3 Oligomerization and structural characterization of the human VASP EVH2 domain

It has been suggested that the C-terminal domain of VASP is necessary for tetramerization of the protein [13], which in turn is essential for proper function. If VASP forms a tetramer, the arrangement of subunits is crucial for models of the interactions of VASP with other proteins. A potential role of EVH2 mediated oligomerization would be in targeting bundled EVH1 domains to several FPxϕP motifs in target proteins or alternatively clustering several proteins containing EVH1 interaction motifs. Therefore the structure of the EVH2 domain is of special interest in the context of VASP function.

The EVH2 domain of human VASP consists of residues 336 to 380. Its sequence is shown in Figure 17, aligned with several related sequences of other proteins.

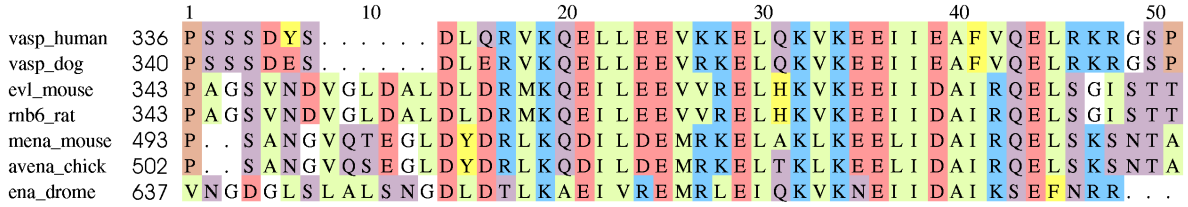

Figure 17
Alignment of EVH2 domains from several species. The domain is highly conserved between species as distant as *Homo sapiens* and *Drosophila melanogaster*. Colours are: Red – acidic; blue – basic; purple – polar; green – hydrophobic; yellow – aromatic; brown – Pro.

A high degree of conservation is observed, pointing to the importance of this domain. Especially the pattern of hydrophilic and hydrophobic residues is well conserved. This pattern is characteristic of an α-helical conformation, suggesting a coiled coil arrangement in the EVH2 tetramer. The sequence homology between EVH2 and other tetrameric coiled coil proteins for which high resolution structures have been solved, is however rather low. Examples are the Mnt repressor from the *Salmonella* phage P22 [35], which forms a dimer of two right handed coiled coil dimers; the NSP4 receptor protein from rotavirus [36], which forms a parallel tetrameric coiled coil; and a fragment of the surface layer protein tetrabrachion from the hyperthermophilic archaeon *Staphylothermus marinus* [37], which also forms a parallel tetrameric coiled coil.

A secondary structure prediction of the VASP EVH2 domain is shown in Figure 18 [38].

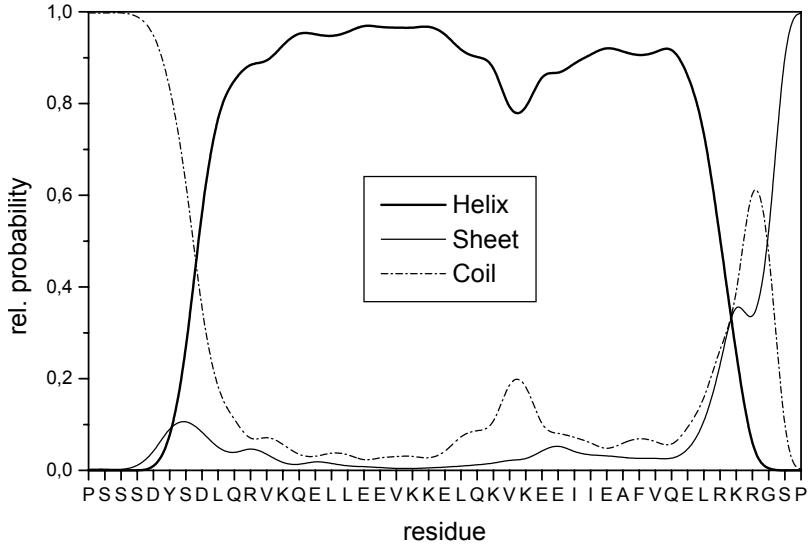

Figure 18
Secondary structure prediction for the VASP EVH2 domain. The major part of this domain is predicted to be α-helical. For this prediction the method by Combet et al. has been used (http://npsa-pbil.ibcp.fr/).

The secondary structure prediction clearly shows the domain to be predominantly α-helical. A helical wheel representation of the sequence shows a putative hydrophobic interface on one side of the helix (Figure 19), while the opposite face is hydrophilic. This is a further hint pointing to a coiled coil arrangement.

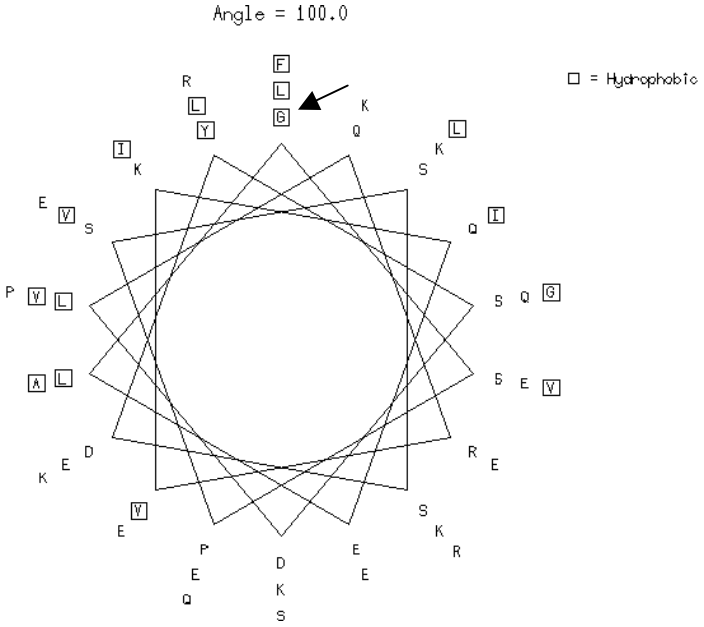

Figure 19
Helical wheel plot of the EVH2 sequence, including two additional residues (Gly and Ser) at the N-terminus (see Materials and Methods). The arrow points to the N-terminus (Figure prepared with Seqlab, GCG Wisconsin package).

3.1 CD Spectroscopy

CD spectroscopy of VASP (336-380) was performed at different pH values and temperatures to investigate the stability and secondary structure of this construct. The CD spectra showed all characteristics of a helical conformation with minima at 209 nm and 222 nm.

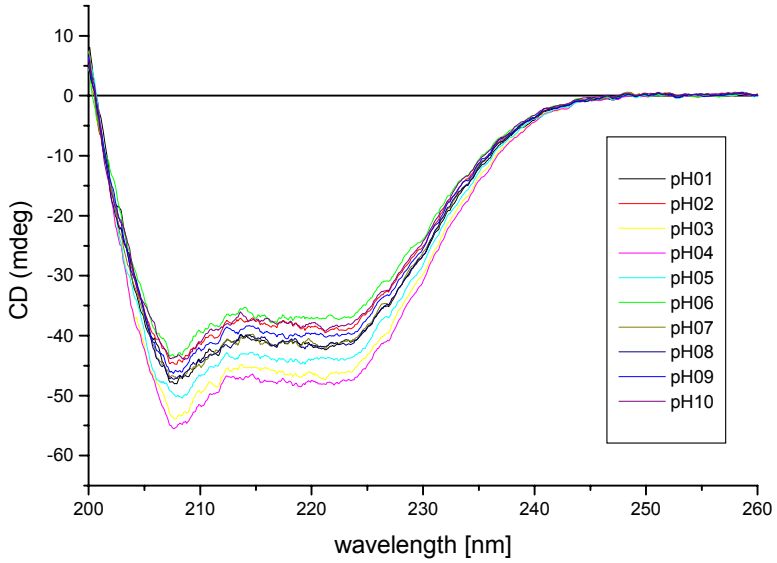

Figure 20
CD spectra of EVH2 at different pH values. The sample concentration was 0,025 mg/ml.

The protein was found to be highly stable, remaining almost fully folded from pH 1 to 10 (at 24°C) and within a temperature range of +4° to +80° C (at pH 6.0). The melting temperature was determined as 120° C (U. Walter, Würzburg, personal communication). Although the strongest signal in the CD spectra was observed at pH = 4, further investigations were carried out at pH = 6, presumably closer to a physiologically relevant environment.

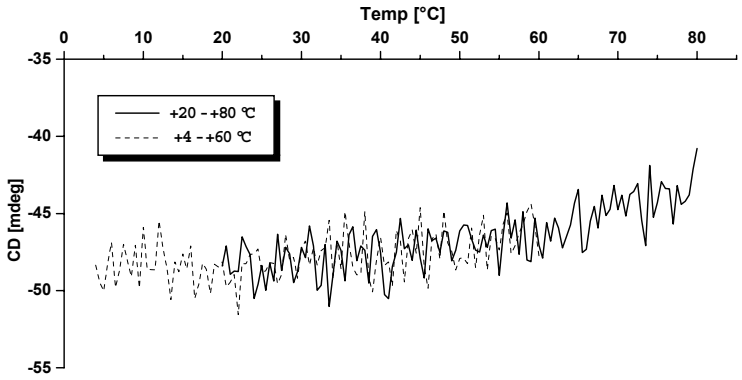

Figure 21
CD temperature scan of EVH2. Two scans were performed and the results overlaid. The most pronounced minimum at 209 nm was monitored at pH = 6.0 (pH set at 20° C).

3.2 NMR Resonance Assignment

The 600 MHz ^{15}N-HSQC spectrum of VASP (336-380) at pH 6.0 and 300K is shown in Figure 22. It was possible to assign 43 out of the 47 residues (two of which are prolines). 3D CBCA(CO)NNH and CBCANNH spectra led to an almost complete set of ^{15}N, ^{1}HN, ^{13}Cα and ^{13}Cβ assignments, with the exception of residues Ser 2 and Lys 30. The assignments were deposited in the BioMagResBank under accession code 5955.

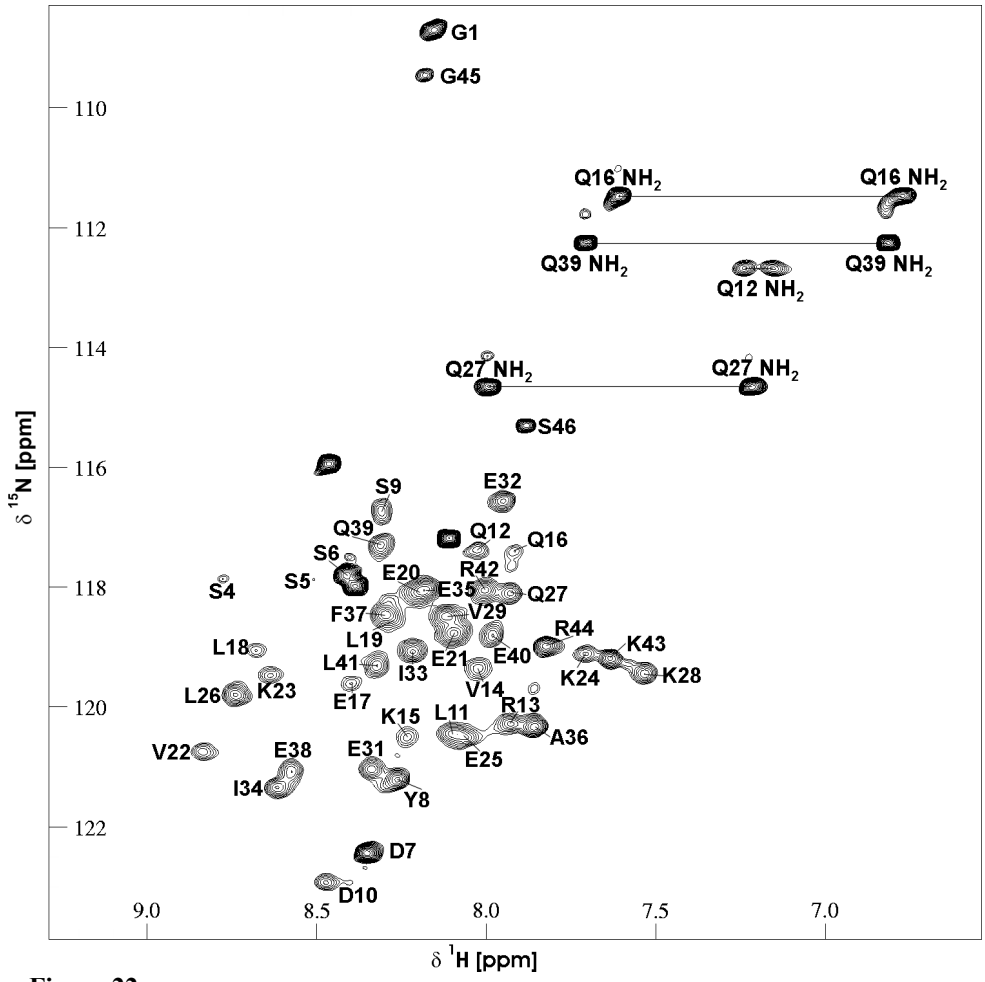

Figure 22
^{15}N-HSQC spectrum of a 5 mM (monomer concentration) uniformly ^{15}N-labelled sample of EVH2 at 300 K. Backbone amides for 43 out of 45 non-proline residues were assigned unambiguously. Peaks connected by a line indicate Gln sidechain NH_2 group resonances.

The presence of a single set of peaks in the HSQC spectrum indicates the magnetic equivalence of the monomeric units of any oligomers present. For a tetramer, this limits the possible symmetry arrangements to either C_4 or C_{22} symmetry groups. In all other arrangements, monomers will

experience different environments and so have distinct chemical shifts, resulting in multiple sets of ^{1}H-^{15}N peaks. For example, the C_2 symmetry of the tetrameric Mnt repressor results in two sets of ^{1}H-^{15}N peaks in the spectra [35]. The spectra clearly show the EVH2 tetramer to possess either C_4 symmetry (a parallel four-helix, coiled coil arrangement) or C_{22} symmetry (a dimer of dimers [39], see also Figure 28, p. 45).

Due to the entirely α-helical nature of the protein the spectra show a considerable amount of overlap. This is exemplified for the ^{15}N-NOESY-HSQC (Figure 23).

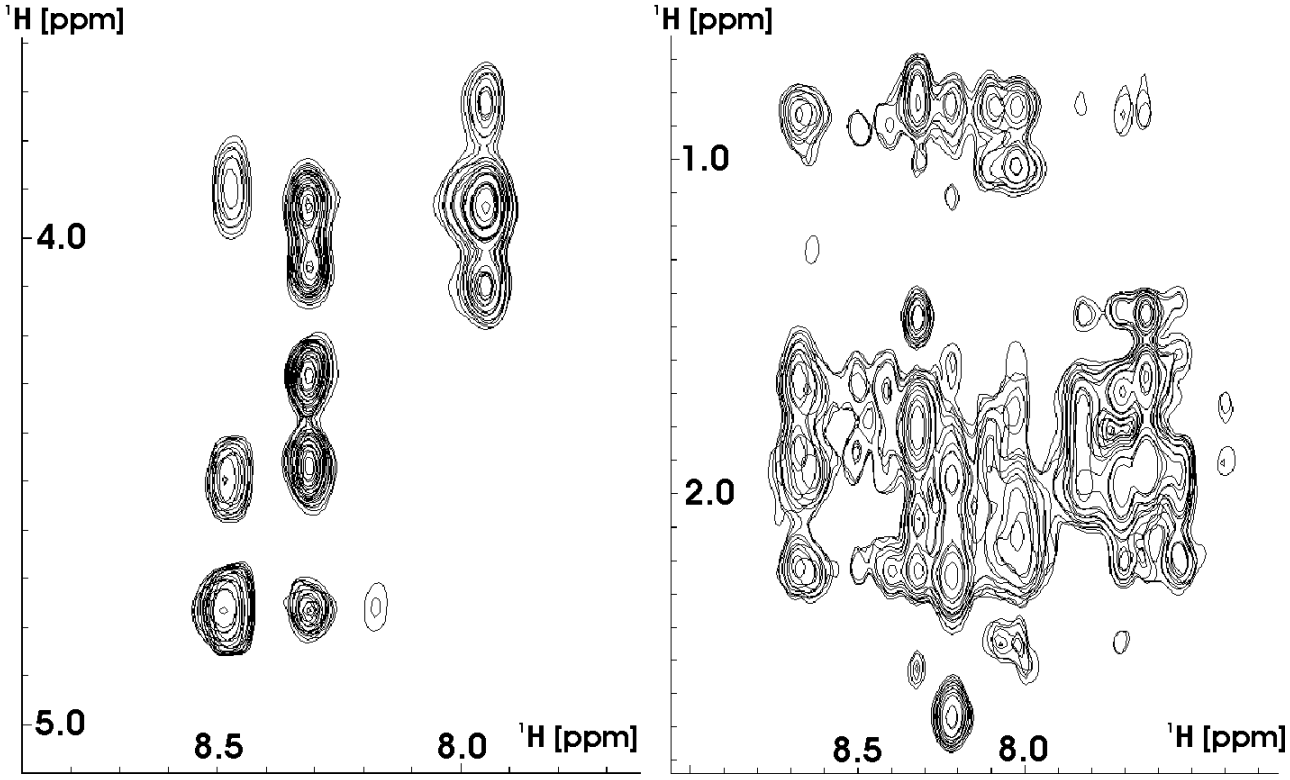

Figure 23
Two planes of the EVH2 ^{15}N-NOESY-HSQC, showing ^{1}H,^{1}Hn NOE correlations. The spectrum plane on the left is well resolved, while that on the right shows strongly overlapping peaks; the right panel is representative of the major part of the NOESY spectra.
The left plane is at a ^{15}N shift of 115.9 ppm, that on the right at a ^{15}N shift of 119.0 ppm.

While the region of the ^{15}N-NOESY spectrum on the left (Figure 23) is well resolved, the region on the right shows strongly overlapping peaks. Strongly overlapping peaks are observed in the major part of both the ^{15}N-NOESY and the ^{13}C-NOESY spectra. This results in a complication of assignments due to degenerate shifts and especially difficulties in integrating the spectra.

The NOE connectivity pattern shown in Figure 24, obtained from the 3D-NOESY-HSQC spectrum, clearly indicates an α-helical structure for VASP(336-380), in agreement with CD data. This is especially evident from the series of $d_{\alpha,N}(i,i+3)$ connections characteristical of α-helices. A closely packed arrangement of monomers was further corroborated by deuterium exchange experiments, which showed a series of slowly exchanging amide hydrogens, also marked on the Figure.

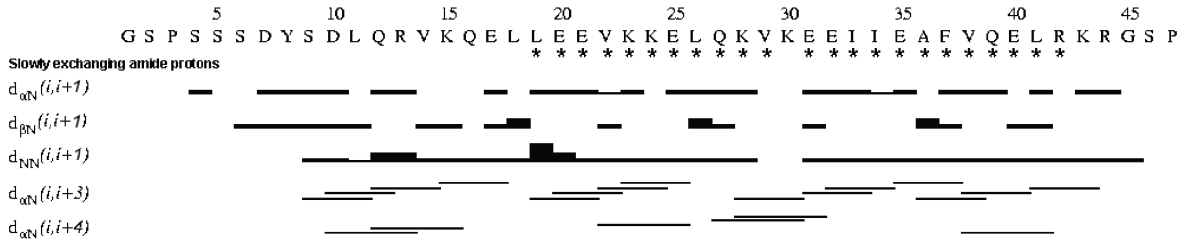

Figure 24
NOE connectivities as derived from the ^{15}N-NOESY-HSQC spectrum of EVH2. Stars indicate amide protons that exchange only slowly (within several hours or days, as opposed to a few seconds for non-protected protons) with protons of the buffer solution and thus may be involved in hydrogen bonds within the protein.

3.3 Analytical ultracentrifugation

The relative equilibrium populations of EVH2 oligomers at a series of different protein concentrations and temperatures were investigated by analytical ultracentrifugation. This technique was used to get information about possible equilibria between monomers and oligomeric species at concentrations too low for NMR investigations. The studies were performed by Dirk Labudde, experiments will therefore be described only briefly. Concentrations studied were 0.0625, 0.125 and 0.25 mM, each at 10°, 20° and 27 °C. From sedimentation velocity (SV) experiments (see Figure 27.a), the sedimentation coefficient, s, was determined as 1.98 S ($1S=10^{-13}$sec), and from sedimentation equilibrium (SE) experiments (Figur 27.b), a molecular mass of 21.9 kDa was calculated for the detected species. This is in excellent agreement with a predicted molecular weight of 21,988 kDa for a ^{15}N-labelled VASP(336-380) tetramer.

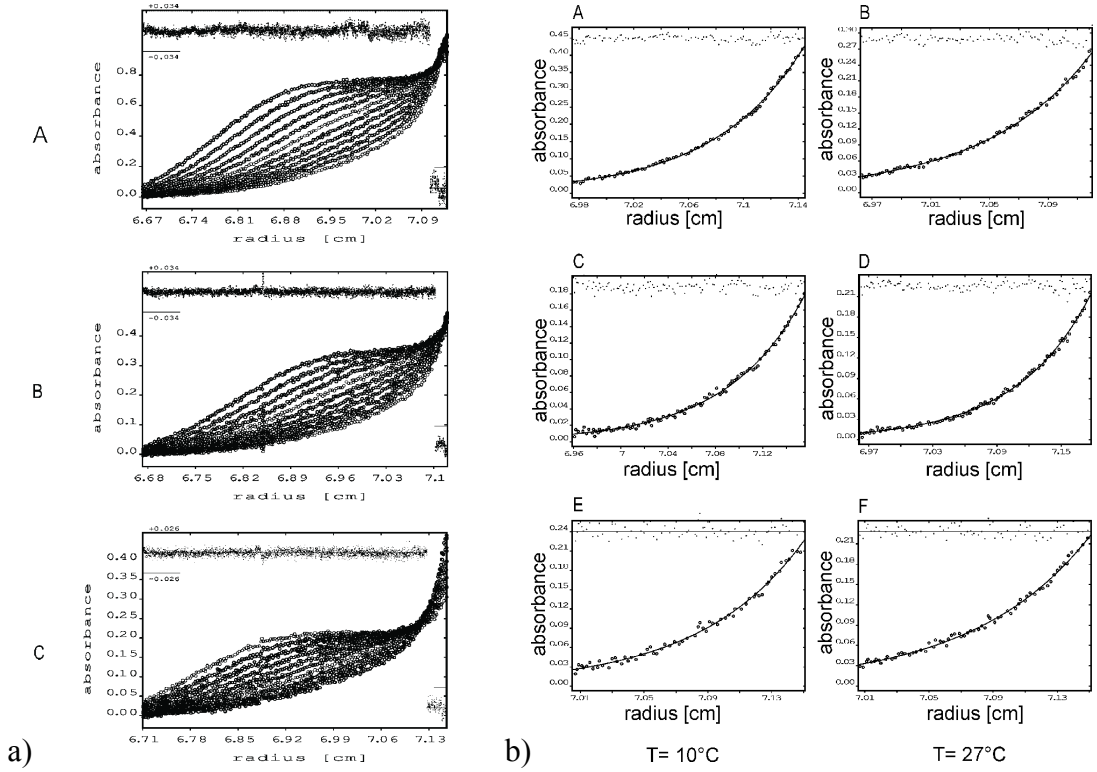

Figure 25
Panel a) Radial absorbance distributions of sedimentation velocity experiments and fitted data of A 0.25 mM, B 0.125 mM and C 0.0625 mM in NMR buffer (see methods). Residuals (indicating the r.m.s.d. of the data to the fit) are shown along the top of each plot are of all datasets and given in twofold amplification and scaled according to the axis marked on the plots. Calculated Svedberg constants (s), for each concentration were (a) 2.01 S (b) 1.96 S and (c) 1.97 S, with an average of 1.98 S (S =10^{-13}s).
Panel b) Radial concentration distribution curves from sedimentation equilibrium experiments. Concentrations are: A,B: 0,25 mM; C,D: 0,125 mM; E,F: 0,0625 mM. SE experiments were performed at 10° and 27° C. Residuals calculated from all datasets are shown above each plot, in twofold amplification and scaled separately according to the axis marked on the plots.

Additionally, these parameters allowed the calculation of the diffusion coefficient (D) using the Svedberg equation:

$$M = \frac{sRT}{D(1-\rho\bar{v})} \qquad \textbf{Equation 1}$$

where s is the sedimentation coefficient, M is the molecular mass, $\bar{v}$ is the partial specific volume of the solute, ρ is the density of the solvent, R the ideal gas constant and T the temperature. The diffusion coefficient (D) was calculated to be 8.703 * 10^{-7} (cm^2/sec). On the basis of Equation 2, it was then possible to specify the frictional ratio, f/f_0, for the unhydrated protein:

$$f/f_0 = 10^{-8}\sqrt[3]{\frac{1-\rho\cdot\bar{v}}{s\cdot D^2\cdot\bar{v}}} \qquad \textbf{Equation 2}$$

where f is the frictional coefficient of the protein and f_0 is the ideal frictional coefficient for a perfectly spherical molecule (defined as 1.0) (equations taken from [41]). A value of 1.32 for the frictional ratio f/f_0 of the protein was obtained, implying that its overall conformation deviates significantly from spherical. Prediction of the molecular dimensions from the molecular mass and s value, yielded possible axis ratios (longest axis/shortest axis) of 6.00, 6.09, and 6.87, for long-rod, prolate and oblate models, respectively. These values are highly dependent on the accuracy of D, and therefore only give an estimate of the molecular shape. Although it is not possible to distinguish which of these three different models corresponds to EVH2, it is clear that the molecule possesses a considerably elongated overall structure.

In the SE experiments, the concentration gradients yielded a MW of 21.9 kDa irrespective of protein concentration and temperature (Figure 27.b), indicating that a single oligomer species is present as opposed to an equilibrium mixture of different oligomers. The average value obtained for the MW corresponds closely to a tetramer of EVH2, and is in excellent agreement with the result obtained from the SV experiments and agrees well with the results of the ^{15}N relaxation experiments. A number of different models were tested in order to determine the nature of the oligomerization equilibrium. Within the concentration range used, no other component than the tetramer could be detected in the SE gradients. This further supports the notion that EVH2 behaves as a single, ideal, tetrameric entity under the range of conditions examined, suggesting a very low dissociation constant of the oligomer.

3.4 ^{15}N relaxation measurements

In order to investigate the self-association behaviour of EVH2 and to study the dependence of oligomerization on the protein concentration, a series of ^{15}N T_1 and T_2 relaxation experiments were used in order to determine the average correlation time of the domain at different protein concentrations. The concentrations studied ranged from 0.25 to 5.0 mM (total protein concentration). The relaxation properties shown in Figure 25 demonstrate clearly that EVH2 is highly structured in the range of residue 9 to 43 with little internal mobility, except at the extreme

N- and C-termini. This was in agreement with heteronuclear ^{15}N-^{1}H NOE experiments, carried out at the same concentrations (data not shown).

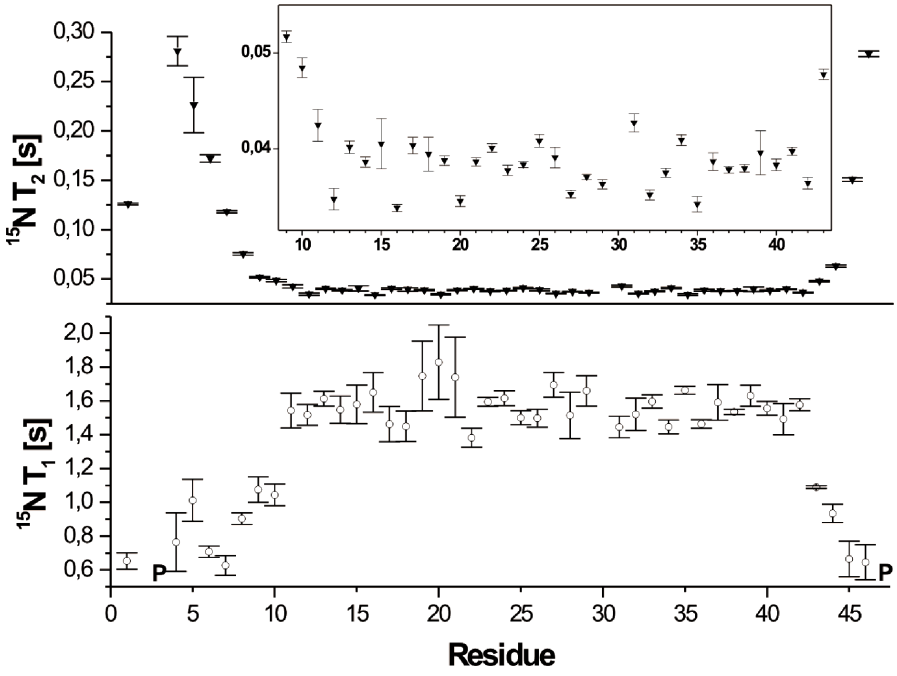

Figure 26
Plot of ^{15}N relaxation time constants, T_1 and T_2, for each residue of EVH2, shown for a protein concentration of 5 mM (Inset: Extension of T_2 for res. 9 to 43). The protein is motionally restricted in the core region (res. 10-44 corresponding to aa 346-377 of VASP numbering) with some flexibility observed at the N- and C-termini.

The average ^{15}N T_1, ^{15}N T_2 and calculated molecular correlation times, τ_C, at total protein concentrations 0.25, 0.625, 1.25, 2.5 and 5.0 mM are shown in Table 3.

Conc. (mM)	T_1 (ms)	T_2 (ms)	τ_C (ns)
5.0	1,521.2 (78.4)	39.6 (0.8)	19.6 (0.8)
2.5	1,513.7 (71.1)	38.8 (0.7)	19.7 (0.8)
1.25	1,481.3 (184.6)	42.7 (3.2)	18.6 (1.7)
0.625	1,433.2 (255.6)	42.7 (3.2)	18.3 (1.0)
0.25	1,083.9 (606.7)	31.6 (11.3)	18.5 (1.9)

The average τ_C of 18.9 ns over the five concentrations measured by NMR can be attributed to a molecular mass of ~ 20-22 kDa within the limits of experimental error (see Figure 26). As there is no precise relationship between correlation time and molecular mass of a protein, this plot can only yield a rough estimate of the mass. The correlation time is also related to overall molecular shape. For an elongated molecule a larger τ_C than for a globular molecule of the same molecular mass

would be expected. The calculated τ_C is in reasonable agreement with the theoretical molecular weight of 21.5 kDa for tetrameric EVH2.

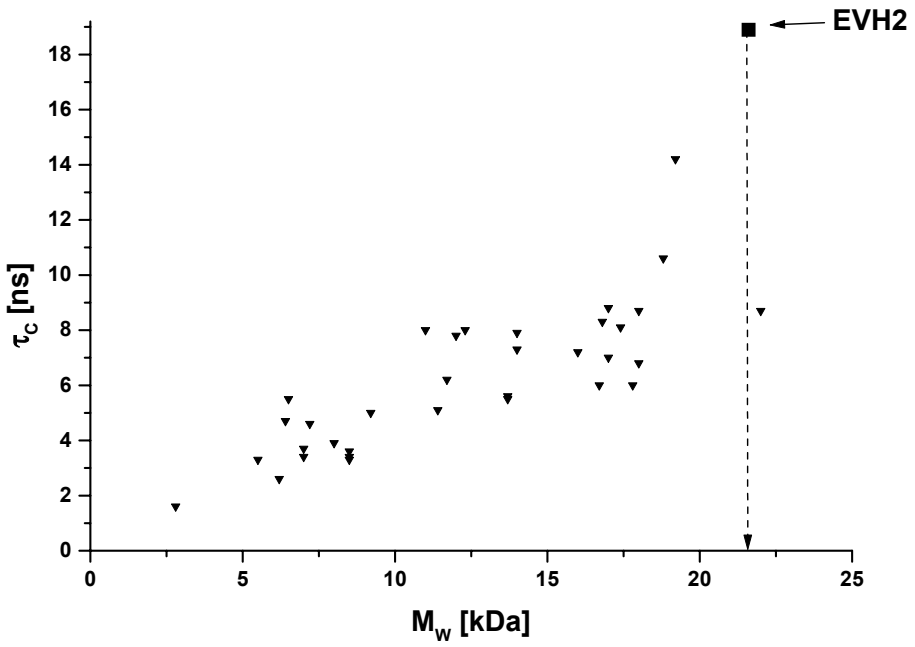

Figure 27
Molecular mass and overall correlation time τc for a number of proteins[40]. The mark for EVH2 has been put at the correlation time derived from relaxation experiments and mass determined by Analytical Ultracentrifugation.

Furthermore, despite a 20-fold difference between the maximum and minimum protein concentrations measured, the overall τ_C remained constant within the limits of experimental error. This data cofirms that EVH2 is fully tetrameric in solution and that this state is independent of the protein concentration in the range 0.25-5.0 mM.

3.5 Symmetry Considerations

Due to the high degree of symmetry of the tetramer and the small size of the protein all NOEs in the spectra are potentially ambiguous. They can either be intramonomer, intermonomer, or a mixture of both possibilities. For globular proteins with an interface area that is small in comparison to the overall protein surface, only a few distance restraints will be ambiguous, and a calculation using ambiguous distance restraints in e.g. ARIA should be feasible. In the case of the EVH2 tetramer a large part of the surface of each monomer is most likely involved in contacts to the other three monomers. Even if *inter*monomer and *intra*monomer contacts could be distinguished, it would still be unclear to which of the other three monomers they point.

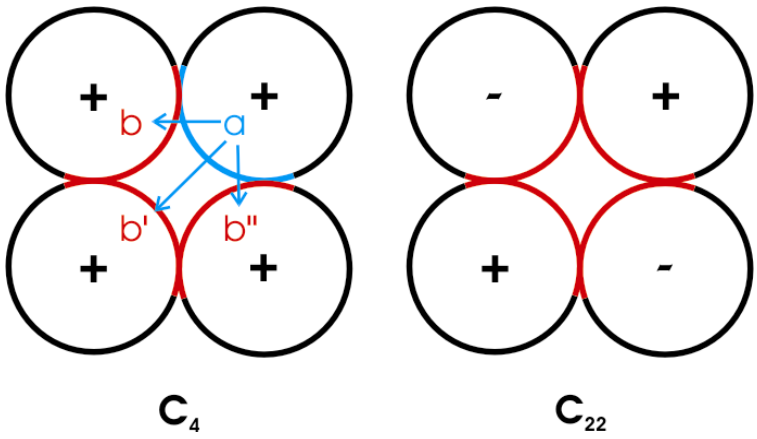

Figure 28
The two possible symmetry arrangements of the EVH2 tetramer and ambiguity of NOEs.

As illustrated in Figure 28, an intermolecular NOE originating on nucleus **a** can go to either nucleus **b**, **b'** or **b"**, or any combination of these possibilities. Hence, every NOE observed is ambiguous as to whether it is intra- or intermonomeric, and intermonomeric NOE's are ambiguous themselves. Based on NOEs alone it would be extremely difficult, maybe impossible to derive a structure for EVH2. To create a sample with differently labelled monomers would only be of limited help. While this approach will solve all ambiguities for a dimeric ensemble, and also to some extent for trimers, it fails with a tetramer based on symmetry considerations.

3.6 Residual Dipolar Couplings

Further experimental restraints were required that would yield unambiguous information on the conformation of a monomer. This was gained via the collection of Residual Diploar Coupling (RDC) data.

Two types of coupling between nuclei are important in solution state NMR spectroscopy. The first one is the scalar coupling, which occurs between two chemically bonded atoms. Its magnitude depends on the nature of the coupling atoms and the involved geminal and dihedral angles.

The second type is the diploar coupling, occuring between any two nuclei that are sufficiently close in space. The magnitude of this coupling is dependent on the orientation of the vector between the two coupled nuclei with respect to the magnetic field B_0. Due to the isotropic brownian tumbling of molecules in solution the dipolar interactions are strictly averaged to zero in traditional solution state NMR experiments.

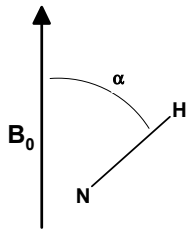

Figure 29
Illustration of the angle α between the magnetic field B_0 and a vector between two coupled nuclei. For dipolar couplings the nuclei of interest need not be connected by chemical bonds.

For measurement of RDC's in solution a small degree of alignment needs to be introduced into the sample, leading to anisotropic tumbling of the molecules with respect to the magnetic field axis B_0. Several strategies have been suggested for this purpose, including the use of bicelles, phage, polyacrylamide gels and lipids doped with Lanthanum ions. In the present study, Pf1 phage was used.

It has been recommended to use five RDC's measured for each peptide plane in a protein to orient all planes with respect to a common alignment frame [42]. These are shown in Figure 30.

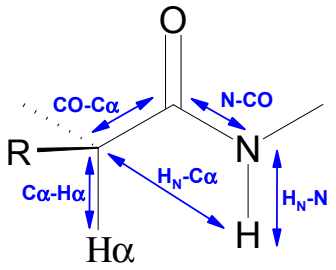

Figure 30
Couplings used in order to define the orientation of a peptide plane

Three types of experiment were used to determine these couplings. The first is based on the 3D HNCO pulse sequence and has been adapted for the measurement of RDCs [43]. While in a standard 3D HNCO experiment the couplings between the H_N and N nuclei and between the CO and Cα nuclei, respectively, are supressed, these couplings are retained in this spectrum (Figure 31, Panel a) and b)).

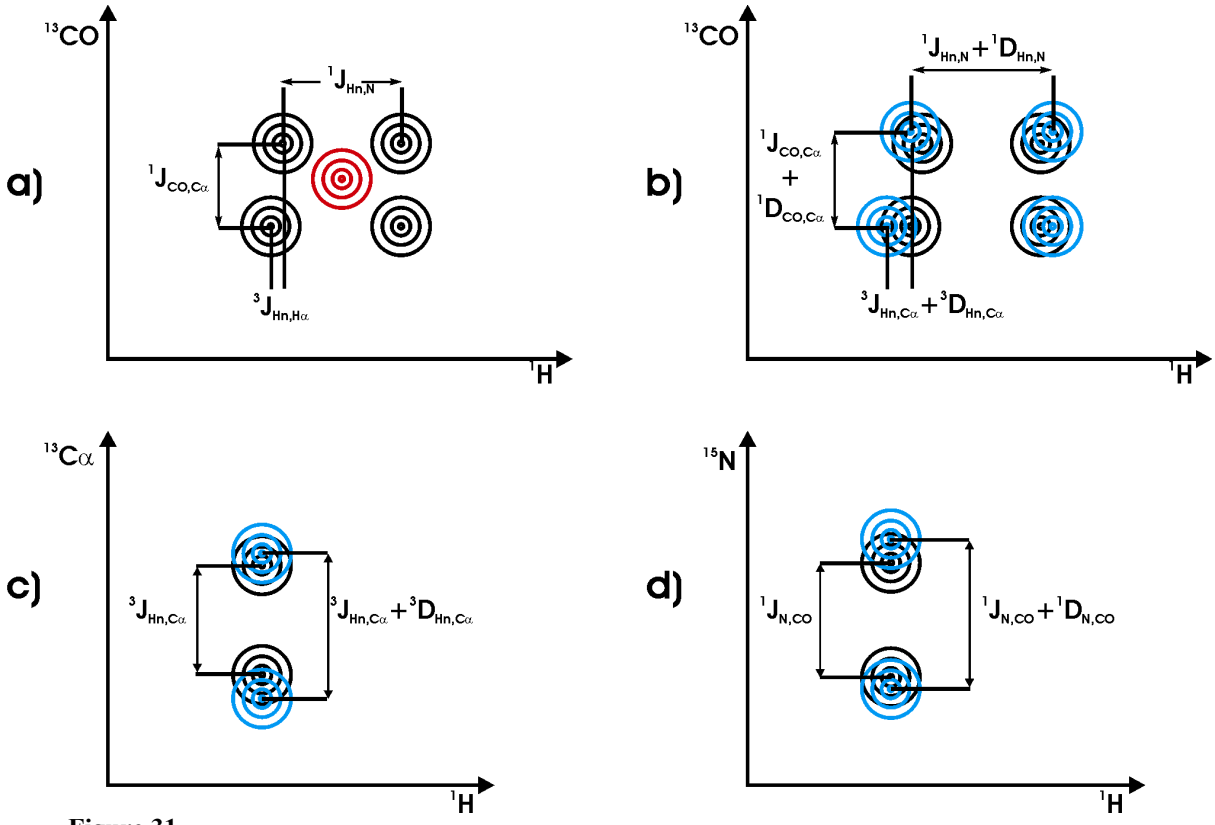

Figure 31
Schematic representation of spectra used in the determination of RDC's. a) **Standard HNCO peak (red), and without decoupling (black). The peak is split into four components due to the scalar coupling between the amide nitrogen and hydrogen atoms ($^1J_{Hn,N}$, ~ 90 Hz) in the proton dimension, and the scalar coupling between the carbonyl carbon and Cα carbon atoms ($^1J_{CO,Cα}$, ~ 55 Hz) in the carbon dimension (^{13}CO). The 3-bond coupling between the amide hydrogen and Cα carbon atoms is also indicated on the plot ($^3J_{Hn,Cα}$, ~ 3 Hz).** b) **Same non-decoupled spectrum from** a). **In a protein sample with partial alignment due to the addititon of phage (See Materials and Methods), the peaks shift by the relevant diploar couplings (cyan spectrum). The diploar couplings are small (several Hertz) and lead only to small changes in peak positions.** c) **HNCA experiment without (black) and with phage (cyan).** d) **N-CO experiment without (black) and with phage (cyan).**

The four peaks of the coupled 3D HNCO spectrum have the same ^{15}N-chemical shift as compared to the decoupled spectrum and thus appear in the same ^{15}N plane of this spectrum. They are split by the $^{1}J_{Hn,N}$ coupling in the proton dimension and by the $^{1}J_{CO,C\alpha}$ coupling in the ^{13}CO dimension (Where J denotes a scalar coupling, D is used for dipolar couplings). The $^{3}J_{Hn,C\alpha}$ coupling is visible as a difference in shift between two peaks almost sharing the same ^{1}H shift. This coupling is very small and was not used due to a relatively large uncertainty of the measurement.

The $^{1}D_{C\alpha,H\alpha}$ coupling is measured in an equivalent fashion, using an HNCA sequence without decoupling (Figure 31, Panel c)). The same is true for the $^{1}D_{N,CO}$ coupling, based also on the HNCO experiment [43] (Figure 31, Panel d)).

The splitting Δ between two peaks in an aligned spectrum corresponds to the sum of the respective scalar and dipolar couplings between the interacting nuclei. The value of the scalar coupling is measured from the unaligned spectrum, hence the diploar coupling D can be calculated according to Equation 3:

$$\mathbf{D = \Delta_{aligned} - \Delta_{unaligned}} \qquad \textbf{Equation 3}$$

The measured couplings were imported directly into the structure calculation program CNS[44], utilizing the SANI module developed by James Choy and Lewis Kay [42]. The required values of the axial and rhombic component of the alignment tensor, A_a and A_r were determined from the couplings using the program "Module" by Martin Blackledge [45]. The order parameter, S, was assumed to be 1 for all calculations.

Using NOE restraints alone did not lead to convergence to a sensible structure, as was expected for the ambiguous NOEs. In an attempt to calculate a structure of the EVH2 *monomer* first, violated NOE restraints were manually weakened to an upper distance limit of 6.0 Å in an iterative manner. After each calculation, the parameters A_a and A_r for RDC's were recalculated based on the calculated structure. Also included in the calculations were dihedral angle restraints derived with the TALOS program [46], and hydrogen bonds measured from D_2O exchange. Structure calculations in ARIA yielded the structure of a helix for each monomer but were unable to correctly orient the four monomers in the tetramer.

3.7 Model of the EVH2 monomer

It was possible to calculate a tentative model of the EVH2 monomer. Analysis of the Ramachandran plot shows 35 of 47 residues (81.4%) in the most favoured region, 7 in the additionally allowed region and 1 in the disallowed region (Serin 4, undefined part of the structure). The remaining four residues are two Glycines and two Prolines, all in the less well defined regions at the two termini of the molecule.

The monomer calculated on this basis is shown below (Figure 32). The model is assumed to be incorrect in certain aspects, mainly the kink seen near the N-terminus at residue 8. Also the sidechain rotamers are expected to be ill-defined.

Still this model has properties that are expected to be true for the tetrameric structure. As predicted from the helical wheel representation of the sequence, most hydrophobic residues cluster along one side of the helix. Although hydrogen bond restraints were used for the final round of calculation, the helix forms on the basis of NOE, dihedral angle and RDC restraints alone and is not introduced artificially by the hydrogen bonds.

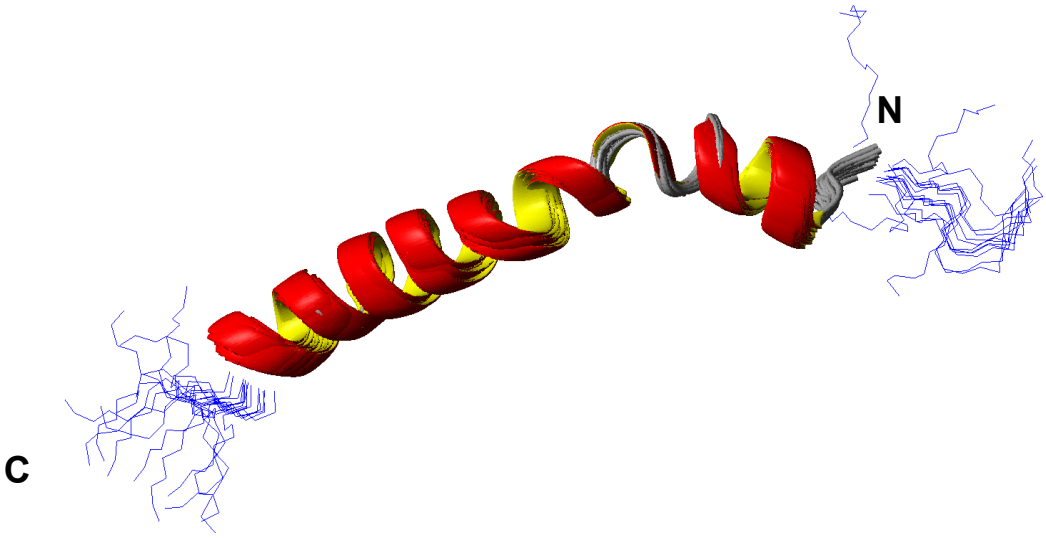

Figure 32
Ensemble of the 15 lowest energy structures of EVH2. Regions of defined secondary structure are shown in ribbon representation (residues 6-43). The termini are undefined and shown only as lines. The picture was created with MolMol[48].

The NOESY spectra do not show any obvious head-to-tail connections betwwen monomers, which would be present in a tetramer of C_{22} symmetry. It was therefore concluded that the human VASP EVH2 domain most likely forms a parallel coiled coil tetramer of C_4 symmetry.

3.8 Discussion

The EVH2 domain of the Ena/VASP proteins may be important for strengthening interactions between EVH1 target proteins via oligomerization. The frequency and distribution of PRMs in the primary sequences of the binding partners of their host proteins, and the oligomerization states of both interaction partners, can provide clues about the possible stoichiometry of the complexes formed. For example, the four EVH1-binding FPxϕP repeats in zyxin and ActA contrast with the single FPxϕP copy in vinculin and other EVH1-binding proteins. The oligomerization state of zyxin is monomeric [49]. This is in contrast to vinculin, which is known to multimerize [50, 51]. The ActA protein has been shown to exist in either monomeric [27] or dimeric states [52]. It has been shown here that the isolated, highly conserved structural element comprising the VASP EVH2 domain forms a tetramer in solution. Binding of tetradentate VASP to EVH1-binding proteins with fourfold repeats of the FPxϕP motif could result in either the formation of a 4:1 complex or a 16:1 complex, depending on the mode of binding.

EVH1-binding proteins with single copy FPxϕP-motifs may either exist as oligomers in their own right (as has been observed for vinculin) or may become oligomerized upon binding. Tetramerization thus renders VASP a potential tetradentate ligand for its different binding partners This could provide an independent mechanism for increasing not only the binding affinity of the otherwise low-affinity interaction with PRM-containing proteins. It could also facilitate the local concentration of effector proteins such as profilin or F-actin during targeted microfilament re-organisation. Evidence in favour of such affinity-modulating effects has been described with an *ena* allele encoding a mutant protein that lacks the C-terminal 49 amino acids. This oligomerization-deficient mutant is impaired in its interaction with zyxin and an Abl-SH3 domain despite the presence of the corresponding binding regions in its amino acid sequence [53]. VASP constructs with the proposed F-actin binding site but lacking the C-terminal 40 amino acids displayed considerably reduced F-actin-binding affinity *in-vitro* and *in-vivo* [13].

The symmetry of the tetrameric VASP may play an important role in complex formation with different interaction partners by imposing proximity and directional constraints on the architecture of the signalling complexes formed. A parallel, coiled-coil arrangement of monomers, with a C_4 rotational axis will align the four VASP molecules so that their EVH1 domains are orientated in the same direction.

The methods used in this work to study VASP oligomerization provided highly complementary information on the thermodynamic stability of VASP(336-380). The ^{15}N relaxation experiments

were valuable for studying the oligomerization behaviour of VASP(336-380) at higher protein concentrations (up to 5 mM), whereas analytical ultracentrifugation provided data at lower concentrations (down to 0.0625 mM). Independent use of these techniques showed clearly that VASP(336-380) exists exclusively as a tetrameric species under all conditions of temperature and protein concentration studied, with no detectable presence of any other oligomeric species at equilibrium. This observation may be of direct importance for understanding the molecular basis of the pronounced dominant negative effects of VASP EVH2 fragments harbouring amino acids 336-380 on VASP-mediated cellular processes such as epithelial or endothelial barrier formation [54, 55]. These dominant negative effects have been attributed to mixed tetramer formation of the N-terminal truncated VASP fragment with wild-type protein thereby perturbing its function. However, formation of mixed tetramers seems to be highly unlikely in light of the thermodynamic stability of tetrameric VASP(336-380) described here.