interaction surfaces exist in FF domains that have evolved to recognize distinct binding motifs.

Splicing of pre-mRNA is catalyzed by the spliceosome, a large ribonucleoprotein complex composed of five U snRNPs and about 100 accessory, non-snRNP splicing factors (1). In early stages of spliceosome assembly, the splicing factor Prp40 associates with the U1 snRNP and plays an important role in bringing the 5' and the 3' splice site into spatial proximity (2). Prp40 is a modular protein comprising an N-terminal WW domain pair followed by several FF domains. While the Prp40 WW domains bind to proline-rich sequences in the branch-point binding protein BBP and the U5 snRNP-associated protein Prp8 (2,3), the region spanning the FF domains has been shown to associate with the 5' end of U1 snRNA (4) and to interact with at least two different proteins: the splicing factor CFI1 / Syt3p (crooked neck-like factor 1 / Synthetic lethal with cdc forty) (5-7), and the C-terminal domain (CTD) of the largest subdomain of RNA polymerase II (RNAPII) (8).

FF domains were first identified in the murine splicing factor FBP11 (formin binding protein 11) as a repeated sequence of about 60 amino acids containing two conserved phenylalanine (F) residues that give name to the domain (9). While protein-interaction modules are commonly found in functionally unrelated proteins, FF domains are primarily present in only three protein families: the splicing factors FBP11 and Prp40, the transcription factor CA150, and p190 RhoGTPase-related proteins (9). With only few exceptions, FF domains are arranged in arrays of up to six domains that in the case of CA150...
seem to create multiple independent binding sites, rather than to confer binding cooperativity (10). Interestingly, Prp40, FBP11 and CA150 have all been shown to recognize phospho-CTD repeats through their FF domains (11). Some progress has been made to understand how FF domains recognize phosphorylated ligands. In this respect, the solution structure of the N-terminal FF domain (FF1) of human Huntingtin yeast partner (HYP/A/FBP11 has shown that the FF domain fold consists of three α-helices arranged as an orthogonal bundle and a $3_0$ helix in the loop connecting the second and the third helix (12). Using chemical shift perturbation experiments Bycroft and coworkers were able to identify the binding site of HYP/A/FBP11 FF1 domain for a peptide corresponding to two doubly phosphorylated RNAPII CTD repeats, leading to the suggestion that FF domains may represent a new class of phosphopeptide binding modules (12). Significantly, the phospho-ligand recognition site described for the HYP/A/FBP11 FF1 domain is only conserved in a particular subset of FF sequences, indicating that different binding motifs may exist for other FF domains.

For the splicing factor Prp40, previous studies have shown that a region including its first FF domain interacts with the protein Clf1. Clf1/Syf3p is an essential and well-conserved multifunctional protein involved in cell-cycle progression, pre-mRNA splicing and initiation of DNA replication in yeast (5-7). Clf1 is composed of fifteen crooked neck-like tetra-trico-peptide repeats (crn-TPR) (5,13) (6). TPR motifs represent a class of ubiquitous protein-interaction modules of partially conserved sequence. While some RNAPII CTD heptapeptides form beta-turns at the central SPTS motif in the presence of the target protein (14-16) and others adopt an extended conformation (17), TPR motifs in general fold as a pair of anti-parallel α-helices. Repeats of TPR motifs often assemble into right-handed helical superstructures creating extensive binding surfaces with diverse specificities (18). Furthermore, in contrast to the RNAPII CTD, phosphorylation does not seem to occur in TPR motifs. The diversity in not only the function but also in the structure of these Prp40 FF domain binding partners raises important questions as to whether distinct binding specificities exist in FF domains and how the FF domain fold enables this potential promiscuity in ligand binding.

To gain more insight into the interaction mode of FF domains, we have determined the solution structure of the first FF domain of Prp40 and characterized its interaction with the first crn-TPR motif (crn-TPR1) of Clf1 by chemical shift perturbation experiments. Interestingly, the interaction surfaces of the Prp40 and HYP/A/FBP11 FF1 domains are adjacent, but structurally separate. NMR binding studies with the Prp40 FF1 domain using phosphorylated and unphosphorylated RNAPII CTD repeats revealed no interaction. Furthermore, no interaction was observed between the C-terminal FF domain (FF4) of Prp40 and the Clf1 crn-TPR1 motif showing that not all FF domains in the splicing factor Prp40 are functionally equivalent. Sequence alignments and phylogenetic reconstructions using the FF domains of Prp40, HYP/A/FBP11, and CA150 proteins reveal that Prp40 and HYP/A/FBP11 are not orthologous proteins and support the different ligand specificities shown by their respective FF1 domains. Taken together, at least two spatially separate interaction surfaces seem to exist in FF domains that recognize distinct binding motifs.

Materials and Methods

Sequence alignment and phylogenetic tree reconstructions - Sequence alignments were performed using T-Coffee (19). For initial alignments all predicted FF domain sequences in the literature (8) and those found in databases, such as (http://smart.embl-heidelberg.de) and (http://www.sanger.ac.uk/Software/Pfam), were used. After an initial alignment, FF domains that did not satisfy our confidence level were manually removed resulting in the final alignment (see Supplementary Fig. 1). Proteins annotated in databases as hypothetical were classified by domain architecture and length of the linkers connecting the WW domains as follows: proteins containing two WW domains separated by a short (~15aa) conserved linker were initially classified as similar to FBP11 proteins, whereas proteins with two or more WW domains separated by long, variable linkers were categorized as CA150-related. Since in both S. cerevisiae (Sc.) and S. pombe (Sp.) there is only a unique Prp40 protein, while all remaining organisms contain at least one CA150-type (transcription factor) and one FBP11-
type (splicing factor) proteins, independence of the Prp40 subclass was maintained for the analysis. As sequence identity between equivalent FF domains within each of the three protein families is higher than 75%, two, three and four representative protein sequences were selected from Prp40, CA150 and FBP11 proteins, respectively, for sequence alignment. Phylogenetic trees were generated with Phylip 3.5 based on the previous alignment. The protist program was used to compute distance matrices from protein sequences employing the Dayhoff PAM matrix and neighbor programs. The final tree is shown in Figure 5A.

The SwissProt/UniProt accession numbers of the protein sequences are: CG3542: Dm. FBP11, Q90WG3: Gg. FBP11 (Formin binding protein 11-related protein), the FF domain containing region is identical to human protein FBN3_HUMAN, (Huntingtin yeast partner A/Huntingtin interacting protein HYP4/FBP11) and due to this we named the sequence as Hs/Gg HYP4/FBP11, Q6NWY9: Hs. HYP4/FBP11 (Huntingtin-interacting protein C/ Formin binding protein 11-related protein), Q8IEF0: Pf. FBP11, O14776: Hs. CA150, Q86MP: Ce. CA150, Q7PMV0: Ag. CA150, O14176: Sp. Prp40, P33203: Sc. Prp40.

Sample preparation - DNA fragments encoding the respective FF domains and TPR motifs were amplified by PCR using genomic S. cerevisiae DNA as template. FF domain constructs correspond to residues 134-189 (FF1) and 121-189 (extended FF1), and residues 488-552 and 465-552 (FF4 and extended FF4) of Prp40 and residues 212-266 of Ypr152. Protein constructs of TPR (TPR1 motif had a protein concentration of ~1 mM for structure determination and were dissolved in 20mM sodium phosphate buffer, 100mM NaCl, 0.02% (w/v) NaN3 in 90% H2O/10% D2O or 100% D2O at pH 6.2.

NMR spectroscopy - Except for chemical shift perturbation experiments, all NMR data were collected at 285K for the Prp40 FF1 domain and at 295K for the Clf1 crn-TPR1 motif on Bruker DRX-500, DRX-600 and DRX-800 NMR spectrometers equipped with triple-resonance z-gradient probes. All spectra were processed with the NMRpipe/NMRDraw package (20) and analyzed with XEASY (21), while 15N relaxation data were analyzed using NMRView (22). 1H, 13C and 15N chemical shifts were assigned by standard methods (23) using 1H,15N-HSQC, HNCA, CBCANH, CBCA(CO)NH, H3^N-J experiments for backbone assignments and 1H-13C-CT-HSQC, HC(CO)NH-, C(CO)NH- and HC(C)H-TOCSY experiments for aliphatic side-chain assignments. For the Prp40 FF1 domain aromatic side-chain assignments were achieved using 2D (H0)C0(C0C)H and (H0)C0(C0C)H experiments (24), 2D homonuclear 1H-TOCSY and NOESY, and 3D 13C-edited NOESY experiments, while the 3D 13C-edited NOESY was sufficient to assign aromatic side-chains in the Clf1 crn-TPR1 motif. Heteronuclear 1H,15N NOE experiments were recorded at 500.13 MHz proton frequency by standard two-dimensional methods (25) using a 1.0 mM 15N-labeled sample. The heteronuclear NOE experiments were run twice in an interleaved fashion with and without (reference experiment) proton saturation during the recovery delay. Errors in the peak intensities were estimated from the average base-line noise in the spectra of repeated experiments. 1H,15N NOEs were determined as the peak intensity ratio between the reference and the saturation experiment and are the average of two measurements.
Structure calculation - Inter-proton distance restraints were derived from fully assigned peaks in 3D $^{15}$N- and $^{13}$C-edited NOE experiments (using mixing times of 140 ms) integrated with the XEASY package. Hydrogen bond restraints were applied as indicated by $^3J$(H$^N$,H$^O$) coupling constants, $^{13}$C$^O$ and $^{13}$C$^N$ secondary chemical shifts and NOE patterns. Restraints for the backbone angles $\phi$ and $\psi$ were derived from the program TALOS (26) and applied where they agreed with experimental $\phi$ angles determined by quantitative $^3J$(H$^N$,H$^O$)-correlation experiments (27). For the dihedral angles predicted by TALOS the upper and lower restraint-boundaries were doubled as compared to those suggested by the program. $^3J$(N,C$^O$) coupling constants were measured by spin-echo difference experiments (28) to restrain the side-chain angle \( \chi_1 \) to 180$^\circ$ $\pm$ 40$^\circ$ for trans-rotamers and to 0$^\circ$ $\pm$ 90$^\circ$ for gauche-rotamers. For structure calculation the programs CNS (29) and ARIA 2.0 (30,31) were used with a mixed torsion and Cartesian angle dynamics simulated annealing protocol. Structures were calculated in 8 iterations producing 30 structures in each of the first 7 iterations and 50 structures in the final iteration. The 20 lowest energy structures of the final iteration were submitted to water refinement. The quality of the 10 lowest energy structures was analyzed after water refinement using the programs CNS and PROCHECK-NMR (32) and the results are summarized in Table I.

Ultracentrifugation experiments on the crn-TPR motif – Equilibrium analytical centrifugation of the crn-TPR motif at starting concentrations of 0.1mM and 0.3mM was performed on a Beckman Optima XL-I centrifuge at 285K and 295K to investigate temperature dependent aggregation of the crn-TPR1 motif. Equilibrium was achieved at 55000 rpm after 12 hours on a Ti60 rotor. Curves were fitted with the ULTRASCAN software to a monomer-dimer equilibrium profile (see Supplementary Fig. 4).

Models of the crn-TPR motif – These models were generated with Insight II using the NMR data to define the boundaries of the helices and their relative orientation (see Supplementary Fig. 5). Available structures of TPR motifs were not used as templates to generate the model since the crn-TPR motif used in this work is quite divergent with respect to the canonical TPR repeats characterized so far. In them, the packing of both helices is normally stabilized by contacts between the highly conserved residues W-XXX-G-XX-L from the first helix that form a hydrophobic pocket and the side-chain of a conserved aromatic residue of the second helix that fits into the pocket. In the case of the crn-TPR1 repeat, the equivalent positions of Trp and Gly are occupied by an Asn and a Glu, respectively, unlikely to form the classical cavity. As such, models that would have been generated only by sequence homology would have given an orientation of the helices different from the one obtained in this work, which is based on NOE data.

Chemical shift mapping experiments - Chemical shift perturbation experiments with the $^{15}$N-labeled Prp40 FF1 domain and unlabeled Clf1 crn-TPR1 were performed at 600 MHz proton frequency on a Bruker DRX-600 NMR spectrometer at 295K using 2D $^1$H,$^{15}$N-HSQC experiments. All samples were dissolved in the same NMR buffer as described above. The $^{15}$N-labeled proteins were 0.5mM in concentration. Unlabeled Clf1 crn-TPR1 was added to the $^{15}$N-labeled Prp40 FF1 domain to a final molar ligand-protein ratio of ~2:1. As a control experiment unlabeled Clf1 TPR1 was added to the $^{15}$N-labeled C-terminal FF4 domain of Prp40 to a final molar ratio of ~3:1 (data not shown).

For binding studies with the $^{15}$N-labeled Prp40 FF1 domain and peptides derived from the RNAPII CTD, a synthetic doubly phosphorylated CTD repeat with the sequence YSpSTpSPS and an unphosphorylated CTD repeat with the sequence (YSPTSPS)$_2$ were purchased from MWG-Biotech AG (Ebersberg, Germany). A (YpSTpSPS)$_2$ peptide was synthesized in house using the F-moc strategy. The FmocSer(PO(OBzl)OH)OH amino acid was purchased from Novabiochem (Darmstadt, Germany). The final peptide was cleaved with 95%TFA/5% H$_2$O and precipitated in cold ether. The crude material was purified by preparative HPLC to 90 % purity (as characterized by HPLC-MS). The peptides were added to the $^{15}$N labeled Prp40 FF1 domain up to a molar peptide: protein ratio of ~7:1. The NMR data corresponding to the (YpSTpSPS)$_2$ titration were acquired on a Bruker DRX-800 NMR spectrometer (see Supplementary
Average chemical shift changes upon ligand binding were calculated with the equation \[ \Delta \delta^{\text{av}} = \left[ \left( \Delta \delta_{\text{H}} \right)^2 + \left( \Delta \delta_{\text{N}} \right)^2 \right]^{1/2} \], where \( \Delta \delta_{\text{H}} \) and \( \Delta \delta_{\text{N}} \) are the linear change along the \(^1\text{H}\) and \(^{15}\text{N}\) axes, respectively, and \( n \) is the ratio of the chemical shift dispersion in \(^{15}\text{N}\) and \(^1\text{H}\). Chemical shift perturbations were determined to be significant if they were > 0.2 ppm.

**Affinity measurements** - The dissociation constant of the Prp40 FF1:Clf1 TPR1 complex was determined as \((150 \pm 20) \mu\text{M}\) by fluorescence spectroscopy, titrating a 60\(\mu\text{M}\) solution of Prp40 FF1 domain with a 1.1mM solution of the Clf1 crn-TPR motif. Both samples were used in the same buffer as for the NMR experiments. The temperature was set to 295K to avoid dimerization of the ligand. The excitation wavelength was 297 nm. Changes in the intensity of the fluorescence signal upon addition of the ligand at 343 nm were used for calculating the dissociation constant.

**Homology models** - FF2 homology models were generated using MODELLER6.1 (33) based on the sequence alignment shown in Supplementary Fig. 1, the lowest energy structure of the Prp40 FF1 domain and the solution structure of the HYPA/FBP11 FF1 domain (PDB entry: 1uzc). Other methods - Figures of 3D structures and surface representations were prepared with MOLMOL (34). Solvent accessible surface areas were calculated using NACCESS (35). NMR data were represented using XWinPlot, version 3.1.

### RESULTS

**Prp40 FF1 domain boundaries** - Since sequence alignments produced consistent domain boundaries for the N- and the C-terminal Prp40 FF domain (hereafter referred to as FF1 and FF4, respectively (8,9,36)), we designed two initial constructs of the Prp40 FF1 and FF4 domain (Prp40 aa134-189 and Prp40 aa488-551, respectively). We also expressed a 55 residue construct that corresponds to the unique FF domain present in Ypr152. The three constructs yielded folded samples as judged from 2D \(^{1}\text{H},^{15}\text{N}\)-correlation (HSQC) spectra. To evaluate furthermore whether the charged and hydrophobic residues preceding the Prp40 FF1 and FF4 domains respectively were important for the domain fold, protein constructs with extended N-termini were prepared, extended FF1, (121-189) and extended FF4, (471-551). However, in the 2D \(^{1}\text{H},^{15}\text{N}\)-HSQC spectra of the extended constructs we did not observed chemical shift changes for the folded residues as compared to the ones in the initial constructs. The additional peaks displayed random coil chemical shifts and besides, heteronuclear NOE experiments showed that they have negative or very small intensities, supporting the conclusion that the N-terminal extensions of both Prp40 FF1 and FF4 domains were unstructured in solution, (HSQC data shown for both FF4 constructs, Supplementary material Fig.2). In general, the flanking regions of FF domains are highly variable in sequence, further indicating that FF domains seem to fold autonomously with about 55 residues. Remarkably, one exception to the rule has been already observed in the HYPA/FBP11 FF1 structure (12), where additional residues located at the N-terminus of the FF domain directly interact with the canonical domain fold. When this ten residue extension is compared to the corresponding region of Prp40 FF1 domain, no residue is conserved, suggesting that only HYPA/FBP11 FF1 related sequences may display these additional contacts. Given the direct evidence for an interaction between the Prp40 FF1 domain and the splicing factor Clf1 we decided to focus on the structural characterization of Prp40 FF1 domain construct without N-terminal extension (aa134-189).

**Description of the Prp40 FF1 domain structure** - Analysis of the NMR spectra of the Prp40 FF1 domain resulted in almost complete assignment of \(^{1}\text{H},^{15}\text{N}\) and \(^{13}\text{C}\) resonances. For structure determination, a total of 1192 non-redundant NOE distance restraints were assigned and applied together with 42 hydrogen bond restraints and 59 dihedral angle restraints derived from \(J\)-coupling experiments and database mining using the program TALOS (26) (Table I). The final ensemble of the 10 lowest energy structures is well defined with a backbone (N, \(^{13}\text{C}^\alpha\), \(^{13}\text{C}^\prime\) atoms) rmsd of 0.20 Å to the mean structure (Fig.1d). The structural statistics for the NMR structure ensemble are summarized in Table I.
The fold of the Prp40 FF1 domain comprises three alpha-helices, α1 (aa 134-146), α2 (aa 154-163), α3 (aa 175-187), and a 3₁₀ helix (aa 167-170) located in the loop that connects the second and the third helix (Fig.1B). As in the previously determined structure of the HYPA/FBP11 FF1 domain (12), an extensive network of semi-conserved aromatic (Phe139, Phe154, Tyr168, Phe182 and Tyr185) and aliphatic residues (Ala135, Ile140, Leu143, Ile157, Leu161, Val171 and Pro175) forms the core of the domain (Fig.1B). The first loop (aa 147-153) is not disordered, in agreement with the large number of medium- and long-range NOEs observed for these residues and as judged from the average ¹H-¹⁵N heteronuclear NOE values (data not shown). Asp149, Ser150, Thr151 and Trp152 form a type I β-turn, with Val148 and Trp152 displaying numerous long-range NOEs to the core of the FF domain. The importance of Val148 for the FF domain fold is underlined by the conservation of aliphatic amino acids at this position in FF domain sequences, while W152 is exclusively present on Prp40 FF1 domains. The loop connecting the second and third helix (loop2) comprises a 3₁₀ helix (residues 167-170) with a conserved DxFY motif whose aromatic residue contributes to the hydrophobic core of the domain. Another interesting feature of this linker is the presence of three successive negatively charged residues (D172-174) that are unique to FF1 and FF4 Prp40 sequences.

Comparison with the structure of the HYPA/FBP11 FF1 domain – The structure of the N-terminal FF domain of human HYPA/FBP11 has been determined by NMR (12). With a backbone rmsd of ~ 0.7 Å (excluding α2 from the fit), the overall fold of the Prp40 FF1 domain is very similar to that described for the HYPA/FBP11 FF1 domain (Fig.1C). The main differences between the Prp40 and HYPA/FBP11 FF1 domain structures reside in the orientation of the second helix (α2). While this helix packs tightly to the core of the FF domain in the HYPA/FBP11 structure, the Prp40 structure is more open. This is probably caused by the bulky side-chain of Trp152, which intercalates between the first and second helix. The corresponding residue in the HYPA/FBP11 FF1 domain is an alanine (Ala409) that is smaller than the tryptophane and requires less space to fit into the structure (Fig.1C). While numerous long-range NOEs were observed from Trp152 to Asn146 (at the C-terminus of α1) and to Leu161 (at the C-terminus of α2) for the Prp40 FF1 domain, none of the Ala409 protons lie within 5 Å of the corresponding residues in the HYPA/FBP11 FF1 domain structure (Lys403 in α1 and Ile418 in α2). On the other hand, the H⁺ proton of Ile418 (α2) contacts many side-chain protons of Leu399 (α1) in the HYPA/FBP11 FF1 domain structure, while no NOEs were observed between the corresponding H⁺ proton of Leu161 and the side-chain of Met142 in the Prp40 FF1 domain. Moreover, the second turn of the second helix is irregular in the Prp40 FF1 domain structure, resulting from an unusual pattern of NOEs involving Ile157 (α2) and Leu161 (N_i+4 and β_i+4, respectively) instead of the regular α1 to N_i+3 / β_i+3 NOE pattern (Supplementary Fig. 3). It is worth noting that Trp152 is only conserved in the FF1 domains of yeast Prp40 (Sc. and Sp.) and that the sequence of the second helix is the least conserved region in FF domain sequences (Supplementary Fig. 1).

Interaction between Prp40FF1 and crn-TPR motifs - Although four proteins comprising TPR motifs are present in the spliceosome, only Clf1 has been shown to interact with Prp40. This interaction has been mapped to the N-terminal eight crn-TPR motifs of Clf1 and a region of Prp40 harboring the first FF domain (5-7). For proteins as Clf1 consisting of multiple structural repeats separated by short linkers, the selection of the correct motif frame is critical, since changes in the motif register will have important structural consequences (37). For construct design we have used the boundaries of Clf1 crn-TPR motifs published by Ben-Yehuda et al. (13). Of all prepared Clf1 TPR motif constructs only those corresponding to the N-terminal TPR (crn-TPR1) motif of Clf1 (aa31-64), and an extended crn-TPR1 motif (aa1-64) yielded sufficient amount of soluble and folded proteins for binding studies. To investigate the solubility and monomeric properties of the (aa31-64) crn-TPR1 motif, sedimentation equilibrium data at different temperatures were acquired. At 295K
almost all residues in the characters of both binding surfaces. Furthermore, the TPR1 binding surface underlining the different charged residues map to the determined Clf1 M170, P175, L176) and a number of negatively Prp40 FF1 domain structure (I158, P166, W169, solvent accessible surface area (SASA) in the most hydrophobic residues with more than 50% binding site of the HYPA/FBP11 FF1 domain, predominantly positively charged RNAPII CTD helix (D173, D174, L176, W177, K178, K179, W169, V171) and the N-terminal half of the /helices, namely L40, L43, R44, Y46, and Q47 in the α1 helix and T52, E53, E55, Y57, L58, and N61 in the α2 helix, with the biggest changes observed for the amides of L43, E53, Y57 and E58. According to the model, the affected residues cluster on one side of the motif (Fig 3C). The crn-TPR1 sequence used in this work is shown as Fig 3D. Unfortunately, additional information about the orientation of the Clf1 crn-TPR1 motif on the Prp40FF1 domain structure could not be obtained
from NOESY spectra of the complex due to the line broadening observed. However, our results demonstrate that the independent N-terminal crn-TPR motif of Clf1 is sufficient for recognizing the Prp40 FF1 domain. This is in agreement with previous studies showing that individual TPR repeats may have different peptide-binding specificities and thus the same binding affinity for their ligands as the full-length protein (38,39).

Interaction studies with the RNAPII CTD - In a previous study the region spanning the FF domains of Prp40 was shown to interact with hyperphosphorylated RNAPII CTD (8). Furthermore, mapping of the interaction surface between the HYPA/FBP11 FF1 domain and a doubly phosphorylated pair of CTD repeats revealed that two positively charged residues in the FF domain binding site may recognize the phosphate groups of the CTD (12). We have therefore performed chemical shift perturbation studies with the $^{15}$N-labeled Prp40 FF1 domain and three synthetic peptides derived from the RNAPII CTD. However, even at high molar ligand:domain ratios we found that the Prp40 FF1 domain did not interact with the phosphorylated CTD peptides (SYpSPTpSPS and (SYpSPTpSPS)$_2$, respectively) nor with the unphosphorylated tandem CTD repeat (YSPTPSYSPSTPS). Titration data for the (SYpSPTpSPS)$_2$ peptide are shown as supplementary Fig 6. It is worth noting that two positive control experiments using the $^{15}$N-labeled Rsp5 WW2 domain and the unlabeled phosphorylated and phosphorylated CTD peptide, respectively, revealed chemical shift changes (3). Furthermore, in contrast to the RNAPII CTD binding site of the HYPA/FBP11 FF1 domain, the corresponding surface area of the Prp40 FF1 domain is predominantly negatively charged and therefore unfavorable for binding the phospho-CTD (Fig. 4A). Although we cannot rule out the possibility that the wrong phosphorylation pattern was used in the synthetic peptides employed in our experiments, the negative surface potential of the FF1 domain suggests that other FF domains in Prp40 are responsible for recognizing the phosphorylated RNAPII CTD.

Domain composition of Prp40, FBP11 and CA150 proteins – Both Prp40 and HYFA/FBP11 associate with the U1 snRNP in yeast and metazoans, respectively. Given that their first FF domains seem to recognize different ligand motifs, we were interested in clarifying whether Prp40 and FBP11 are orthologous proteins. The overall sequence conservation of FF domains, as already mentioned, is relatively low leading to discrepancies in the literature regarding the number and location of the yeast Prp40 FF domains. To clarify both questions, we have combined the knowledge of the structure-determining residues of Prp40 and FBP11/HYPA FF1 domains with sequence alignments and phylogenetic tree reconstruction tools using representative sequences from the three functionally related FBP11, Prp40 and CA150 protein families.

In agreement with previous predictions (8), our sequence analysis showed that CA150 proteins comprise six FF domains (8), while Sp. Prp40 contains five well-defined FF domains. As predicted by Bedford et al. (9), Sc. Prp40 includes only four FF domains. However, we cannot exclude the possibility that two additional FF domains may exist in Prp40 as previously suggested (8), but that their sequences have diverged too much to be detected with the level of confidence used in our analysis. In contrast to CA150 proteins, the number of FF domains in FBP11 proteins seems to depend on the organism; human and chicken FBP11 proteins display six, the Drosophila counterpart five and Plasmodium only three.

Based on the multiple sequence alignment (Supplementary Fig. 1) we generated a neighboring tree with the Phylip package (Fig. 5A). If Prp40 and FBP11 proteins were orthologous their FF domains should form a cluster within the phylogenetic tree. If on the other hand they were different proteins, some Prp40 FF domains should group together with FBP11 FF domains and others with CA150 FF domains. Indeed, we find that in general domains belonging to a certain subclass are grouped together (for instance, all FBP11 FF1 domains form a cluster as shown in Fig. 5A). Prp40 FF domains, on the other hand, cluster with either CA150 or FBP11 FF domains, depending on the particular FF domain analyzed. For instance, the Sc. Prp40 FF1 domain clusters with CA150 FF1 domains and is clearly distinct from FF1 domains in FBP11 proteins. This behavior once more supports the different binding specificities
observed for Prp40 and HYPA/FBP11 FF1 domains. Furthermore, the Prp40 FF2 domain seems to be more closely related to CA150 FF5 domains and FBP11 FF4 domains than to the corresponding CA150 and FBP11 FF2 domains (Fig 5A). Taken together, we conclude from this analysis that despite their presumably similar functions in spliceosome assembly FBP11 and Prp40 are not orthologous proteins.

DISCUSSION

The sequence alignment generated with sequences of three FF domain containing proteins, Prp40, FBP11 and CA150 reveals that positions contributing to the secondary structure are always occupied by hydrophobic residues with little preference for any particular hydrophobic amino acid. Only a pair of positively charged residues are conserved at the C-terminus of FF domains in most sequences. The most divergent part of the sequence is localized in helix2 and in loop2 that connects helix2 to helix3. Within loop2 we notice that half of the FF domain sequences contain a DxRY/F motif. The side-chain of the aromatic residue forms part of the protein core and thus explains its presence in many FF sequences. In Prp40 the ligand binding site identified in the NMR titrations is mostly formed by the DxRY/F motif, encircled by negatively charged and neutral residues together with two aromatic residues, W169 and W177, only present in the Prp40 FF1 domain. We have also observed that FF4, which does not contain the motif, is unable to interact with the crn-TPR repeat used in this work. It seems thus possible that other ligands yet to be identified may exist as targets for the FF domains that lack the DxRY/F motif, for instance Prp40 FF2 and FF4, several FBP11 FF domains and members of the p190 RhoGAP family (the latter not included in the alignment given as Supplemental Fig.1). Indeed, the p190-A RhoGAP region comprising the FF domains has been recently described to interact with the transcription factor TFII-I and this interaction is regulated by tyrosine phosphorylation on the first FF domain (40). Interestingly, the phosphorylated tyrosine is located in the loop2 region. Thus, probably the FF domains that lack the DxRY/F motif may have a more flexible loop2, even with the $3_{10}$ helix unformed. This flexibility, specially in the case of p190-A RhoGAP FF1 domain, may allow the kinase to access and phosphorylate the tyrosine. Since tyrosine phosphorylation has an inhibitory role in the interaction with the transcription factor TFII-I, it may indicate that the tyrosine is directly involved in ligand binding. If this is the case, both p190-A RhoGAP FF1 and Prp40 FF1 binding sites may be localized on the same part of the surface. Certainly structural information on RhoGAP FF domains and on the complexes they form will clarify this hypothesis. Moreover, the study of the potential effects that phosphorylation may have in the domain structure will also be very valuable to obtain a detailed description of the role of loop2 in the interaction.

TPR motifs are not the unique targets that can be recognized by FF domains containing the DxRY/F motif. Recently, a number of transcription and splicing factors were identified as potential interaction partners for CA150 (10,11). In particular, the Tat-specific factor 1 (Tat-SF1) encompasses multiple weak binding sites for the CA150 FF domains that conform to the consensus motif (D/E)$_{2/5}$F/W/Y-(D/E)$_{2/5}$ (10). The RNAPII CTD comprises up to 52 heptapeptide repeats with the consensus sequence YSPTSPS (41). Phosphorylation of CTD repeats at positions 2 and 5 is thought to create a “CTD-code” that regulates interactions with a variety of transcription and splicing factors, including CA150 (42-44). Interestingly, both these serines are phosphorylated during the M phase of the cell cycle to inhibit RNA splicing and promote gene silencing (45). Phosphorylation at these serine positions in a tandem CTD repeat creates a motif that resembles that of (D/E)$_{2/5}$F/W/Y-(D/E)$_{2/5}$ motif in the sense that an aromatic residue is surrounded by negatively charged residues. This similarity may provide a basis for the involvement of FF domain containing proteins in linking mRNA splicing to transcription. While no structural information is available for the binding mode of (D/E)$_{2/5}$F/W/Y-(D/E)$_{2/5}$ motifs, the phospho-CTD binding site of the HYPA/FBP11 FF1 domain has been mapped by chemical shift perturbation experiments (12). The phospho-CTD binding site includes two lysine residues, often present in FBP11 and CA150 FF domains, suitable for forming salt bridges with the phosphate groups of the tandem CTD repeat. With the exception of
the FF2 domain these lysines are rare in Prp40 FF domains.

The differences in charge distribution of the phospho-peptide binding site in the Prp40 and HYPA/FBP11 FF1 domains led us to calculate pK\textsubscript{a} values for all FF domains included in the sequence alignment shown in Supplementary Fig. 1. Whereas the Prp40 FF1 domain possesses an overall pK\textsubscript{a} of 4.7, the FBP11 and CA150 FF1 domains have pK\textsubscript{a} values ranging from 8.8-9.9 (Fig. 5B). In contrast to the negatively charged FF1 domain, the Prp40 FF2 has an overall pK\textsubscript{a} of 9.9 and thus seems to be more favorable for an interaction with the phospho-CTD than the Prp40 FF1 domain. It is also of note that CA150 FF domains are almost exclusively positively charged. In FBP11 proteins, however, only the FF1 domains has a pK\textsubscript{a}>8.0, while the FF2, FF3 and FF4 domains are negatively charged (with the exception of the Pf FF4), and the FF5 and FF6 domains have a neutral pK\textsubscript{a}. This might suggest that in contrast to CA150 proteins only the FF1 domains of FBP11 proteins can interact with the phospho-CTD, while other FBP11 FF domains may rather recognize splicing factors, as the Prp40 FF1 domain.

Prp40 was originally discovered as a protein associated with U1 snRNA (4). Indeed, a suppressor mutation in the Prp40 FF2 domain (S240F) was found to rescue otherwise lethal mutations in the 5’ end of U1 snRNA (4). To our surprise, sequence analysis and a homology model of the Prp40 FF2 domain revealed that the S240F suppressor mutation maps to the crn-TPR binding site of the HYPA/FBP11 FF1 domain (Fig. 4A-C). In contrast to the Prp40 FF1 domain, the binding pocket of the Prp40 FF2 domain is highly positively charged with three lysine residues (K228, K247 and K248) in close proximity to the S240F mutation. Furthermore, four hydrophobic residues (F204, P235, Y238 and H239) surround the S240F mutation potentially creating a favorable environment for negatively charged molecules, as phosphoproteins or even RNA (in the case of a direct interaction with the 5’ end of U1 snRNA). Supporting this hypothesis is the recent finding that an ortholog of CA150, that of the dipteran Chironomus tentans protein hrp130, was shown to associate with nascent pre-mRNA (46). However, it is still unclear whether this interaction involves the FF domains or other regions of the protein. If the RNA interaction turns out to be the case, some FF domains may contain a new RNA binding motif. Taken together, our analysis of sequences and surface potentials of FF domains provides a first step towards understanding the binding versatility of these domains. Distinct ligand binding sites and specificities even within a single FF domain may enable scaffolding proteins such as Prp40 to mediate multiple interactions with unrelated binding partners and thereby enhance the assembly of large ribonucleoprotein complexes as the spliceosome. Clearly, further studies, in particular of FF domain interactions, will be necessary to elucidate possible individual roles of FF domains and to understand the principles governing their binding specificities.
FOOTNOTES

Deposition of assignments and coordinates - The $^1$H, $^{15}$N and $^{13}$C chemical shift assignments of the Prp40 FF1 domain have been deposited in the BioMagResBank database (http://www.bmrb.wisc.edu) under the accession number 6850, while the protein coordinates for the solution structure of the Prp40 FF1 domain are available in the protein data bank (http://www.rcsb.org) under accession number 2B7E.

Acknowledgements - We acknowledge J. Castresana for his help with the FF domain sequence and phylogenetic analysis, G. Stier for the modified pET24d vector, His-tagged TEV protease and experimental advice. We are grateful to P. Uetz for critical reading of the manuscript. S.W. acknowledges the Boehringer Ingelheim Fonds, P.M-M. the IRB and X. R. the IRB and EMBL for pre-doctoral fellowships. At present, S.W. is a recipient of an Emmy-Noether-fellowship of the German Research Foundation (DFG). A.G. was funded by the HFSPO and the Structural Biology program of the EMBL. This work was supported by a Human Frontier Science Program Research Grant (RG0234/2000M) and by the Spanish Ministerio de Educación y Ciencia, grants GEN2003-20642-C09-04/NAC and BFU2005-06276/BMC.

Abbreviations used - aa, amino acid; BBP, branch-point binding protein; Clf1, Crooked neck-like factor1; CTD, carboxy-terminal domain; FBP11, formin binding protein 11; GST, Glutathione-S-transferase; HSQC, heteronuclear single quantum coherence; HYPA, Huntingtin yeast partner A; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RNAPII, RNA polymerase II; Prp40, pre-mRNA processing protein 40; rmsd, root mean square deviation; snRNP, uridine-rich small nuclear ribonucleoproteins; TEV, tobacco etch virus; crn-TPR, crooked neck tetratrico-peptide repeat.

Key words - Prp40, FF domain, NMR structure, Clf1, TPR motif, CTD repeat.

This paper is dedicated to Professor Manuel Grande Benito on the occasion of his 60th birthday.
Table I: Structural statistics of the 10 lowest energy structures of the Prp40 FF1 domain.

| Number of structural restraints |         |         |
|--------------------------------|---------|---------|
| All                            | 1192    |         |
| Sequential ( |i-j| = 1) | 201     |         |
| Medium range ( 2 ≤ |i-j| ≤ 4) | 235     |         |
| Long range ( |i-j| > 4)  | 91      |         |
| Intraresidual                  | 665     |         |
| Unambiguous                    | 1178    |         |
| Ambiguous                      | 14      |         |
| Hydrogen bonds                 | 42      |         |

| R.m.s. deviation from experimental restraints(a) | <SA>(b) |
|------------------------------------------------|---------|
| All distance restraints (Å)                      | 3.83x10^{-2} ± 1.03x10^{-3} |
| Hydrogen bonds (Å)                              | 3.69x10^{-2} ± 2.20x10^{-3} |
| Dihedral angles (º) (24 φ and 26 θ)             | 0.19 ± 0.04 |

| R.m.s. deviation from idealized covalent geometry |         |
|--------------------------------------------------|---------|
| Bond lengths (Å)                                 | 3.48 x10^{-3} ± 1.06x10^{-4} |
| Bond angles (º)                                  | 0.45 ± 1.18x10^{-2} |
| Improper dihedral angles (º)                     | 0.15 ± 6.31x10^{-3} |

| Average atomic r.m.s. deviation from the mean structure (Å) |         |
|--------------------------------------------------------------|---------|
| Residues in secondary structural elements (N, C', C")        | 0.20 ± 3.06x10^{-2} |
| Residues in secondary structural elements (all heavy atoms)  | 0.49 ± 3.06x10^{-2} |
| All residues (N, C', C")                                     | 0.38 ± 9.06x10^{-2} |
| All residues (all heavy atoms)                               | 0.66 ± 6.50x10^{-2} |

| Structural quality |         |
|--------------------|---------|
| E_{L-J}(c)         | -567.97 ± 3.99 |
| Residues in most favored regions of Ramachandran plot(d)     | 91.0% |
| Residues in additionally allowed region                      | 9.0% |

(a) No distance restraint was violated by more than 0.3Å and no dihedral angle restraint was violated by more than 5 degrees.
(b) <SA> refers to the ensemble of the ten structures with the lowest energy.
(c) E_{L-J} is the Lennard-Jones energy calculated using the CHARMM PARMALLH6 parameters. E_{L-J} was not included in the target function during the structure calculation.
(d) Excluding glycine and proline residues.
FIGURE LEGENDS

Figure 1. Solution structure of the Prp40 FF1 domain.
A. Stereoview of the best-fit backbone (N, Cα, C') superposition of the 15 lowest energy structures after water refinement.
B. Ribbon representation of the lowest energy structure with secondary structure elements shown in blue. Conserved aromatic residues in the core of the domain are depicted in dark red. Residue numbers of labeled side-chains correspond to their position in the full-length Prp40 protein.
C. Ribbon representation of the superimposed structures of the Prp40 FF1 domain (blue) and the HYPA/FBP11 FF1 domain (gold) (PDB entry 1uzc).

Figure 2. Mapping of the binding surface of the Prp40 FF1 domain
A. Overlay of the 1H,15N-HSQC spectra of the 15N-labeled Prp40 FF1 domain in the absence (reference spectrum in black) and presence of increasing amounts of unlabeled crn-TPR1 motif. Spectra in dark blue, cyan, orange, and red correspond to molar protein:ligand ratios of 1:0.5, 1:1, 1:1.5 and 1:2, respectively. Labels indicate residues affected by ligand binding.
B. Ribbon representation of the Prp40 FF1 domain displaying amide atoms of residues (represented as blue spheres) that are affected by ligand binding (as defined in C).
C. Bar representation of the chemical shift changes corresponding to the data shown in Fig. 2A. Chemical shift changes were calculated as the difference between the value obtained at the 1:2 protein:peptide ratio and the equivalent chemical shift corresponding to the free domain, using the equation \( \Delta \delta^\text{av} = \left[ (\Delta \delta_{1H})^2 + (\Delta \delta_{15N})^2/2 \right]^{1/2} \). Residues that get broadened or that disappear upon ligand binding are represented with blue stars while residues that change more than 0.2 ppm are represented as blue bars. These two types of residues were considered to be affected upon ligand binding and thus were displayed in Fig 2B.
D. Comparison of the binding sites of the Prp40 FF1 domain and the HYPA/FBP11 FF1 domain. Left hand side panel: HYPA/FBP11 FF1 binding site (as described in (12)), right hand side panel: Prp40 FF1 binding site (obtained in this work). N and C termini as well as elements of secondary structure are labeled in both figures.

Figure 3. Interaction between the 15N-labeled crn-TPR1 motif and the Prp40 FF1 domain.
A. Two views of the model of the crn-TPR1 motif built with Insight II that best represents the experimental data. The length and orientation of both helices (represented in gray) was defined using a set of medium and long-range NOEs. The latter are represented as dotted lines that connect the side chains of residues (displayed in dark blue) giving rise to the NOEs.
B. Overlay of the 1H,15N-HSQC spectra of the crn-TPR1 motif in the absence (in black) and after addition of the unlabeled Prp40 FF1 domain molar ligand:protein ratios of 1:0.5 (orange), 1:1.5 (violet) and 1:2 (cyan).
C. Ribbon representation of the crn-TPR1 model displaying residues whose amides experience chemical shift changes upon ligand binding. Residues with weak changes are labeled in black while medium changes are in orange.
D. Amino acid sequence of the crn-TPR1 repeat used in this work.

Figure 4. Comparison of the Prp40 FF1 and FF2 and the HYPA/FBP11 FF1 electrostatic surface potential.
All surface representations on the left hand side are shown in an orientation identical to that in Figure 2B, the right hand side ones are rotated by 90°.
A. Electrostatic surface potential of the Prp40 FF1 domain.
B. Electrostatic surface potential of the HYPA/FBP11 FF1 domain. The RNAPII CTD binding site is indicated by a black line in the panel on the right hand side.
C. Electrostatic surface potential of the Prp40 FF2 domain model. The S240F suppressor mutation is highlighted in green.

Figure 5. Domain composition of Prp40, FBP11 and CA150 proteins.
A. Neighbor-joining phylogenetic tree of FF domains in the three related proteins Prp40, FBP11 and CA150. Species are named as follows: Ag.: Anopheles gambiae, Ce.: Caenorhabditis elegans, Dm.: Drosophila melanogaster, Gg.: Gallus gallus, Hs.: Homo sapiens, Pf.: Plasmodium falciparum, Sc.: Saccharomyces cerevisiae and Sp.: Schizosaccharomyces pombe. FF domains of proteins with fewer than six FF domains may have two numbers: The first number corresponds to the position of the FF domain in the protein sequence (e.g. Prp40 FF4), while the second number indicates the similarity to other FF domains based on their phylogenetic relationship (e.g. Prp40 FF4_6). The scale bar represents a distance of 0.2 substitutions per site.
B. pKa values of FF domains in the three related proteins Prp40, FBP11 and CA150. White, gray, and black boxes were used for neutral (6.0<pKa<8.0), positively charged (pKa>8.0), and negatively charged (pKa<6.0) FF domains, respectively. Open boxes indicate absent domains.
References

1. Burge, C. B., Tuschi, T., and Sharp, P. A. (1999) in *The RNA World*, pp. 525-560, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA
2. Abbovich, N., and Rosbach, M. (1997) *Cell* 89, 403-412
3. Wiesner, S., Stier, G., Sattler, M., and Macias, M. J. (2002) *J. Mol. Biol.* 324, 807-822
4. Kao, H. Y., and Siliciano, P. G. (1996) *Mol. Cell. Biol.* 16, 960-967
5. Chung, S., McLean, M. R., and Rymond, B. C. (1999) *RNA* 5, 1042-1054.
6. Vincent, K., Wang, Q., Jay, S., Hobbs, K., and Rymond, B. C. (2003) *Genetics* 164, 895-907
7. Wang, Q., Hobbs, K., Lynn, B., and Rymond, B. C. (2003) *J. Biol. Chem.* 278, 7875-7883
8. Morris, D. P., and Greenleaf, A. L. (2000) *J. Biol. Chem.* 275, 39935-39943.
9. Bedford, M. T., and Leder, P. (1999) *Trends Biochem. Sci.* 24, 264-265.
10. Smith, M. J., Kulkarni, S., and Pawson, T. (2004) *Mol. Cell. Biol.* 24, 9274-9285
11. Carty, S. M., Goldstrohm, A. C., Sune, C., Garcia-Blanco, M. A., and Greenleaf, A. L. (2000) *Proc. Natl. Acad. Sci. U S A* 97, 9015-9020
12. Allen, M., Friedler, A., Schon, O., and Bycroft, M. (2002) *J. Mol. Biol.* 323, 411-416
13. Ben-Yehuda, S., Dix, I., Russell, C. S., McGarvey, M., Beggs, J. D., and Kupiec, M. (2000) *Genetics* 156, 1503-1517
14. Fabrega, C., Shen, V., Shuman, S., and Lima, C. D. (2003) *Mol. Cell* 11, 1549-1561
15. Meinhart, A., and Cramer, P. (2004) *Nature* 430, 223-226
16. Noble, C. G., Hollingworth, D., Martin, S. R., Ennis-Adeniran, V., Smerdon, S. J., Kelly, G., Taylor, I. A., and Ramos, A. (2005) *Nat. Struct. Mol. Biol.* 12, 144-151
17. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., and Noel, J. P. (2000) *Nat. Struct. Biol.* 7, 639-643
18. D'Andrea, L. D., and Rega, L. (2003) *Trends Biochem. Sci.* 28, 655-662
19. Notredame, C., Higgins, D. G., and Heringa, J. (2000) *J. Mol. Biol.* 302, 205-217
20. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* 6, 277-293
21. Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biomol. NMR* 5, 1-10
22. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* 4, 603-614
23. Sattler, M., Schleucher, J., and Griesinger, C. (1999) *Prog. NMR Spectrosc.* 34, 93-158
24. Yamazaki, T., Forman-Kay, J., and Kay, L. E. (1993) *J. Am. Chem. Soc.* 115, 11054-11055
25. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* 33, 5984-6003
26. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* 13, 289-302
27. Kubinowa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) *J. Biomol. NMR* 4, 871-878
28. Hu, J., and Bax, A. (1997) *J. Biomol. NMR* 9, 323-328
29. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D* 54, 905-921
30. Folmer, R. H., Hilbers, C. W., Konings, R. N., and Nilges, M. (1997) *J. Biomol. NMR* 9, 245-258
31. Nilges, M., and O’Donoghue, S. I. (1998) *Progr. NMR Spectr.* 32, 107-139
32. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* 8, 477-486
33. Fiser, A., and Sali, A. (2003) *Methods Enzymol.* 374, 461-491
34. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics* 14, 51-55
35. Hubbard, S. J., and Thornton, J. M. (1993) *Department of Biochemistry and Molecular Biology, University College London, UK.*
36. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000) *Nucleic Acids Res.* 28, 231-234
37. Pascual, J., Castresana, J., and Saraste, M. (1997) *Bioessays* 19, 811-817
38. Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000) *Cell* 101, 199-210
39. Brinker, A., Scheufler, C., von der Mulbe, F., Fleckenstein, B., Hermann, C., Jung, G., Moarefi, I., and Hartl, F. U. (2002) *J. Biol. Chem.* 22, 19265-19275
40. Jiang, W., Sordella, R., Chen, G. C., Hakre, S., Roy, A. L., and Settleman, J. (2005) *Mol. Cell.* 17, 23-35
41. Corden, J. L. (1990) *Trends Biochem. Sci.* 15, 383-387
42. Kobor, M. S., and Greenblatt, J. (2002) *Biochim. Biophys. Acta* 1577, 261-275
43. Buratowski, S. (2003) *Nat. Struct. Biol.* 10, 679-680
44. Hampsey, M., and Reinberg, D. (2003) *Cell* 113, 429-432
45. Xu, Y. X., Hirose, Y., Zhou, X. Z., Lu, K. P., and Manley, J. L. (2003) *Genes Dev* 17, 2765-2776
46. Sun, X., Zhao, J., Kyllberg, K., Soop, T., Palka, K., Sonnhammer, E., Visa, N., Alzhanova-Ericsson, A. T., and Daneholt, B. (2004) *Chromosoma* 113, 244-257
Figure 1

A. 

B. 

C. 

N-terminus

C-terminus

α1

α2

α3

3_10

A409

W152

Y185

F182

F154

β39

Y168

N

C
Figure 3

A. Helix1 Helix2

B. 

\[
\begin{array}{c}
15N \text{ ppm} \\
8.6 8.4 8.2 8.0 7.8 \\
115 \quad 120 \quad 125 \\
\end{array}
\]

C.

D.

GAMGSTNIDILDLEELREYQRKRTYEYEGYKRNRLD
Figure 4

A.

B.

C.
Figure 5

A.

B.
FIGURE LEGENDS FOR SUPPLEMENTAL MATERIAL

Figure 1. Multiple sequence alignment of the FF domains in the three related protein families Prp40, FBP11, and CA150. Boxes within and on the top of the alignment indicate secondary structure elements. The four FF domains present in Prp40 are highlighted in red, while the S240F suppressor mutation in the Sc. Prp40 FF2 domain is shown in green. Highly conserved residues are marked in black, while semi-conserved residues are shown in gray. Rows are labeled by protein names, domain number and species (Ag.: Anopheles gambiae, Ce.: Caenorhabditis elegans, Dm.: Drosophila melanogaster, Gg.: Gallus gallus, Hs.: Homo sapiens, Pf.: Plasmodium falciparum, Sc.: Saccharomyces cerevisiae and Sp.: Schizosaccharomyces pombe). FF domains of proteins with less than six FF domains may have two numbers: The first number corresponds to the position of the FF domain in the protein sequence (e.g. Prp40 FF4), while the second number indicates the similarity to other FF domains based on their phylogenetic relationship (e.g. Prp40 FF4_6) (see Fig. 5). Stars on top of the alignment indicate the two conserved Phe residues that give the name to the domain, while solvent accessible residues found in the binding surfaces of the HYPA/FBP11 and Prp40 FF1 domain are indicated by open and filled circles, respectively.

Figure 2. HSQCs corresponding to the two constructs prepared for FF4 domain. Panel A displays the short construct, showing a well-dispersed pattern of signals. Panel B shows the data corresponding to the N-terminus extended construct. The new signals are poorly dispersed, characteristic of disordered regions.

Figure 3. Unusual NOE pattern of the α2 helix of the Prp40 FF1 domain. Strips of the aliphatic regions of the 15N-HSQC- (A) and 13C-HMQC-NOESY (B), respectively, for Ile157, Ile158 and Leu161 showing the unusual Hn(I157) to Hn(L161) and Hn(I157) to Hn(L161) NOEs labeled in orange. Sequential and medium-range NOEs are indicated by lines.

Figure 4. Sedimentation equilibrium data for the crn-TPR1 motif at 285K and 295K. The data was fitted using the ULTRASCAN II software assuming an ideal, reversibly self-associating monomer-dimer system (monomer shown as dotted line, dimer as solid line). While at the lower temperature of 285K there is ~50% dimer population at the NMR concentrations used (~1mM) a ten degree increase of temperature moves this equilibrium position such that the 50% dimer population would only be present at concentrations higher than 1M.

Figure 5. NMR data used to generate a structural model of the Clf1 crn-TPR1 motif. Secondary structure elements (dark grey boxes) derived from 13Cα and 13Cβ secondary chemical shifts (top panel), short to medium range NOEs and heteronuclear {1H}-15N NOEs (bottom panel).

Figure 6. Chemical shift perturbation studies with the (SYpSPTpSPS)2 peptide and the 15N-labeled Prp40 FF1 domain. 1D 1H- (left panels) and 2D 1H,15N-NMR spectra of the Prp40 FF1 domain in the absence (reference spectra in black) and presence of unlabeled (SYpSPTpSPS)2 peptide at molar protein:peptide ratios of 1:1 (cyan), 1:3 (red) and 1:7 (green). While proton signals from the peptide increase in the 1D spectra with growing peptide concentrations, no chemical shift changes are observed for the FF domain even at a large excess of peptide. At the bottom of the figure we represent the chemical shift changes calculated as the difference between the value obtained at the 1:7 protein:peptide ratio and the equivalent chemical shift corresponding to the free domain, using the equation Δδν = [(Δδν)2 + (Δδ15N/n)2]1/2.
Supplemental Material Figure 4

Self association profile for crn-TPR1 at 285K

Self association profile for crn-TPR1 at 295K

% of Total concentration

Total Molar Concentration

Optical Density Difference

Residuals

Radius/Radius_0 (prot)
Supplemental Material Figure 6

1D experiments

HSQC experiments

Supplemental Material Figure 6
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
Alexander Gasch, Silke Wiesner, Pau Martin-Malpartida, Ximena Ramirez-Espain, Lidia Ruiz and Maria J. Macias

J. Biol. Chem. published online October 27, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M508047200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2005/11/10/M508047200.DC1