Interfacial water molecules in SH3 interactions: a revised paradigm for polyproline recognition

Jose M. MARTIN-GARCIA1, Javier RUIZ-SANZ and Irene LUQUE2
Department of Physical Chemistry and Institute of Biotechnology, Faculty of Sciences, University of Granada, 18071 Granada, Spain

In spite of its biomedical relevance, polyproline recognition is still not fully understood. The disagreement between the current description of SH3 (Src homology 3) complexes and their thermodynamic behaviour calls for a revision of the SH3-binding paradigm. Recently, Abl-SH3 was demonstrated to recognize its ligands by a dual binding mechanism involving a robust network of water-mediated hydrogen bonds that complements the canonical hydrophobic interactions. The systematic analysis of the SH3 structural database in the present study reveals that this dual binding mode is universal to SH3 domains. Tightly bound buried-interfacial water molecules were found in all SH3 complexes studied mediating the interaction between the peptide ligand and the domain. Moreover, structural waters were also identified in a high percentage of the free SH3 domains. A detailed analysis of the pattern of water-mediated interactions enabled the identification of conserved hydration sites in the polyproline-recognition region and the establishment of relationships between hydration profiles and the sequence of both ligands and SH3 domains. Water-mediated interactions were also systematically observed in WW (protein–protein interaction domain containing two conserved tryptophan residues), UEV (ubiquitin-conjugating enzyme E2 variant) and EVH-1 [Ena/VASP (vasodilator-stimulated phosphoprotein) homology 1] structures. The results of the present study clearly indicate that the current description of polyproline-rich sequence recognition by protein–protein interaction modules is incomplete and insufficient for a correct understanding of these systems. A new binding paradigm is required that includes interfacial water molecules as relevant elements in polyproline recognition.

Key words: binding energetics, polyproline-rich ligand recognition, protein–protein interaction module, SH3 domain, water-mediated interaction.

INTRODUCTION

The recognition of proline-rich sequences by SH3 (Src homology 3) domains is central to the proper functioning of the cell. These domains act as mediators of transient protein–protein interactions in signal transduction cascades and are also implicated in the allosteric regulation of the proteins that contain them [1,2]. Because of their central role in the regulation of cell proliferation and their implication in the development of important diseases, such as cancer, AIDS or inflammatory processes, inhibitors of SH3 interaction have been proposed as potential therapeutic agents [3]. Nonetheless, in spite of the wealth of structural and functional information collected over the last two decades, a full understanding of the binding affinity and specificity of SH3 domains remains elusive.

SH3 domains recognize proline-rich sequences that typically contain the \( \phi Pp\phi P \) motif (where \( \phi \) and \( p \) are frequently hydrophobic and proline residues respectively) [4]. SH3 ligands bind in a PPII (left-handed polyproline II helix) conformation to a flat and hydrophobic surface in the domain consisting of three shallow pockets. Each of the \( \phi P \) moieties packs tightly into one hydrophobic pocket formed by highly conserved aromatic residues on the surface of the domain. Additional interactions, responsible mostly for binding specificity, are established between residues flanking the core motif and a third pocket delimited by the RT and n-Src loops, whose sequences vary among different SH3 domains [5]. The pseudo-symmetry of the PPII conformation allows SH3 ligands to bind in a forward (N-to-C terminal in class I ligands) or reverse (C-to-N terminal in class II ligands) orientation with respect to the binding site. The orientation of the ligand is frequently determined by a basic amino acid located two residues N-terminal from the \( \phi Pp\phi P \) motif in class I ligands and two residues C-terminal from the motif in class II ligands [5,6].

In contrast with this description, which depicts a mostly hydrophobic interaction, all thermodynamic studies of SH3 interactions have revealed a thermodynamic signature (markedly negative binding enthalpies partially compensated by unfavourable entropic contributions) opposed to what would be expected for an interaction driven by the hydrophobic effect [7]. This thermodynamic behaviour, which cannot be easily rationalized in terms of direct interactions between hydrophobic surfaces, reveals an underlying complexity in the recognition of proline-rich ligands by SH3 domains. Additional factors, such as the redistribution of the conformational ensembles of domains and ligands or the modulation of SH3 dynamics, have been proposed to contribute to the observed thermodynamic signature [8–10]. Interfacial water molecules have also been proposed to play a role in the determination of the SH3 ‘anomalous’ thermodynamic signature [11]. In fact, the recognition of peptide ligands by the SH3 domain of the Abl tyrosine kinase was shown to occur via a dual binding mechanism that combines the stacking of hydrophobic surfaces with the establishment of an intricate network of water-mediated hydrogen bonds, mediating the interactions between the peptide ligand and protein residues in the periphery of the canonical hydrophobic binding site. These water-mediated interactions, which defined an extended and more
polar binding interface, were found to contribute substantially to the extremely negative binding enthalpy ($\sim 92 \text{kJ}\cdot\text{mol}^{-1}$) characteristic of the binding of p41 to Abl-SH3 [12]. Whether this dual binding mechanism is a specific feature of Abl-SH3 or a general property of SH3 domains remains to be elucidated.

Considering the wealth of high-quality structural information available for SH3 domains, the present study provides a detailed analysis of the SH3 structural database aimed at investigating the relevance of water-mediated interactions in proline-rich ligand recognition by SH3 domains. Our analysis revealed that buried well-ordered water molecules are systematically found at the binding interface in SH3 complexes, mediating peptide recognition by the establishment of several interactions with both protein and ligand. Different hydration patterns have been identified and their sequence and structure determinants investigated. Additionally, tightly bound structural water molecules were also frequently found at highly conserved hydration sites in free SH3 domains. These water molecules, generally conserved in SH3 complexes, seem to be an integral part of the domain and should be included for a complete definition of the SH3-binding site for docking or structure-based design. On the basis of these results we propose a revision of the current binding paradigm for proline-rich ligand recognition by SH3 domains, as the current model is incomplete and clearly insufficient to fully understand these interactions. Analysis of the structural information available for other proline-rich recognition modules indicates that the SH3 dual binding mechanism seems to be extensive to the WW (protein–protein interaction domain containing two conserved tryptophan residues), EVH-1 [Ena/VASP (vasodilator-stimulated phosphoprotein) homology 1] domains. These results reveal a universal role for interfacial water molecules in the recognition of proline-rich sequences by protein–protein interaction modules, with important implications for binding specificity and rational design.

**EXPERIMENTAL**

A total of 98 crystal structures of SH3 domains solved at a resolution higher than 2.6 Å (1 Å = 0.1 nm) were selected from the PDB and analysed for the presence of interfacial water molecules. To account for possible deviations in the solvation patterns due to poor resolution, a parallel analysis of a more restrictive set of 71 structures, with resolutions below 2.0 Å, was also conducted.

The presence of interfacial water molecules was investigated for all selected structures. For SH3 complexes, buried water molecules mediating the interactions between the protein domain and the ligand were selected. These criteria were shown previously to reasonably approach the situation in solution [13]. Accordingly, water molecules simultaneously located within 7 Å of the ligand and the SH3 domain were identified, and their solvent ASA (accessible surface area) was calculated using the Lee and Richards algorithm [14] as described previously [11]. All water molecules with ASA values greater than 20 Å² were discarded and removed from the structure. ASAs were recalculated, and the set of occluded water molecules was iteratively refined until no further changes in accessibility were observed. Similarly, for the free SH3 domains, all water molecules within 5 Å of any protein residue and ASA values less than 20 Å² were selected for analysis.

For all selected water molecules, the interactions established by the identified water molecules with the protein or ligand were evaluated considering a maximum cut-off distance of 3.5 Å and a minimum angle of 90° for hydrogen bond formation. B factors were also analysed in order to establish the degree of mobility of the selected water molecules with respect to the protein atoms.

To facilitate the comparison between different crystal structures, B factors were normalized following a procedure described previously [15], to have a distribution of zero mean and unit variance, according to:

$$B_{\text{norm}} = \frac{B - \langle B \rangle}{\text{sd}(B)}$$

where (B) and sd(B) are the mean value and the S.D. of the distribution of B factors for protein atoms within each crystal structure respectively. According to this normalization, a $B_{\text{norm}}$ value of 0 would indicate a water molecule with the average mobility of the protein atoms, and positive and negative values will correspond to higher or lower mobility values with respect to the protein average.

**RESULTS AND DISCUSSION**

**Identification of interfacial water molecules in the SH3 structural database**

For the analysis of interfacial water molecules, two structural SH3 databases were generated: a database containing a total of 98 structures of SH3 domains with resolutions below 2.6 Å and a more restrictive database of 71 SH3 structures with resolutions below 2.0 Å. The identities and PDB codes of the structures in both databases are shown in Supplementary Table S1 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm). As summarized in Supplementary Table S2 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm), these structures correspond to a total of 44 different SH3 domains contained in proteins of diverse functions and architectures from several organisms. These include free domains as well as complexes with small peptide ligands from natural targets, rationally designed peptides, full-length protein partners and intramolecular interactions. Thus this set of structures can be considered a representative selection, including a wide spectrum of SH3 domains with different binding modes and specificities, from which general conclusions about the role of water molecules in proline-rich ligand recognition can be drawn.

For all of the selected structures, the presence of interfacial water molecules, i.e. buried water molecules mediating the interactions between the protein and the ligand at the binding interface, was investigated (see the Experimental section for details). Strikingly, interfacial water molecules were invariably found in all 37 SH3 complexes studied, occupying, as shown in Figure 1(A), different hydration sites throughout the binding interface. Tightly bound water molecules were also found at the binding site of more than 90% of free SH3 domains. A total of 263 interfacial water molecules were identified in the 98 analysed structures (an average of 2.7 water molecules per structure). The overall properties of these water molecules are shown in Figure 1(B). In general terms, most identified water molecules are significantly buried from the solvent ($\text{ASA}_{\text{average}} = 9.2 \pm 6.8 \text{Å}²$), with more than 30% of them being characterized by ASA values below 5 Å². These molecules are also characterized by small B factors ($B_{\text{norm,average}} = 0.6 \pm 2.1$), similar to the average values for protein atoms, and, in some cases, close to the structure minimum. Nonetheless, as will be discussed later, the mobility of water molecules was found to vary significantly between the different hydration sites. Finally, these molecules are implicated in the establishment of multiple specific polar interactions with the protein and/or the ligand, with an average of 3.3 ± 1.0 hydrogen bonds per water molecule. These values are very similar to those derived for interfacial water molecules ($\text{ASA}_{\text{average}} = 10.4 \pm 12.1$).
A Revised paradigm for polyproline recognition by protein interaction models

Figure 1  Interfacial water molecules in SH3 complexes

(A) Cartoon representation of 28 superposed structures of SH3 complexes. SH3 domains are coloured in a rainbow scheme (blue, N-terminus, and red, C-terminus), whereas ligands are represented in light pink. Buried interfacial water molecules are shown as purple non-bonded spheres. (B) Properties of the 263 water molecules identified at the SH3-binding interface. Columns indicate the number of identified water molecules with a given range of values of normalized B factors, solvent accessibility and number of hydrogen bonds.

Å², $B_{\text{norm,average}} = 0.4 \pm 1.6$ and an average of three interactions) from the statistical analysis of a dataset of 392 high-resolution protein–ligand complexes covering a wide range of protein–ligand-binding sites with different shapes and chemical properties [15], and significantly different from those determined in the same study for surface water molecules, which are characterized by higher ASA and $B_{\text{norm}}$ values (ASA$_{\text{average}} = 38.8 \pm 14.6$ Å² and $B_{\text{norm,average}} = 1.6 \pm 1.1$) and an average of one hydrogen bond per water molecule [15–17]. In summary, tightly bound water molecules are systematically found at the SH3-binding site in both SH3 complexes and free domains, indicating that the Abl-SH3 dual binding mechanism, according to which the stacking of hydrophobic surfaces is complemented by a network of water-mediated hydrogen bonds, is indeed of general applicability to proline-rich ligand recognition by SH3 domains.

As shown in Figure 1(A), interfacial water molecules were found at different hydration sites throughout the binding interface, mediating the interactions between the ligand and SH3 residues in the periphery of the canonical binding site. Some of these hydration sites, especially those at the polyproline recognition region, are highly conserved among the different SH3 complexes, even though a higher variability in water configurations was observed in the vicinity of the n-Src and RT loops. In any case, as described for Abl-SH3 complexes [11,12], water-mediated interactions were established between the ligand and residues in the periphery of the canonical binding site, different from those defining the hydrophobic pockets for proline recognition. This highlights the importance of considering interfacial water molecules for a correct definition of the complete binding interface.

The presence of interfacial water molecules provides a second level of interaction, in addition to the stacking of hydrophobic residues, that results in extended and more polar interaction surfaces, in better agreement with the markedly favourable binding enthalpies characteristic of SH3 interactions [7]. Thus, together with other effects associated to the redistribution of the protein and ligand conformational ensembles upon binding described previously [8,10], the analysis in the present study clearly reveals the presence of interfacial water molecules as a universal factor contributing to the ‘anomalous’ thermodynamic signature for proline-rich ligand recognition by SH3 domains. In this context, the current description of SH3 complexes, based primarily on the establishment of direct, mostly hydrophobic, interactions between the ligand and the SH3 domain appears insufficient for a complete understanding of these systems, and needs to be replaced by a new binding paradigm that incorporates interfacial water molecules as relevant elements in the mechanism of polyproline recognition by SH3 domains.

Structural water molecules in SH3 domains

Although the presence of most water molecules appeared to be associated to ligand binding, buried water molecules were observed in 93% of the free SH3 structures, indicating that most SH3 domains contain structural water molecules at the binding site, of relevance for ligand recognition. Two highly conserved hydration sites were identified in free SH3 domains that were also occupied in many of the SH3 complex structures: one located at the base of the $3_{10}$ helix and the second at the n-Src loop.

Structural water molecules at the $3_{10}$ helix

In previous studies, the Abl-SH3 domain was shown to contain a structural water molecule located at the base of the $3_{10}$ helix.
helix, present in all high-quality crystal structures of this domain independently of the ligation state. Molecular dynamics simulations had revealed that this tightly bound water molecule, characterized by residence times over 1 ns, was coordinated by backbone atoms of the residues in the \( \beta_0 \) helix region and, less frequently, by the side-chain atoms of a highly conserved asparagine residue at position 114, two positions C-terminal from the invariable proline of the \( \beta_0 \) helix [12]. On the basis of these results it was hypothesized that equivalent water molecules would probably be observed in other SH3 domains. Analysis of the SH3 structural database indicates that this is indeed the case. As illustrated in Figures 2(A) and 2(B), the \( \beta_0 \) helix hydration site was occupied in 77% of the SH3 structures with resolutions below 2.0 Å, with similar values for free domains (74%) and complexes (81%). Equivalent results (73.3% and 72.5%) were obtained for the 2.6 Å database. The identity and properties of water molecules at this site are summarized in Supplementary Tables S3, S4 and S5 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm) for free structures, peptide complexes and intramolecular interactions respectively. As is characteristic of highly immobilized molecules, water molecules at this site are generally characterized by low solvent accessibilities (\( \text{ASA}_{\text{average}} = 10.9 \pm 7.3 \text{ Å}^2 \)) and low crystallographic B factors (\( B_{\text{norm,average}} = -0.2 \pm 1.7 \)), close to the protein average, that are of similar magnitude in free domains and in complex structures. It is interesting to point out that, as reflected in Supplementary Tables S3, S4 and S5, in many cases the B factors of water molecules at this site are close to the minimum values in the structure. It has been reported in the literature that these extremely low B factors are frequently associated with the establishment of a high number of specific polar interactions. Indeed as summarized in Supplementary Tables S6 and S7 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm), water molecules at this site are generally implicated in at least three highly optimized hydrogen bonds with protein atoms in one of the most stable regions of the SH3 domain [18], maintaining a hydrogen bond configuration very similar to that described previously for Abl-SH3 structures [11,12]. From these data, it can be concluded that the water molecule at the base of the \( \beta_0 \) helix is an invariant element in most SH3 structures that seems to be an integral part of the domain structure. It is notable that, in most cases, the absence of this structural water in free SH3 domains is associated with the presence of elements that block the hydration site (residues with bulky side chains, such as proline or valine, at position 114 or anomalous conformations of SH3 loops) or with very short side chains (glycine or alanine) in the \( \beta_0 \) helix (see Supplementary Figure S1 at http://www.BiochemJ.org/bj/442/bj4420443add.htm). The latter would significantly increase the solvent accessibility and the mobility of the water molecule, hindering its detection by X-ray crystallography. Even though this structural water is maintained in most complexes, in some rare cases it is displaced by residues in the ligand. This displacement is also generally associated with anomalous non-PPII conformations in the proline-rich region.

In summary, the structural water molecule at the base of the \( \beta_0 \) helix, present in most SH3 domains independently of its ligation state, should be incorporated into binding site descriptions for docking or rational design. This tightly bound water may constitute a good anchor site against which to design new interactions with the ligand, with minimal enthalpy/entropy compensation effects and, thus, significant impact on the binding affinity [19,20].
Structural water molecules at the n-Src loop

A second structural water molecule was observed in a high percentage of free domains and complexes (59% and 55% respectively) at the n-Src loop region (see Figures 2C and 2D). The identity and properties of water molecules at this site are shown in Supplementary Tables S8 and S9 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm). In summary, waters at this site are fully buried (ASA average $= 10.4 \pm 12.1$ Å$^2$) and characterized by low B factors ($B_{w,\text{average}} = -0.3 \pm 1.3$), in some cases well below the average for protein atoms, both in complexes and in free SH3 domains. Although not directly involved in the recognition of the ligand, these water molecules mediate the interactions between residues within the n-Src loop, establishing an average of $4.1 \pm 1.0$ hydrogen bonds (see Supplementary Tables S10 and S11 at http://www.BiochemJ.org/bj/442/bj4420443add.htm) for free domains and complexes respectively and, thus, stabilizing its conformation. In fact, the presence of this water molecule is related to the length of the n-Src loop, being found in 85% of the structures of SH3 domains characterized by n-Src loops with lengths of between 5 and 8 residues and absent in domains with shorter or longer loops. In the short loops there is insufficient space for a water molecule, whereas the longer ones are stabilized by direct contacts between loop residues. A stabilizing role for buried water molecules has been described previously for different types of protein loops, such as the reactive loop in chymotrypsin [21], hairpin structures in lecints [22], twisted $\beta$-turns in MHC class-I molecules [23], or the $\Omega$-loop in class-I $\beta$-lactamases [24,25]. Albeit not directly implicated in the interactions with the ligand, this structural water molecule at the n-Src loop may play an important role in SH3 binding by regulating the dynamic properties of the n-Src loop, known to be important for the conformational equilibrium of SH3 domains and, thus, for binding specificity [7,26].

Hydration patterns in SH3 complexes

In addition to the structural water molecules at the 3$_{10}$ helix and n-Src loop regions, other water molecules were consistently found at the binding interface in SH3 complexes originating highly variable hydration patterns.

Polyproline-recognition region

A third conserved hydration site was identified, occupied by buried water molecules that mediate the interactions between the proline-rich region of the ligand and residues at the 3$_{10}$ helix of the domain. In previous studies, Abl-SH3 complexes were shown to be characterized by the presence of one water molecule in this region, coordinated almost exclusively by the carbonyl oxygen of proline residues in the ligand [11,12]. This hydration site was not occupied in the free Abl-SH3 domain. In our structural database, water molecules were identified in this region in approximately one third of the complex structures (32% and 28% for the 2.0 Å and 2.6 Å databases respectively), whereas no water molecules were found in this region for the free domains. The identity and properties of these water molecules are summarized in Supplementary Tables S4 and S5 for the inter- and intra-molecular complexes respectively. On average, water molecules at this site are characterized by a low solvent accessibility (ASA average $= 6.5 \pm 0.4$ Å$^2$) and B factors close to the protein average ($B_{w,\text{average}} = 0.2 \pm 1.4$), but higher than those of structural water molecules at the 3$_{10}$ helix and n-Src loop. Water molecules at this site are implicated in an average of $3.1 \pm 1.0$ hydrogen bonds with atoms from both protein and ligand as well as with the structural water molecules at the base of the 3$_{10}$ helix (see Supplementary Tables S12 and S13 at http://www.BiochemJ.org/bj/442/bj4420443add.htm). As illustrated in Figure 3(A), a significant variability exists in the position of this third hydration site when compared with the highly conserved site at the base of the 3$_{10}$ helix. Nonetheless, the pattern of interactions is very similar in all complexes (see Supplementary Table S12). A detailed inspection of the structures reveals that the presence of this third water molecule is determined primarily by two factors: the length of the side chain at position 114 (Abl numbering) in the 3$_{10}$ helix and the nature of the central amino acids in the $\phi_3$P$_1$$\psi_3$$\phi_3$P$_3$ core motif of the ligand. As reflected in Supplementary Table S14 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm), most SH3 complexes contain an asparagine at position 114 and a minimum of one non-proline amino acid in the central positions of the core motif. In these cases, as illustrated in Figure 3(B), the carbonyl oxygen of the residue at position $p_1$ in class I ligands or $\phi_3$ in class II ligands is at an adequate distance to establish a direct hydrogen bond with the Asn$_{114}$ side chain (3.0 $\pm$ 0.2 Å on average). This interaction has been described previously to support a highly conserved hydrogen bond between the $\phi_3$ residue in the core motif and the side chain of an adjacent tyrosine residue (Tyr$_{116}$ in Abl-SH3) [5]. Consequently, for these complexes, only the structural water molecule at the base of the 3$_{10}$ helix is observed. The presence of additional water molecules in the polyproline-recognition region is invariably associated with an increase in the distance between the side-chain atoms of the residue at position 114 and the peptide ligand. In the database, this occurs when: (i) a shorter amino acid, such as serine or threonine, is found at position 114 in the SH3 domain, precluding the establishment of a direct hydrogen bond with the carbonyl oxygen of residue $p_1$ in the ligand. In this case, two additional water molecules are consistently observed (see Figure 3C); (ii) the two central positions of the core motif ($\phi_3$$\psi_3$) are occupied by proline residues, imposing a rigid PPII conformation that displaces the carbonyl oxygen at $p_1$ and impedes the establishment of a direct hydrogen bond with Asn$_{114}$ (Figure 3D). This is the case for the Abl-SH3–p41 complex; (iii) the ligand presents non-canonical sequences, such as RxxK, and does not adopt a PPII conformation. As illustrated in Figure 3(E), the hydration pattern in this case is highly variable, going from the Abl-SH3 configuration to more complex combinations implicating three or more interfacial waters.

The present study provides a rationalization for the puzzling thermodynamic results obtained for the binding of a set of p41-related peptides to Abl-SH3 [11]. The high-affinity p41 ligand (APSYSPPPPP) was developed by rational design from the natural sequence 3BP1 by substituting leucine and serine residues at solvent-exposed positions by proline, in order to pre-shape the ligand to the PPII conformation and, thus, minimize the conformational entropic penalty upon binding [27]. Nonetheless, the thermodynamic analysis revealed that the substitution of a leucine residue at position 8 by proline (APSYSPLPP$\rightarrow$APSYSPPPPP) is entropically unfavourable and that, unexpectedly, the associated gain in binding affinity originated from a more favourable enthalpic contribution. The analysis in the present study provides additional support to the initial hypothesis relating this thermodynamic behaviour to the modulation of the water-mediated interaction network, since, as discussed above, the introduction of proline residues in the core motif statistically favours the occupation of the polyproline hydration site. The presence of this additional water molecule would probably translate into more negative binding enthalpies and less favourable entropic contributions. In this context, it becomes apparent that discounting interfacial water molecules...
Figure 3  Variability in hydration patterns at the polyproline-recognition region

Cartoon representation of the SH3 complex structures. SH3 domains are coloured in a rainbow scheme (blue, N-terminus, and red, C-terminus), ligands are shown in light pink and buried interfacial water molecules are depicted as purple non-bonded spheres. Hydrogen bonds are represented as broken black lines. The average distance between side chain atoms at position 114 at the 310 helix and the carbonyl oxygen of the residue at position p3 in the ligand is shown (d). (A) Interfacial water molecules at the polyproline-recognition region in all SH3 complexes. (B) Complexes with an asparagine residue at position 114 and one non-proline residue in the central position of the φ1P2p3φ4P5 core motif in the ligand. (C) Complexes with short side chains at position 114. (D) Complexes with an asparagine residue at position 114 and two proline residues at positions p3 and φ4 in the core motif. (E) Complexes with non-canonical RxxK binding motifs. May lead to serious problems in structure-based design with these systems.

Specificity region

In addition to the conserved hydration site at the polyproline-recognition region, interfacial water molecules were also observed with high frequency (85%) mediating the interactions between the RT and n-Src loops and the specificity region of the ligand. The hydration pattern in this region varied greatly among the different complexes, such that interfacial waters were found in different numbers and in a wide variety of configurations, apparently very dependent on the sequence of both ligands and loops. In general, water molecules at the specificity region are characterized by higher B factors than those found at the polyproline-recognition region, with average values of 1.3 ± 2.1 and 1.4 ± 0.9 for the normalized B factors (Bnorm,average) for the n-Src and RT loops respectively. Nonetheless, these water molecules are still fully buried (ASAaverage,n-Src = 1.0 ± 0.7 Å² and ASAaverage,RT = 1.2 ± 0.6 Å²) and implicated in an average of 2.9 ± 1.0 hydrogen bonds with atoms in the protein and the ligand (see Supplementary Table S15 at http://www.BiochemJ.org/bj/442/bj4420443add.htm). Consequently, even though waters at the specificity region are more mobile, probably due to the higher flexibility of the protein regions that coordinate them, their properties are still significantly different from those of surface waters.

Although it is difficult to establish specific patterns of hydration in these regions, water molecules in the RT loop frequently assist the interactions between charged residues in the peptide ligand and the protein. This explains why no such positions are occupied in the Abl-SH3 complexes, the ligands of which lack charged residues. On the other hand, in the Abl-SH3–p41 complex, a cluster of interfacial water molecules stabilizes the n-Src loop in an anomalous conformation, closing it tightly upon the ligand. Although the conformation of the n-Src loop in most complexes is more opened, interfacial water molecules were also frequently found in this region. The variety of hydration patterns observed at the n-Src and RT loops underlines the relevance of water-mediated interactions in SH3 recognition. Interfacial waters appear to act as adaptors that facilitate the recognition of different sequences, contributing to the versatility and promiscuity of these
revised paradigm for polyproline recognition by protein interaction models

449

Figure 4 Water-mediated interaction patterns in WW and UEV domains

Cartoon representation of WW and UEV complex structures. WW and UEV domains are shown in white, ligands are depicted as marine blue sticks and relevant protein residues are represented as cyan sticks. Interfacial waters are shown as non-bonded spheres (waters in complex structures are coloured in blue tones, whereas waters in free structures are coloured in red, orange, yellow and purple). Hydrogen bonds are depicted as broken lines following the same colour scheme. (A) Class IV Pin1-WW domain, free (PDB codes 1PIN, 2Q5A, 2F21, 2ITK and 1ZCN [37–39]) and in complex with a phosphoserine peptide (PDB code 1F8A [40]). (B) Class I dystrophin-WW domain, free (PDB code 1EG3 [41]) and in complex with a β-dystroglycan peptide (PDB code 1EG4 [41]). (C) Class II FE65-WW domain, free (PDB code 2IDH [31]) and in complex with proline-rich peptide from Mena (PDB code 2OEI [31]). (D) Tsg101-UEV domain, free (PDB codes 3OBS and 2FOR [30,42]) and in complex with proline-rich viral late domain sequences (PDB codes 3OBQ, 3OBU and 3OBX [30]). Hydrogen bond patterns common to most structures (including the free domains) are shown as broken marine blue lines, whereas interactions specific to particular structures are shown in their respective colours.

It is interesting to point out that, in spite of the variability among different complex structures, no significant differences have been found between the hydration patterns observed for class I and class II ligands. Conserved water molecules in the polyproline-recognition region were equally present in complexes with both classes of ligands and a similar variability in water configurations at the specificity region was observed in both cases. This indicates that interfacial water molecules do not seem to play a significant role in determining peptide binding orientation.

Water-mediated interactions in proline-rich ligand recognition by other protein–protein interaction modules

In addition to SH3 domains, other polyproline recognition modules (WW, UEV, EVH1 and GYF domains) that share some features of the SH3-binding mode have been identified [5]. These domains present also a common thermodynamic signature for ligand recognition, characterized by negative binding enthalpies partially compensated for by unfavourable entropic contributions [30–32]. This raises the question about the generality of the SH3 dual binding mechanism for proline-rich ligand recognition by other protein–protein interaction modules. To investigate the relevance of water-mediated interactions in these systems, our analysis was extended to WW, UEV and EVH-1 domains.

A total of thirteen structures of five different WW domains were analysed. The identity and properties of the observed binding site water molecules are summarized in Supplementary Table S16 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm). Interfacial water molecules mediating the interaction between the peptide ligand and the WW domain were found in three of the four different complex structures available. Interestingly these three domains [dystrophin-WW, FE65-WW and PIN1 (peptidyl-prolyl cis/trans isomerase 1)-WW] belong to different specificity classes and, thus, recognize different core motifs in their ligands with different binding modes. The class I WW domain of dystrophin binds to ligands with the PPxY motif that fit into an SH3-like hydrophobic φP pocket and a tyrosine-binding pocket. As illustrated in Figure 4B, binding site water molecules in the φP pocket of the free protein are displaced upon binding of the ligand, whereas new interfacial waters are incorporated in the complex. A similar situation occurs in the complex of class II FE65-WW (see Figure 4C) that has two φP pockets in its binding site and recognizes PPLP motifs, very similar to SH3 ligands. In both cases, a tyrosine residue constituting one of the hydrophobic φP pockets is also implicated in the coordination of a water molecule that mediates its interaction with the carbonyl oxygen of a proline residue in the ligand. The case of class IV Pin1-WW, which recognizes phosphorylated serine or threonine residues in ligands with the phospho(S/T) motif, is particularly interesting. In this system, in addition to other interfacial water molecules in

© The Authors Journal compilation © 2012 Biochemical Society
the complex, there seems to be a structural water molecule, reminiscent of that at the 3_10 helix in SH3 domains. This water, coordinated by residues in the third β strand, is present in most structures of the free domain and remains in the complex structure mediating the interactions with a tyrosine side chain in the ligand (see Figure 4A). An equivalent water molecule is also found in the structure of the PIN1 homologue ESS1 (PDB code 1YW5 [33]).

As summarized in Supplementary Table S17 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm), the structural information for UEV domains is limited to six structures of Tsg101 (tumour susceptibility gene 101)-UEV and a close homologue Vps (vacuolar protein sorting)-UEV. As shown in Figure 4(D), several binding site water molecules, observed in the structures of the free Tsg101-UEV domain, are conserved in the complexes and implicated in the recognition of the peptide ligand. Additionally, as shown in Supplementary Figure S2 (at http://www.BiochemJ.org/bj/442/bj4420443add.htm), the structures of free EVH-1 domains are characterized by two conserved hydration positions at the interaction site for LIN domains, although no water molecules are found at the polyproline-binding region (see Supplementary Table S18 at http://www.BiochemJ.org/bj/442/bj4420443add.htm). Nonetheless, in one of the two complex structures of Mena-EVH-1, one water molecule is found mediating the interaction with the peptide ligand. No interfacial water molecules were observed in the crystal structures available for the SMY2 suppressor of MYO2 (myosin-2)-66 protein-GYF domain (PDB codes 3K3V and 3FMA [34]).

In summary, even though the structural information is too scarce to derive any conclusions for EVH-1 and GYF domains, these results seem to point towards an SH3-like dual binding mechanism for the recognition of proline-rich ligands by both WW and UEV domains. The actual relevance of this second level of interactions in all families of polyproline-recognition domains will need to be confirmed as the structural database increases. Although limited, the information currently available clearly indicates that the tendency of proline-rich sequences to be highly hydrated, illustrated by the solid and extensive hydration structure of collagen triple helices [35,36], has been exploited by nature to increase the adaptability and plasticity of the different families of protein modules for the recognition of their proline-rich targets. In the light of this evidence, a revision of the binding paradigm for polyproline recognition is necessary.

AUTHOR CONTRIBUTION
Irene Luque and Javier Ruiz-Sanz designed the research, José Martín-García performed the experiments; Irene Luque, Javier Ruiz-Sanz and Jose Martín-García analysed the data; and Irene Luque wrote the paper.

ACKNOWLEDGEMENTS
We thank Andrés Palencia and Ana Cámara-Artigas for many helpful discussions and J.C. Martin-García for critically reading the manuscript prior to submission.

FUNDING
This work was supported by the Spanish Ministry of Science and Technology [grant numbers BIO2008-0517-C02-01 and BIO2009-13261-C02-01], FEDER (Fondo Europeo de Desarrollo Regional) Funds and the Andalusian Government [grant number CVI-5915]. J.M.M.-G. was supported by a predoctoral research contract from the Spanish Ministry of Science and Technology.

REFERENCES
Revised paradigm for polyproline recognition by protein interaction models


Received 20 June 2011/23 November 2011; accepted 24 November 2011
Published as BJ Immediate Publication 24 November 2011, doi:10.1042/BJ201111089