Chapter 5

Solution structure of the Prp40 WW domain pair

† Until recently, structure determination by NMR spectroscopy relied exclusively on short-range restraints such as inter-proton distances shorter than 5 Å and torsion angles, derived from NOEs and three-bond scalar coupling constants, respectively (Wüthrich, 1986). For elongated and/or multi-domain proteins, however, conformationally highly restrictive NOEs between protons remote in the sequence are often sparse, especially in interconnecting regions. In the absence of interdomain NOEs, high-resolution NMR structures of multi-domain proteins have therefore been precluded. Nonetheless, many eukaryotic proteins encompass cooperating domains and/or domain repeats fine-tuning their functions and confering multiple binding sites (Pawson, 1995; Andrade *et al.*, 2001). Detailed knowledge of the individual domain structures as well as their relative orientation and interplay is, hence, key to understand the functions of modular proteins. Only since a few years distance-independent restraints, such as residual dipolar couplings (RDCs) (Tolman *et al.*, 1995; Tjandra & Bax, 1997) and rotational diffusion anisotropy (Fushman *et al.*, 1999; Tjandra *et al.*, 1997), allow to determine high-resolution NMR structures with low NOE densities, which has considerably increased the potential of NMR spectroscopy in structure determination. RDCs and rotational diffusion anisotropy depend on the average orientation of bond vectors (*e.g.* the ^{1}H–^{15}N bond vectors) relative to an external tensor (the alignment and diffusion tensor, respectively) and therefore provide *a priori* long-range restraints for the structural refinement. In this chapter, the solution structure of the WW domain pair of the *Saccharomyces cerevisiae* (*Sc.*) splicing factor Prp40 (pre-mRNA processing protein) is presented. Using ^{1}H–^{15}N RDCs in combination with "classical" short-range restraints, not only the structure of the individual domains were determined accurately and precisely, but also their relative orientation despite the lack of inter-domain NOEs.

Prp40 is a splicing factor associated with the U1 snRNP (uridine-rich small nuclear ribonucleoproteins) and involved in the early steps of yeast mRNA splicing. This modular

†This chapter is based on *Wiesner, S., Stier, G., Sattler, M. and Macias, M. J. (2002). J. Mol. Biol. 234: 807–822* with kind permission from Elsevier.

protein comprises two WW domains (Bork & Sudol, 1994) and six consecutive FF domains (Bedford & Leder, 1999). The region spanning both WW domains has been implicated in cross-intron bridging (Abovich & Rosbach, 1997). In the so-called commitment complex, Prp40 contacts the 5' splice-site (ss) and interacts with the branch-point binding protein BBP (also referred to as Msl5 and ySF1) bringing the 5' ss and the branch-point in spatial proximity. Furthermore, the Prp40 WW domains have been shown to interact with the U5 snRNP core component Prp8 (Abovich & Rosbach, 1997) and phosphorylated repeats of the C-terminal domain (CTD) of the RNA polymerase II largest subunit (Morris & Greenleaf, 2000). The latter interaction suggests a direct function of Prp40 in coupling transcription to splicing. Given the involvement in this complex interaction network, the three-dimensional structure of the Prp40 WW domains should provide insight into how the structures enable the apparent binding promiscuity of the two Prp40 WW domains and whether the relative orientation of the WW domains sterically allows or hinders a simultaneous interaction with two ligands or even cooperative binding.

5.1 Resonance assignment and secondary structure

The recombinant protein used for the structural studies comprised the 75 N-terminal residues of the Sc. Prp40 protein containing the two WW domains and their 10-residue linker (Figure 5.1 and 5.2). The Prp40 WW domains domains share 37% sequence identity and in particular the "YYYN" motif located in the central β-strand of the WW domains, which is found in many other WW domains. Nevertheless, the NMR data allowed the unambiguous assignment of both WW domains as well as the linker, which adopts secondary structure in solution. Due to the good dispersion of the H^N and N^H resonances, the backbone assignment was straightforward for the whole sequence, except for the above-mentioned repeated "YYYN" motif (see Fig. 5.1). The central tyrosine of this motif has nearly identical C^α, H^N

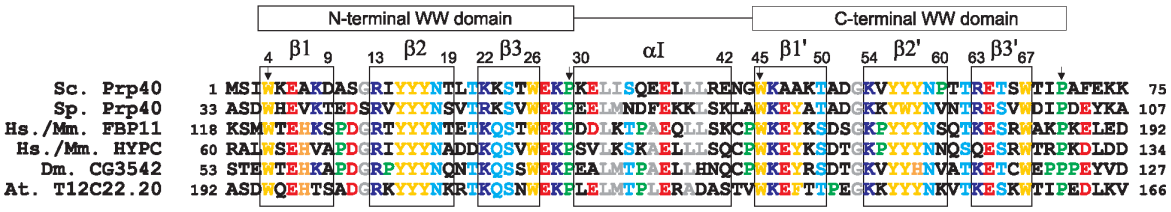

Figure 5.1: Multiple sequence alignment of the regions spanning the tandem WW domains in Prp40 and its orthologues. The WW domain boundaries are indicated at the top and by arrows pointing to the conserved residues at the N- and C-termini of the WW domains, while secondary structure elements are boxed. Conserved aromatic, aliphatic, hydrophilic, positively and negatively charged residues are marked in yellow, grey, cyan, blue and red, respectively, while prolines are marked in green. Rows are labelled by species (Sc.: Saccharomyces cerevisiae, Sp.: Saccharomyces pombe, Hs.: Homo sapiens, Mm.: Mus musculus, Dm.: Drosophila melanogaster and At.: Arabidopsis thaliana) and protein names (Prp40: pre-mRNA processing protein 40, FBP11: formin binding protein 11, HYPC: huntingtin yeast partner C) or gene names. The SwissProt accession numbers are P33203, O14176, O75400/Q9R1C7, O75401/Q9WVC9, Q9VK5 and Q9LPD8, respectively.

and N^H chemical shifts in both domains and three out of the four residues in the first strand opposed to the "YYYN" motif are also identical (Trp4/Trp45, Lys5/Lys46 and Ala7/Ala48, respectively). To solve this problem standard backbone experiments had to be combined with the use of H^N–H^N and H^α–H^α NOEs (see 5.2) and the measurement of hydrogen bonds across the β-strands (Cordier & Grzesiek, 1999). The first tyrosine (Tyr15/Tyr56) of the "YYYN" motif could be unambiguously assigned based on the residue preceding the tyrosine (C^β 36.38 ppm for Ile14 and C^β 30.62 ppm for Val55). The alanine (Ala7/Ala48) opposed to the first tyrosine (Tyr15/Tyr56) was then assigned by the hydrogen bond across the β-strands. The second tyrosine (Tyr16/Tyr57) was assigned using the H^α–H^α and H^N–H^N NOEs across the β-strands (H^α 4.36 ppm for Glu6 and 4.82 ppm for Ala47), while for the assignment of the third tyrosine (Tyr17/Tyr58) the H^α–H^α and H^N–H^N NOEs to the third strand were used. Finally, the asparagines following the triple tyrosine sequences (Asn18/Asn59) were assigned using the H^N–H^N NOEs to the third strand, since the H^α chemical shifts of the opposite tryptophans (Trp4/Trp45) are almost identical.

Compared to values present in the BioMagResBank (http://www.bmrb.wisc.edu) anomalously low chemical shifts were observed for the α-protons of residues preceding the conserved C-terminal prolines (Lys28 and Ile69), all protons of the prolines themselves (Pro29 and Pro70) and one of the β-protons of Asn18 and Asn59, respectively. This effect arises from the spatial proximity of the constitutive N-terminal Trp and the C-terminal Pro in WW domains. These unique chemical shifts have been observed for other WW domains (Macias *et al.*, 1996, 2000; Kanelis *et al.*, 1998) and serve hence as a useful indicator of tertiary structure in samples containing WW domains.

The secondary C^α chemical shifts, the $^3J(H^N,H^\alpha)$ couplings and the characteristic NOEs (Fig. 5.2) confirm not only that the secondary structure of each WW domain is a triple stranded β–sheet, but show also that the linker connecting the WW domains adopts an α-helical fold. Accordingly, the heteronuclear $\{^1H\}$-^{15}N NOEs of the linker residues indicate that the linker is rigid and adopts secondary structure in solution (Fig. 5.2).

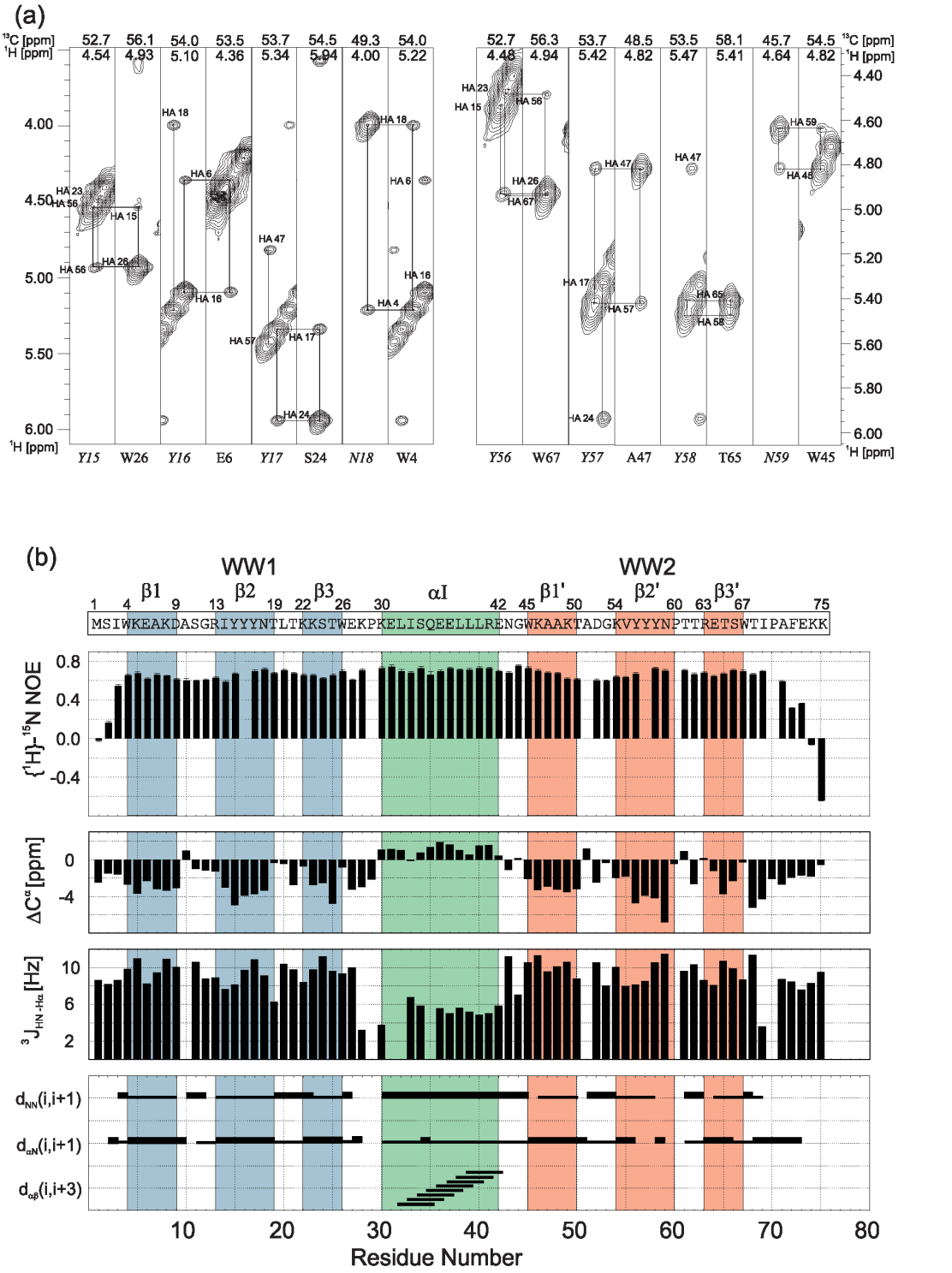

Figure 5.2: (a) Characteristic NOEs in the β-strands of the Prp40 WW domains. Strips from the 3D ^{13}C-HMQC-NOESY displaying the H$^{\alpha}$-H$^{\alpha}$ NOEs used for the assignment of the YYYN motif (italic residue names).(b) Summary of heteronuclear $\{^1$H$\}$-^{15}N NOEs, C$^{\alpha}$ secondary chemical shifts, ^3J(H$^{\mathrm{N}}$H$^{\alpha}$) couplings and sequential connectivities (from top to bottom) used to define the secondary structure of the Prp40 WW domain pair. Strong and weak NOE intensities for the sequential d_{NN}, $d_{\alpha N}$ and $d_{\alpha\beta}$ connectivities are indicated by thick and thin horizontal bars, respectively. β-strands are coloured in blue for WW1 and in red for WW2, while the α-helical linker is coloured in green.

5.2 Structure determination and assessment

The use of "classical" short-range restraints, such as NOEs, hydrogen bonds and dihedral angles, in the structure refinement led to a structure ensemble that was well defined for the individual WW domains (Fig. 5.3 (a) and (b)). However, the exclusive use of short-range restraints did not allow us to define the relative domain orientation precisely, since only few NOEs connecting the linker to the respective domain were observed and inter-domain NOEs were completely absent. To obtain long-range restraints for the structure calculation ^{1}H–^{15}N residual dipolar coupling constants (RDCs) were measured dissolving the tandem Prp40 WW domains in a liquid crystalline medium (Tjandra & Bax, 1997) (Fig. 5.4(a)). Especially in the case of multi-domain proteins, however, a potential pitfall of this method is that the domains can tumble individually in solution precluding the use of a unique alignment tensor of the whole protein to define the relative domain orientation by RDCs.

To ensure that the tandem WW domains of Prp40 tumble as a single folded unit, a detailed analysis of ^{1}H–^{15}N RDCs and ^{15}N relaxation data was carried out for the Prp40 WW domain pair and independently for the individual domains. ^{15}Nrelaxation experiments were performed to identify regions of high mobility and to investigate whether the obtained ^{15}N R_2/R_1 ratios (Fig. 5.4(c)) and heteronuclear $\{^{1}$H$\}$-^{15}N NOEs (Fig. 5.2) are different for each Prp40 WW domain indicating individually tumbling domains. However, the ^{15}N R_2/R_1 ratios and heteronuclear $\{^{1}$H$\}$-^{15}N NOEs obtained are very similar for both domains and even the linker. Furthermore, the average diffusion tensor parameters of the individual domains and the whole WW domain pair, respectively, were derived independently with the program TENSOR v2.0 (Dosset *et al.*, 2000) using the measured ^{15}N relaxation rates and the seven lowest-energy structures refined with RDCs. In all cases, an axially symmetric diffusion tensor was obtained yielding virtually identical values for the parallel and perpendicular diffusion coefficients $(D_{\parallel}, D_{\perp})$ and the polar angles (ϕ, θ) relating the diffusion tensor to the NMR structure frame. The results are summarised in Table 5.1 and confirm independently the domain orientation obtained with RDCs, since the ^{15}N R_2/R_1 ratios were not incorporated in the structure refinement. The ^{15}N relaxation rates and the obtained diffusion tensor parameters were further used to assess the correlation between the ^{15}N R_2/R_1 ratios observed and those back-calculated from the seven lowest-energy structures refined with RDCs. As shown in Fig. 5.4(d), the observed ^{15}N R_2/R_1 ratios agree well with the back-calculated ^{15}N R_2/R_1 ratios.

To ensure furthermore that the Prp40 WW domain pair orients in the liquid crystalline medium with an alignment tensor common to the whole protein, the experimental backbone ^{1}H–^{15}N RDCs were fitted to the structures of the isolated domains and the whole WW domain pair, respectively, using the program MODULE (Dosset *et al.*, 2001). As was the case for the diffusion tensor parameters, the alignment tensor parameters are virtually identical

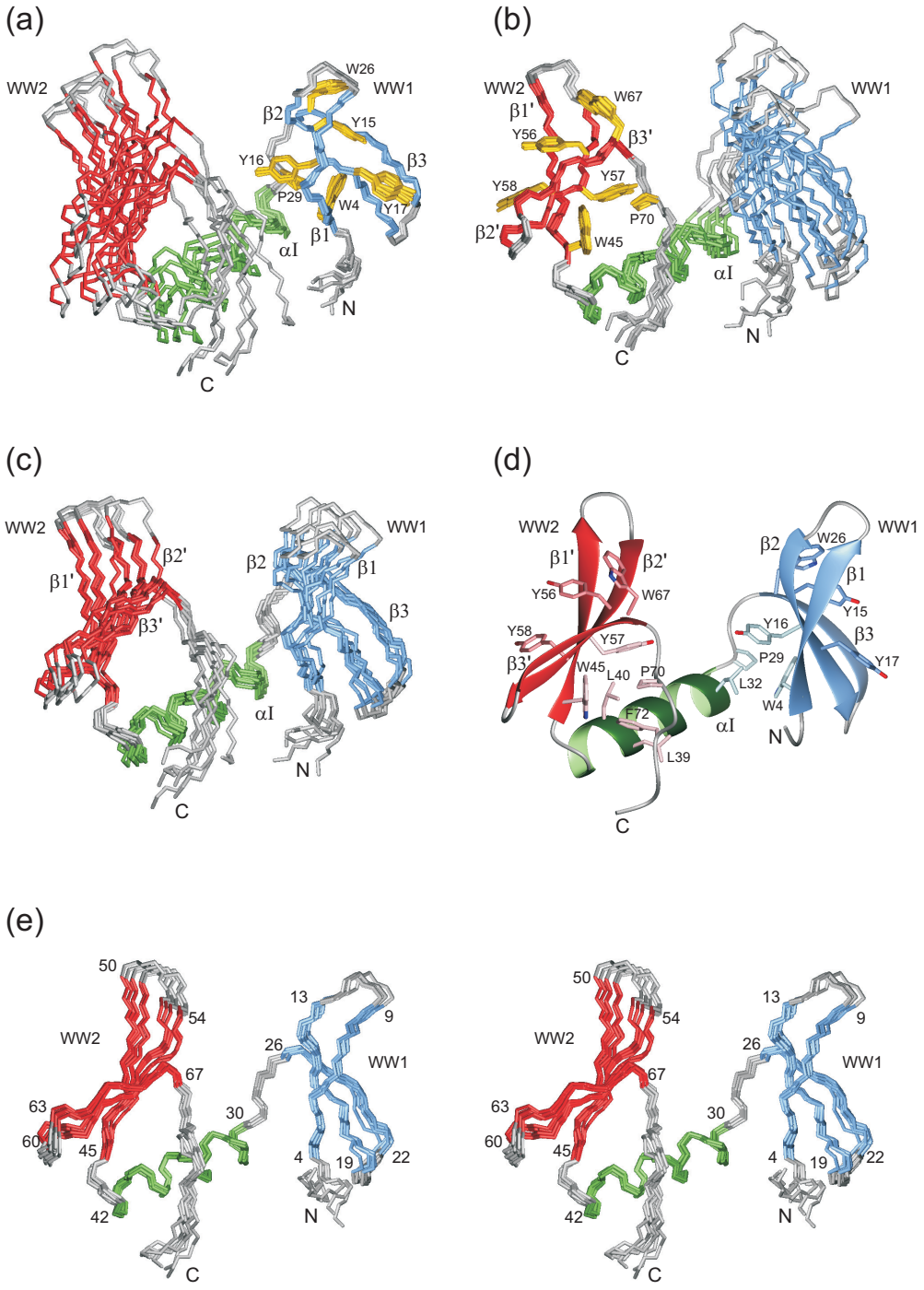

Figure 5.3: Structural alignment of the 7 lowest energy structures (of 20 calculated) of the tandem WW domains of Prp40 refined with and without residual dipolar coupling (RDC) restraints. In all figures, β-strands are coloured in blue for WW1 and in red for WW2, while the α-helical linker is coloured in green. Backbone superposition of residues 4–30 (WW1) (a) and residues 45–71 (WW2) (b) for the structure ensemble refined without RDCs. Characteristic side-chains are shown in yellow. (c) Backbone superposition of residues 4–71 (WW1–2) for the structure ensemble refined without RDCs. (d) Ribbon representation of the lowest energy structure refined with RDCs showing the residues forming the hydrophobic cores on the concave and convex surfaces of the WW domains. (e) Stereoview showing the best-fit backbone (N, C^{α}, C') superposition of residues 4–71 (WW1–2) for the structure ensemble refined with RDCs.

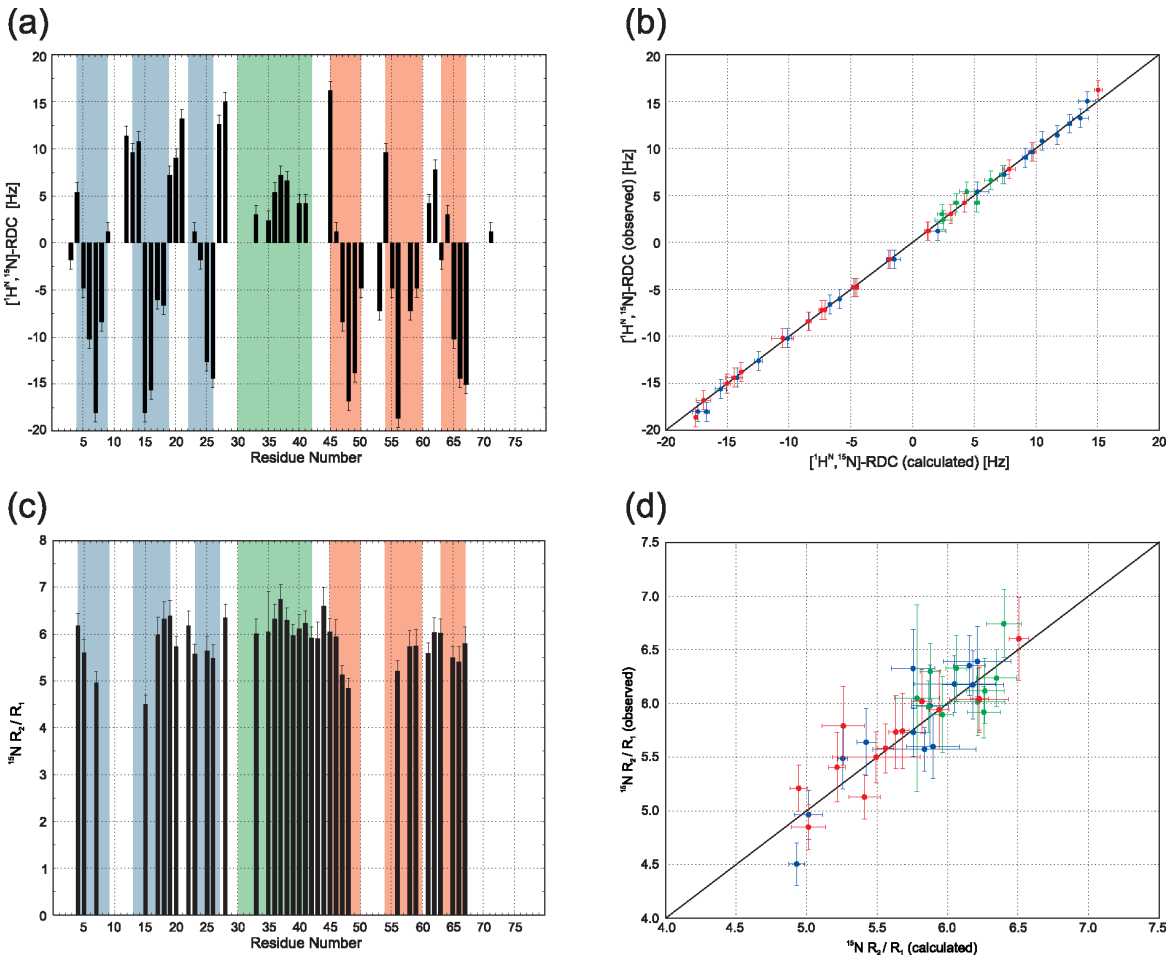

Figure 5.4: Structure assessment of the tandem WW domains of the splicing factor Prp40. The WW domain boundaries are indicated by grey shaded boxes (shown at the top), while secondary structure elements are boxed in blue for the N-terminal WW domain (WW1), in green for the linker region and in red for the C-terminal WW domain (WW2) in panel (a) and (c). In panel (b) and (d), blue symbols represent residues in WW1, green symbols residues in the linker and blue symbols residues in WW2. (a) Observed [^{1}H–^{15}N]-residual dipolar couplings (RDCs) versus residue number. (b) Correlation between observed [^{1}H–^{15}N]-RDCs and [^{1}H–^{15}N]-RDCs back-calculated from the seven lowest energy structures. (c) ^{15}N R_2/R_1 ratios versus residue number. (d) Correlation between observed ^{15}N R_2/R_1 ratios and ^{15}N R_2/R_1 ratios back-calculated from the seven lowest energy structures using an axially symmetric diffusion tensor and taking only rigid residues into account.

for the individual domains and the whole protein (see Table 5.1). In addition, the correlation between the experimental backbone ^{1}H–^{15}N RDCs and those back-calculated from the refined structure ensemble was assessed. As can be seen in Fig. 5.4(b), the experimental backbone ^{1}H–^{15}N RDCs agree well with the back-calculated values not only for the individual WW domains, but also for the WW domain pair. All results obtained are thus consistent with a single alignment and diffusion tensor, respectively, for the whole protein. Moreover, these results confirm that the lack of inter-domain NOEs is not due to domain motions, but due to an inter-domain distance larger than 5 Å.

Table 5.1: Mean diffusion and alignment tensor parameters.

	Diffusion tensor parameters*			Alignment tensor parameters				
Domain	$D_\perp/D_\parallel$ †	φ[°]	ϑ[°]‡	D_a[Hz]	D_r[Hz]§	α[°]	β[°]	γ[°]¶
WW1-2	1.36±0.05	-79±7	51±7	-8.0±0.3	-4.0±0.5	129±4	71±13	124±8
WW1	1.58±0.10	-75±7	56±10	-8.0±0.4	-3.9±0.4	130±2	72±12	124±9
WW2	1.38±0.06	-76±12	58±10	-7.9±0.4	-4.1±0.4	129±3	71±14	124±7

*not used in structure refinement.

†Ratio of the parallel and perpendicular diffusion axes defined as $D_\parallel=D_{zz}$ and $D_\perp=D_{yy}=D_{xx}$, respectively, for a prolate rotator approximation.

‡Polar angles $\{\varphi,\vartheta\}$ describing the orientation of the principal axis of the diffusion tensor $\tilde{D}$ with respect to the NMR structure frame.

§Axial and rhombic component of the alignment tensor $\tilde{A}$ defined as $D_a = c_{\mathrm{HN}} A_a$ and $D_r = c_{\mathrm{HN}} A_r$ (see (3.6)). The dipolar interaction constant c_{HN} equals $\left[-\mu_0/(8\pi^2)\right] \hbar\gamma_{\mathrm{H}}\gamma_{\mathrm{N}}\langle r_{\mathrm{HN}}\rangle^{-3}$.

¶Euler angles α, β, γ describing the orientation of the principal axis of the alignment tensor with respect to the NMR structure frame.

The use of a unique alignment tensor in the structure calculation led to a well-defined structure ensemble for the whole protein as shown in Fig. 5.3(e). Clearly, the ensemble of structures is better defined after the incorporation of RDCs in the structure calculation as reflected by a decrease of the backbone root mean square (r.m.s.) displacement of residues 4–71 from 1.14 Å to 0.55 Å (Fig. 5.3(c) and (e), Table 5.2). It is noteworthy that the final domain orientation is very similar in calculations with and without RDCs and that none of the calculated structures has inter-domain distances below 5 Å. This indicates that the use of RDCs does not induce an artificial orientation of the domains, but solely increases the resolution of the final set of structures. Table 5.2 summarises the structural statistics for the seven lowest energy structures analysed with CNS (Brünger *et al.*, 1998) and PROCHECK-NMR (Laskowski *et al.*, 1996).

5.3 Description of the structure

The Prp40 WW domains each fold as a triple-stranded curved anti-parallel β-sheet and are connected by a well-defined, stable α-helical linker. The boundaries of the secondary structure elements are depicted in Fig. 5.1 and 5.3(e) and range from residue 4/45–9/50 for the first β-strand, from residue 13/54–19/60 for the second β-strand, from residue 22/63–26/67 for the third β-strand and from residues 30–42 for the α-helical linker. The three-dimensional structures of the individual Prp40 WW domains are very similar to previously determined WW domain structures with an r.m.s. deviation of the β-sheet backbone atoms of 0.89 Å for WW1 and WW2 to the solution structure of the FBP28 WW domain (Macias *et al.*, 2000)

and 0.72 Å for WW1 and 0.63 Å for WW2 to the crystal structure of the dystrophin WW domain (Huang *et al.*, 2000).

On the convex surface of each WW domain, the three highly conserved residues Trp4/45, Tyr16/57 and Pro29/70, form a hydrophobic core (Fig. 5.3 (d)). This core is stabilised by additional hydrophobic interactions with residues C-terminal of each WW domain (Leu32 and Phe72, respectively). Furthermore, the second WW domain is stabilised by contacts to Leu39 and Leu40, located close to the C-terminus of the helical linker. These interactions restrict the relative orientation of the WW domains, but do not lead to a well-defined ensemble of structures without long-range restraints such as RDCs. Nevertheless, the aforementioned interactions appear to be sufficient to keep the whole molecule rigid enough to tumble and align as one entity. Interestingly, the sequence of the linker is almost as conserved as the sequence of the WW domains themselves in all mammalian Prp40 orthologues (Fig. 5.1). Since in particular the linker residues Leu32 and Leu40, which fold back into the hydrophobic cores on the convex surface of each WW domain, are conserved, the orthologous Prp40 WW domains including their linkers can be expected to adopt a very similar fold.

On the concave surface of each WW domain, an aromatic pocket is formed by Tyr15/56, Tyr17/58 and Trp26/67. It is known that these residues form together with Thr19/P60, Lys22/Arg63 and Ser24/Thr65 the binding pocket for the proline-rich targets of WW domains (Fig. 5.3(d)) (Macias *et al.*, 1996; Huang *et al.*, 2000; Kanelis *et al.*, 2001; Pires *et al.*, 2001). Interestingly, the binding pockets of the Prp40 WW domains face opposite sides in the structure, instead of forming a continuous binding pocket. This suggests that the Prp40 WW domains bind their ligands in an individual, non-cooperative fashion, provided that no dramatic domain rearrangements occur upon ligand binding. The stable linker could function in keeping the binding site in place such that two different proteins could interact with the Prp40 WW domains simultaneously as required for the formation of large protein-protein assemblies. The accessibility of both binding sites supports furthermore the bridging function of the Prp40 WW domains in the splicing complex.

Table 5.2: Structural statistics.

	#*	⟨SA⟩ r.m.s. deviation without $^1D_{\mathrm{HN}}$	⟨SA⟩† r.m.s. deviation with $^1D_{\mathrm{HN}}$
R.m.s.deviation (Å) from experimental distance restraints‡			
All distance restraints (Å)	1982	0.0153±0.0016	0.0200±0.0021
Unambiguous NOEs (Å)	1938	0.0145±0.0018	0.0184±0.0023
Hydrogen bond restraints (Å)	44	0.0331±0.0036	0.0554±0.0039
R.m.s. deviation (°) from torsion angle restraints§			
Dihedral angles (62 ϕ, 61 ψ, 8 χ_1)	131	0.545±0.074	0.856±0.0046
R.m.s. deviation (Hz) from experimental residual dipolar coupling restraints			
D^1_{HN}	50	–	0.516±0.022
R.m.s. deviation from idealised geometry			
Bond lengths (Å)		0.00158±0.00011	0.00211±0.00011
Bond angles (°)		0.3297±0.0130	0.4318±0.0163
Improper dihedral angles (°)		0.1737±0.0011	0.3317±0.0251
Coordinate precision of N, C^α, C' / all heavy atoms (Å)¶			
Residues 4–71 (WW1-2)		1.14±0.48 / 1.32±0.43	0.55±0.14 / 0.85±0.14
Residues 4–30 (WW1)		0.30±0.06 / 0.65±0.04	0.21±0.05 / 0.66±0.08
Residues 45–71 (WW2)		0.23±0.10 / 0.54±0.14	0.23±0.06 / 0.59±0.05
Ramachandran plot (%)‖			
Residues in most favoured regions		91.3±2.7	89.3±2.1
Residues in additionally allowed regions		8.7±2.7	10.7±2.1
Residues in generously allowed regions		0.0±0.0	0.0±0.0
Residues in disallowed regions		0.0±0.0	0.0±0.0

*# 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 0.4 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 glycines, prolines and flexible residues at the termini and in the first loop of each WW domain (1–2, 10–11, 51–52, 72–75).

5.4 Discussion and conclusions

The potential difficulty in accurately establishing long-range order from the "classical" short-range NMR restraints has already been realised when the methodology for protein structure determination by NMR spectroscopy was developed (Wüthrich *et al.*, 1982; Havel & Wüthrich, 1985). Due to the high proton density in proteins and the typical organisation of compact, global proteins around hydrophobic cores that bring many protons from remote parts of the amino-acid sequence close together in space, there is usually sufficient information to determine long-range order from short-range restraints. However, if NMR spectroscopy is applied to deuterated, elongated or multi-domain proteins, establishing long-range order and relative domain orientations has proven more difficult (Brüschweiler *et al.*, 1995; Tjandra *et al.*, 1997; Braddock *et al.*, 2002). The development of methods to measure residual dipolar couplings (Tolman *et al.*, 1995; Tjandra & Bax, 1997) has recently opened the way to determine average domain orientations in solution.

However, residual dipolar couplings alone do not define a unique orientation of a bond vector within its alignment frame. There are at least four equivalent orientations which agree with each measured value. This four-fold degeneracy is inherent to the orientation of any three-dimensional structure relative to a molecular alignment tensor and derives from simple symmetry operations (180° rotations around the alignment tensor axes A_{xx}, A_{yy} and A_{zz}). This degeneracy can be eliminated by measuring an additional set of RDCs in another alignment medium or by applying the RDCs at a relatively late stage in the struture calculation, when already pre-folded structures exist. The latter strategy was applicable to the Prp40 WW domain pair, since despite the absence of inter-domain NOEs numerous NOEs connected the respective WW domain to the α-helical linker. These NOEs were sufficient to restrict the domain orientation to vary only to a small degree around an average structure. It is also noteworthy that this average structure fell within the structural ensemble determined using long-range information from RDCs. This underlines that the incorporation of the long-range information from RDCs did not produce an artificial domain orientation.

In a number of studies RDCs have been used to determine the relative orientation of domains in multi-domain proteins by NMR spectroscopy. In most of these studies, however, previously determined coordinates from crystal or solution structures were used to assess the agreement of the measured RDCs with the protein coordinates by order matrix analyses (Losonczi *et al.*, 1999; Fischer *et al.*, 1999) or to orient the domains by hinge rotations such that the coordinates fitted to the measured RDCs (Skrynnikov *et al.*, 2000; Goto *et al.*, 2000). In another study, RDCs were successfully applied to refine the subdomain orientation of the ribosomal protein S4 Δ41 (Markus *et al.*, 1999). However, in this study also a number of interdomain NOEs was included in the structure refinement. Furthermore, direct refinement against a large number of experimental RDCs of a highly deuterated protein has

been found to produce structures with large numbers of violations specifically in the dipolar coupling restraints (Mueller *et al.*, 2000). The large numbers of violations was attributed to the comparatively small number of observed NOEs, which led to a probably extremely complex eneregy surface over which refinement occured due to the degeneracy of the RDC data. This problem could be solved by an altered incorporation of the RDCs in the structure refinement. First, preliminary NMR structures were calculated based on short-range restraints, from which the molecular alignment tensor was determined using the measured RDCs. Subsequently, the average structure was rotated into its alignment frame and the orientations of the internuclear vectors compared to the degenerate possible orientations obtained from the RDCs alone. The most likely orientation was then extracted and used as orientational restraint in the structure calculation.

In this study it is has been shown that including long-range restraints derived from RDCs in the structure calculations not only the structures of the individual Prp40 WW domains were precisely and accurately defined, but also their relative orientation despite the absence of inter-domain NOEs. In the free form, the binding surfaces of the Prp40 WW domains point away from each other, suggesting that the domains can interact simultaneously with two proline-rich target sites as present in the branch-point binding protein BBP and the U5 core component Prp8. It is interesting to note that in almost all cases where crystal structures of the multi-domain proteins studied by NMR were available, the domain orientations were different. While in solution the protein can interact with the alignment medium such that the domain orientation is altered, crystal contacts or flash freezing of the crystals can force protein domains into artificial orientations. However, the complex behaviour of modular proteins can only be understood in terms of the relative arrangement of the cooperating domains. It is therefore important to keep in mind the methodological limitations in the determination of domain orientations by solution state NMR and X-ray crystallography.

5.5 References

Abovich, N. & Rosbach, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell,* **89**, 403–412.

Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001). Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134** (2–3), 117–31.

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

Bork, P. & Sudol, M. (1994). The WW domain: a signalling site in dystrophin? *Trends Biochem. Sci.* **19** (12), 531–3.

Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D. & Clore, G. M. (2002). Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature,* **415**, 1051–6.

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.

Brüschweiler, R., Liao, X. & Wright, P. (1995). Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science,* **268**, 886–9.

Cordier, F. & Grzesiek, S. (1999). Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{N,C'}$ scalar couplings. *J. Am. Chem. Soc.* **121**, 1601–1602.

Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR,* **16**, 23–28.

Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. (2001). A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR,* **20** (3), 223–31.

Fischer, M. W. F., Losonczi, J. A., Weaver, J. L. & Prestegard, J. H. (1999). Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry,* **38**, 9013–9022.

Fushman, D., Xu, R. & Cowburn, D. (1999). Direct determination of interdomain orientation on ligation: Use of the orientational dependence of 15N NMR relaxation in Abl SH(32). *Biochemistry,* **38**, 10225–10250.

Goto, N. K., Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. (2000). What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. *J. Mol. Biol.* **308**, 745–64.

Havel, T. F. & Wüthrich, K. (1985). An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J. Mol. Biol.* **182**, 218–94.

Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M. & Eck, M. (2000). Stucture of a WW domain containing fragment of dystrophin in complex with b-dystroglycan. *Nat. Struct. Biol.* **7**, 634–38.

Kanelis, V., Farrow, N. A., Kay, L. E., Rotin, D. & Forman-Kay, J. D. (1998). NMR studies of tandem WW domains of Nedd4 in complex with a PY motif-containing region of the epithelial sodium channel. *Biochem. Cell Biol.* **76**, 341–350.

Kanelis, V., Rotin, D. & Forman-Kay, J. D. (2001). Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat. Struct. Biol.* **8** (5), 407–12.

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.

Losonczi, J. A., Andrec, M., Fischer, M. W. F. & Prestegard, J. H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Res.* **138**, 334–342.

Macias, M. J., Gervais, V., Civera, C. & Oschkinat, H. (2000). Towards an understanding of beta-sheet structures: design of a prototype WW domain. *Nat. Struct. Biol.* **7**, 375–379.

Macias, M. J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. & Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature,* **382**, 646–649.

Markus, M. A., Gerstner, R. B., Draper, D. E. & Toshia, D. A. (1999). Refining the Overall Structure and Subdomain Orientation of Ribosomal Protein S4 Δ41 with Dipolar Couplings Measured by NMR in Uniaxial Liquid Crystalline Phases. *J. Mol. Biol.* **292**, 375–387.

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.

Mueller, G. A., Choy, W. Y., Yang, D., Forman-Kay, J. D., Venters, R. A. & Kay, L. E. (2000). Global folds of proteins with low densities of NOEs using residual dipolar couplings: Application to the 370-residue maltodextrin-binding protein. *J. Mol. Biol.* **300**, 197–212.

Pawson, T. (1995). Protein modules and signalling networks. *Nature,* **373**, 573–80.

Pires, J. R., Taha-Nejad, F., Toepert, F., Ast, T., Hoffmuller, U., Schneider-Mergener, J., Kuhne, R., Macias, M. J. & Oschkinat, H. (2001). Solution structures of the YAP65 WW domain and the variant L30K in complex with the peptides GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *J. Mol. Biol.* **314** (5), 1147–56.

Skrynnikov, N. R., Goto, N. K., Yang, D., Choy, W. Y., Tolman, J. R., Mueller, G. A. & Kay, L. E. (2000). Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. *J. Mol. Biol.* **295** (5), 1265–1273.

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.

Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A. & Clore, G. M. (1997). Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat. Struct. Biol.* **4**, 443–9.

Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. USA,* **92**, 9279.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids.* John Wiley & Sons, Inc., New York.

Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982). Sequential assignment as a basis for the determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J. Mol. Biol.* **115**, 311–9.