1 General Introduction

1.1 Protein-Protein Interactions Mediated through Non-Catalytical Protein Domains

Signal transduction events in eukaryotic cells involve the reversible assembly of large multiprotein complexes. These signaling 'machines' integrate and transmit the information required to steer important biological processes, such as cytoskeletal rearrangements, patterns of gene expression, cell cycle progression and programmed cell death. Most of the involved protein-protein interactions are arranged through specific, non-catalytical protein domains, most of them recognizing exposed sequences or structured motifs in their ligands. Densely stacked protein interaction domains are often used repeatedly in distinct proteins to mediate a wide range of regulatory processes [1]. These interaction domains (e.g. Src homology 2 (SH2), Src homology 3 (SH3), WW domains (name refers to two signature tryptophan residues)) with 40-150 amino acids can be considered as autonomous due to the fact that they fold and recognize their ligands independently of the other parts of the protein.

The interactions between the domains and their ligands are mediated through complementary structural features ('epitopes') in both partners. Most domains recognize linear peptides, and in those conserved residues of a small core-motif. SH3, WW, *Drosophila Enabled* (Ena)/Vasodilator-stimulated phosphoprotein homology 1 (EVH1) and ubiquitin E2 variant (UEV) domains recognize proline-rich segments [2-4], SH2 and phosphotyrosine binding (PTB) domains bind to phosphotyrosine-containing peptide sequences [5-7], forkhead associated (FHA) and WD40 (name refers to conserved W and D residues) domains interact with phosphoserine- and phosphothreonine-containing epitopes [8]. PDZ (postsynaptic density 95, discs large, zonula occludens-1) domains mostly recognize the C-termini of proteins [9, 10]. For particular peptide recognition modules within the same family, binding-partner specificity is determined by key residues flanking the core binding motif [11, 12]. The interaction between the domains and their ligands is always reversible and, as a rule, they show low affinities typically ranging from 1 µM to 500 µM [2, 13]. In addition to the proteins recognizing exposed features of their

binding partners, sometimes in post-translationally modified form, some of them can undergo homo- or heterotypic dimerization or oligomerization. This is a notable feature of domains that regulate apoptotic signal cascades, but also of α-sterile motive (SAM) and PDZ domains [14-16].

Besides protein-protein interacting domains, most signaling proteins contain catalytically active modules such as kinase or phosphatase domains as well as domains interacting with other cellular components like DNA or phospholipids. This modularity of protein structure is an evolutionary advantage which allows the recombination of different protein domains to proteins with new functions [17]. Usually, a more developed organism shows a higher number of domain family members, e.g. *S. cerevisiae* contains 24 SH3 domains, whereas more than 300 are present in the *H. sapiens* proteome. A similar situation is observed for PDZ domains (Table 1.1). Furthermore, the domain composition of proteins in a higher organisms is more divergent [18, 19].

Table 1.1 Species Distribution of PDZ Domains.

Species	PDZ domains	Proteins
Archaea	17	15
Bacteria	558	435
Eukaryota	3033	1784
Viruses	0	0
S. cerevisiae	3	2
A. thaliana	26	23
C. elegans	141	102
D. melanogaster	233	141
M. musculus	753	425
H. sapiens	781	430

Footnotes: From the SMART database (December 2003)

Complete genome sequences have revealed thousands of these domains, requiring improved methods for identifying their physiologically relevant binding partners. Bioinformatics-based methods of sequence alignment, profiling and Hidden Markov Models dramatically ameliorated the recognition of modular signaling domains from primary structure. In contrast, little progress has been made in developing accurate methods that predict short linear sequence motifs that these domains recognize within

proteins in a genome-wide manner. However, this precise knowledge regarding the preference of the binding motif of one of the protein domains enables the identification of new interacting partners within the proteome and the possibility to suggest cellular functions [20, 21].

1.2 Binding Specificities of PDZ Domains

PDZ domains were discovered as sequence repeats apparent in the three proteins postsynaptic density 95 (PSD-95), discs large (Dlg) and zonula occludens-1 (ZO-1). They have been firts referred to as DHR (discs large homology repeat) domains or GLGF repeats (after the highly conserved four-residue GLGF sequence within the domain) [22-24]. Shortly thereafter, the acronym PDZ (from the initial letters of PSD-95, Dlg and ZO-1) was proposed and adopted by the scientific community [25]. With the sequences of several genomes now available, we know that PDZ-containing proteins are widespread in metazoa, plants, bacteria and vertebrates [26], but they are surprisingly rare in yeast and viruses.

PDZ domains occur in one or multiple copies in mainly cytoplasmic proteins. PDZ domain-containing proteins can be classified into three principal families according to their modular organization. The membrane associated guanylate kinases (MAGUKs), including the proteins PSD-95, Dlg and ZO-1, which contain one or three PDZ domains, one SH3 domain, and a guanylate kinase domain (GuK), make up the first family. The second family comprises proteins consisting largely of PDZ domains. The number of PDZ domains can vary from two to more than ten in those proteins. The only interaction domains, which are found in such high numbers in proteins, are the WD40 or leucine-rich repeats [LRR] domains. However, oligomerization of the named domains is required for ligand binding, whereas PDZ domains function independently. The third family encompasses multi-domain proteins without a GuK domain [27].

PDZ domains are ~90 residues β-barrel structures flanked by α-helices [28] and mostly binding to the carboxyl termini (C-termini) of their interaction partners. Peptide ligand binding takes place in an elongated groove on the PDZ surface as an additional antiparallel β-strand, which interacts with the second β-strand (βB) (for nomenclature see Materials

and Methods Chapter 2.1.1) and the second helix (αB). As shown in Figure 1.1 for the third PDZ domain of the synaptic protein PSD-95 [29], the free carboxylate group at the end of the peptide ligand interacts through a carboxylate-binding loop preceding βB which containing the conserved motif (R/K)xxxGψGψ. This binding mode is called henceforth 'canonical'. The N- and C-termini of the PDZ domain are close to each other on the opposite side of the domain relative to the peptide-binding groove, an arrangement common to protein interaction modules.

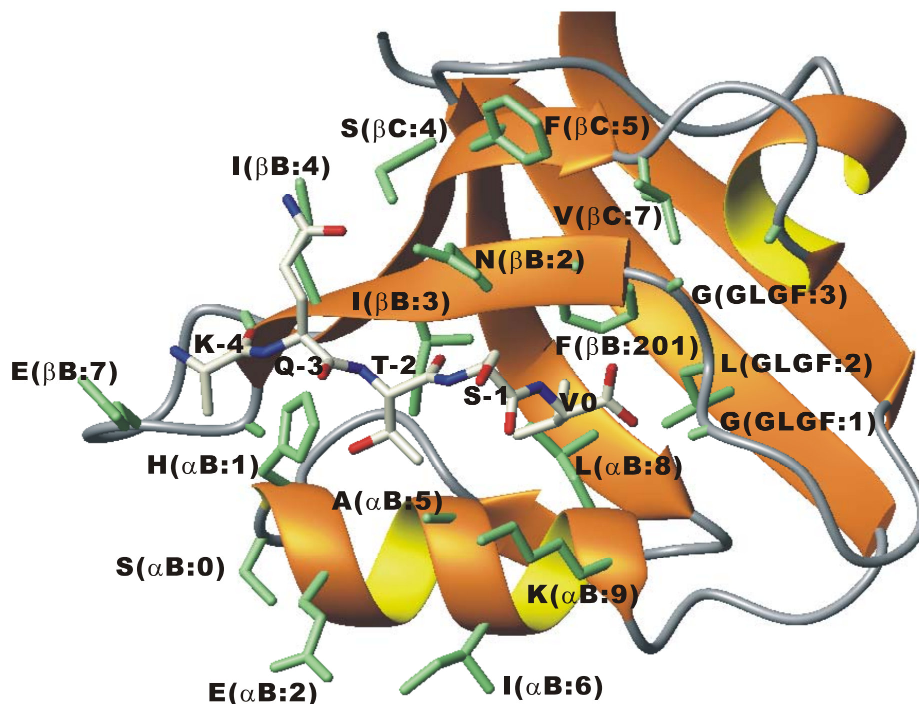

Figure 1.1 Ribbon Model of the Third PSD-95 PDZ Domain.
Crystal structure (PDB: 1BE9) of the third PDZ domain from the synaptic protein PSD-95 in complex with the C-terminal peptide (sequence KQTSV$_{COOH}$) derived from cysteine-rich interactor of PDZ three (CRIPT) [29].

The first studies with PDZ domains rapidly clarified which C-terminal residues were crucially required for protein interaction. Deletion of the C-terminal residue or mutation drastically reduced binding affinity, as did any mutation at position -2 (position 0 refers to the C-terminal residue, preceding residues are numbered as -1, -2. and so on). Songyang *et al*. [30] confirmed these findings using a degenerate peptide library and described two major classes of PDZ domains: Class I domains bind to peptides with the

consensus x(S/T)xΦ_{COOH} (for peptide nomenclature see Materials and Methods Chapter 2.1.1), whereas class II domains recognize the motif xΦxΦ_{COOH}.

The residues at positions 0 and -2 of the canonical peptide ligands have been described as playing the most critical role for the specificity and the affinity of the interactions. However, from the structural studies it is known that all residues between ligand positions 0 and -3 are involved in contacts to the PDZ domain [29, 31-33]. Accordingly, residues in positions -1 and -3 and residues down-stream of ligand position -3 also influence the individual PDZ domain's ligand propensity [32-34].

Some PDZ domains interact with lipids or form complexes with other PDZ domains (without C-terminal ligand recognition). The best examples of PDZ-PDZ interaction involve the PSD-95 (second PDZ) and syntrophin PDZ domains, both of which bind to the nNOS PDZ domain in a "head-to-tail" fashion [35]. The nNOS PDZ domain has a 30-residue C-terminal extension, that folds into a β-hairpin structure (β-finger) and docks into the binding groove of the syntrophin or PSD-95 PDZ domains [16, 36]. Alternatively, a PDZ domain can directly form a complex with itself through homodimerization [37], as observed for the glutamate receptor-interacting protein (GRIP) (Figure 1.2).

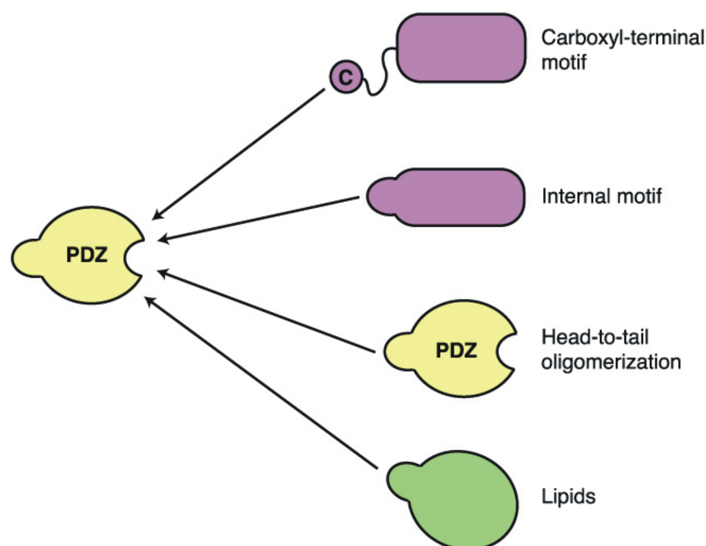

Figure 1.2 Possible PDZ Interaction Modes.
PDZ domains participate in at least four different types of interaction: Recognition of carboxyl-termini of proteins, recognition of internal motifs (β-finger), PDZ-PDZ dimerization (without C-terminal ligand recognition), and recognition of lipids.

PDZ domain proteins are frequently associated with the plasma membrane, a compartment where high concentrations of phosphatidylinositol 4,5-bisphosphate (PIP_2) can be found. Using gel-filtration assays and surface plasmon resonance (SRP), Zimmermann *et al*. [38] recently demonstrated direct interactions between PIP_2 and a subset of PDZ domains (syntenin, CASK, Tiam-1).

An interesting similarity exists among the PDZ, PTB and pleckstrin homology (PH) domains. Firstly, all three domains fold in a very similar fashion, despite the absence of sequence homology and the lack of any common peptide binding specificity [39]. Secondly, all three domains have the ability to interact with peptides and lipids, albeit with different preferences. A seductive idea is that these domains arose from a common ancestral domain and acquired increased binding specificities towards peptides and lipids during evolution through a diversification of the primary peptide sequence.

Almost all C-terminal peptides known to associate with PDZ domains can be grouped within ligand classes defined by Songyang *et al.* [30], although a growing number of 'exceptions' are being discovered. For instance, the neural nitric oxide synthase (nNOS) PDZ domain (class III) prefers peptides with the motif $G(E/D)xV_{COOH}$ [40] and the amyloid β A4 precursor protein-binding family A member 1 (APB1 or Mint-1) PDZ domain binds to the $(E/D)xW(C/S)_{COOH}$ peptide sequence [41].

PDZ domains vary in their stringency of specificity. For example, the PDZ domain of the protein interacting with C kinase 1 (PICK1) interacts with a multitude of peptides including those derived from the protein kinase C (class I, peptide: $QSAV_{COOH}$) and the ERB2 receptor (class II, peptide: $DVPV_{COOH}$) [42, 43]. Thus, a single PDZ domain can show both class I and class II specificity; the three-dimensional structural basis of this promiscuity remains to be determined.

1.3 Functions of PDZ Domains

PDZ domains have emerged as modules for establishing specific protein-protein interactions, important not only in the clustering of membrane proteins, but also in linking signaling molecules in a multiprotein complex at specialized membrane sites.

1.3.1 PSD-95 - NMDA

A paradigm for PDZ domain function is PSD-95, which contains three PDZ domains, a SH3 and an inactive GuK domain. PSD-95 binds to subunits of different channels, and thereby induces channel aggregation. Each individual PDZ domain of PSD-95 has distinct binding specificities, leading to the formation of clusters that contain heterogeneous groups of proteins. Thus, the ability of the third PSD-95 PDZ domain to bind the cell-adhesion molecule neuroligin may direct the N-methyl-D-aspartate (NMDA) receptor and the K^+ channel to specific synaptic sides. The channels may interact on their part with the first and second PDZ domains [44]. PDZ interactions may therefore both coordinate the localization and the clustering of receptors and channels, and provide a bridge to the cytoskeleton or to intracellular signaling pathways. By organizing such protein networks, PDZ scaffolding proteins increase the efficiency and specificity of signal transduction.

1.3.2 Syntrophin - nNOS

In addition to their interaction with the C-termini of ion channels and receptors, PDZ domains also form apparently specific associations with other PDZ domains. The second PDZ domain of PSD-95 and the single PDZ domain of syntrophin bind the unique C-terminal PDZ domain of nNOS. These interactions, respectively, bring nNOS in close proximity to the NMDA receptor channel in neuronal synapses and mediate the association of nNOS with sarcolemmal membranes in skeletal muscle [35]. The binding of nNOS to the NMDA receptor couples NMDA receptor mediated Ca^{2+}-influx to the activation of nNOS [45].

On the other hand, association with syntrophin in muscle cells localizes nNOS to the dystrophin complex [46], coupling NO production to muscle contraction. The resulting NO exerts a protective effect by increasing blood flow to match the heightened metabolic load of contracting muscle. Loss of this response in Duchenne muscular dystrophy may contribute to muscle degeneration [47].

1.3.3 Multimerization and Linkage of PDZ Domains

PDZ-containing proteins often-self associated to form multimers, which is of obvious utility for the assembly of macromolecular complexes. PDZ proteins can multimerize via other mechanism than the canonical binding of C-termini. PSD-95, for example, multimerizes by its N-terminal region [48]. It appears that also other PDZ domains show the propensity to undergo multimerization. For instance, the GRIP and the AMPA binding protein (ABP) (containing six or seven PDZ domains and no other recognizable domain) form homomultimers and heteromultimers through their PDZ domains [49, 50]. The inactivation-no-afterpotential-Drosophila (INAD containing five PDZ domains) multimerizes via PDZ-3 and PDZ-4, apparently without interfering with PDZ-ligand binding [51]. Thus distinct surfaces of the PDZ domain may be used for multimerization and peptide binding. This mode of interaction differs from the β-finger-involving interaction occuring in the association of nNOS with syntrophin or PSD-95, which is competitively inhibited by the peptide ligand [35].

1.3.4 Regulation of PDZ Domain Interaction

In recent years, many further PDZ domains were found, many structures were solved and a multitude of new PDZ domain interacting proteins were discovered. Nevertheless, the picture about the regulation of PDZ domain interaction is not yet complete.

One mechanism already investigated involves phosphorylation of the residues in position -2 or -3. Serine at position -2 of the inward rectifier K+ channel Kir2.3 (IRK4) falls within a consensus sequence for protein kinase A (PKA); phosphorylation of this site by PKA abolishes IRK4 interaction with the PSD-95 PDZ domain [52]. Phosphorylatable residues need not be at -2 position to affect PDZ binding. For instance, serine at position -3 of the AMPA receptor subunit GluR2 C-terminus can by phosphorylated by protein kinase C (PKC), and this modification prevents the GluR2 binding to the PDZ domain protein GRIP [53]. Phosphorylation of residues near the C-terminus is likely to be a common mechanism for a negative regulation of PDZ interaction.

Conversely, agonist-dependent activation of cell surface receptors is sometimes required to promote interaction with a PDZ protein, such as Na(+)/H(+) exchange

regulatory cofactor (NHERF) which interacts only with the activated β_2-adrenergic receptor [54].

1.4 PDZ Domain-Containing Proteins Involved in Human Diseases

The structural features of PDZ domains allow them to mediate specific protein-protein interactions that underlie the assembly of large protein complexes involved in signaling or subcellular transport. Not surprisingly, the disruption of these interactions can play a role in human diseases. Mutations in the gene encoding harmonin, a PDZ-containing protein, caused Usher syndrome type 1C, an autosomal recessive disorder characterized by congenital sensorineural deafness, vestibular dysfunction, and blindness [55, 56]. This was the first observed mutation in a PDZ-encoding gene linked to human disease. Subsequently, mutations in the periaxin gene, which also encodes a PDZ-containing protein, have been identified as a cause of Dejerine-Sottas neuropathy, a severe demyelinating form of peripheral neuropathy [57, 58]. Nevertheless, a human disease is not described since today, which could be direct related to mutations within the PDZ domain.

1.5 PDZ Domains Used in this Work

In this work, we used three different PDZ domains derived from three different proteins to analyze their binding specificity and selectivity. We choose these PDZ domains due to their different binding preferences reported in recent publications. For example, the AF6 and the ERBIN PDZ domain are known to bind both class I and II ligands whereas α-syntrophin interacts only with class I ligands.

1.5.1 The AF6 Protein and its PDZ Domain

The *AF6* gene was found to be fused to the *ALL-1* gene in a subset of acute lymphoblastic leukemias caused by chromosomal (t 6;11) translocation events - hence the name ALL-1 fused gene on chromosome 6 (AF6) [59]. There are three splice variants of the human AF6 protein [SP: P55196]: one with 1816 amino acids [SP: P55196-1], one with 1611

[SP: P55196-2] and one with 1743 amino acids [SP: P55196-3]. The rat protein afadin, a homologue of the human AF6 protein shows two splice variants (l-afadin [SP: O35889] and s-afadin [SP: O35890]), whereas the homologous proteins canoe [SP: Q24279] from *D. melanogaster* and Ce-AF6 [SP: Q9XY66] from *C. elegans* were found only in one version.

AF6 contains two N-terminal Ras-assiociation domains (RA), one forkhead-associated (FHA) and one dilute (Dil) motif, and a PDZ domain followed by an extended C-terminal tail interspersed with proline-enriched patches, making it a representative adaptor protein. AF6 fulfils its cellular functions by establishing interactions with cellular targets. AF6's cytoplasmic binding partners, such as the Ras subfamily (small monomeric GTPases, including, H- / K- / N-Ras, Rap1 and M-Ras) and the tight junction protein ZO-1 [60-62], interact with the two N-terminal RA domains. In contrast, the interactions with known integral membrane components such as a subset of ephrine receptor protein-tyrosine kinases [63, 64], the poliovirus receptor-related protein PRR2/nectin [65], the junctional adhesion molecule (JAM) [66] and the breakpoint cluster region protein (BCR) [67] are mediated by the AF6 PDZ domain. Furthermore, Su *et al.* [62] recently reported the interaction of the AF6 PDZ domain with an internal peptide motif of the GTPase-activating protein-related domain (GRD) of the GTPase-activating protein SPA-1. Although these features imply that AF6 may function by integrating the signals related to cell adhesion and cytoskeletal reorganization, its exact functions remain to be elucidated.

The importance of AF6 for the generation and/or maintenance of cell-cell junctions is demonstrated by the phenotype of AF6-deficient mice. The absence of AF6 results in embryonic lethality due to a disorganization of cell-cell junctions and to defects in the polarity of the embryonic ectoderm [68, 69].

Taken together, little is known about the regulation of AF6. In one case, the interaction of AF6 with the EPB3 tyrosine kinase receptor is increased after activation of the receptor by its ligand [63]. In the other case, the efficient binding of the C-terminus of BCR to the AF6 PDZ domain depends on the previous phosphorylation of AF6 through BCR [67]. Only after the formation of the AF6/BCR complex, AF6 interacts further with Ras, thereby disconnecting Ras from downstream signaling. This mechanism may be responsible for

maintaining the cells in a non-proliferative state in which the BCR kinase is constitutively active.

1.5.2 The ERBIN Protein and its PDZ Domain

ERBIN (ERB2 interacting protein) [SP: Q9NR18] was originally identified as a protein that interacts with the receptor protein tyrosine kinase ERB2 which is involved in cell proliferation and differentiation [70]. ERB2 is an epidermal growth factor receptor-related tyrosine kinase that is amplified in a high percentage of metastatic breast tumors and which is a causal factor in the development of some forms of breast cancers [71]. ERBIN plays a role in the localization of ERB2 at the basolateral membrane of epithelial cells as shown by mutation experiments [72]. Recent studies have shown that ERBIN is also present in high concentrations at neuronal postsynaptic membranes and neuromuscular junctions, where it also interacts with ERB2 [73]. Deletion of the ERB2 PDZ domain-binding motif results in an aberrant accumulation of ERB2 on the apical plasma membrane, presumably due to the inability to interact with a PDZ domain-containing protein that is necessary for the targeting of the receptor to, or retention at, the basolateral membrane [70].

In addition to this interaction, ERBIN is known to bind to p120-like catenin (δ-catenin, ARVC and p0071) in a PDZ-specific manner [74, 75]. ERBIN belongs to a new protein family named LAP (Leucine-rich repeat and PDZ-containing) proteins, which contain characteristically 16 leucine-rich repeats (LRR), a short LAP-specific domain at their amino termini and either one or four PDZ domains at their carboxyl termini. The LAP proteins are a family of scaffolding proteins that are involved in the formation of membrane-proximal complexes and the maintenance of epithelial and neuronal cell shape and polarity [76]. For example, in *D. melanogaster*, mutation of the scribble protein [SP: Q86QS7] results in a loss of epithelial cell polarity, changing morphology and uncontrolled, tumor-like growth [77]. Loss of function of LET-413 [SP: O61967] (homologue of human ERBIN) or scribble is embryonic lethal and results in disorganization of apical determinants [77, 78].

Putative SH3 and WW binding sites in ERBIN have also been identified [70]. Furthermore, there are two potential immunoreceptor tyrosine-based inhibitory motifs (ITIM) located before the PDZ domain, ^{1227}ANYSQI, and ^{1250}IDYLML, which might

attract, when phosphorylated, the tyrosine protein phosphatase SHP-1, even though Ala^{1227} is not typical for an immunoreceptor tyrosine-based inhibitory motif [79].

ERBIN has the potential to interact in a PDZ-independent manner with the N-terminal region of the epidermal bullous pemphigoid antigen-1 (eBPAG1) and the cytoplasmic domain of the integrin β4 subunit [80]. The ERBIN interaction with ERB2, which in turn associates with the integrin β4 subunit, suggests that ERBIN provides a link between hemidesmosome assembly and ERB2 receptor signaling. A recent paper by Huang *et al.* [73] shows that ERBIN suppresses extracellular signal-regulated kinase (ERK) activation by interfering with the activation of Raf-1 by Ras-GTP *via* its LRRs. This raises the intriguing possibility that ERBIN could shift Ras signalling from ERK pathways to other effectors and could function as a switch between different Ras effectors rather than as a mere inhibitor.

1.5.3 The SNA1 Protein and its PDZ Domain

The syntrophins are a family of intracellular peripheral membrane proteins that are components of the dystrophin-associated protein complex (DAPC) in skeletal muscle. The dystrophin complex appears to link the cytoskeleton to the extracellular matrix in skeletal muscle and stabilize the sarcolemma. The three known syntrophin isoforms, α1, β1 and β2 are encoded by separate genes and are differentially expressed. Each syntrophin has two pleckstrins homology (PH) domains followed by a C-terminal syntrophin-unique (SU) domain [81]. Inserted into a loop of the first PH domain is a single PDZ domain. Interest in the syntrophins was originally triggered by the observation that syntrophin is associated with dystrophin and other members of the dystrophin protein family, including utrophin and dystrobrevin. Mutations in the dystrophin gene lead to Duchenne and Becker muscular dystrophies [82].

The PDZ domain of α-syntrophin (SNA1) [SP: Q61234] is known to bind to the PDZ-containing N-terminal region of the neuronal nitric oxide synthase (nNOS), thereby targeting the enzyme to the sarcolemma [35]. Biochemical experiments have shown that the PDZ domain of SNA1 can bind likewise to the soluble α-2 chain of the guanylate cyclase (CYG4) [34], stress-activated protein kinase-3 (SAPK3) [83], glutaminase L [84], and aquaporin-4 [85].

Gee *et al.* [86] demonstrated that the PDZ domain of all three syntrophins bind to voltage-gated sodium channels CIN4 and CIN5 in a stable complex with dystrophin. Because dystrophin binds to actin, syntrophin may act as an adaptor that links voltage-gated sodium channels to the DAPC and therefore to the actin cytoskeleton and the extracellular matrix.

1.6 Characterization of Protein-Protein Interactions

1.6.1 NMR Spectroscopy to Study Protein/Ligand Interactions

The interactions between pairs of molecules are crucial in many molecular recognition processes in biology, the main examples being found in complexes of proteins with other proteins (e.g. in signaling processes), with small molecules (e.g. enzyme-substrate or protein-drug interactions), and with nucleic acids (e.g. protein-RNA complexes, transcription factors with DNA). Like many other biophysical techniques, nuclear magnetic resonance (NMR) provides information about protein-ligand interactions, ranging from structures to dynamics, kinetics and thermodynamics. About 17% of the structures deposited in the Protein Data Bank (PBD) have been solved by NMR spectroscopy, proving its importance with regards to structural proteomics for small- and medium-size proteins (below 30 kDa) [87]. Recent developments in NMR technology, together with the routine use of stable isotopes (^{13}C, ^{15}N, ^{2}H), have enabled the use of a much greater variety of NMR-based techniques for studying protein/ligand interactions. One example is the application of NMR as a method for high-throughput screening in drug development programs in the pharmaceutical industry.

The first step in any study of protein-ligand interactions by NMR is to estimate the rate of exchange between the free and complexed forms, since the choice of NMR experiments most appropriate is dependent on the exchange regime. On the NMR time scale, exchange regimes are broadly classified as slow, intermediate and fast.

Several main approaches are commonly used to extract structural information from protein-ligand systems. The *chemical shift changes* are one example for such NMR experiments. These are broadly used to locate ligand binding sites on proteins [88] or to determine the moieties within a ligand interacting with the protein. These experiments are

best carried out using isotope-editing methods, where only protons attached to either ^{13}C or ^{15}N are detected. This method can be used irrespective of whether the protein structure is known, as long as sequence-specific resonance assignments are available.

NMR studies of protein/ligand interactions are hydrogen exchange rates, surface mapping by paramagnetic agents, determination of protein or ligand ionization states and protein dynamics.

1.6.2 Other Methods to Study Protein/Ligand Interactions

The knowledge of accurate molecular structures is a prerequisite for rational drug design and for structure-based functional studies to aid the development of effective therapeutic agents and drugs. NMR, X-ray crystallography or electron microscopy can reliably provide the answer to many structure-related questions, from global folds to atomic details of bonding.

X-ray crystallography exploits the fact, that X-rays are diffracted by crystals. X-rays have the proper wavelength (in the Ångström range, $\sim 10^{-8}$ cm) to be scattered by the electrons of an atom of comparable size. Based on the diffraction pattern obtained from X-ray scattering of the periodic assembly of molecules or atoms in the crystal, the electron density can be reconstructed. Additional phase information must be extracted either from the diffraction data or from supplementing diffraction experiments to complete the reconstruction (the phase problem in crystallography). A model is then progressively built into the experimental electron density, refined against the data and the result is a quite accurate molecular structure. In contrast to NMR, theoretically no size limitation exists for the molecule or complex of interest.

Electron microscopy (EM) has established itself as a powerful method to study structure and function of biological macromolecules. The method exploits the fact that images of unstained molecules represent two-dimensional projections of molecules in their native state. From such projections, the structure can be obtained by combining data of different views and by using different reconstruction techniques depending on the nature of the specimen. However, unstained biological samples are extremely susceptible to radiation damage inflicted by the incident electron beam. The impact of radiation damage can partly be reduced by rapidly freezing (vitrifying) the specimen slowing down the diffusion of

radicals produced by electron irradiation. Although there are no upper limits to the mass of the particle, the need to identify individual particles in the noisy raw images impose a lower size cutoff of approximately 250 kDa of molecular weight. Smaller complexes or individual protein molecules need to be in the form of 'two-dimensional crystals' or to posse helical symmetry in order to make the analysis feasible. A three dimensional model could be then calculated (through Fourier transformation) from a set of collected two-dimensional projections (EM pictures) from different views of the sample. In contrast to the diffraction pattern resulting from X-ray, EM images contain both amplitude and phase information.

1.6.3 Prediction of Protein/Ligand Interactions

The genome sequencing projects have provided a large number of coding sequences for which we have little or no functional information [18, 19]. In fact, the function of 30-35% of encoded proteins remains elusive. Large-scale studies of protein-protein interaction networks have been carried out in several organisms to understand their role in the cellular context. Currently, much effort is devoted toward studying gene, and hence protein, function and regulation by analyzing mRNA expression profiles, gene disruption phenotypes, two-hybrid interactions, and protein subcellular localization [90-94]. The analysis of these large networks may enable the development of new drugs that could specifically interrupt or modulate protein interactions contributing to malfunction in human diseases.

A large number of interactions in the cell are mediated by modules, which are found frequently but in different combinations in corresponding proteins. Typically, these modules mediate protein-protein interactions through the recognition of exposed features in their target proteins (for details see Chapter 1.2). Several approaches, based upon the screening of repertoires of combinatorial peptides (phage display, peptide arrays), have been developed to investigate the recognition specificity of these domain families. Furthermore, new developed databases of protein interactions, like DIP (Database of Interacting Proteins) [95], BIND (Biomolecular Interaction Network Database) [96] and MINT (Molecular Interaction database) [97], give the opportunity to exchange information of postulated or found interactions (reviewed in [98]). Additionally, a number of

computational methods have been developed for the prediction of protein interactions from given genomic information, extending into the prediction of the residues that participate in the interacting surface (reviewed in [99]).

For the prediction of PDZ domain binding specificity, only two different methods are currently known. The recently developed algorithm *iSPOT* (Specificity Prediction of Target) [100] is based on a statistical method that, by taking into account the frequency with which residue x in the domain binding surface faces residue Y in a collection of ligand peptides at any of the contact positions, permits the evaluation of the likelihood that any domain binds to any peptide. The other approach used for computer-aided determination of PDZ binding specificity is based on the algorithm called PERLA (Protein Engineering Rotamer Library Algorithm) [101, 102], which enables the identification and sorting of amino acid sequences that have optimal stability for desired three dimensional structure [103]. Nevertheless, these both methods depend on the availability of structural information of at least one domain/peptide complex and/or of a collection of experimentally determined ligands for the respective domains. This means, that both methods are only able to determine putative interactions based on the known PDZ domain/ligand complexes. For example, the *iSPOT* database relates the AF6 PDZ domain to a class II PDZ domain, despite its ability to bind class I ligands.

1.7 Objectives of this Work

PDZ domains are protein interaction modules that play a key role in cellular signaling. Most of them bind specifically to the C-termini of their interaction partners, which are often belonging to the families of receptors or ion channels. As representative examples, we analyze here the specificity of the AF6, the ERBIN and the α1-syntrophin (SNA1) PDZ domains.

The major focus of this thesis is to obtain a description of the individual PDZ domain/ligand specificity-relationships. For a better understanding of these ligand specificities, it is necessary to begin with the investigation of the structure (by modeling and NMR titrations) and the binding properties of the respective PDZ domains (by dissociation constant measurements and application of peptide libraries). Based on the

obtained PDZ domain/ligand interaction data, we have the opportunity to rationalize why PDZ domains bind to a diverse set of peptide sequences and furthermore, to explain the difference of the class-specific PDZ domain interactions. As exciting application of this study, the design and development of a PDZ domain super-binding and/or intervening ligands will be conceptually feasible.

Therefore, one objective of this thesis is to develop an improved strategy to generate cellulose membrane-bound peptides with free C-termini via SPOT synthesis for PDZ domain screening. The method previously published by Hoffmüller *et al.* [104] needs to be improved to obtain shorter reaction times, together with high coupling efficiencies of the different chemical compounds with regard to automating the SPOT synthesis process.

Altogether, the main goal of the presented work is to give a better understanding of the molecular mechanisms of the PDZ domain interplay in a biological context. Thus, the identification of novel PDZ domain/ligand interaction contributes not only to the studies of the particular pathway, but also too much wider fields of biomedical research. A large set of PDZ domain/ligand interactions intends to provide a foundation for the search of such interplay by both experimental and computational means.