1 Introduction

A major part of early biochemical research was concerned with the study of proteins, often purified in tedious processes following a certain activity through steps of chromatography and centrifugation. Only in the 1940 did attention turn to genes and gene expression. For more than half a century, gene expression and its regulation have been the main focus of biochemical and biomolecular research. For some time now the centre of attention has been shifting back to gene products - proteins - as the real executives of most processes of life. As we are in the so called post-genomic era, with genomes of a growing number of species, including humans, deciphered, efforts turn to proteomics and structural genomics, with the focus on elucidating the structures of proteins and understanding their function and interaction.

Humans have used proteins, microorganisms to be more exact, for literally millenia in such ancient tasks as brewing beer (*Saccharomyces cerevisiae*) and making cheese (*Lactobacillus*), which for a long time were two of only a few ways of preserving food and drink. Without knowing about proteins and the real nature of the active agents in these "biotechnological" processes, humans have always made use of proteins.

Research in the late 19th century revealed proteins at the bottom of fermentation (e.g. for cheese, bread, tea etc.), alcoholic beverages and basically every process involving the turnover of substances. Hence proteins which are active as *enzymes* were discovered and intensely studied. Only much later were proteins discovered which possessed no enzymatic activity. These include structural proteins which form tissues, hair and muscle, e.g. keratin, silk or collagen. Other proteins are involved in signal transduction within cells, or from one cell to another, apart of nerve cells they are the hot wires of signalling. Soon a relatively small set of protein-protein interaction domains emerged. Well known examples include the interactions of SH3-, WW- [1] and Ena-VASP homology 1 (EVH1) [2] domains which interact with sequences containing PxxP, PPxY/PLPP and FPxϕP sequence motifs, respectively. The binding affinities of these interactions are generally in the range of 5-100 μM. Finely-tuned interactions of this type are vital for the accurate regulation of important signal transduction events involved in proliferation, migration, or differentiation of cells in the adult organism or during its development [3]. These domains are utilized in a broad variety of signalling processes in all organs and tissues, during all stages of the life span of an organism.

The Ena/VASP proteins have emerged as key components in remodelling processes of the cytoskeleton and the overall cell shape in a variety of cell types and cellular processes. These proteins contain an N-terminal EVH1 domain.

Ena/VASP proteins were first described in the context of platelet aggregation [4, 5]. A protein was detected in blood platelets that was phosphorylated in response to vasodilating agents like sodium nitroprusside, nitroglycerol and prostaglandins. Therefore the protein was named VASP (Vasodilator stimulated phosphoprotein) [6].

Homologues of VASP have since been found in a broad variety of species, including man, mouse, rat, dog, *Drosophila*, *Dictyostelium* and others [3]. They share a tripartite domain organization [7] as illustrated in Figure 1. An N-terminal EVH1 domain (Ena/VASP homology) is followed by a low complexity region, a proline-rich region, and a C-terminal EVH2 domain.

Figure 1
Domain structure of Ena/VASP proteins. The numbers correspond to the human VASP protein (SwissProt P50552).

The human VASP EVH1 domain binds peptides containing the minimal binding motif FPxϕP, where x is any amino acid and ϕ is hydrophobic [8]. It belongs to the PH superfamily of protein domains. A detailed discussion of EVH1 domains follows below.

The central proline rich region contains four copies of the sequence motif GPPPPP, one directly following the EVH1 domain, the three others in close succession starting at residue 168. It was shown that the latter region interacts with the protein profilin [9, 10] with an affinity of 0.5 µM [11]. Profilin is thought to recruit actin monomers to sites of actin polymerization via this interaction with VASP.

The EVH2 domain has a dual function. The region between residues 221 and 336 contains binding sites for both G- and F-actin. Starting at residue 233, human VASP contains the sequence motif KLRK necessary for binding G-actin. Mutations in this motif drastically reduce the affinity for G-actin and abolish the actin nucleating activity of VASP [12]. Residues 258 to 277 of human VASP have been shown to be necessary for binding to F-actin *in vitro* [13].

The C-terminal region of the EVH2 domain (residues 336 - 380 of human VASP) forms a coiled coil domain, necessary for tetramerization of VASP [13, 14]. The oligomerization state is required

for the actin polymerization activity of VASP, as constructs lacking the coiled coil domain fail to induce polymerization [12]. It has been shown that the Listerial ActA protein may interact with up to four VASP EVH1 domains simultaneously, hence the oligomerization state and symmetry of the oligomer are of interest with respect to the biological function of VASP.This domain will be discussed in detail in Chapter 3.

1.1 EVH1 domains are protein-protein interaction modules

EVH1 domains occur in a number of functionally diverse multi-domain proteins. A selection of these proteins is shown in Figure 2:

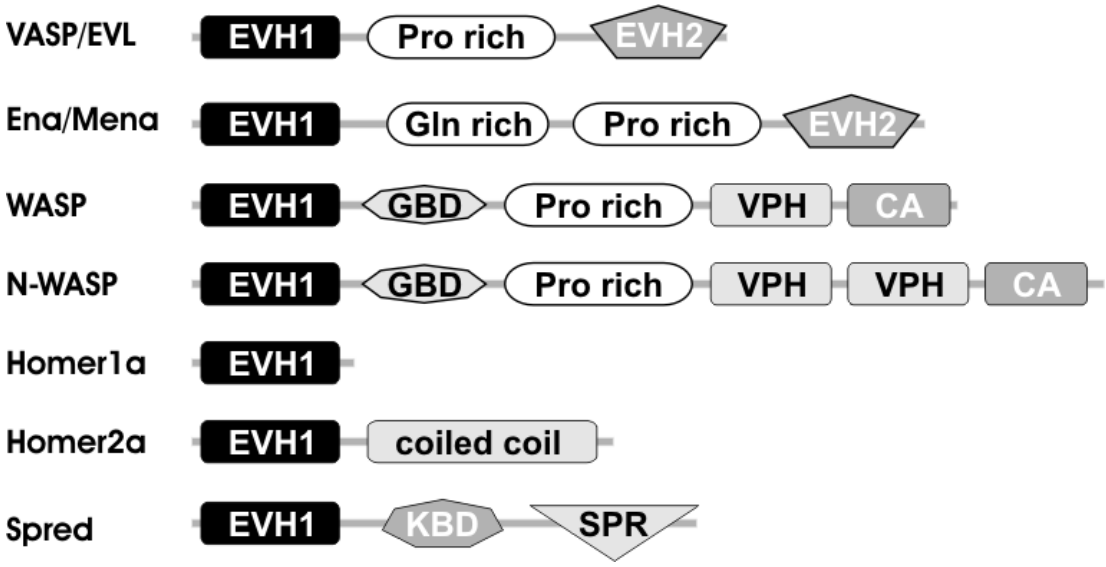

Figure 2
A few proteins containing EVH1 domains. The diagram is representative of the domain organization of these proteins, it is not drawn to scale. Domains are: EVH1/EVH2 - Ena/VASP homology 1/2; GBD - GTPase binding domain; VPH - verprolin homology; CA - cofilin homology and acidic; KBD - c-Kit binding domain; SPR - Sprouty related.

Ena/VASP proteins are associated with dynamic membrane regions such as filopodia and cell-cell junctions. They colocalize with actin filaments and are suggested to link signals from the extra cellular space via integrins to the actin cell skeleton [15].

The WASP family of proteins are believed to regulate actin assembly downstream of Cdc42 and Phosphatidylinositol-4,5-bisphosphate signalling pathways [16, 17].

Homer/Vesl proteins are enriched in neuronal tissue and are implicated in memory formation [18]. They bind to C-terminal regions of several receptor proteins.

Spred proteins are the latest members of the EVH1 family. They are involved in the Ras/Raf signalling cascade. Their role will be discussed below.

EVH1 domains act as protein-protein interaction domains, binding to peptides containing proline rich motifs (PRM's) (Interaction partners for Spred EVH1 domains have not been identified to date.). Protein-protein interactions mediated by PRM's and PRM-binding domains are becoming increasingly recognized as important players in signal transduction pathways [1, 19]. The high degree of conservation of the PRM's can be rationalized on the basis of high resolution structural data available for many of these complexes. These allow the detailed study of interactions between domain and ligand. From these studies it has become clear why proline is a prominent amino acid in these ligands [20].

1.2 Proline is special

Proline is the only natural N-substituted amino acid and its unusual backbone structure gives rise to a unique mechanism for highly specific recognition [20]. Stretches of several Pro residues adopt a special type of secondary structure, the PPII (polyproline II) helix, characterized by backbone angles $\phi = -78°$ and $\psi = +146°$. Every fourth Pro residue is exactly superimposed onto the first one when looking along the helix axis (Figure 3).

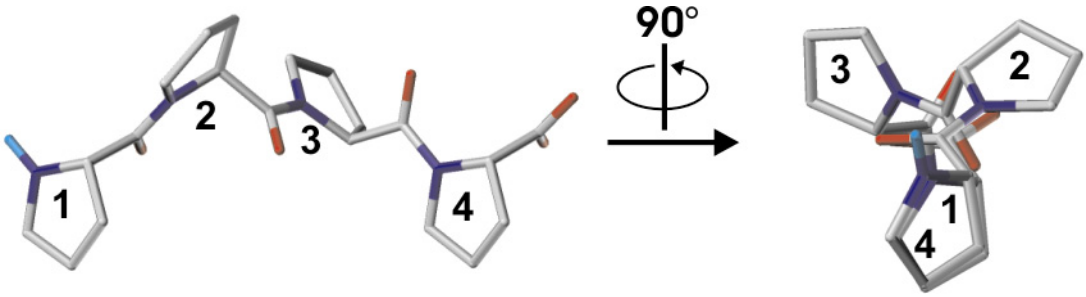

Figure 3
Structure of the canonical PPII helix, sideview and along the helix axis

This structure, along with the typical topology of binding interfaces for PRMs, has prompted the development of a model for these binding sites, described in detail in a recent review (Ball et al., Angewandte, in press). In short, the mode of binding of an EVH1 domain to the PRM can be

described using an "umbrella" model, where the stem of the umbrella is formed by a hydrogen bond formed between the carbonyl group of a residue corresponding to Pro 2 and an appropriate donor on the interacting domain surface, often the sidechain amine of a Trp residue.
The low affinities observed make these protein-protein interaction modules difficult targets for drug design. However, the importance of these interactions has already triggered several attempts to overcome the principal problems involved in designing ligands for protein surfaces [21]. Novel ligands for the human VASP EVH1 domain have been developed in the context of this thesis. This work is described in Chapter 2.

1.3 Spred proteins contain a new class of EVH1 domains

Recently a new family of proteins containing EVH1 domains was described by Yoshimura and coworkers [22], called Spred (Sprouty-related with EVH1 domain). The EVH1 domain in this protein was first identified in a study on the murine proteins Spred-1 and Spred-2, both bearing a C-terminal domain relating them to the human protein Sprouty-4.
Several Spred proteins have been identified to date. They consist of three distinct domains, namely the N-terminal EVH1 domain, a KBD (kinase binding) domain and the C-terminal SPR (sprouty related) domain (see Figure 2). The KBD domain binds the tyrosine kinase c-Kit, it is not related to other tyrosine kinase binding domains such as SH2, PTB or c-Met and was first described in Spred [22]. The SPR domain is necessary for the localization of Sprouty [23]. The Spred proteins are implicated in the regulation of the Ras/Raf signalling pathway. This pathway regulates proliferation and differentiation of cells in response to extracellular signals.
Spred has been shown to assocoiate with Ras and to inhibit the activation of MAP kinase by suppressing the phosphorylation and activation of Raf, which leads to an enhancement of the interaction between Ras and Raf. Both the EVH1 and SPR domain are necessary for proper function of Spred. All Spred proteins that have been described so far localize to the plasma membrane [24]. A specific interaction partner for a Spred EVH1 domain has not been identified to date.
The determination of the solution structure of the human Spred2 EVH1 domain is described in Chapter 4.