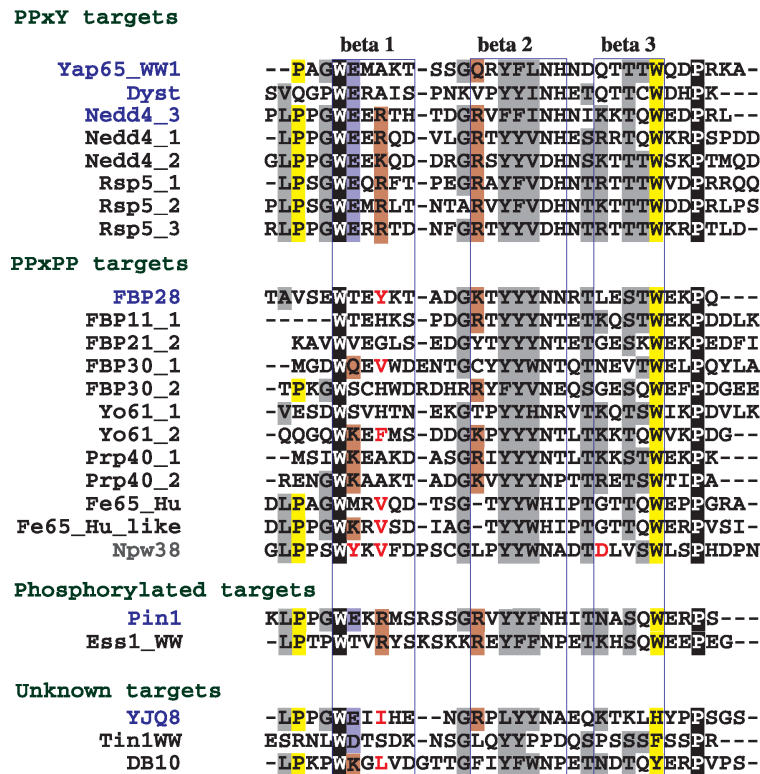# Chapter 4

WW and SH3 domains, two different scaffolds to recognise proline-rich ligands

* Due to its small size and compact fold, the WW domain became an attractive model for studies of protein stability and design (Koepf *et al.*, 1999; Ibragimova & Wade, 1999; Jiang *et al.*, 2001; Jager *et al.*, 2001). Specific residues have been identified that play a critical role in the structure and function of the domain and also in modulating its stability. In fact, the WW domain is the first protein module that has been successfully designed de novo, demonstrating the significant insight we already have regarding its fold (Macias *et al.*, 2000). Besides, the WW domain sequence is well conserved in length, even in its loops, which is a remarkable feature of this domain, compared with others, making protein modeling a useful tool for generating three-dimensional representations of their sequences. Nevertheless, attempts to predict binding targets for a specific WW domain sequence, or even for one of its subgroups or classes, with a good probability have not been made so far. Based on the pattern of semi-conserved residues, WW domain sequences have been classified into three groups as described previously (Macias *et al.*, 2000). Group I contains the C-terminal tryptophan and the N-terminal proline; Group II sequences lack the N-terminal proline and finally Group III with sequences without the second tryptophan. In another classification, based on the ligand predilection, WW domains were divided into two major and two minor groups (Sudol & Hunter, 2000). One major group (Group I) binds polypeptides with the minimal core consensus PPxY, whereas the other binds ligands with the PPLP motif usually embedded in a long stretch of prolines (Group II). Group III WW domains select poly-P motifs flanked by R or K, whereas Group IV WW domains bind to short sequences with phospho-S or phospho-T followed by P, in a phosphorylation-dependent manner (Sudol & Hunter, 2000). A sequence alignment of some selected WW sequences combining binding preferences and sequence conservation is shown in Fig. 4.1.

```
PPxY targets
                         beta 1      beta 2      beta 3
   Yap65_WW1   --PAGWEMAKT-SSGQRYFLNHNDQTTTWQDPRKA-
        Dyst   SVQGPWERAIS-PNKVPYYINHETQTTCWDHPK---
     Nedd4_3   PLPPGWEERTH-TDGRVFFINHNIKKTQWEDPRL--
     Nedd4_1   -LPPGWEERQD-VLGRTYYVNHESRRTQWKRPSPDD
     Nedd4_2   GLPPGWEEKQD-DRGRSYYVDHNSKTTTWSKPTMQD
      Rsp5_1   -LPSGWEQRFT-PEGRAYFVDHNTRTTTWVDPRRQQ
      Rsp5_2   PLPSGWEMRLT-NTARVYFVDHNTKTTTWDDPRLPS
      Rsp5_3   RLPPGWERRTD-NFGRTYYVDHNTRTTTWKRPTLD-

PPxPP targets
       FBP28   TAVSEWTEYKT-ADGKTYYYNNRTLESTWEKPQ---
     FBP11_1   -----WTEHKS-PDGRTYYYNTETKQSTWEKPDDLK
     FBP21_2    KAVWVEGLS-EDGYTYYYNTETGESKWEKPEDFI
     FBP30_1   --MGDWQEVWDENTGCYYYWNTQTNEVTWELPQYLA
     FBP30_2   -TPKGWSCHWDRDHRRYFYVNEQSGESQWEFPDGEE
      Yo61_1   -VESDWSVHTN-EKGTPYYHNRVTKQTSWIKPDVLK
      Yo61_2   -QQGQWKEFMS-DDGKPYYYNTLTKKTQWVKPDG--
     Prp40_1   --MSIWKEAKD-ASGRIYYYNTLTKKSTWEKPK---
     Prp40_2   -RENGWKAAKT-ADGKVYYYNPTTRETSWTIPA---
     Fe65_Hu   DLPAGWMRVQD-TSG-TYYWHIPTGTTQWEPPGRA-
Fe65_Hu_like   DLPPGWKRVSD-IAG-TYYWHIPTGTTQWERPVSI-
       Npw38   GLPPSWYKVFDPSCGLPYYWNADTDLVSWLSPHDPN

Phosphorylated targets
        Pin1   KLPPGWEKRMSRSSGRVYYFNHITNASQWERPS---
     Ess1_WW   -LPTPWTVRYSKSKKREYFFNPETKHSQWEEPEG--

Unknown targets
        YJQ8   -LPPGWEIIHE--NGRPLYYNAEQKTKLHYPPSGS-
      Tin1WW   ESRNLWDTSDK-NSGLQYYPPDQSPSSSFSSPR---
        DB10   -LPKPWKGLVDGTTGFIYFWNPETNDTQYERPVPS-
```

Figure 4.1: Multiple sequence alignment of selected WW domain sequences generated with ClustalX (Thompson *et al.*, 1994). The strictly conserved residues are boxed in black and the semi-conserved residues, previously used for classifying the WW sequences (Macias *et al.*, 2000), are boxed in yellow. Other conserved residues are shown in gray, negatively charged residues in blue and positively charged residues in brown boxes. Residues shown in red correspond to the suggested second binding site, as explained in the text.

In this Chapter, the structural characteristics of WW domain-ligand complexes determined so far will be elaborated and compared to SH3 domains. On the basis of four WW domain structures in complex with different peptides and two structures of free WW domains (Macias *et al.*, 1996; Verdecia *et al.*, 2000; Huang *et al.*, 2000; Kanelis *et al.*, 2001; Pires *et al.*, 2001), a three-dimensional structure has been modeled for the Npw38 WW domain that allows us to compare binding properties of WW and SH3 domains.

4.1 WW domain as a phosphate-dependent SH3 domain?

WW domains have the ability to bind proline-rich cores and/or phospho-S-P/ phospho-T-P-containing motifs (Sudol & Hunter, 2000). It is interesting that such a small and well-conserved module has a surprisingly large repertoire of potential ligands. The dissociation constants (K_d) for WW-ligand complexes lie in the high nM to low μM range for proline-rich ligands, and in the low μM range for phospho-SP- or phospho-TP-containing ligands (Sudol & Hunter, 2000). Phosphorylation of the terminal tyrosine in the ligand PPxY for Yap65

WW domain abolishes the binding in vitro (Chen & Sudol, 1995) and in vivo (Ilsley *et al.*, 2001), suggesting that this modification could represent a negative regulation mechanism for a large subset of WW domains. Little is known about the regulation of other WW domain complexes whose ligands do not contain phosphorylatable residues. In its function, the WW domain embodies elements of SH3 and SH2 domains by recognising proline-rich ligands and being in some cases regulated by phosphorylation. In this sense, one could consider WW domains as phosphorylation-dependent SH3 domains. As discussed later on in this chapter, similarities in the binding site between WW and SH3 domains seem to be more extensive than originally expected, at least for a subgroup of WW sequences.

4.2 What do we learn from the structures?

4.2.1 WW domains binding PPxY cores

Most of the structural studies performed on WW complexes refer to WW sequences that bind to the PPxY motif, namely YAP65 (wild type (Macias *et al.*, 1996) and L30K mutant (Pires *et al.*, 2001)), dystrophin (Huang *et al.*, 2000) and Nedd4 (Kanelis *et al.*, 2001), highlighted in blue in Fig. 4.1. The structures of these complexes are compared in Fig. 4.2.
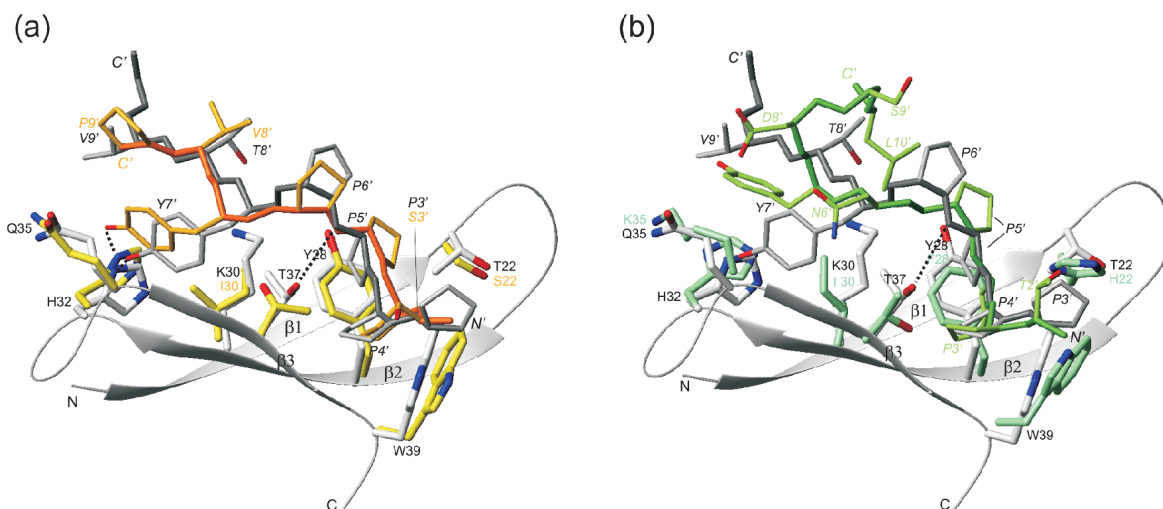
Figure 4.2: Structure comparison of complexes formed by WW domains binding to ligands displaying the PPxY motif. The YAP65 WW domain (L30K) in complex with the polypeptide PPPPYTV (depicted in gray) is compared to (a) the dystrophin WW domain (shown in yellow) in complex with the SPPPYVP peptide from β-dystroglycan (in orange) and (b) to the Nedd4 WW:TPPPNYDSL complex (in green). Hydrogen bonds are indicated by dashed lines. Coloured residue label correspond to residues that are different in the respective complexes, otherwise only the residues of the YAP65 WW complex are denoted. The Figures have been prepared using the program MOLMOL (Koradi *et al.*, 1996).

In the central strand and contributing to the binding site, the three WW domain sequences have at position 28 either tyrosine or phenylalanine, an aliphatic residue at position 30 (L/I/V, respectively) and a histidine at position 32. As it is shown in Figure 4.2, the three

domains (YAP65 (L30K), dystrophin and Nedd4) interact with the peptide in a similar manner. The tyrosine (Y7') of the peptide contacts residue 30 (I/K respectively) and the histidine 32 located in the central strand as well as the hydrophilic residue 35 (either Q/K) found at the beginning of the third strand. The two structural prolines of the peptide (P4' and P5') are packed between the highly conserved aromatic residues Y/F28 and the W39. Compared to the YAP65 and dystrophin complexes, the Nedd4 WW domain binds a peptide, which has a C-terminal, helical extension in addition to the PPxY motif. The additional residues extend the contact interface between both proteins to the first strand of the WW domain.

When the three complexes YAP65 (L30K), Nedd4 and dystrophin are carefully compared, minor differences can be discovered. These differences concern the position of the two prolines in the peptide (P4' and P5') that contact the aromatic binding pocket, the orientation of the peptide tyrosine ring (Y7') and the distance from the peptide backbone to the highly-conserved T37. As it is shown in Fig. 4.1 and 4.2, most of the residues localized in the last two strands and especially those contributing to the binding site are highly conserved. Therefore the first strand and the loop1 that harbor the most significant sequence divergences must be responsible for the observed differences in the peptide binding. The last residue of the first strand, residue 22 (T/S/H, respectively) is in close contact with the second interacting proline (P5') of the peptide and it is not conserved. While the dystrophin WW domain binds the second proline (P5') of the PPxY motif quite parallel to the Y28 ring, in the YAP65 (L30K) complex it appears more perpendicular to it. In the Nedd4 complex however, P5' is also parallel to the aromatic ring, but shifted away from it. Furthermore, the preceding proline (P4') is buried deeper between the aromatic rings of F28 and W39 than in the other two complexes. These differences in the structures can be related to the variable surface available at the binding site, in particular that around the aromatic ring of Y/F28. The small serine in the dystrophin structure allows the peptide to bulge out, while the additional methyl group of the threonine leaves less space for P5' to pack against the Y/F28 ring and consequently leads to a rotation of the P5' in the YAP65 structure (Pires *et al.*, 2001). The even bulkier histidine ring in the Nedd4 complex finally forces the peptide to move away from the aromatic binding site. The resulting weakened interaction between the P5' and the F28 rings could be compensated by the extra contacts from the helical extension of the peptide to the first strand.

In the YAP65 (L30K) complex, the hydroxyl-group of T37 in the third strand points towards the carbonyl group of P5' (marked with a dashed line in Fig. 4.2), at a distance close enough to be involved in a hydrogen bond. In the dystrophin and Nedd4 complexes, the hydroxyl group is in the proximity of this carbonyl, but only a different rotamer of T37 (in which the carbon beta and the hydroxyl will change their positions) will allow such a hydrogen bond to occur. Mutations of this threonine in the YAP65 WW domain only allow serine at this position with all other amino acids leading to reduce ligand-binding affinity (Toepert *et al.*, 2001). This fact reinforces the functional role of this residue. Remarkably,

most WW sequences display a S or T at this position (see Fig. 4.1), suggesting a structural relevance for this hydrogen bond (Pires *et al.*, 2001).

For the three above-mentioned structures and especially for the YAP65 WW complex, where systematic substitution analyses on both, peptide and protein sequences, have been performed (Pires *et al.*, 2001; Toepert *et al.*, 2001), replacement of the tyrosine residue (Y7') of the peptide always decreases the binding affinity, even if it is mutated to phenylalanine. The reason could be that the tyrosine hydroxyl group and the histidine (H32) located at the end of the second strand of the WW domain are involved in a hydrogen bond, as observed in the X-ray structure of the dystrophin complex (Kanelis *et al.*, 2001). A similar decrease in affinity is observed when the H32 is mutated to Ala in the YAP65 WW domain (Macias *et al.*, 1996), supporting the idea that such a hydrogen bond is present in all complexes between PPxY motifs and WW domains. This hydrogen bond will be abolished if the tyrosine hydroxyl group gets phosphorylated, being a way of negative regulation of the peptide-protein interaction (Chen & Sudol, 1995; Sudol & Hunter, 2000; Ilsley *et al.*, 2001). Finally, mutations of the prolines in the peptide either reduce the binding affinity or abrogate the binding completely, probably due to losses of the poly-proline type II population of the peptide prior to binding (Chen & Sudol, 1995; Pires *et al.*, 2001; Ilsley *et al.*, 2001).

4.2.2   The Pin1 and C-terminal repeat of the RNA polymerase II complex

A surprise for the WW domain family came from the description of the prolyl-isomerase Pin1 WW (previously included in Group I since its sequence displays the appropriate semi-conserved residues (Macias *et al.*, 2000)), with a phosphorylated peptide that does not contain the PPxY motif (Verdecia *et al.*, 2000). The binding specificity between Pin1WW domain and its target peptide is achieved by interactions between the phosphate group of one phosphorylated serine in the peptide and arginine and serine side-chains of the domain located in the first strand and in the $\beta1$-$\beta2$ loop. Interestingly, the peptide does not adopt the poly-proline type II conformation observed in the other complexes, but binds to the domain in an extended conformation and acquiring a reverse orientation with respect to the other complexes. All WW domains present in the prolyl-isomerase family ranging from yeast to humans, display a high number of conserved positively charged residues and serines in the binding site, all well suited to interact with negatively charged groups such as phospotyrosines/serines. In addition, they exhibit an extra residue in the first loop that folds as a type II β-turn and it allows additional contacts from the peptide to the protein backbone. Since the residues involved in the recognition are conserved in all Pin1 sequences, the interaction with phosphorylated peptides as well as the peptide orientation seem to be a general feature of the prolyl-isomerase family. In fact, this sequence is so far the only WW domain well documented to recognise phosphorylated peptides (Sudol & Hunter, 2000; Sudol *et al.*, 2001). Nedd4 WW domains

were also shown to bind to phosphorylated proteins and peptides, but more work has to be done to document this interaction in vitro and in vivo (Sudol & Hunter, 2000). Some Rsp5 WW domains bind phosphorylated sequences (Chang *et al.*, 2000), though with lower affinity than for the un-phosphorylated sequence. In particular, titration experiments performed on the second WW domain of Rsp5 and the sequence used for the Pin1-peptide complex revealed a six-fold decrease in affinity when compared to the equivalent un-phosphorylated sequence, (Civera, C., Wiesner, S., Chang, A., Sudol, M. and Macias, M.J., unpublished results).

4.2.3   Modeling of aromatic-rich WW sequences

The second group of WW sequences (Macias *et al.*, 2000) contains the formin-binding proteins (FBP) and other related splicing factors such as the U1 snRNP Prp40 protein. Since the cellular role of several of these proteins is well characterized, the targets that interact with the WW domains are also known. All these domains shared the presence of three or more aromatic residues in the binding site, reducing the room left for the peptide to bind. As a consequence of this, most of these WW sequences prefer to bind highly proline-rich peptides without aromatic residues. For instance, the FE65 WW and FBP WW domains 21 and 28 bind a PPLP motif (Chan *et al.*, 1996; Ermekova *et al.*, 1997; Bedford *et al.*, 1997). The transcription factor Npw38 preferentially recognizes a short proline-rich sequence surrounded by an arginine residue, known as the PGR motif (Komuro *et al.*, 1999*a,b*). Although the structure of one representative sequence has been solved, namely that of the FBP28 WW (Macias *et al.*, 2000), other three-dimensional structures of complexes belonging to this group have not been determined so far. We have therefore tried to illustrate the interaction by modeling one complex. However, it is evident that high-resolution structures are required to obtain a detailed description of these interactions. We have chosen to model the Npw38 WW domain because it is the first WW domain shown to interact with sequences containing charged amino acids in its peptide-target. In addition, we tried to understand whether the arginine in the peptide sequence could play a role in domain recognition. Furthermore, the target peptide for the Npw38 WW domain also binds to SH3 domains allowing us to compare the binding site of WW and SH3 domains (Craggs *et al.*, 2001). The dual ability of some peptides to interact with both domain types has also been previously described for the formins (Bedford *et al.*, 1997), but unfortunately three-dimensional structures of complexes showing how two different domains could bind the same ligand are not available.

Considering the high number of prolines, the Npw38 WW target peptide, RLLPPGPPP, has been modeled to adopt a poly-proline type II helical conformation for all residues but the arginine. For the homology modeling of the Npw38 WW domain we used the WW structures of dystrophin, Nedd4, Pin1 and FBP28 as structural templates and the program MODELLER (Sali & Blundell, 1993). The orientation of the peptide in the Npw38 WW complex was then

modeled by superimposing the homology model of the Npw38 WW domain to the dystrophin and Pin1 WW domain, respectively. The Npw38WW:RLLPPGPPP models were then energy minimized using the CHARMm program (Brooks *et al.*, 1983) in QUANTA (Molecular Simulations Inc., 1997). Fig. 4.3 shows the Npw38WW:RLLPPGPPP models with both possible peptide orientations.
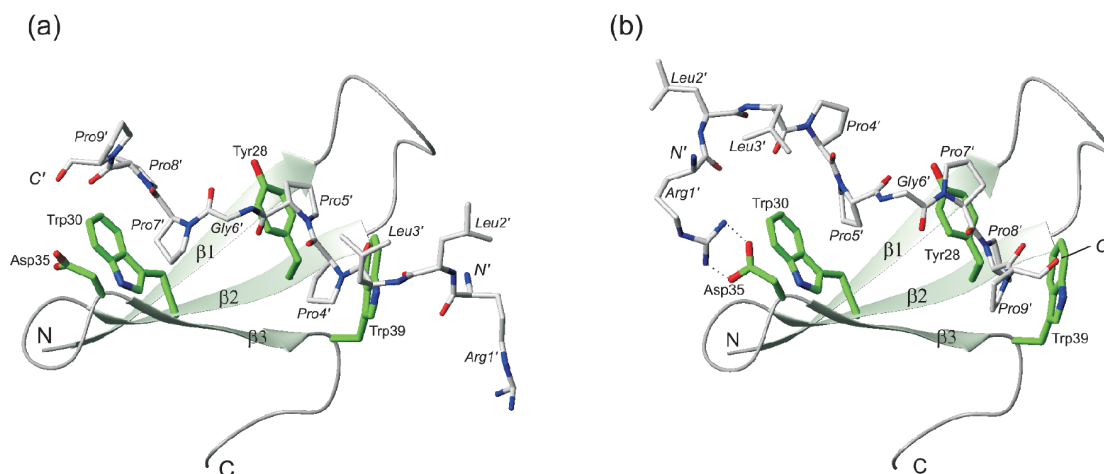
Figure 4.3: Models of the Npw38 WW domain in complex with the nonapeptide RLLPPGPPP. (a) The ligand is oriented as in the dystrophin WW: β-dystroglycan complex. (b) The ligand adopts the same orientation as in the Pin1 WW domain complex with a phosphorylated C-Terminal Domain repeat from the RNA polymerase II. The salt-bridge between the peptide and the WW domain is indicated by dashed lines.

As can be observed in Fig. 4.3, both orientations of the PPGPPP motif result in identical residues interacting with the aromatic rings. This is due to the symmetry of the poly-proline helix and to the presence of two proline-binding pockets on the domain surface. Considering also the N-terminal peptide sequence, RLLP, the peptide orientation as observed in the Pin1 WW complex seems more favorable. The interaction-surface is larger and a salt-bridge can be formed between R1' and D35, while L3' and P5' can contact the W30 ring. In addition, the Gly6' kinks the poly-proline type II helix allowing the two following prolines (P7' and P8') to intercalate in the first binding pocket formed by the Y28 and W39 rings. The opposite orientation, however, only shows contacts from the prolines to the aromatic rings, but no interactions from L2' and R1' to the domain. Therefore, according to our model, the predicted interactions between the R1' and L2' may control the binding orientation and can select a reverse orientation relative to the one observed for the PPxY complexes. Interestingly, by using peptide libraries the Npw38 WW domain has been shown to bind the PPLPP motif, as many formin binding proteins do, and even PPPPPP. This suggests that the interaction between the prolines and both binding pockets is very stable by itself. Hence, flanking residues could be important to define the specificity and thus the final peptide orientation (Zarrinpar & Lim, 2000).

4.3   Comparison between SH3 and WW binding sites: two different scaffolds for binding left-handed poly-proline type II helix ligands

The left-handed poly-proline type II helix, PPII (ϕ =-78°, ψ =+146°), is found in proline-rich peptides although a quarter of the PPII helices found in globular proteins do not contain prolines (Adzhubei & Sternberg, 1993). Proline-rich sequences are frequently found in inter-domain regions of multi-domain proteins and also on the surface of many globular proteins. One reason for the abundance of prolines in proteins is the intrinsic stability of the PPII helix being more resistant to proteolysis. In addition, proline residues have a flat hydrophobic surface well suited for van der Waals interactions with aromatic residues (Brooks *et al.*, 1983; Yu *et al.*, 1994; Musacchio *et al.*, 1994*a,b*). In particular, the stacking of prolines with aromatic rings and the presence of a hydrogen bond between a backbone carbonyl and the indole proton of a tryptophan seem to be conserved for all protein domains that interact with proline-rich peptides, such as SH3, EVH1, profilin, and WW (Zarrinpar & Lim (2000) and references therein). It is remarkable how well the two prolines from β-dystroglycan bound to the dystrophin WW domain and the C4-C7 residues of the Abl-SH3 complex superimpose when the tryptophans of each domain are compared (Huang *et al.*, 2000). In addition, the carbonyl groups of prolines are good hydrogen bond acceptors, increasing the binding affinity by means of backbone hydrogen bonds from the peptide to the protein side-chains.

If the number of consecutive prolines in a peptide is equal or longer than four, the peptide will no longer be extended in solution, but it will adopt a PPII conformation prior to binding to the target protein (Williamson, 1994). As a result of this, the entropic cost of the interaction is highly reduced by the helical conformation and the interaction kinetics is accelerated. One of the most characteristic features of the poly-proline helix, however, is its symmetry. The left-handed PPII has three residues per turn arranged in a three-fold symmetry axis. As previously shown for a peptide with several consecutive prolines (Lim *et al.*, 1994), Pro(i) and Pro(i+4) are pointing to the same side of the helix, in a way that their rings will occupy similar positions in space independently from the peptide orientation. Even more intriguing, the peptide carbonyls of equivalent residues are able to make hydrogen bonds to the same protein residues in both peptide orientations. Thus the presence of the two binding pockets in the SH3 domain allows the peptide to choose between both possible orientations, with the final orientation being defined by interactions between specific residues.

A different scenario, however, is observed in WW domains that interact with the PPxY motif. Since the binding surface contains only a single binding pocket to allocate the proline-rich part of the peptide, the specific interactions with the tyrosine ring (Y7' in Fig. 4.2) and its hydroxyl group define the ligand orientation. Interestingly, the WBP1 (WW domain

binding protein), that interacts with the YAP65 WW domain (Macias *et al.*, 1996), has several copies of the PPxY motif and one YPPPP sequence, raising the question whether the WW domain can also bind the reversed motif. Although the symmetry of the PPII helix will maintain contacts between the Y7' ring and the two prolines in both PPPPY and YPPPP, the second and third proline of the PPPPY motif bind to the aromatic protein-binding site, while for the YPPPP motif it would be the first and the second proline in order to maintain the tyrosine-histidine hydrogen bond. Hence both peptides are not equivalent and so far experimental data only support binding of the PPPY motif and not of the YPPPP motif (Chen *et al.*, 1997)). Remarkably the YPPPPY peptide recovers the binding (Espanel & Sudol, 2001) suggesting that it behaves as the PPxY motif itself. Therefore, it is very unlikely that WW sequences can bind PPxY proline-rich motifs in both orientations. Although these WW domains interact with the proline-rich part in a similar way as SH3, profilin and EHV1 domains do, the specific interactions with the remaining ligand residues are achieved in a very peculiar manner that has only been observed for WW domains so far.

For the PPLPP-binding WW sequences, the situation can be quite different from the one described above. First their peptide-targets include the PxxP motif, previously described as the SH3 binding motif. Moreover, this group of WW sequences has two potential binding pockets formed by the two sets of aromatic clusters, also very similar to the SH3 binding site, and a third interaction site responsible for the specificity (according to our model). If the aromatic residues present in the Sem5 SH3 and in the Npw38 WW domain are superimposed and the peptide bound to the SH3 domain is depicted as in Fig. 4.4(a), both binding sites are very similar. Even more surprising is the fact that not only the residues but also the size of the binding site are conserved. Thus, if our model is correct, we would conclude that Nature has achieved (either by conservation or convergence of the binding sites through evolution) two scaffolds, namely the SH3 and a sub-set of WW sequences, which can accomplish a similar function. However, three-dimensional structures of complexes between WW sequences and their targets as well as SH3 domains in complex with the same peptides are required to fully prove this model of binding site conservation. In addition, these new complexes will help to understand in more detail the role of the variable residues, especially those located in the connecting loops, that regulate the specificity of the interactions between this set of WW domains and proline-rich peptides (Zarrinpar & Lim, 2000).
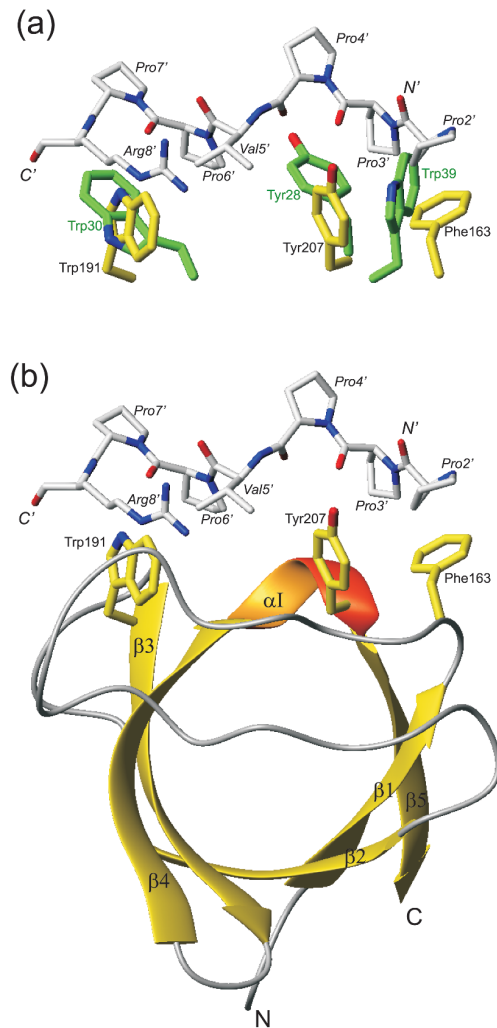
Figure 4.4: Comparison of the binding sites of SH3 and WW domains. (a) Superposition of the side-chains in the binding site of the C-terminal Sem5 SH3 domain (colored in yellow) in complex with the PPPVPPR peptide and the side-chains of the Npw38 WW domain model (shown in green). (b) Ribbon representation of the complexed C-terminal Sem5 SH3 domain in the same scale and orientation.

4.4   Do we need extra interactions from neighboring domains to acquire specificity, to consolidate the WW domain structure or both?

Some WW domains have been found embedded in larger structures (dystrophin and Pin1 WW domains) or requiring the presence of the ligand to increase their stability (YAP65 WW domain). In the dystrophin:β-dystroglycan complex the EF hands and the N-terminal helix present in the protein surround the WW domain and probably maintain it folded, explaining why the independently expressed dystrophin WW domain does not bind β-dystroglycan (Huang *et al.*, 2000). The ligand also interacts with the first EF hand, while the WW domain binds the PPxY motif. Since the β-dystroglycan has two PPxY motifs, the extra interactions

with the EF hand can facilitate the selection of the correct target and increase the overall binding affinity. Hence, the resulting ligand orientation is probably originating from a dual optimization of both the structural stability and of the binding affinity. A similar situation is observed in the Pin1 complex. There, the multiple contacts between the prolyl-isomerase catalytic part and the WW domain are probably important for increasing domain stability. This is supported by the fact that upon ligand binding, the relative orientation of both domains changes in order to enlarge the binding area.

On the other hand, a quite different scenario is observed for the Nedd4 complex. In this case, at least 9 residues from the peptide are involved in the binding and despite the presence of longer tails at both ends of the domain, all contacts between the peptide and the WW domain are made through the standard domain-binding sites. Other examples for independently folding WW domains are the FBP28 WW domain and seven WW domains from *Saccharomyces cerevisiae* (YJQ8 (Macias *et al.*, 2000) and Macias, M.J., unpublished results), which have been shown to be folded and stable in solution without additional contacts to other parts of the protein or ligands, in good agreement with their definition as autonomous units with characteristic structure and function.

As we have seen, some WW domains may require the presence of additional contacts either to improve their structural stability or to increase their binding affinity through cooperativity. Such a situation, nevertheless, should be regarded as exceptional and taken with caution. If protein domains are supposed to be "the building blocks of the protein world" (Dawson & Kent, 2000), "evolutionarily mobile, and thus can occur in otherwise non-homologous proteins but still carrying a certain conserved function" (Doolittle, 1995), they should be capable to function independently. The WW scaffold is probably the shortest protein domain found in Nature, but its binding site has for example the size of an SH3 domain (Fig. 4.4). As suggested for SH3 domains, we cannot exclude that both, SH3 and WW domain, may carry "crude specificity constraints" and that other modular domains present in their interacting proteins will tune the interactions (Cohen *et al.*, 1995). Only more structures of new complexes involving WW domains, together with biochemical experiments will allow us to discern between what is the rule and what are its exceptions.

4.5   References

Adzhubei, A. A. & Sternberg, M. J. (1993). Left-handed polyproline II helices commonly occur in globular proteins. *J. Mol. Biol.* **229**, 472–93.

Bedford, M. T., Chan, D. C. & Leder, P. (1997). FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. *EMBO J.* **16**, 2376–2383.

Brooks, B. R., Bruccoleri, R. E., Olafsson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.

Chan, D. C., Bedford, M. T. & Leder, P. (1996). Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and functionally resemble SH3 domains. *EMBO J.* **15**, 1045–54.

Chang, A., Cheang, S., Espanel, X. & Sudol, M. (2000). Rsp5 WW domains interact directly with the carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **275**, 20562–71.

Chen, H. I., Einbond, A., Kwak, S. J., H., L., Koepf, E., Peterson, S., Kelly, J. W. & Sudol, M. (1997). Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J. Biol. Chem.* **272**, 17070–7.

Chen, H. I. & Sudol, M. (1995). The WW domains of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl. Acad. Sci. USA,* **92**, 7819–7823.

Cohen, G. B., Ren, R. & Baltimore, D. (1995). Modular binding domains in signal transduction proteins. *Cell,* **80**, 237–48.

Craggs, G., Finan, P. M., Lawson, D., Wingfield, J., Perera, T., Gadher, S., Totty, N. F. & Kellie, S. (2001). A nuclear SH3 domain-binding protein that colocalizes with mRNA splicing factors and intermediate filament-containing perinuclear networks. *J. Biol. Chem.* **276**, 30552–60.

Dawson, P. E. & Kent, S. B. (2000). Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **69**, 923–60.

Doolittle, R. F. (1995). The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **64**, 287–314.

Ermekova, K. S., Zambrano, N., Linn, H., Minopoli, G., Gertler, F., Russo, T. & Sudol, M. (1997). The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of Drosophila enabled. *J. Biol. Chem.* **272**, 32869–77.

Espanel, X. & Sudol, M. (2001). Yes-associated protein and p53-binding protein-2 interact through their WW and SH3 domains. *J. Biol. Chem.* **276**, 14514–23.

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.

Ibragimova, G. T. & Wade, R. C. (1999). Stability of the beta-sheet of the WW domain: A molecular dynamics simulation study. *Biophys. J.* **77**, 2191–2198.

Ilsley, J., Sudol, M., Espanel, X. & Winder, S. (2001). The WW domain: Linking cell signalling to the membrane cytoskeleton. *Cell Sign.* **13**, 625–632.

Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. (2001). The folding mechanism of a beta-sheet: the WW domain. *J. Mol. Biol.* **311**, 373–393.

Jiang, X., Kowalski, J. & Kelly, J. W. (2001). Increasing protein stability using a rational approach combining sequence homology and structural alignment: Stabilizing the WW domain. *Prot. Sci.* **10**, 1454–1465.

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.

Koepf, E. K., Petrassi, H. M., Sudol, M. & Kelly, J. W. (1999). WW: An isolated three-stranded antiparallel beta-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* **8**, 841–53.

Komuro, A., Saeki, M. & Kato, S. (1999*a*). Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-rich motif containing glycine and arginine. *J. Biol. Chem.* **274**, 36513–9.

Komuro, A., Saeki, M. & Kato, S. (1999*b*). Npw38, a novel nuclear protein possessing a WW domain capable of activating basal transcription. *Nucleic Acids Res.* **27**, 1957–65.

Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis macromolecular structures. *J. Mol. Graphics,* **14**, 51–55.

Lim, W. A., Fox, R. O. & Richards, F. M. (1994). Stability and peptide binding affinity of an SH3 domain from the Caenorhabditis elegans signaling protein Sem-5. *Protein Sci.* **3**, 1261–6.

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.

Macias, M. J., Wiesner, S. & Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters,* **513**, 30–37.

Musacchio, A., Saraste, M. & Wilmanns, M. (1994*a*). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat. Struct. Biol.* **1** (8), 546–51.

Musacchio, A., Wilmanns, M. & Saraste, M. (1994*b*). Structure and function of the SH3 domain. *Prog. Biophys. Mol. Biol.* **61**, 283–97.

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.

Sali, A. & Blundell, T. (1993). Comparative protein modelling by satisfaction of spacial restraints. *J. Mol. Biol.* **234**, 779–815.

Sudol, M. & Hunter, T. (2000). NeW Wrinkles for an Old Domain. *Cell,* **103** (7), 1001–1004.

Sudol, M., Sliwa, K. & Russo, T. (2001). Functions of WW domains in the nucleus. *FEBS Lett.* **490** (3), 190–195.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL_W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–80.

Toepert, F., Pires, J., Landgraf, C., Oschkinat, H. & Schneider-Mergener, J. (2001). Synthesis of an Array Comprising 837 Variants of the hYAP WW Protein Domain. *Angew. Chem. Int Edit.* **40**, 897–900.

Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. & Noel, J. P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* **7**, 639–643.

Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. *Biochem. J.* **297**, 249–260.

Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W. & Schreiber, S. L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell,* **76**, 933–45.

Zarrinpar, A. & Lim, W. A. (2000). Converging on proline: the mechanism of WW domain peptide recognition. *Nat. Struct. Biol.* **7**, 611–3.