4 Structure of the human Spred2 EVH1 domain

Sequence alignment of the human Spred2 EVH1 domain with other EVH1 domains clearly identifies it as belonging to this family of domains, while at the same time showing some striking differences (Figure 33). The residues forming the canonical binding site in Class I and II EVH1 domains are highlighted with arrows.

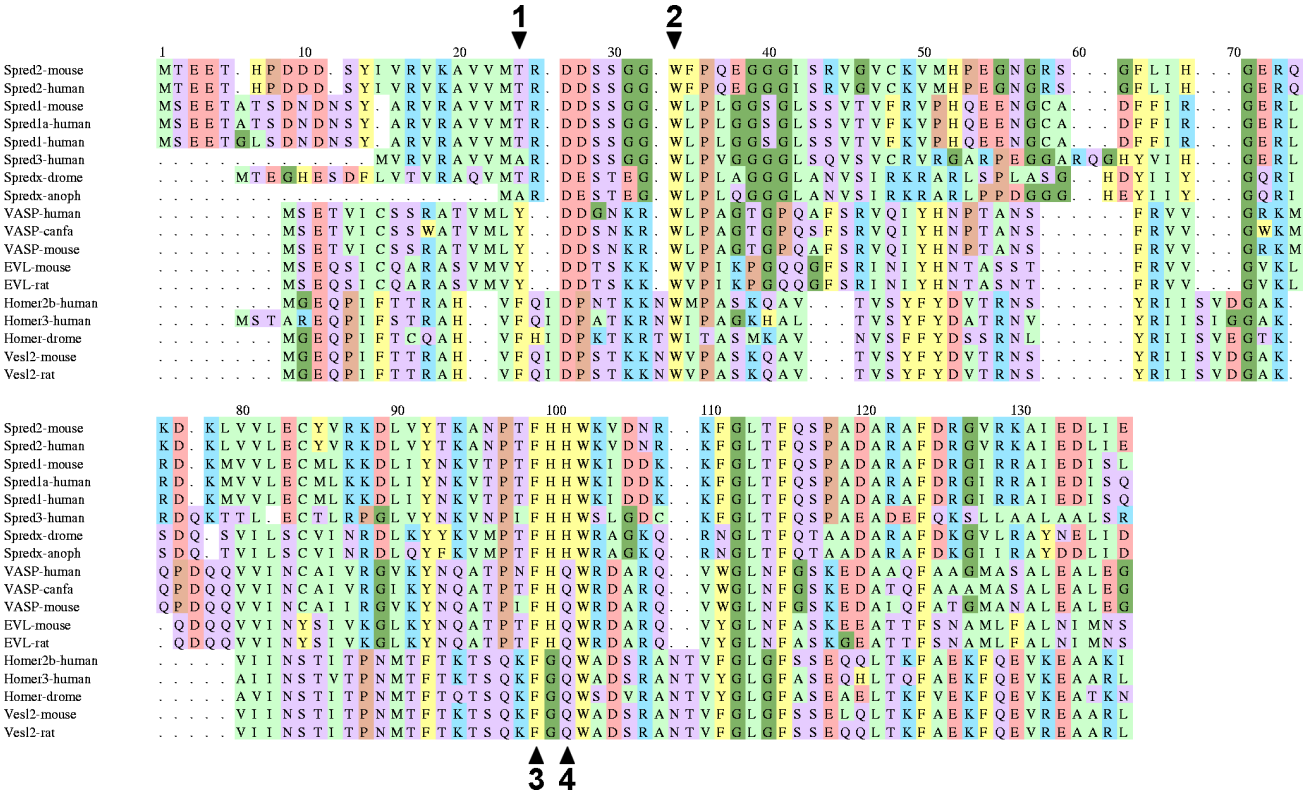


Figure 33
Alignment of the known Spred EVH1 domains with selected sequences of class I and class II EVH1 domains. Residues forming the canonical EVH1 binding epitope are numbered 1-4 and marked with arrows (See also Table 2, p. 68). Data base accession codes are: Spred2-mouse Q924S7, Spred2-human AAP49415, Spred1-mouse Q924S8, Spred1a-human Q8N256, Spred1-human AAP59414, Spred3-human sequence kindly provided by Prof. Yoshimura, Kyushu University, Japan, see also ref. [24], Spred4-drome Q9V756, Spred4-anoph EAA00355, VASP-human P50552, VASP-canfa P50551, VASP-mouse P70460, EVL-mouse P70429, EVL-rat O08719, Homer2b-human O95349, Homer3-human O95350, Homer-drome O96607, Vesl2-mouse O89025, Vesl2-rat O88802.
The alignment was performed manually. Amino acids are coloured as: Red – acidic; blue – basic; purple – polar; green – hydrophobic; yellow – aromatic; brown – Pro.

Three Spred proteins have been described in the literature, Spred1 (mouse), Spred2 (mouse), Spred3 (human) [22, 24]. Five other homologues were then identified via a BLAST search in SwissProt/Trembl. All of them bear an N-terminal EVH1 domain, shown in the alignment. For comparison, several class I and class II EVH1 domains have been included.

The Spred proteins themselves can be divided into four separate subfamilies so far: Spred1, Spred2, Spred3 and the domains of Drosophila and Anopheles. Strikingly, human and mouse Spred2 EVH1 domains are identical, as are the human Spred1a and mouse Spred1 EVH1 domains. The human Spred1 EVH1 domain is also very closely related to these, showing two conservative mutations in comparison with the other two domains. The human Spred3 EVH1 domain shows several much less conservative mutations and an insertion. The two Spred4 EVH1 domains, so far only found in Drosophila and Anopheles, differ in a number of non-conservative mutations from the other Spred proteins.

4.1 Resonance assignment

The Spred2 EVH1 domain was expressed as a GST-fusion protein in E. coli and purified from the soluble fraction of the cell lysate. After cleavage of the GST moiety the EVH1 domain was stable and folded.

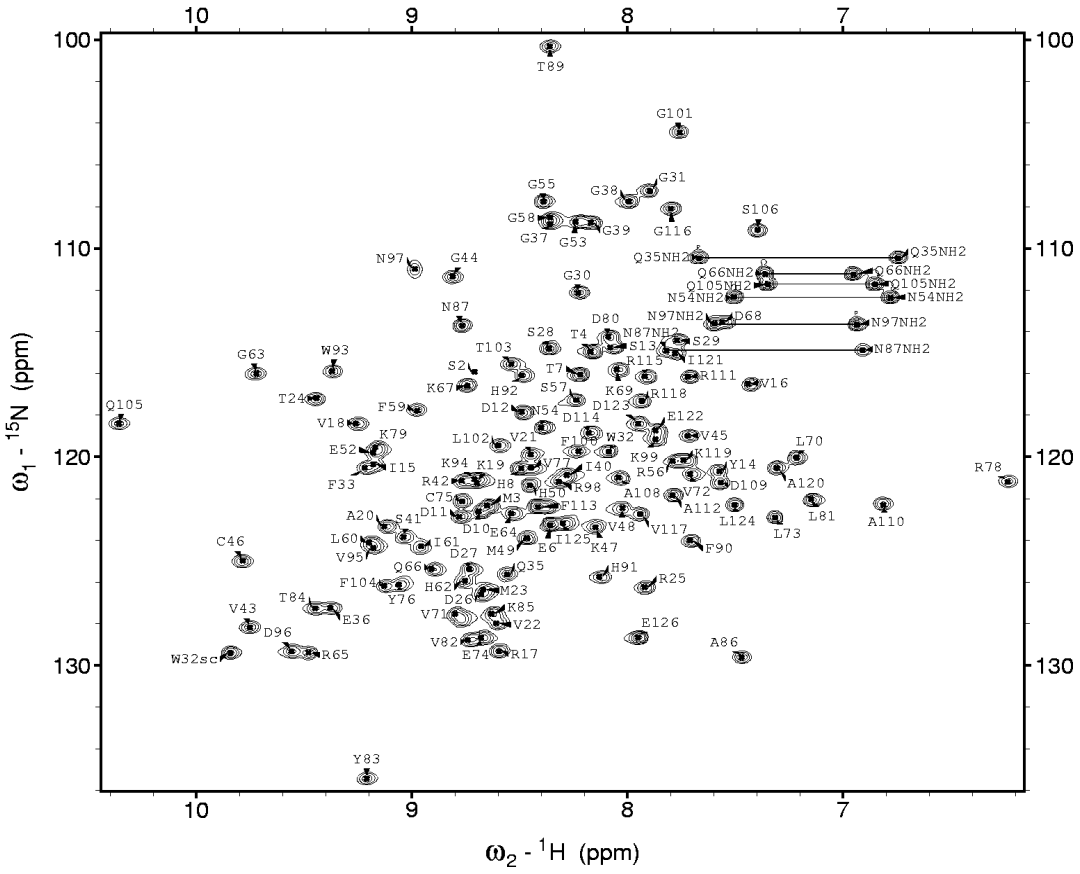


Figure 34
^{1}H-^{15}N-HSQC spectrum of Spred2, recorded at 300 K on a Bruker DRX 600 spectrometer. The protein concentration was 1.0 mM.

Figure 34 shows the ^{15}N-HSQC of the Spred2 EVH1 domain. The resonaces are well resolved, most striking are the unusual ^{15}N-shifts of Thr89 and Tyr83. The highfield shift of Thr89 may be rationalized based on the orientation of the aromatic ring of Phe 90 in the 3D-structure, which causes shielding of the backbone nitrogen atom of Thr 89. The lowfield shift of the Tyr 83 backbone nitrogen atom may be due to unshielding effects from its own aromatic ring.

An almost complete backbone assignment was achieved using the CBCANNH and CBCACONNH pair of 3D spectra. Gly 1 is not visible in the spectra, Glu 5 and Glu 6 are completely overlapped in the ^{1}H-^{15}N-dimensions. The use of Pro selective experiments proved to be especially useful in the backbone assignment and helped to untangle uncertainties arising from the many Gly residues in the protein. The four spectra link a residue preceeding Pro to that following Pro, thus bridging the gap usually created by proline due to its lack of a backbone H_N atom. The spectra and assignments are shown in Figure 35.

Sidechain resonances were assigned mainly from a CCCONH and HCCCONH pair of spectra for aliphatic sidechains. Aromatic sidechains were assigned using 2D NOESY, TOCSY and DQF-COSY spectra, acquired in 100% D_2O. The sidechain assignment is ~ 95% complete. Details of NMR data acquisition are described in Materials and Methods. A list of assigned nuclei is given in Appendix 6.3.1. The assignments were deposited in the BioMagResBank under accession code 5939.

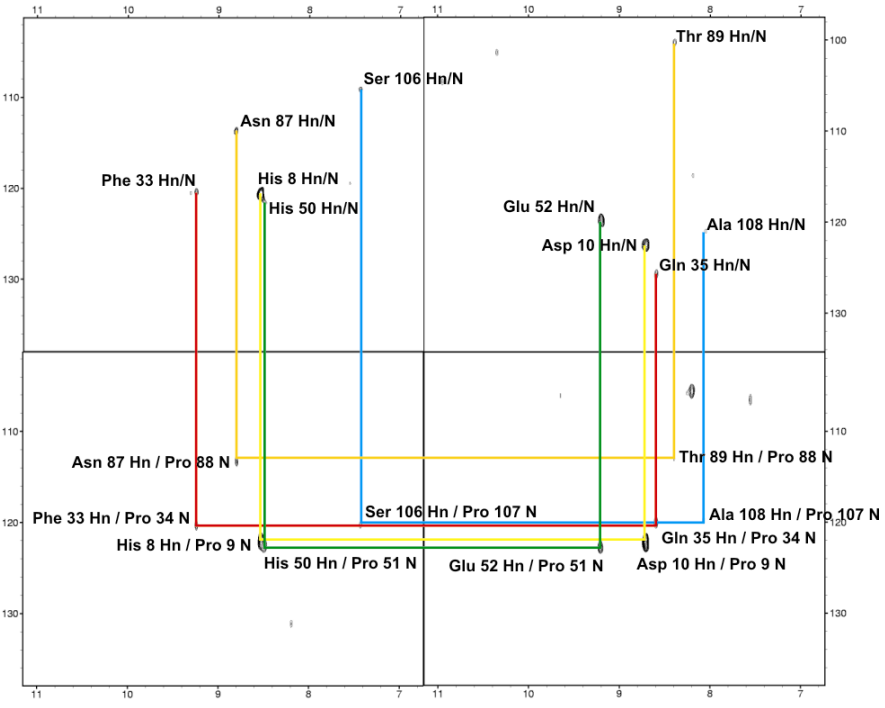


Figure 35
Pro selective experiments. The four panels show counterclockwise from top left the P(i-1), P(i-1,np), P(i+1,np) and P(i+1) experiments [56].

4.2 ^{15}N relaxation measurements

The relaxation times T_1 and T_2 were measured for individual residues and are shown in Figure 36. The total correlation time of the Spred2 EVH1 domain was calculated on the basis of average T_1 and T_2 times as approximately 8.8 ns, in accordance with a monomer of MW 14.193 Da [57].
All residues of the Spred2 EVH1 domain show a very similar behaviour, with the exception of the first seven N-terminal residues. These are also less well defined by NOE's in the spectra and are apparently more flexible.

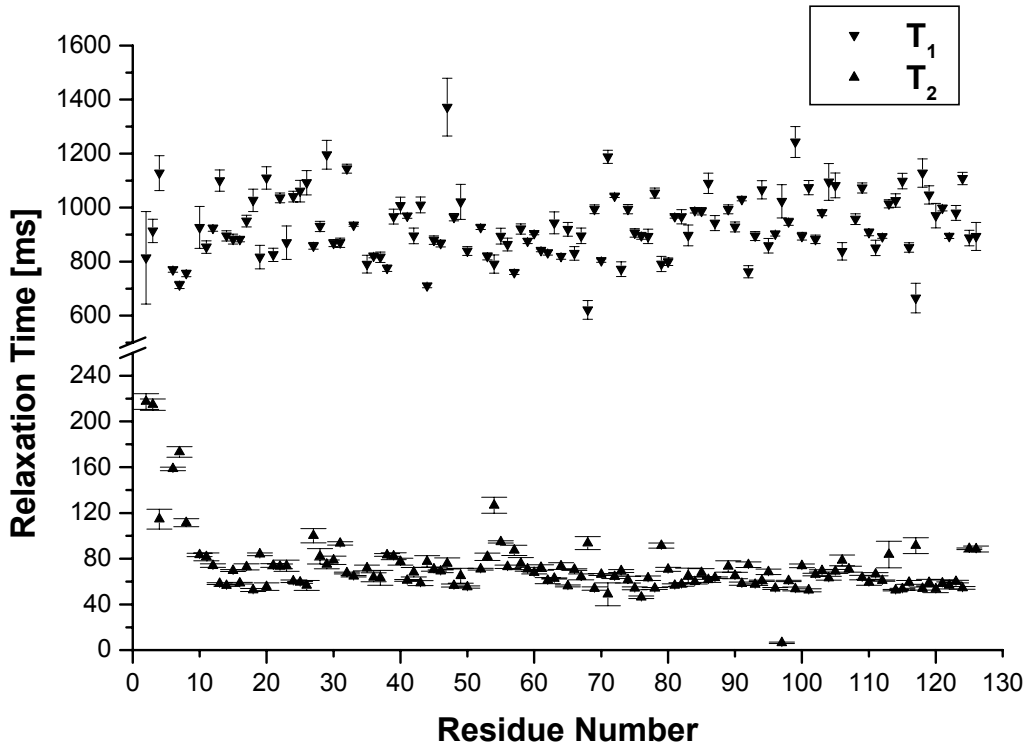


Figure 36
Relaxation properties of the backbone amide groups of Spred2. Plotted are the T_1 and T_2 relaxation times for all non-proline residues that have been assigned. The protein concentration was 1mM. An overall correlation time of 8.8 ns was calculated, corresponding to a monomer.

The sequence clearly forms a stable domain, with little internal mobility. Three regions in the domain exhibit a slightly different behaviour, especially visible from T_2 relaxation times. Small increases in T_2 can be observed around residues 29, 36 and 53. These regions are also less well defined in the structure and reflect increased flexibility of these parts of the protein.

4.3 Dihedral Angle restraints

Based on the chemical shifts of H_N, N, CO, Cα and Cβ nuclei for all assigned, non-Proline residues, a prediction of the mainchain dihedral angles ϕ and ψ was performed using the program TALOS [46]. Several values for each residue were predicted by TALOS, the prediction was checked visually in TALOS based on the distribution of predicted values in a Ramachandran plot. Only angles with closely matching prediction values were used in further calculations. Figure 37 shows the predicted values and error margins.

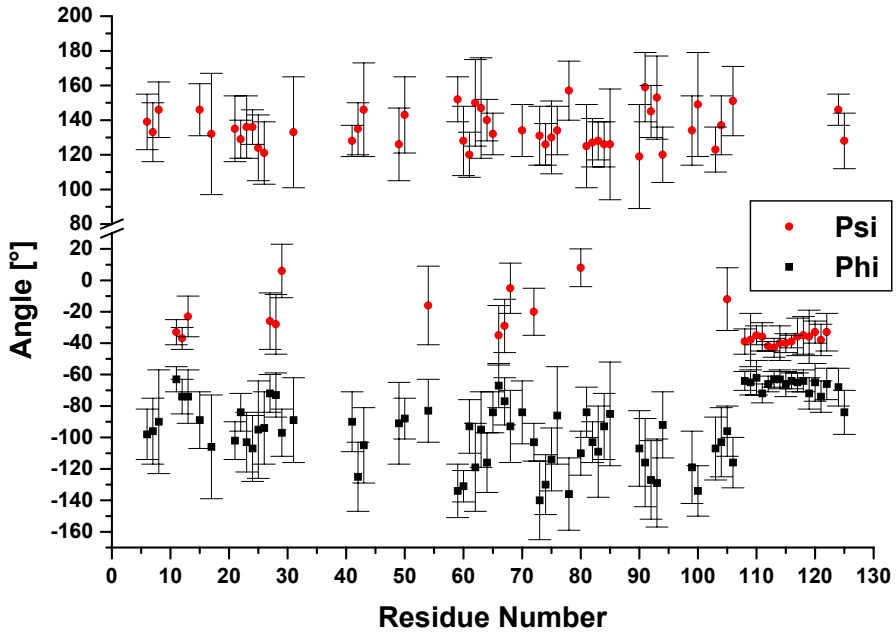


Figure 37
TALOS predictions based only on chemical shifts for ϕ and ψ angles.

While TALOS predictions are quite accurate for the α-helical part of the protein, larger errors were observed for most other resiudes. The predicted values were included in the calculations with error margins of at least ± 30° (error margins were increased to ± 30° if TALOS predictions had lower error) in order to avoid biasing the NOE assigments when using the automatic assignment algorithm CANDID in CYANA [58] (see below).

In order to obtain experimental restraints for ϕ angles, $^3J_{HnH\alpha}$ couplings were measured. The strategy for the measurement is based on a series of 2D J-modulated ^{15}N-^1H-COSY experiments [59] modified by Peter Schmieder (not published). The experiments were recorded with different J evolution delays, resulting in a damped cosine type modulation of peak intensities. Example curves are shown in Figure 38.

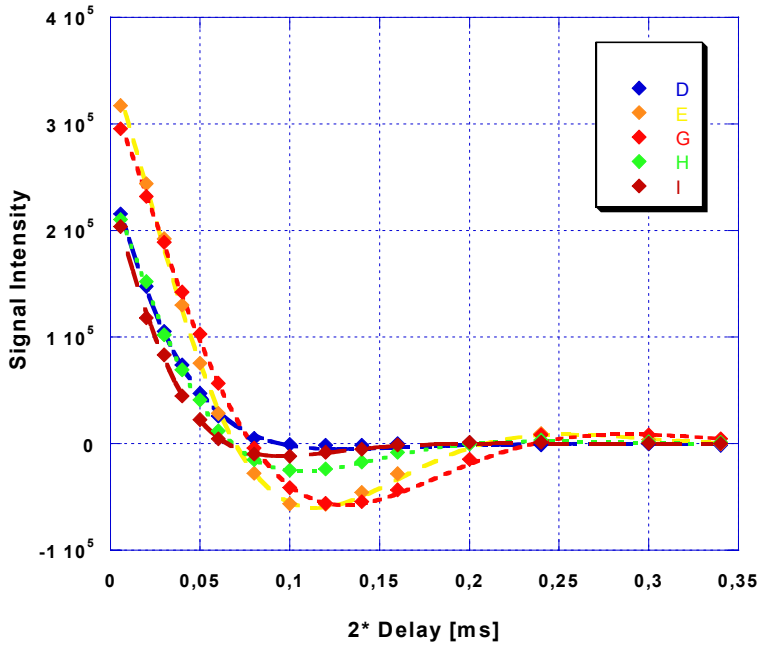


Figure 38
Data and curve fits of several residues for the $^3J_{HnH\alpha}$ coupling. Data for residues 3, 4, 6, 7 and 8 are shown.

Fitting these data results in values for the $^3J_{HnH\alpha}$ couplings. The fitting procedure is described in Chapter 7. Due to the degenaracy of the Karplus curve (Figure 39), only values of $^3J_{HnH\alpha}$ greater than 7 Hz were taken into account. These were included in the calculations as angle restraints corresponding to a β-strand.

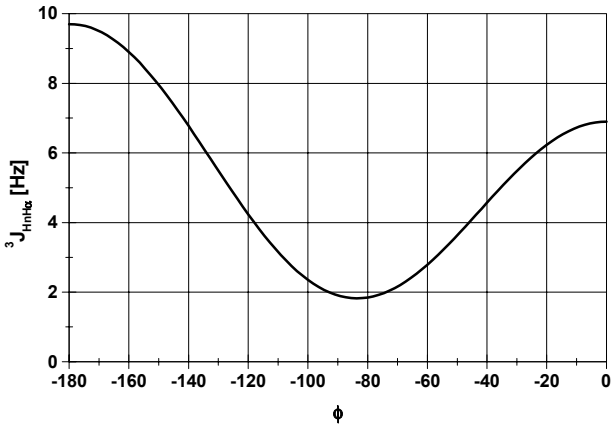


Figure 39
Karplus curve showing the relation between dihedral angle ϕ and $^3J_{HnH\alpha}$. The curve was derived using: $^3J_{HnH\alpha} = A\cos^2\theta + B\cos\theta + C$ [60], with A=6.4, B=-1.4, C=1.9 [61]

4.4 CYANA structure calculation

After manual resonance assignment the structure of the huamn Spred2 EVH1 domain was calculated using Peter Güntert's program CYANA [58]. This program uses a list of resonance assignments (e.g. chemical shift values for as many nuclei in the protein as can be achieved) and peak lists of NOESY spectra to perform an automated assignment of these NOESY spectra. Traditionally this task was performed in a manual, user driven manner and was time consuming, taking several month at least. CYANA assigns NOEs based on the chemical shift lists, evaluates these assignments and calculates structures based on the resulting set of distance restraints, also including dihedral angle and possibly H-bond restraints. The resulting ensemble of structures is included in the next cycle in the evaluation of assignments, all NOESY peaks are taken into account in each cycle. This procedure is repeated six times. The progression of the structure calculation through the seven cycles in a CYANA run is depicted in Figure 40.

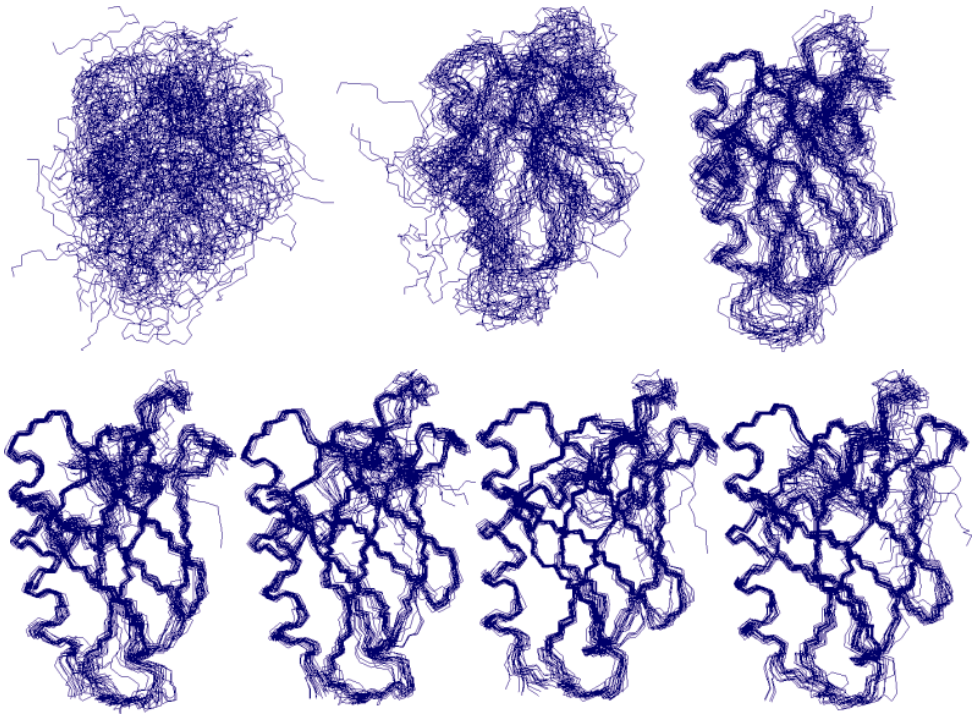


Figure 40
Ensemble of the 20 structures with lowest target function from each of the seven CYANA cycles. Backbone rmsd's are: 4.27, 2.28,1.15, 0.62, 0.44, 0.52 Å.

Traditional structure calculation programs such as XPLOR and CNS [44] use a forcefield for refinement of the structure and the resulting overall energy as a measure for convergence and

quality of structures. CYANA performs torsional angle dynamics based on the algorithm of the program DYANA [58] and uses a target function rather than energies. This function is defined as:

$$tf = (d - b)^2 \qquad \textbf{Equation 4}$$

where d is the actual distance and b the upper distance bound (e.g. from an NOE restraint). All bond lengths and geminal angles are fixed to their ideal values. For this calculation no manual NOE assignments were used. The target function and backbone RMSD converged very well, especially the rmsd exhibits only small changes after cycle four (Figure 41).

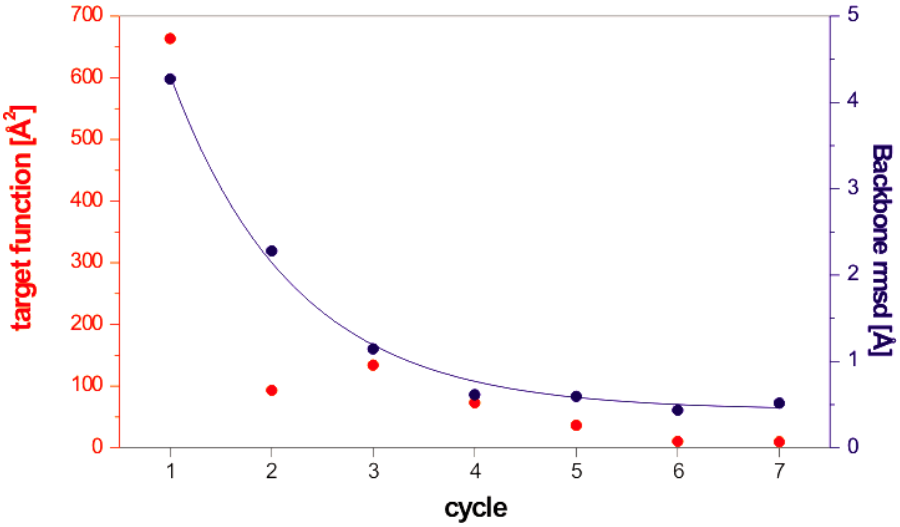


Figure 41
Development of target function and backbone rmsd during the seven cycles of the CYANA structure calculation. The backbone rmsd has been fitted to a decaying exponential function simply to illustrate the convergence behaviour.

The final ensemble of 20 structures had a backbone RMSD of 0.56 Å for the well ordered regions (residues 7-26, 31-34, 39-49 and 57-126). A summary of constraints and structural characteristics is given in Table 4.

A total of 1162 unique NOE distance restraints were found by CYANA. Almost a third (391) of these are long range (> i,i+4) restraints, which are essential for defining the global fold of the protein. 147 torsional angle restraints for backbone ϕ and ψ angles were included, derived both from TALOS predictions and measurements of $^{3}J_{Hnh\alpha}$ couplings. 26 hydrogen bond restraints have been used. Of these, 16 help to define the α-helix, 10 were assigned iteratively in β-sheets over several CYANA runs.

Only small violations were observed for the final ensemble of 20 structures. Most residues fall in the most favoured region of the Ramachandran plot (~73 %), with only 0.9 % in the disallowed region. The structure was deposited in the PDB under accession code 1Q8E.

A general lack of restraints was observed for the N-terminus. These residues have only few, mostly intraresidue peaks in the NOESY spectra. Pronounced dips in the distribution were observed around residues Ser 29, Gly 37 and Gly 55, corresponding to a reduced definition of structures in the final ensemble. This corresponded well to the increased flexibility in these regions observed in the relaxation experiments. Residues Phe 59 and Leu 60 at the centre of β-strand 3 show the largest number of long range restraints and are apparently very important for defining the hydrophobic core of the protein. A general correspondence of regions of *secondary* structure with an elevated number of restraints cannot be observed.

Most restraints are sequential (352), followed by intra-residue (249). 170 restraints spanning three or four residues were assigned. The remaining restraints are long range (391), these are distributed throughout the entire sequence and are sufficient for a good definition of the protein fold.

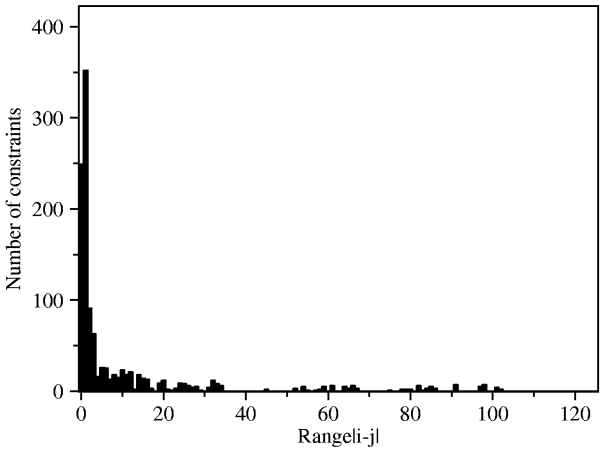


Figure 42
Distribution of NOEs as assigned by CYANA. The panel shows the number of NOEs as a function of range.

NOE upper distance limits	
Total	1162
Long range (> (i,i+4))	391
Torsion angle constraints	147
Hydrogen bond contraints	26
CYANA target function value (Å^2)	9.67
Distance constraint violations	
Number > 0.5 Å	1
Maximum (Å)	0.54
Torsion angle constraint violations	
Number > 6°	2
Maximum (°)	6.09
Hydrogen bond constraint violations	
Number > 0.2 Å	1
Maximum (Å)	0.25
Ramachandran plot analysis (all residues) (%)[a]	
Most favored regions	72.9
Additionally allowed regions	22.4
Generously allowed regions	3.7
Disallowed regions	0.9
RMS Deviation for N, Cα, C' (Å)[b]	
Secondary structure[c]	0.50
Well ordered region[d]	0.56
All residues	0.89

Table 4
Details of CYANA structure calculation. Shown are average values for 20 conformers with the lowest CYANA target function, representing the NMR solution structure.
[a]Using PROCHECK-NMR[47]
[b]Calculated with MolMol[48]
[c]β-sheet for 17-26, 38-50, 56-66, 71-76, 81-86, 90-94, 98-105; helix for 9-12 (3_{10}), 107-125 (α)
[d]Including residues 7-26, 38-50, 56-126

4.5 Structure analysis

The Spred2 EVH1 domain clearly belongs to the PH (pleckstrin homology) family of folds like all other known EVH1 domains the structure of which has been determined. The domain organization is illustrated in Figure 43. One turn of a 3_{10}-helix close to the N-terminus is followed by seven stretches of β-strand, which form three β-sheets. Strands 1 and 2, 3 and 4 and 5 to 7 form antiparallel β-sheets, with strand 7 having contacts to both strand 1 and 6. The C-terminus is formed by an extended α-helix.

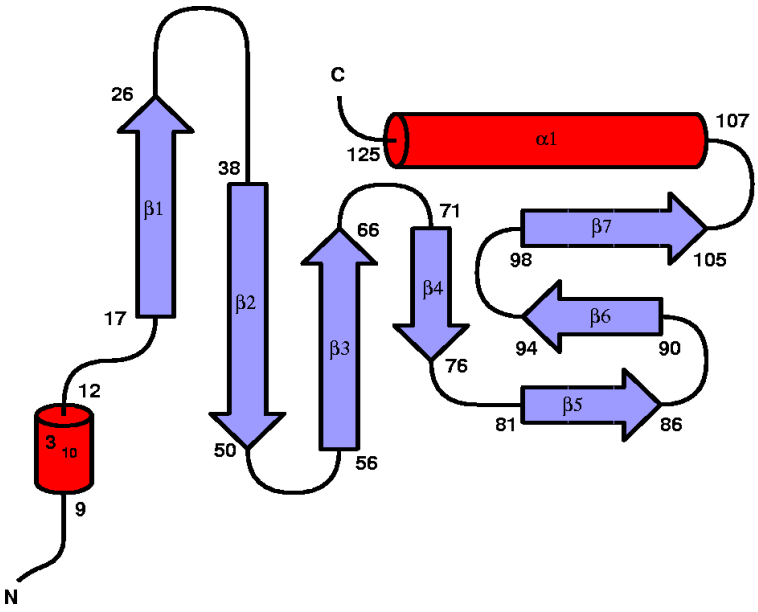
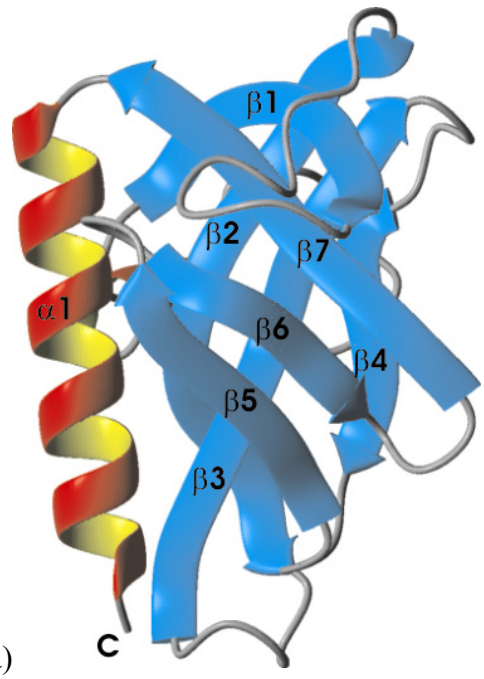


Figure 43
Secondary structurediagramm of the Spred2 EVH1 domain. The arrangement of β-strands corresponds to the interactions of these strands in the domain, it is not representative of hydrogen bonding networks.

The 3D-structure consists of an all-antiparallel β-sandwich closed by an α-helix on one side, as depicted in the ribbon representation in Figure 44.a. The ensemble of 20 structures with the lowest target function values is shown in Figure 44.b. Three more flexible loops are coloured blue. These have also been identified from relaxation measurements (see Figure 36 above).



a)

Figure 44
Ribbon representation of the Spred2 EVH1 domain structure as determined with CYANA. The N-terminus is not visible in this orientation.

Figure 45 shows an overlay of the structures of the Spred2 EVH1 and VASP EVH1 domains. While the overall fold is similar, the structures differ in many details and can only loosely be superimposed.

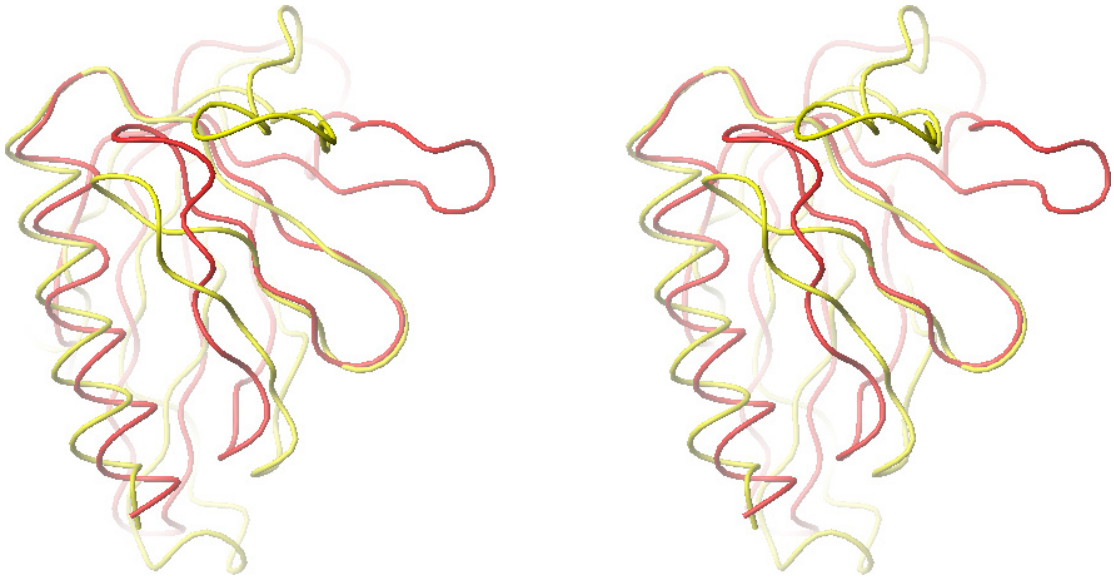


Figure 45
Superposition of Spred2 EVH1 domain (yellow) onto the human VASP EVH1 domain (red). The overall fold is similar, especially β-strands 6 and 7 and the α-helix can be nicely overlaid. The largest difference is observed for the loop connecting β-strands 1 and 2, depicted more clearly in Figure 46.
(Stereo picture, side-by-side)

Figure 46 shows a superposition of the β-sheet forming the binding surface of the Spred2 and VASP EVH1 domains.

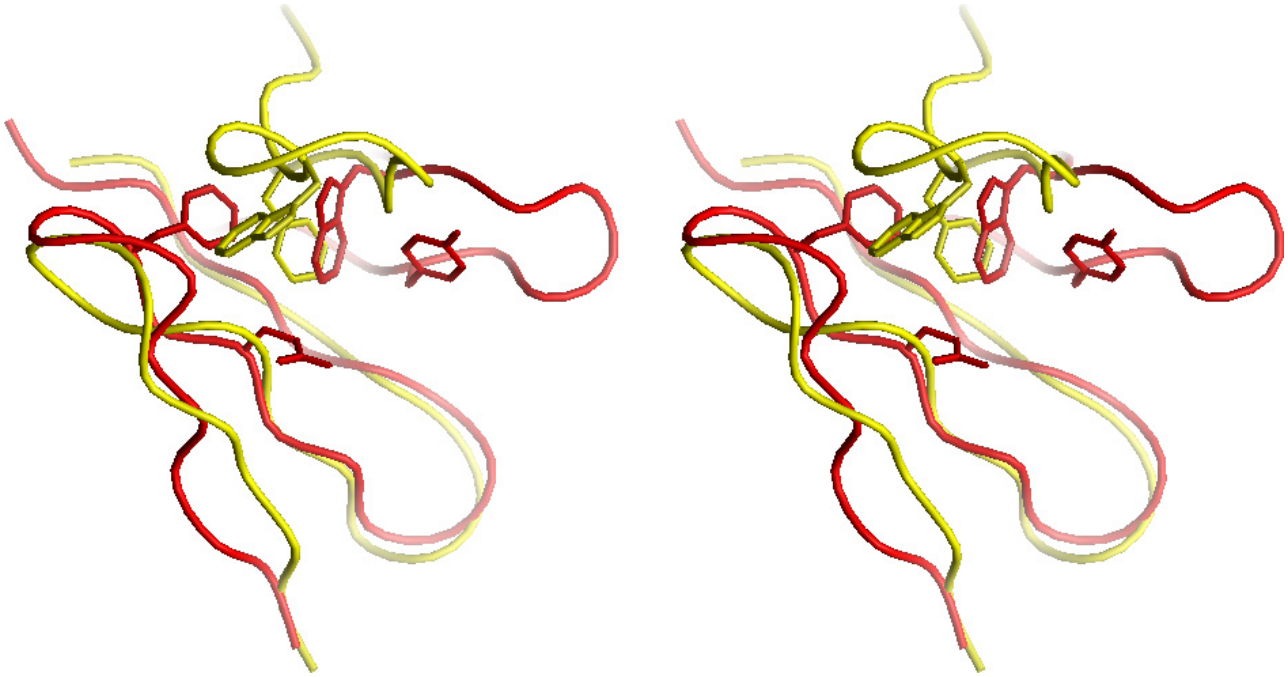


Figure 46
The β-sheet adjacent to the α-helix in the two EVH1 domains has been overlaid for this figure. While the sheet has a similar structure in these domains, the loop on top of it shows a considerably different orientation. (Spred2 yellow (residues Val77-Ser106), VASP red (residues Ile66-Ser95)); in brackets: residues of β-sheet overlaid in this figure)
(Stereo picture, side-by-side)

While the β-sheet of the Spred2 EVH1 domain is very similar to that of the VASP EVH1 domain, the loop containing residues 32 and 33 of the Spred2 EVH1 domain and 16 and 23 of the VASP EVH1 domain, respectively, is in a distinctly different positon in the Spred2 EVH1 domain and causes a different topology of the potential binding interface of the this domain.

All class I and class II domains share a conserved binding motif, the most important residues for this motif occur in very similar positions in the different sequences of these domains. The residues are specified in Table 5.

	Protein (PDB-code)	Hydrophobic	Trp	Phe	Gln
Class I	VASP (1EGX)	Tyr 16	Trp 23	Phe 79	Gln 81
	Evl (1QC6)	Tyr 16	Trp 23	Phe 78	Gln 80
Class II	Homer2b (1DDW)	Phe 16	Trp 24	Phe 74	Gln 76
	Vesl (1I2H)	Ile 18	Trp 26	Phe 76	Gln 78
Class III	Spred2 (1Q8E)	Phe 33	Trp 32	-	-

Table 5
Residues important for ligand binding in class I and II EVH1 domains and putatively homologues residues in the Spred2 EVH1 domain (Compare also Fig. 33, p. 56).

There are three completely conserved residues in Class I and II EVH1 domains, these are Trp 23, Phe 79 and Gln 81 (VASP numbering) and a conserved hydrophobic residue in position 16, which is either Tyr, Phe or Ile. The sequence alignment shown in Figure 33 implies a conservation of these residues in Spred2, as well. This is only true on the sequence level. The comparison of the 3D structures as shown in Figure 46 identifies residues Trp 32 and Phe 33 as corresponding to Trp 23 and Phe 16 of the human VASP EVH1 domain. While in Class I and II domains these two residues are proximate via a loop in the structure and thus form part of the binding epitope, this motif is formed by resiudes 32 and 33 in the Spred2 EVH1 domain, which are immediate sequence neighbors. A residue corresponding to Phe 79 of VASP cannot be identified in the structure of the Spred2 EVH1 domain, as there is no residue of Spred in a similar spacial position at all (Figure 47). Also, a hydrogen bond acceptor similar to Gln 81 cannot be identified in a straightforward manner. Similar to the canonical binding interface for proline residues, the aromatic sidechains of the two residues Trp 32 and Phe 33 of the Spred2 EVH1 domain show an orthogonal orientation towards each other. Thus they might form a binding site for one proline residue. So far, no interaction partner for the Spred2 EVH1 domain has been identified, and the function of this domain is unknown.

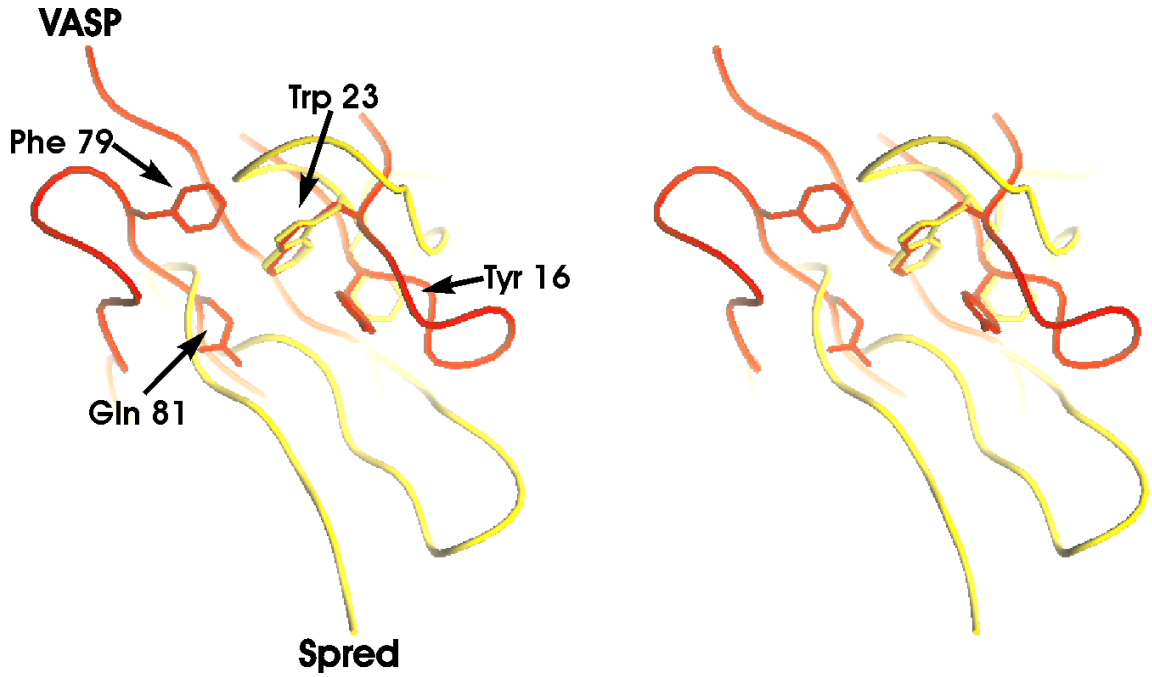


Figure 47
Binding epitope of the VASP EVH1 domain, overlaid with the potential binding epitope of the Spred2 EVH1 domain. The backbone of the Spred2 EVH1 domain is shown in yellow, omitting hydrogens for clarity. Labels correspond to VASP sidechains, see also Table 2. (Spred2 yellow, VASP red)
(Stereo picture, side-by-side)