PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Solution structure of the fourth FF domain of yeast Prp40 splicing factor

Roman Bonet,[1] Lidia Ruiz,[1] Begoña Morales,[1] and Maria J. Macias[1,2]*

[1] The Institute for Research in Biomedicine Barcelona (Protein NMR Group), Barcelona Science Park, Baldiri Reixac10-13, E-08028 Barcelona, Spain

[2] Institució Catalana de Recerca i Estudis Avançats ICREA, Passeig Lluis Companys 23, E-08018 Barcelona, Spain

Key words: FF domains; NMR structures; Prp40 splicing factor; phospho-CTD repeats.

INTRODUCTION

Prp40 protein was originally identified as a suppressor of 5′ end U1 RNA point mutations.[1] Prp40 is a U1 snRNP-associated protein that participates in the early steps of yeast premessenger RNA splicing. Prp40 associates with the branch-binding point protein to bring the 5′ splicing site and the intron branch point into spatial proximity.[2] Additionally, Prp40 has been implicated in the binding to the phosphorylated C-terminal domain (herein referred as phospho-CTD) of RNA polymerase II through regions involving the WW and FF domains.[3] However, a subsequent study on the structure of the Prp40 WW domain pair also showed that, in the absence of additional FF domains, the WW domains do not interact with the phospho-CTD repeats.[4]

The solution structure of the first FF domain of Prp40 has been determined.[5] That study also examined the binding of Prp40FF1 to the splicing factor Clf1 and to a pair of bisphospho-CTD repeats. The binding site for the association with the first TPR motif of Clf1 involves the helices α2, the 3_{10}, and the N-terminal half of α3. In contrast, no interaction was detected for the Prp40FF1 domain with the phospho-CTD repeats and for the Prp40FF4 domain with the TPR motif of Clf1.[5]

Recently, other ligand partners have been identified for Prp40 FF domains, namely Snu71, a component of U1 snRNP, and Luc7, a splicing factor associated with U1 snRNP involved in the 5′ splicing site recognition.[6–8] Furthermore, it was shown that only the region comprising the first two FF domains of Prp40 is critical for yeast viability, whereas the deletion of the region including FF3 and FF4 results in a slow-growth phenotype.[6] These results are in agreement with a previous report that showed that a deletion of the FF1 domain and a fragment of FF2 of Prp40 caused yeast lethality.[9]

To further study the remaining FF domains present in Prp40, we cloned and produced constructs corresponding to the FF2, FF3, and FF4 domains. Of these, only the constructs corresponding to FF4 gave a folded and stable sample. Indeed, several constructs spanning FF2 and FF3 and even that of the FF1-2 pair were either partially structured or unstable after refolding from inclusion bodies.

Here we report the solution structure of the Prp40FF4 domain. Furthermore, prompted by the observation that the charge distribution of the FBP11FF1 region involved in the interaction with the bisphospho-CTD repeats[10] is partially conserved in Prp40FF4, we also examined whether this domain also interacted with the phospho-CTD repeats, but no binding was detected under our experimental conditions.

METHODS

Sample preparation

The Prp40 FF constructs corresponded to the following residues: 200 to 260 (FF2), 346 to 412 (FF3), 488 to 552

*Correspondence to: Maria J. Macais, Institució Catalana de Recerca i Estudis Avançats ICREA, Passeig Lluis Companys 23, E-08018 Barcelona, Spain.
E-mail: maria.macias@irbbarcelona.org
Received 27 April 2009; Revised 11 June 2009; Accepted 2 July 2009
Published online 20 July 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22547

(FF4), and 465 to 552 (extended FF4). The domains were purified as described previously.[11] The Prp40FF4 sample for NMR experiments was concentrated to 0.5 m*M* in 20 m*M* sodium phosphate buffer, 130 m*M* NaCl, 0.03% (w/v) NaN_3 in 90% H_2O, 10% D_2O, or 100% D_2O at pH 5.2.

NMR spectroscopy

Triple resonance spectra were recorded at 285K on either Bruker DRX-600 or Bruker DRX-800 NMR spectrometers. Homonuclear experiments 2D TOCSY (65-millisecond mixing time) and 2D NOESY (120-millisecond mixing time) were acquired on a Bruker- DRX-800 NMR spectrometer. All spectra were processed with the NMRPipe/NMRDraw software[12] and analyzed with CARA.[13] For the NMR titrations with the phospho-CTD, ^{15}N-Prp40FF4 short and extended constructs were prepared at 0.2 m*M*, and unlabeled ligand was added to a final ratio of 1:3. Heteronuclear ^{15}N-NOE values were measured on a 0.5-m*M* ^{15}N-labeled sample at 285K as described.[14]

Structure calculation

Distance restraints derived from fully assigned peaks in NOESY experiments were used for structure calculation together with $^3J(H^N, H^\alpha)$ and hydrogen bond restraints obtained from HNHA and D_2O exchange experiments, respectively. The structures were calculated with the programs CNS[15] and ARIA 2.0.[16] The 15 lowest-energy structures were analyzed with PROCHECK-NMR,[17] and the statistics from the analysis are shown in Table I. MOLMOL[18] and PyMOL[19] were used to visualize the structures and generate the figures.

Homology models

Prp40 FF2 and FF3 homology models were generated by MODELLER9v6[20] with the multiple template mode using several known FF structures and a multiple structure-based alignment of the FF family.

Peptide synthesis

The sequence SYp**S**PTpSPSYpSPTpSPSY corresponding to a doubly phosphorylated pair of repeats of the C-terminal domain of RNA polymerase II was manually synthesized using Fmoc solid-phase peptide synthesis and Rink amide matrix.[21] The peptide was cleaved from the resin with a TFA 95%/H_2O 2.5%/TIS 2.5% mixture, precipitated in cold ether, and purified by HPLC in a C4 reverse phase column (Vydac) using a water-acetonitrile gradient.

Table I
Structural Statistics for the 15 Lowest Energy Structures of Prp40FF4

Restraints used for the calculation ⟨SA⟩[a]	
Intraresidual	0
Sequential (|i-j| = 1)	846
Medium range (1< |i-j| ≤ 4)	782
Long range (|i-j| > 4)	936
Unambiguous	All
Ambiguous	0
Dihedrals	82
Hydrogen bonds	62
All	2564
Restraint per residue ratio	37.1
RMSD (Å) from experimental[b]	
NOE:	$0.01216 \pm 3.4 \times 10^{-4}$
Bonds (Å)	$0.01309 \pm 5.1 \times 10^{-4}$
Angles (°)	21.21 ± 1.11
Coordinate Precision (Å)[c]	
Backbone secondary structure elements	0.26
Heavy atoms in secondary structure elements	0.87
Heavy atoms all residues	1.06
CNS potential energy (kcal mol^{-1})	
Total energy[d]	-2327.6 ± 413.6
Electrostatic	-2862 ± 64.2
van der Waals	-393.9 ± 35.7
Bonds	39.2 ± 4.4
Angles	184.1 ± 6.2
Structural quality (%residues) 15 best structures	
In most favored region of Ramachandran plot	86.9
In additionally allowed region	10.8

[a] ⟨SA⟩ refers to the ensemble of the ten structures with the lowest energy.
[b] R.m.s. deviation between the ensemble of structures ⟨SA⟩ and the lowest energy structure.
[c] No distance restraint in any of the structures included in the ensemble was violated by more than 0.3Å.
[d] $E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy calculated using the CHARMM-PARMALLH6 parameters. $E_{L\text{-}J}$ was not included in the target function during the structure calculation.

RESULTS AND DISCUSSION

We determined the solution structure of the yeast Prp40 FF4 domain using standard multidimensional heteronuclear NMR spectroscopy. The final ensemble of the 15 lowest energy conformers is shown in Figure 1(A), and the structural statistics are summarized in Table I.

Like other previously described FF domains,[5,10,11] Prp40FF4 folds as a compact four-helical bundle, with a α1-α2-3_{10}-α3 topology [Fig. 1(B)]. The helices comprise residues Glu489-Thr507 (α1), Trp519-Leu526 (α2), Tyr532-Gly536 (3_{10}), and Asp539-Phe549 (α3), respectively. Each helix is connected to the other helices through long-range NOEs. For instance, Tyr532, located in the 3_{10} helix, is in contact with residues in α1 (Phe500 and Leu503), α2 (Ser523), and α3 (Arg542). The loop connecting the first two alpha helices exhibits distinct features in Prp40FF4 compared with other FF structures. A superimposition of Prp40FF4 to Prp40FF1 and URN1FF shows that the loop in Prp40FF4 presents an extra turn as a consequence of the insertion of five amino acids in the FF4 sequence [superimposition of structures and sequences

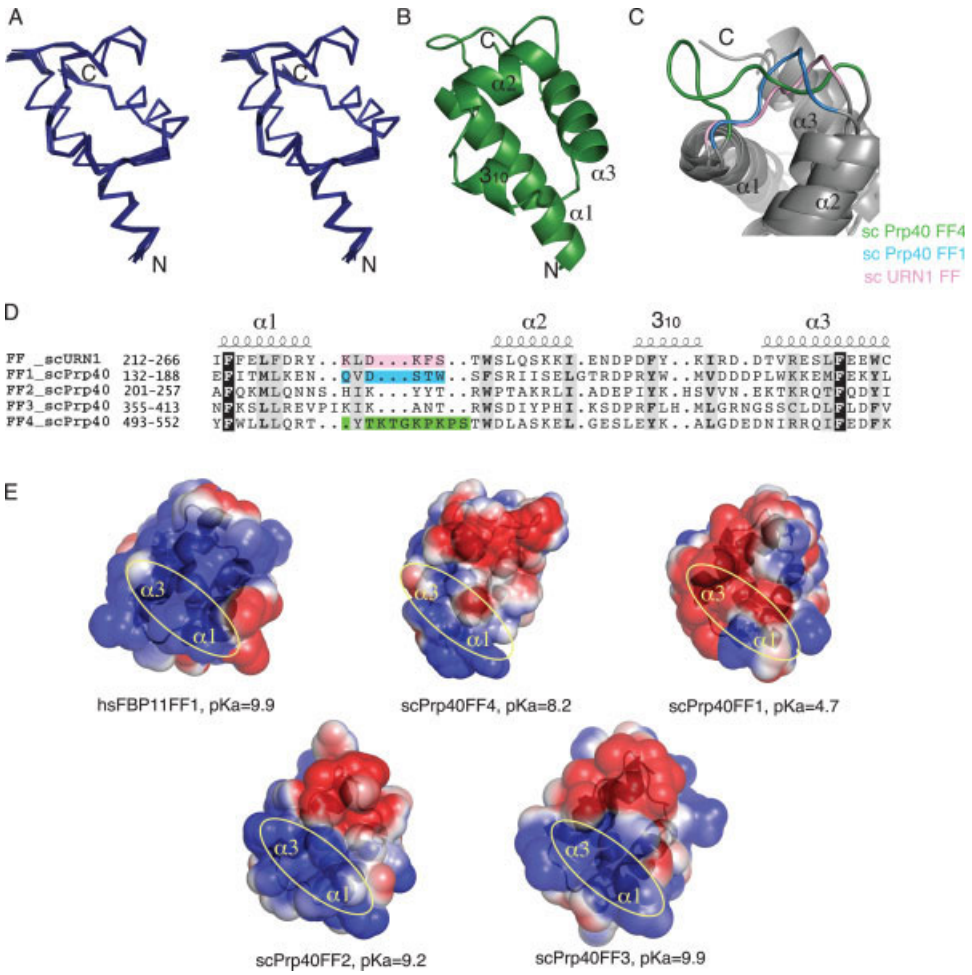

Figure 1

A: Overlay of the backbone 15 lowest-energy conformers of the Prp40FF4 domain after water refinement. **B**: Ribbon representation of the lowest-energy structure of Prp40FF4. **C**: An expanded view of the superimposition of Prp40FF4 to Prp40FF1 and URN1FF structures in the region of the loop connecting α1 and α2 helices. The loop is colored in green for Prp40FF4, in light blue for Prp40FF1, and in pink for URN1FF. **D**: Sequence alignment of *Saccharomyces cerevisiae* FF domains. The alignment was generated with ClustalX[22] and edited manually. Conserved residues are shaded in gray. The region corresponding to the first loop is shaded in colors according to the superimposed FF domains of the previous figure. **E**: Electrostatic surface plots of the Prp40 FF1, FF4, and FBP11FF1 structures and the Prp40 FF2 and FF3 models calculated with Adaptative Poisson-Boltzmann Solver.[23] Positive and negative surfaces are drawn in blue and red, respectively. The region corresponding to the phospho-CTD binding surface on the FBP11FF1 domain is marked with a yellow circle on the domains. pKa values for each domain are indicated.

shown in Fig. 1(C,D), respectively]. Although the loop is longer than in other structures, its flexibility is not increased, as shown by heteronuclear NOE experiments and the numerous NOE peaks observed with residues in the alpha helices (i.e., Lys515 side-chain with the rings of Trp501 [in α1] and Phe546 [in α3] or Pro516 with Ala522 [in α2] and Phe546). Remarkably, the hydroxyl groups of Thr507 and Tyr508 (in the beginning of the loop) participate in a network of NOEs with methyl groups of Leu503 and Leu504 and side-chains of Pro516, Glu525, and Leu526, respectively. Accordingly, both side chains are well defined in the structure.

For the binding experiments, we also prepared a second construct of the Prp40FF4 domain that included a predominantly positively charged 23-amino acid extension at the N-terminus. Like Prp40FF1, Prp40FF4 (long and short constructs) showed no binding to a pair of bisphospho-CTD repeats under our experimental conditions. The phospho-CTD binding site in FBP11FF1 was located in a positively charged region comprising the N-terminal of helices α1 and α3.[10] In Prp40FF1, this region is mainly negatively charged and therefore unfavorable for the phospho-CTD binding, whereas in Prp40FF4 the region comprising the N-terminal of helices α1 and α3 is rather positively charged with two small negatively charged patches [Fig. 1(E)]. In contrast, homology models of FF2 and FF3 domains [Fig. 1(E)] show a charge distribution in the potential binding site

for the bisphospho-CTD highly similar to that of FBP11FF1. Thus, regions of Prp40 including FF2 and/or FF3 may be involved in the interaction with the phospho-CTD repeats of RNA polymerase II.

Nevertheless, as discussed in the study by Ester and Uetz,[6] data available to date on FF domain interactions with distinct ligands do not allow us to formulate confident predictions about FF domain binding specificities. More studies on their interactions are required to gain further insight into FF binding modes.

Accession numbers

The chemical shifts have been submitted to the BioMagResBank (BMRB accession number 16176) and the protein coordinates to the Protein Data Bank (PDB ID code 2kfd).

ACKNOWLEDGMENTS

We thank G. Stier for the pETM30 vector and the TEV protease clone and F. Blanco from the CIC bioGUNE for acquisition of the HCCH-TOCSY and ^{13}C-NOESY experiments. We also thank Tanya Yates for revising the manuscript. R.B. acknowledges a pre-doctoral fellowship from IRB Barcelona. This work was financed by a BFU2005-06276 grant (M.J.M.), from the Spanish Ministry of Education and Science.

REFERENCES

1. 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;16:960–967.
2. Abovich N, Rosbach M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 1997; 89:403–412.
3. Morris DP, Greenleaf AL. The splicing factor. Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem 2000;275:39935–39943.
4. Wiesner S, Stier G, Sattler M, Macias MJ. Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. J Mol Biol 2002;324:807–822.
5. 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;281:356–364.
6. Ester C, Uetz P. The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. BMC Biochem 2008;9:29.
7. Fortes P, Bilbao-Cortes D, Fornerod M, Rigaut G, Raymond W, Seraphin B, Mattaj IW. Luc7p, a novel yeast U1 snRNP protein with a role in 5′ splice site recognition. Genes Dev 1999;13:2425–2438.
8. Gottschalk A, Tang J, Puig O, Salgado J, Neubauer G, Colot HV, Mann M, Seraphin B, Rosbash M, Luhrmann R, Fabrizio P. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. RNA 1998;4:374–393.
9. Murphy MW, Olson BL, Siliciano PG. The yeast splicing factor Prp40p contains functional leucine-rich nuclear export signals that are essential for splicing. Genetics 2004;166:53–65.
10. Allen M, Friedler A, Schon O, Bycroft M. The structure of an FF domain from human HYPA/FBP11. J Mol Biol 2002;323:411–416.
11. 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;73:1001–1009.
12. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
13. Bartels C, Xia T-H, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 1995;5:1–10.
14. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and a phosphopeptide-complexed Src Homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 1994;33: 5984–6003.
15. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 1998;54:905–921.
16. Habeck M, Rieping W, Linge JP, Nilges M. NOE assignment with ARIA 2.0: the nuts and bolts. Methods Mol Biol 2004;278: 379–402.
17. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
18. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis macromolecular structures. J Mol Graphics 1996;14:51–55.
19. DeLano WL. The PyMOL molecular graphics system. Palo Alto, CA: DeLano Scientific; 2002. USA. http://www.pymol.org.
20. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci 2007; Chapter 2:Unit 2.9.
21. Wellings DA, Atherton E. Standard Fmoc protocols. Methods Enzymol 1997;289:44–67.
22. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;24:4876–4882.
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;98:10037–10041.