Chapter 7

Determination of the solution structure and ligand binding site of the Prp40 FF1 domain

FF domains were first discovered as repetitive sequences present in the formin binding protein 11 (FBP11), the mammalian counterpart of the U1 snRNP associated *Sc.* splicing factor Prp40 (Bedford & Leder, 1999). FF domains consists of approximately 60 residues and encompass two conserved phenylalanines that give the name to the domain. The sequences of FF domains display a pattern of conserved hydrophobic residues typical for α-helical structures. Accordingly, three α-helical regions are recognised by secondary structure prediction programs for FF domain sequences. However, no three-dimensional structures of FF domains have been reported so far. A notable sequence feature of many FF domains is a conserved L/FxxΦΦ motif (where Φ is a long-chained hydrophobic residue, in most cases a leucine) in the C-terminal half of the first putative helix, which resembles a so-called NR box. NR-boxes occur in transcription coactivators and are recognised by nuclear receptors (Heery *et al.*, 1997). Apart from hypothetical proteins and protein fragments, all of the about 40 FF domain containing proteins known at present entail more than one copy of the domain, often up to six copies. With the exception of the small GTPase related proteins, all FF domain containing proteins are involved in splicing or transcription processes. Most of the two latter encompass two or three WW domains at the N-terminus accompanied by six C-terminal FF domains.

The yeast splicing factor Prp40 was originally identified as a protein associated with the U1 snRNA. However, based on the sequence similarity Prp40 does not contain any known RNA recognition motif. Nevertheless, a point mutation (S240F) in the second FF domain of Prp40 suppresses otherwise lethal mutations in the 5' end of U1 RNA (Kao and Siliciano, 1996). While the Prp40 WW domains have been implicated in cross-intron bridging through interactions with proline-rich regions of the branch-point binding protein BBP (also known as Msl5 and ySF1) and the U5 snRNP core component Prp8, the Prp40 FF domains are believed to interact indirectly with Mud2, the yeast orthologue of U2AF65. U2AF65 is a splicing

factor associated with both mBBP and the 3' splice site and required for commitment complex formation (Berglund *et al.*, 1998). The indirect interaction between Prp40 and Mud2 is presumably mediated by Clf1/Syf3p (Crooked neck-like factor 1/Synthetic lethal with Cdc forty) (Chung *et al.*, 1999; Ben-Yehuda *et al.*, 2000). Clf1 is a conserved multi-functional protein involved in pre-mRNA splicing (Chung *et al.*, 1999; Russell *et al.*, 2000; Uetz *et al.*, 2000), cell-cycle progression (Russell *et al.*, 2000) and in the initiation of DNA replication (Zhu *et al.*, 2002). *In vitro* binding and immunoprecipitation studies have shown that Clf1 interacts with the N-terminal half of Mud2 as well as with a region of Prp40 harbouring the N-terminal FF domain (FF1) (Chung *et al.*, 1999). Furthermore, Prp40 and a transcription factor with almost identical domain architecture, CA150, have been shown to interact with phosphorylated repeats of the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (Morris & Greenleaf, 2000; Suñé *et al.*, 1997).

Clf1 consists of 16 tetratrico peptide repeat (TPR) motifs (Chung *et al.*, 1999; Ben-Yehuda *et al.*, 2000). TPR motifs consist of 34 amino-acids with significant variations in the motif-defining residues (Lamb *et al.*, 1995; Blatch & Lässle, 1999). TPR motifs fold as two anti-parallel α-helices connected by a short turn (Blatch & Lässle, 1999). However, one of the TPR motifs of the peroxin PEX5 has also been found to form a single extended helix instead of the canonical helix hairpin structure (Kumar *et al.*, 2001). This suggests that some TPR motifs might exist in an open and a closed conformation and use a "jack-knife" mechanism to switch between these conformations. In general, TPR motifs mediate protein-protein interactions. TPR motif-containing proteins are thus thought to serve as scaffolds for the assembly of large protein complexes. This function arises from the structural arrangement of repeated TPR motifs, which form a right-handed super-helix. This super-structure forms a channel that serves in some instances as amphipathic binding site for target proteins (Das *et al.*, 1998; Scheufler *et al.*, 2000). However, in the case of the NADPH oxidase subunit $p67^{phox}$, this groove is not used for molecular recognition by the TPR repeats (Lapouge *et al.*, 2000) suggesting that not all TPR repeats may function in a similar manner. The TPR motif is found in a large variety of proteins, present in all three kingdoms of life, but is most common to eukaryotes.

To gain structural information about FF domains in general and the role of the Prp40 FF domains in the yeast splicing commitment complex in particular, the solution structure of the first Prp40 FF domain (FF1) was determined by heteronuclear multi-dimensional NMR spectroscopy. In addition, the interaction between the Prp40 FF1 domain and a TPR motif of Clf1 was studied using chemical shift mapping experiments.

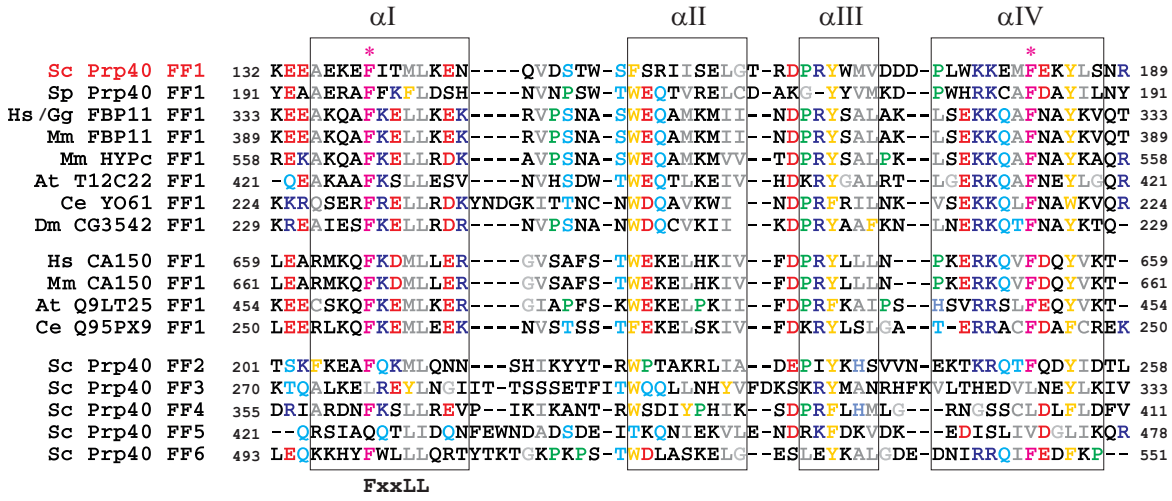

Figure 7.1: Multiple sequence alignment of FF domains. Stars on top of the alignment highlight the conserved phenylalanines (in magenta) giving the name to the domain, while secondary structure elements are boxed. Conserved aromatic, aliphatic, hydrophilic, positively and negatively charged residues are marked in orange, grey, cyan, blue and red, respectively, while prolines are marked in green. Abbreviations used are: Sc.: Saccharomyces cerevisiae, Sp.: Saccharomyces pombe, Hs.: Homo sapiens, Mm.: Mus musculus, Dm.: Drosophila melanogaster and At.: Arabidopsis thaliana) for species and Prp40: pre-mRNA processing protein 40, FBP11: formin binding protein 11, HYPC: huntingtin yeast partner C for protein names. For proteins with unknwon functions the gene names are used. Top: FF1 domains in Prp40-like proteins. Middle: FF1 domains in proteins similar to the transcription factor CA 150. Bottom: Remaining five FF domains of Prp40.

7.1 Determination of the solution structure of the Prp40 FF1 domain

7.1.1 The Prp40 FF domain boundaries

Due to discrepancies in the literature regarding the number and location of FF domains in the Prp40 protein, different protein constructs corresponding to four different putative FF sequences from the two yeast proteins Prp40 and Ypr152 were expressed and purified. As judged from two-dimensional ^{1}H–^{15}N correlation spectra, the constructs containing the Prp40 FF1 domain (residues 134–189, see Fig. 7.1) and that of the single Ypr152 FF domain (residues 212–266) were folded (Fig. 7.2(a) and (b)). In contrast, constructs of the second Prp40 FF domain (FF2) were unfolded (data not shown), even when slightly longer C-terminal boundaries were used (residues 200–260) than those predicted by the SMART database (http://smart.embl-heidelberg.de). Since the linker connecting the FF1 and FF2 domains comprises only twelve residues, a construct spanning both FF domains (residues 134–260) was designed with the hope that the presence of the linker could induce the folding of the FF2. However, instead of a completely folded sample a protein was obtained that was folded in the FF1 domain and unfolded in the FF2 domain (data not shown).

Since the linkers connecting FF domains encompass in general charged and hydrophobic residues, they could contribute to the structure of FF domains. To address this question,

a short and a long construct (residues 488–552 and 465–552, respectively) of the C-terminal Prp40 FF domain (FF6) were designed. Although it cannot be excluded that these additional residues might have a functional role, they do not adopt a three-dimensional structure in solution when linked to an individual FF domain (Fig. 7.2(c) and (d)). The thus determined boundaries of the Prp40 FF domains were confirmed by the expression of the single FF domain present in the yeast protein Ypr152, for which the selected 55-residues construct was structured (Fig. 7.2(b)). In conclusion, FF domains consist ot about 55 residues, which in general fold autonomously. However, the Prp40 FF2 domain might require the presence of flanking residues or domains or even ligands to adopt tertiary structure in solution. Given that for the Prp40 FF1 domain biochemical data support an interaction with the TPR motifs present in Clf1, the construct containing exclusively the Prp40 FF1 domain was chosen for structural characterisation.

7.1.2 Resonance assignment and description of the structure

The recombinant protein used for the structural studies comprised residues 134–189 of the *Sc.* Prp40 protein. Backbone and side-chain assigments were obtained using a set of standard heteronuclear multi-dimensional NMR experiments performed on ^{13}C,^{15}N and/or ^{15}N labelled samples. Proton-proton distance restraints were obtained from ^{15}N- and ^{13}C-edited three-dimensional NOESY spectra, while distance independent projection angle restraints were derived from ^{1}H–^{15}N residual dipolar coupling constants (RDCs). The RDCs were measured by soaking an aqueous solution of the Prp40 FF1 domain in a mechanically stressed poly-acrylamide gel (Sass *et al.*, 2000; Ishii *et al.*, 2001). Attempts to solubilise the protein in other alignment media were unsuccessful, since the protein displayed a high tendency either to interact with bicelles (Tjandra & Bax, 1997) or to precipitate in the presence of n-alkyl alcohol mixtures (Rückert & Otting, 2000).

In agreement with H^N-H^α J-couplings, the pattern of characteristic NOEs and ^{1}H–^{15}N RDCs, the FF domain adopts a helical fold consisting of four α-helices: the first α-helix ranges from residues 135–146 (αI), the second from residues 154–162 (αII), the third from residues 166–171 (αIII) and the forth from residues 175–187 (αIV) (Fig. 7.3). Apart from αIII, all helices are amphipathic and form an extensive network of alternating positively and negatively charged residues pointing towards the solvent. In contrast, in αIII only Tyr168 is oriented towards the core of the domain, while Pro166, Trp169 and Met170 are solvent exposed. The helices are connected by short loops, only the loop between αI and αII is with seven residues relatively long (residues 147–153). Nevertheless, loop I is not disordered as indicated by $\{^{1}$H$\}$-^{15}N heteronuclear NOEs. Residues 147–149 in loop I exhibit strong H^α(i)-H^N(i+1) NOEs and ϕ,ψ angle-pair characteristic of a β-strand, while residues 160–162 induce a turn in the loop such that several residues (Val148, Ser150 and Thr151) within loopI contact

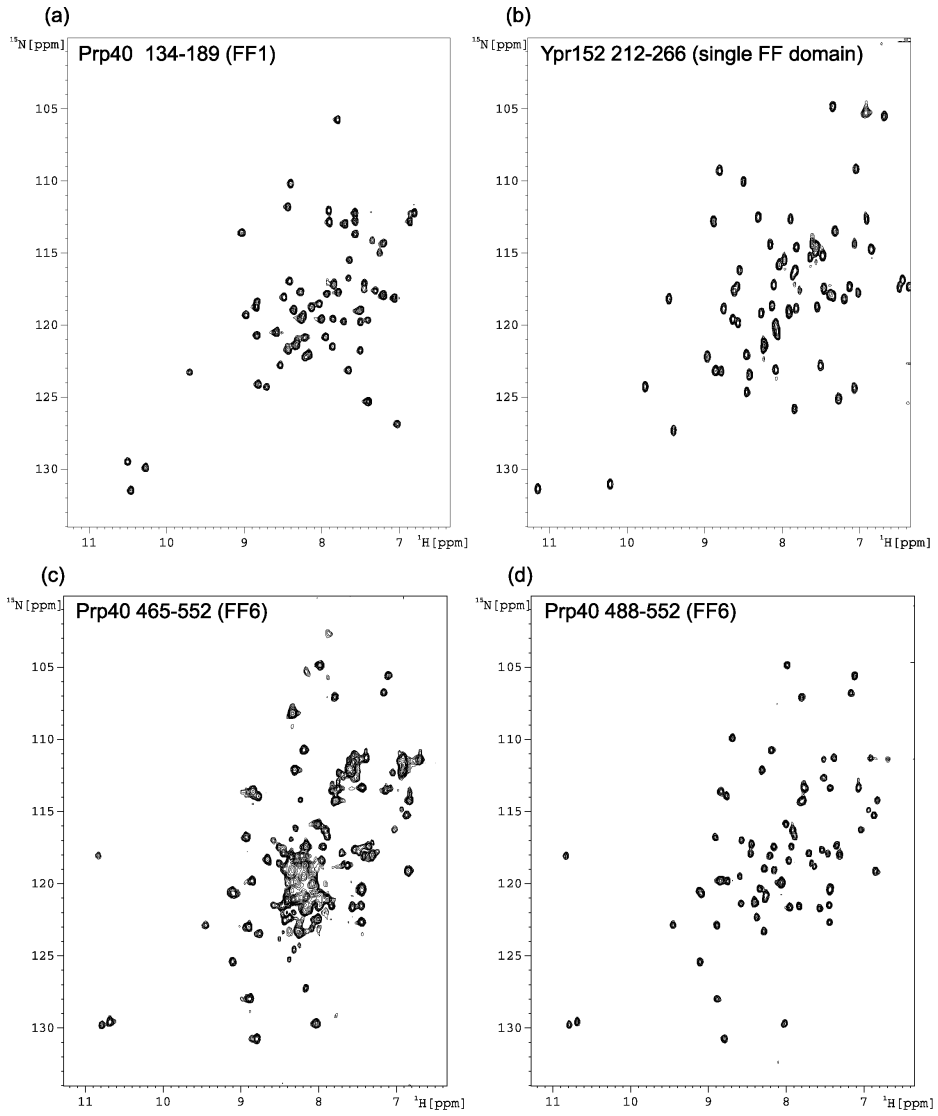

Figure 7.2: ^{1}H–^{15}N correlation spectra of FF domain constructs used to define the boundaries of FF domains.

to the indole ring of Trp152. The proximity of this aromatic ring is reflected in the chemical shifts of *e.g.* Val148, whose alpha and methyl protons resonate at 1.8 ppm and 0.5 ppm, respectively. While loopII (residues 163–165) contains hydrophilic and charged residues, loop III (residues 172–174) consists of three successive aspartate, which are flexible as indicated by below-average $\{^{1}$H$\}$-^{15}N heteronuclear NOEs.

The N- and C-terminus of the Prp40 FF domain face opposite sides of the domain as often found for domain repeats such as EGF modules, immunoglobulin (Ig) domains or spectrin repeats. In the core of the FF domain, an extensive network of aromatic contacts is formed by Phe139, Phe154, Tyr168, Trp177, Phe182 and Tyr185. This core is stabilised by additional hydrophobic interactions to Ala135, Ile 140, Met142 and Leu 143 in αI, Val148 in

loopI, Ile157 and Leu161 in αII, Val171 in αIII and Met181 in αIV (Fig. 7.3(b)). The presence of the aromatic ring network induces some peculiar chemical shifts in the aliphatic residues in the core. For instance, one of the β-protons of Ile143 resonates at 0.1 ppm. With a root mean square (r.m.s.) deviation of 0.145 Å of the backbone heavy atoms of residues 135–187 the structure of the whole domain is well-defined. Only loop II seems to be slightly more disordered, a behaviour that is as often observed for loop regions. ^{15}N relaxation data indicate a stable fold for the whole domain apart from loop II and III, which exhibit below-average $\{^1$H$\}$-^{15}N heteronuclear NOE values and exchange-broadened ^{1}H,^{15}N signals.

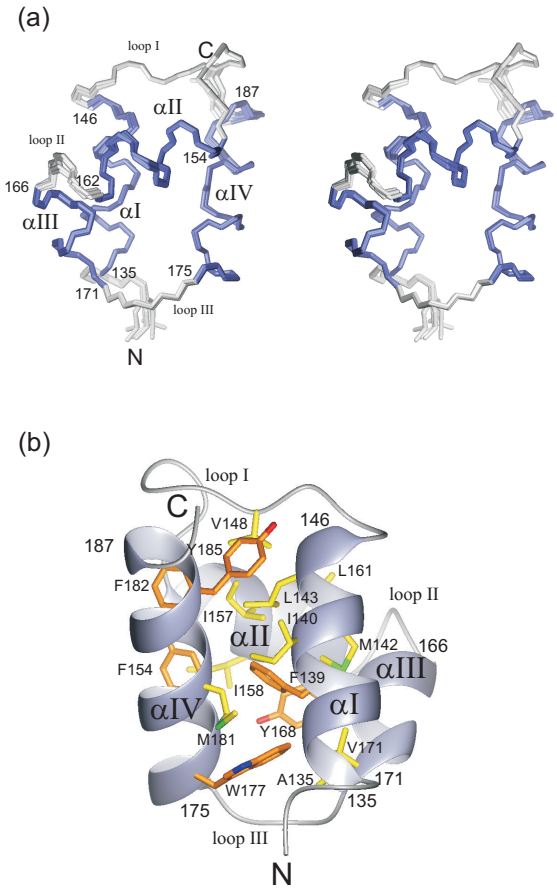

Figure 7.3: Three-dimensional structure of the Prp40 FF1 domain. α-helices are depicted in blue. (a) Stereoview of the backbone of the seven lowest energy structures. (b) Ribbon representation of the lowest-energy structure of the Prp40 FF1 domain highlighting the residues forming the hydrophobic core. Aromatic residues are shown in orange, while aliphatic residues are shown in yellow. The structure is rotated by 180° with respect to those shown in (a).

Fig. 7.3(a) shows a stereoview of the backbone of the structure ensemble. A ribbon representation highlighting the secondary structure elements and the residues forming the protein core are shown in Fig. 7.3(b). Table 1 summarises the structural statistics for the seven lowest energy structures analysed with CNS (Brünger *et al.*, 1998) and PROCHECK-NMR (Laskowski *et al.*, 1996).

Table 7.1: Structural statistics.

	#*	⟨SA⟩† r.m.s. deviation
R.m.s.deviation (Å) from experimental distance restraints‡		
All distance restraints (Å)	1646	0.0235±0.0027
Unambiguous NOEs (Å)	1594	0.0208±0.0031
Hydrogen bond restraints (Å)	52	0.0653±0.0035
R.m.s. deviation (°) from torsion angle restraints§		
Dihedral angles (47 ϕ, 47 ψ, 20 χ_1)	114	1.107±0.043
R.m.s. deviation (Hz) from experimental residual dipolar coupling restraints		
D^1_{HN}	20	1.37±0.02
R.m.s. deviation from idealised geometry		
Bond lengths (Å)		0.0029±0.0001
Bond angles (°)		0.5262±0.0082
Improper dihedral angles (°)		0.5141±0.0147
Coordinate precision of N, C$^\alpha$, C' / all heavy atoms (Å)¶		
Residues 135–187		0.145 / 0.473
Ramachandran plot (%)‖		
Residues in most favoured regions		74.9
Residues in additionally allowed regions		22.6
Residues in generously allowed regions		0.6
Residues in disallowed regions		2.0

*# refers to the number of restraints used.

†⟨SA⟩ is the ensemble of the 7 lowest energy structures out of 20 calculated refined with and without residual dipolar coupling restraints ($^1D_{\mathrm{NH}}$).

‡No distance restraint was violated by more than 0.3 Å.

§No dihedral angle restraint was violated by more than 5°.

¶The coordinate precision is given as the Cartesian coordinate r.m.s. deviation of the 7 lowest energy structures with respect to their mean structure.

‖Excluding glycine and proline residues and flexible residues at the termini.

7.1.3 Structure based sequence analysis of FF domains

The overall sequence conservation of FF domains is relatively low. Therefore, sequence alignments of FF domains are not always straightforward leading to discrepancies in the literature about the number and location of FF domains in the Prp40 protein. While at first four FF domains were suggested to be present in Prp40 (Bedford & Leder, 1999), five FF domains are detected by the SMART and Pfam databases and in other studies six FF domains are suggested (Morris & Greenleaf, 2000). Knowing the structure determining residues in the Prp40 FF1 domain, the sequence alignment of the FF domain family can now be carried out more accurately yielding six FF domains for Prp40 (Fig. 7.1) and its orthologues in *C. elegans*, mouse and humans. This alignment reveals that positions contributing to the protein core and secondary structure are occupied by hydrophobic residues, however with little preference for any particular hydrophobic amino-acid (Fig. 7.1). Even the two phenylalanines that give the name to the domain are not always conserved explaining the difficulties to align and detect FF domains exclusively based on the domain sequence.

Another interesting feature revealed by the three-dimensional structure of the Prp40 FF1 domain is that the FxxΦΦ box, previously suggested to be a recognition sequence for nuclear receptors (Bedford & Leder, 1999), is buried. Phe139, Met142 and Leu143 contribute to the protein core suggesting that the role of these residues is rather structural than functional. Even within the loops hydrophobic residues are conserved which stabilise the protein core (Val148) or induce turns between the α-helices such as the Asp-Pro pair preceeding αIII and αIV. The N- and the C-terminus are surrounded by charged residues of loop III and loop II, respectively. Intriguingly, the linkers connecting the Prp40 FF domains consist in general mostly of charged amino-acids, which could serve to link the domains by salt-bridges from the linker to the respective domain terminus.

7.2 NMR studies of the TPR1 repeat of the splicing factor Clf1

7.2.1 Selection of the TPR repeats for interaction studies

Although four proteins containing TPR motifs are part of the yeast spliceosome, Clf1 is the only protein reported to interact with Prp40 so far (Chung *et al.*, 1999; Ben-Yehuda *et al.*, 2000). Since the binding mode seems to vary among different TPR motif proteins, not only the binding site of the Prp40 FF1 domain was characterised in this study, but also that of the N-terminal Clf1 TPR motif. It is well known that the selection of the correct repeat frame can be difficult for proteins consisting of multiple repeats separated by only short likers. In addition, proteins containing helical repeats are difficult to produce in the mg-amounts re-

quired for structural studies, as they have in some cases a high tendency to aggregate. For the interaction study, the boundaries of the Clf1 TPR repeats were therefore chosen based on the careful sequence alignment described by Ben-Yehuda *et al.* (2000). Since the N-terminal half of Clf1 seems to be sufficient for the interaction with Prp40 (Chung *et al.*, 1999) and in order to increase the likelihood of obtaining soluble proteins, seven different Clf1 constructs were prepared spanning different numbers of TPR motifs: residues 1–64 (extended TPR1), residues 31–64 (TPR1), residues 65–98 (TPR2), residues 31–98 (TPR1–2), residues 235–304 (TPR7–8), residues 31–268 (TPR1–7) and residues 31–304 (TPR1–8). However, of these only Clf1 1–64 and Clf1 31–64 yielded sufficient amounts of soluble, folded protein and were therefore used to perform the binding experiments. The same residues of the Prp40 FF1 domain were found to be involved in the interaction with the TPR1 and the extended TPR1 motif of Clf1 indicating that the shorter TPR1 construct is sufficient for the binding. Therefore, only the construct containing the TPR1 motif sequence (residues 31–64) was characterised in more detail.

7.2.2 Resonance assignment and secondary structure of the Clf1 TPR1

In order to identify the residues of the Clf1 TPR1 motif that interact with the Prp40 FF1 domain, the backbone chemical shifts of the Clf1 TPR1 motif were assigned using a set of standard heteronuclear multi-dimensional NMR experiments performed on ^{13}C,^{15}N- and/or ^{15}N-labelled samples. Based on $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts and H^{N}-H^{α} J-coupling constants the Clf1 TPR1 motif consists of two anti-parallel α-helices (residues 37–48 and 52–59, respectively) connected by a short positively charged turn (Fig. 7.4(b)). However, the experimental data are consistent with a C-terminal α-helix that is one turn shorter than in the canonical TPR fold, which would correspond to residues 52–63 of the Clf1 protein. Although no structural data for an isolated TPR motif are available so far, it is known that in tandem TPR motifs the helix pair is not only stabilised by contacts between the helices within the motif, but also by interactions with the N-terminal α-helix of the successive TPR motif. It is therefore likely that the C-terminal TPR1 α-helix needs to be stabilised by the successive α-helix of the TPR2 motif or at least by residues in the linker connecting the TPR1 and TPR2.

The resonance assignment was more difficult than expected for a 33-residue protein, since not only the dispersion of amide resonance is intrinsically worse for α-helical proteins than for β-sheets, but also and even more importantly because in an isolated anti-parallel helix-pair a typical hydrophobic protein core is missing. Furthermore, internal motions changing the relative helix orientation could account for the relatively low dispersion of the resonances in the ^{1}H,^{15}N correlation spectra. In particular, residues 48–51 exhibit extensive overlap of the amide resonances. Therefore no assignment was possible for residues 49 and

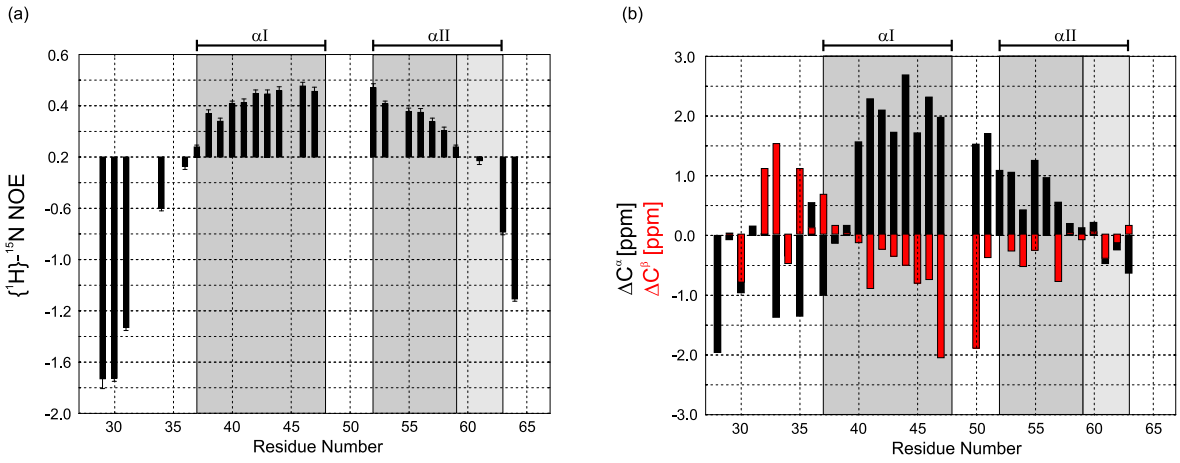

Figure 7.4: Secondary structure elements of the Clf1 TPR1. The α-helices are indicated by shaded boxes. The dark boxes correspond to the α-helices experimentally determined, while the light box indicates the missing last turn of αII. (a) $\{^1H\}$-^{15}N heteronuclear NOE. (b) $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts shown in black and red, respectively.

50. Although $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts and $\{^1H\}$-^{15}N heteronuclear NOEs do not directly indicate the location of the linker (Fig. 7.4), the fact that the resonances of residues 48–51 are not as well resolved as those of the rest of the protein indicates that these residues form the linker connecting the α-helices. This is also in agreement with sequence alignments of the Prp40 TPR1 motif with other TPR motifs of known structure. Interestingly, the sequence of the linker (RRKR) is a nuclear localisation sequence (Dingwall & Laskey, 1991). If this turn is solvent exposed in the complete Clf1 protein this sequence could indeed be recognised by nuclear import factors. Both helices are highly charged. While the first α-helix and the N-terminal half of the second helix are negatively charged, the C-terminal half of the second helix is positively charged.

7.3 Interaction between the Prp40 FF1 domain and the TPR1 motif of Clf1

Chemical shift perturbation experiments were performed to map the residues involved in the interaction between the Prp40 FF1 domain and the N-terminal tetratricopeptide repeat (TPR1) of the spliceosomal protein Clf1.

7.3.1 Identification of the Prp40 FF1 binding surface

Upon step-wise addition of the unlabelled TPR1 motif to the ^{15}N-labelled Prp40 FF1 domain, a specific set of amide resonances of the FF1 domain shifted and disappeared gradually in the ^1H–^{15}N-correlation spectrum. However, the majority of the peaks remained unchanged in

position and intensity indicating that the interaction with the TPR1 motif results in only a local perturbation of the FF1 structure. The perturbed amide resonances can hence be used to map the location of the TPR1 motif binding site onto the three-dimensional structure of the FF1 domain. The loss in signal intensity can be explained by a broadening of the amide resonances due to chemical exchange of the TPR1 motif on and off the FF1 domain. This indicates that the interaction occurs on the intermediate exchange time scale of the chemical shift. Accordingly, raising the temperature at which the binding studies were performed from 7 °C to 22 °C and hence changing the equilibrium to the unbound state some amide resonances (Ser159, Gly162, Trp169, Val171, Asp174 and Lys179) no longer vanish, but are shifted. The ^{1}H–^{15}N-correlation spectra recorded at 22 °C are shown in Fig. 7.5(a) and have been mapped onto the FF1 domain surface in Fig. 7.6. Most of the residues that disappear from the spectrum are located in and close to αIII. The interaction involves almost all residues from the C-terminal half of αII to the N-terminal half of αIV (residues Ile158–Lys179), while in αI only the C-terminal residues (Asn146) participates in the binding. To show that only specific residues are involved in the interaction a surface representation of the structure rotated by 180° is shown in Fig. 7.6 (right panels), while residues involved in the binding are coloured in orange for the residues with disappearing peaks in the ^{1}H,^{15}N-correlation spectra and in yellow for the shifting peaks (left panels).

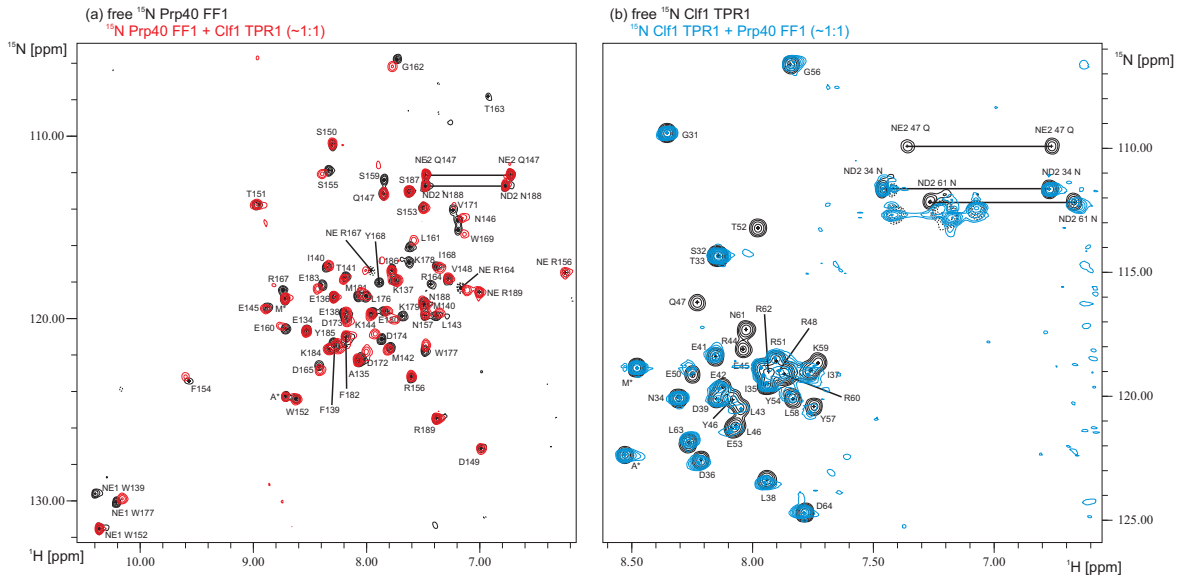

Figure 7.5: Chemical shift mapping experiments. Superposition of the ^{1}H–^{15}N correlation spectra for the interaction of the Prp40 FF1 domain with the N-terminal TPR motif of Clf1. The free domains are shown in black (reference spectra). In both panels, residues marked with a star correspond to residues resulting from the TEV protease cleavage site. (a) Titration of the ^{15}N-labelled FF1 domain with the unlabelled Clf1 TPR1 motif. (b) Titration of the ^{15}N-labelled Clf1 TPR1 motif with the unlabelled FF1 domain.

In contrast to the Prp40 FF1, no interaction was observed between the ^{15}N-labelled Prp40 FF6 (residues 488–552) and the unlabelled TPR1 motif of Clf1 under virtually identical

experimental conditions. This is in agreement with a previous study (Chung *et al.*, 1999), in which a region harbouring the first FF domain of Prp40 was found to be responsible for the interaction with Clf1.

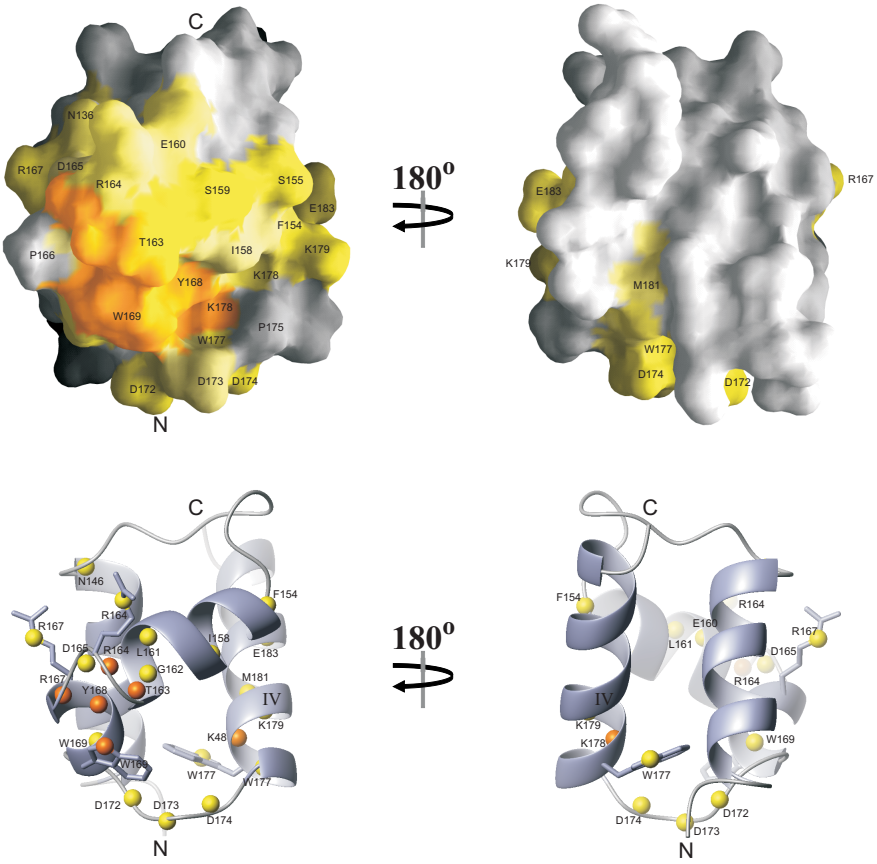

Figure 7.6: Mapping of the chemical shift changes onto the structure of the Prp40 FF1 domain. Top: Surface representation of the Prp40 FF1 domain. Residues disappearing from the ^{1}H–^{15}N correlation spectra intermediate exchange (on the NMR time scale) with the Clf1 TPR1 motif are coloured in orange. Residues in fast exchange with the Clf1 TPR1 motif are coloured in yellow with the darkness of the colour increasing with the size of the chemical shift change observed. Bottom: Ribbon representation of the Prp40 FF1 domain highlighting the ^{15}N atoms involved in the binding. Orange spheres indicate nitrogens in intermediate with the Clf1 TPR1 motif, while yellow ones indicate nitrogens in fast exchange with the Clf1 TPR1 motif.

7.3.2 Mapping of the Clf1 TPR1 binding site

To determine the residues of the TPR1 motif involved in the interaction with the Prp40 FF1 domain in more detail, chemical shift mapping experiments were performed using a ^{15}N-labelled TPR1 sample corresponding to residues 31–64 of Clf1. As observed for the ^{15}N-labelled Prp40 FF1 domain, a set of specific resonances disappeared gradually from the ^{1}H,^{15}N-correlation spectra upon step-wise addition of the unlabelled FF1 domain indicative of intermediate ligand exchange on the NMR time scale. Since the residues involved in the interaction are located in both α-helices and the binding pocket of the Prp40 FF1 domain has

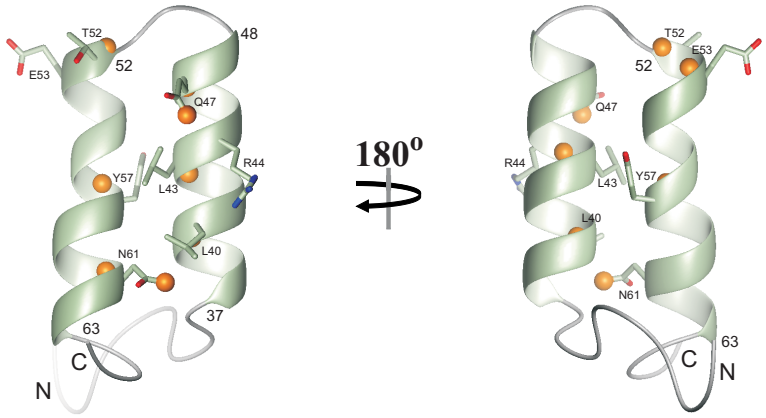

Figure 7.7: Mapping of the chemical shift changes onto the modeled structure of the Clf1 TPR1 motif. Orange spheres indicate nitrogen atoms involved in the interaction with the Prp40 FF1 domain.

about the length of the anti-parallel TPR helices, it is obvious that the Clf1 TPR1 domain folds indeed as an anti-parallel helix pair. In contrast, in an conformation consisting of one extended α-helix as found for the TPR3 motif of Pex5 (Kumar *et al.*, 2001) only about 10 helical residues could be accommodated in the Prp40 FF1 binding site. To interpret these results in more detail, a homology model of the Clf1 TPR1 motif was made based on the experimental NMR data and known structures of TPR motifs forming two anti-parallel α-helices (Das *et al.*, 1998; Rice & Brünger, 1999; Taylor *et al.*, 2001). The fact that backbone amide resonanes of both α-helices disappear from the ^{1}H,^{15}N-correlation spectrum upon ligand binding, but only those side-chain amide-resonances (Gln46 and Asn61) pointing towards one of the large surfaces built by the helices, indicates that the surface shown in the right panel of Fig. 7.7 points towards the FF1 binding surface. The TPR helices could pack around the second FF1 helix such that the otherwise solvent exposed L40 and L43 of the TPR motif contact hydrophobic side-chains of Ile158, Thr163, Tyr168 and Trp169 in the FF1 binding site. However, as this scenario is based on a structural model of the Clf1 TPR1 motif, there is no concrete evidence for the correctness of this interpretation.

7.4 Concluding remarks and discussion

To identify structurally similar proteins, the three-dimensional structure of the Prp40 FF1 domain was submitted to the DALI server (http://www.ebi.ac.uk/dali) (Holm & Sander, 1993). The search yielded 20 proteins, 12 of which were nucleic acid binding proteins. Nevertheless, the structural similarities were almost insignificant. This is also reflected by the fact that the binding sites of the DNA or RNA binding proteins exhibit no sequence similarity with the corresponding residues of the Prp40 FF domains. The fold of the Prp40 FF domain can hence be regarded as a novel α-helical protein fold.

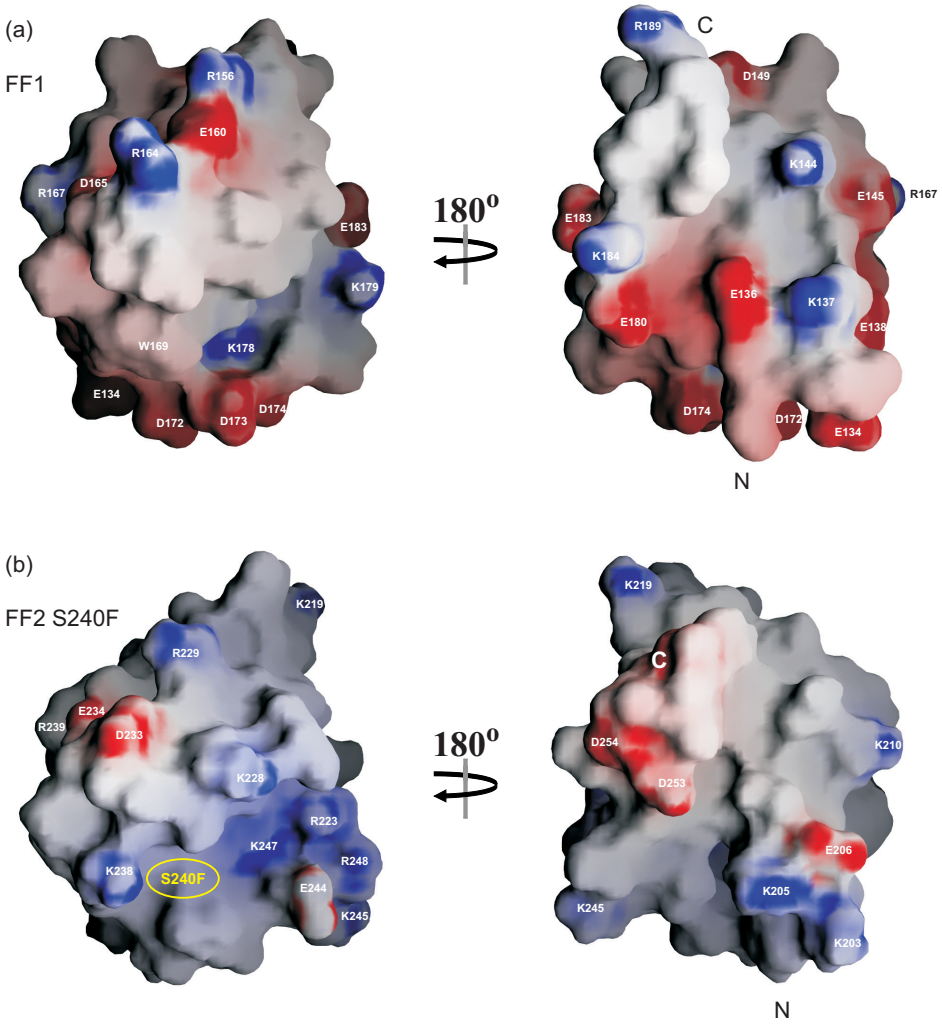

Figure 7.8: Comparison of the electrostatic surface potential of the Prp40 FF1 and FF2 domain. Blue surfaces indicate a posite surface potential, while red regions surface patches with negative surface potential. (a) Surface representation of the Prp40 FF1 domain and (b) of the Prp40 FF2 domain. The U1 snRNA suppressor mutation is denoted by the yellow ellipse.

Nonetheless, Prp40 has originally been identified as a U1 snRNA associated protein (Kao & Siliciano, 1996). Since a point mutation in the second FF domain of Prp40 (S240F) has been found to supress otherwise lethal mutations in the 5' end of U1 snRNA, a homology model was prepared based on the structure of the Prp40 FF1 domain. Intriguingly, mapping the S240F mutation onto the surface of the FF2 model structure reveals that the mutation lies within the binding site determined for the Prp40 FF1 domain. Moreover, a comparison of the FF1 and FF2 binding sites shows that in contrast to the FF1 domain the FF2 binding site is highly positively charged as would be expected for RNA recognition. Unfortunately protein constructs of the Prp40 FF2 were unstructured and could require the presence of U1 snRNA, neighbouring Prp40 domains and/or associated proteins for folding. Moreover, this positively charged binding surface could be responsible for the recognition of phoshorylated

sequences as those contained in the CTD of RNA polymerase II, which has been found to interact with Prp40 *in vitro* (Morris & Greenleaf, 2000). However, further experiments will have to confirm these suggestions.

In summary, this study has revealed that FF domains can mediate protein-protein interactions and that the binding surface is constituted of the second α-helix as well as residues in spatial proximity to αIII. As shown in Fig. 7.8(b), the same binding site could however serve for the binding of RNA or phosphorylated sequences.

7.5 References

Bedford, M. T. & Leder, P. (1999). The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem. Sci.* **24** (7), 264–5.

Ben-Yehuda, S., Dix, I., Russell, C. S., McGarvey, M., Beggs, J. D. & Kupiec, M. (2000). Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in *Saccharomyces cerevisiae*. *Genetics,* **156**, 1503–17.

Berglund, J. A., Abovich, N. & Rosbach, M. (1998). A cooperative interaction between U2Af65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & Dev.* **12**, 858–67.

Blatch, G. L. & Lässle, M. (1999). The tetratrico peptide repeat: a structural motif mediating protein-protein interactions. *Bioessays,* **21**, 932–9.

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. & Warren, G. L. (1998). Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D,* **54**, 905–921.

Chung, S., McLean, M. R. & Rymond, B. C. (1999). Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA,* **5** (8), 1042–54.

Das, A., Cohen, P. & Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* **17**, 1192–9.

Dingwall, C. & Laskey, R. A. (1991). Nuclear targeting sequences: a consensus? *Trens Bioch. Sci.* **16**, 178–81.

Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature,* **387**, 733–6.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

Ishii, Y., Markus, M. A. & Tycko, R. (2001). Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. *J. Biomol. NMR,* **21**, 141–51.

Kao, H. Y. & Siliciano, P. G. (1996). Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* **16** (3), 960–7.

Kumar, A., Rach, C., Hirsh, I. S., Turley, S., deWalque, S., Michels, P. M. & Hol, W. G. J. (2001). An unexpected conformation for the third TPR motif of the peroxin PEX5 from *Trypanosoma brucei*. *J. Mol. Biol.* **307**, 271–82.

Lamb, J. R., Tugendreich, S. & Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* **20**, 257–9.

Lapouge, K., Smith, S. J. M., Walker, P. A., Gamblin, S. J., Smerdon, S. J. & Rittinger, K. (2000). Structure of the TPR domain of $p67^{phox}$ in complex with Rac·GTP. *Mol. Cell,* **6**, 899–907.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR,* **8** (4), 477–86.

Morris, D. P. & Greenleaf, A. L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **275** (51), 39935–43.

Rice, L. & Brünger, A. T. (1999). Crystal structure of the vesicular transport protein Sec17: Implications for SNAP function in SNARE complex disassembly. *Mol. Cell,* **4**, 85–95.

Rückert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7.

Russell, C. S., Ben-Yehuda, S., Dix, I., Kupiec, M. & Beggs, J. D. (2000). Functional analyses of interacting factors involved in both pre-mRNA splicing and cell cycle progression in *Saccharomyces cerevisiae*. *RNA,* **6**, 1565–72.

Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR,* **18**, 303–9.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U. & Moarefi, I. (2000). *Cell,* **101**, 199–210.

Suñé, C., Hayashi, T., Liu, Y., Lane, W., Young, R. & Garcia-Blanco, M. (1997). CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* **17**, 6029–39.

Taylor, P., Dornan, J., Carrello, A., Minchin, R. F., Ratajczak, T. & Walkinshaw, M. D. (2001). Two structures of cyclophilin 40: folding and fidelity in the TPR domains. *Structure,* **9**, 431–8.

Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science,* **278**, 1111–1114.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A. & Judson, R. S. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature,* **403**, 623–7.

Zhu, W., Rainville, I. R., Ding, M., Bolus, M., Heintz, N. H. & Pederson, D. S. (2002). Evidence that the pre-mRNA splicing factor Clf1p plays a role in DNA replication in *Saccharomyces ceerevisiae*. *Genetics,* **160**, 1319–33.