UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1741

Exploring the Role of the PDZ Domain in a Supramodule

LOUISE LAURSEN

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2021

ISSN 1651-6206
ISBN 978-91-513-1181-4
urn:nbn:se:uu:diva-438160

Dissertation presented at Uppsala University to be publicly examined in A1:107a, BMC, Husargatan 3, Uppsala, Thursday, 20 May 2021 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Per Hammarström (Linköping University, Department of Chemistry).

Abstract
Laursen, L. 2021. Exploring the Role of the PDZ Domain in a Supramodule. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1741. 67 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1181-4.

The postsynaptic density (PSD) is a large, dense and membraneless compartment of proteins associated below the postsynaptic membrane bilayer, and which constantly undergoes morphological alteration in response to synaptic activity. Formation of PSD is associated with liquid-liquid phase separation of scaffold proteins in complex with other PSD proteins. PSD-95, one of the most abundant scaffold proteins, contains five domains: PDZ1, PDZ2, PDZ3, Src homology 3 (SH3), and guanylate kinase-like (GK) domain. The domains are functionally divided in two supramodules: PDZ1-PDZ2 and PDZ3-SH3-GK (PSG). Multi-domain proteins are characterized through their isolated domains in most studies and represented by "beads on a string" model, which means that the function of a single domain is independent of the context. In this thesis, the properties of PDZ3 and PSG are compared to elucidate how and when PSD-95 can been characterized by the simple "beads on a string" model. Kinetic characterization of CRIPT binding to PDZ3 showed a two-state mechanism, but a more complex mechanism involving two conformational states upon binding to PSG. The results were consistent with recent structural findings of conformational changes in PSD-95, altogether showing that conformational transitions in supertertiary structures can shape the ligand-binding energy landscape and modulate protein-protein interactions. Next the allosteric networks in a PDZ:ligand complex were experimentally mapped, both in isolation and in the context of a supramodular structure. Data showed that allosteric networks in a PDZ3 domain has high dependency on the supertertiary structure. Furthermore, equilibrium and kinetic folding experiments were applied to demonstrate that the PDZ3 domain folds faster and independently from the SH3-GK tandem, which folds as one cooperative unit. However, concurrent folding of the PDZ3 domain slows down folding of SH3-GK by non-native interactions, resulting in an off-pathway folding intermediate. Finally, the interactome of PSG in PSD was mapped. PDZ3 and PSG show high specificity for peptides with type I PBM. Interestingly, two proteins called SynGap and AGRB1 only bind with high affinity to PSG and forms concentration dependent liquid droplets. The results show how context in terms of supertertiary structure alter affinity and function, and suggest a model for how PSD anchor to the postsynaptic membrane. Altogether, the findings in the thesis show that binding energy landscape, interactome, allosteric network, folding mechanism and phase separation are dependent on the context, which suggest that we need to be careful in interpretation of data obtained from isolated domains in multi-domain proteins.

Keywords: PSD-95, PSG, PDZ3, supramodule, liquid liquid phase separation, allostery, folding, binding, evolution, post synaptic density, SynGap, CRIPT, AGRB1

Louise Laursen, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-1181-4
urn:nbn:se:uu:diva-438160 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-438160)

Supervisor:
Per Jemth, Professor
Department of Medical Biochemistry and Microbiology
Uppsala University
Sweden

Co-Supervisor:
Ylva Ivarsson, Professor
Department of Chemistry
Uppsala University
Sweden

Co-Supervisor in PDZnet:
Søren Østergaard, Principal Scientist
Novo Nordisk
Denmark

Opponent:
Per Hammarström, Professor
Department of Chemistry
Linköping University
Sweden

List of Papers

The PhD thesis is based on the following papers.

Paper I:
Functional interplay between protein domains in a supramodular structure involving the postsynaptic density protein PSD-95
Laursen L, Karlsson E, Gianni S, Jemth P. J Biol Chem. 2020;295(7):1992-2000.

Paper II:
Supertertiary protein structure affects an allosteric network
Laursen L, Kliche J, Gianni S, Jemth P. Proc Natl Acad Sci USA. 2020; 117(39): 24294-304.

Paper III:
Divergent evolution of a protein-protein interaction revealed through ancestral sequence reconstruction and resurrection
Laursen L, Calyseva J, Gibson TJ, Jemth P. Mol Biol Evol, 2021. 38(1): p. 152-167.

Paper IV:
Dissecting inter-domain cooperativity in the folding of a multi-domain protein (submitted)
Laursen L, Gianni S, Jemth P

Paper V:
Determinants of affinity, specificity and phase separation in a supramodule from Postsynaptic density protein 95 (manuscript)
Laursen L, Inturi R, Østergaard S, Jemth P

Review:
Seeking allosteric networks in PDZ domains
Gautier C*, Laursen L*, Jemth P, Gianni S. Protein Eng Des Sel. 2018.

* These authors contributed equally to the work

Contents

Abbreviations

X	Any amino acid
Φ	Hydrophobic amino acids
AA	Amino acids
AGRB1	Adhession GPCR receptor B1
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ASR	Ancestral sequence reconstruction
BLAST	Basic local alignment search tool
CASK	Calcium/calmodulin dependent serine protein kinase
CBP	CREB-binding protein
CC	Coiled coil
C. elegans	*Caenorhabditis elegans*
CRIPT	Cysteine-rich interactor of PDZ3
Dlg	Discs large
DMC	Double mutant cycle
D. melanogaster	*Drosophila melanogaster*
D. ponderosae	*Dendroctonus ponderosae*
FL	Full length
GK	Guanylate kinase like
GKAP	Guanylate kinase-associated protein
H. sapiens	*Homo sapiens*
H. vulgaris	*Hydra vulgaris*
IDP	Intrinsically disordered protein
INSY2	Inhibitory synaptic factor 2A
LLPS	Liquid liquid phase separation
L. loa	*Loa loa*
MAGUK	Membrane associated guanylate kinases
ML	Maximum likelihood
MSA	Multiple sequence alignment
NMDA	N-methyl-D-aspartate
NMR	Nuclear magnetic resonance
O. bimaculoides	*Octopus bimaculoides*
PALS1	Proteins associated with Lin Seven 1
PBM	PDZ binding motif
PDZ	PSD-95/Dlg/ZO-1
PPI	Protein protein interaction
PRS	Perturbation response scanning
PSD	Postsynaptic density

PSG	PDZ3-SH3-GK
PTP	Protein tyrosine phosphatase
S. kowalevskii	*Saccoglossus kowalevskii*
SCA	Statistical coupling analysis
SH3	SRC homology 3
SynGap	Synaptic Ras GTPase-activating protein

Introduction

Logic will get you from A to B. Imagination will take you anywhere. A quote that describes the essential in basic research. Curiosity is the desire to know, thus cultivates the seeds for creativity. My PhD thesis represents a general comparison of single and multi-domain proteins to reveal if multi-domain proteins can be described as "beads on a string" represented by the function of a single domain or if protein context affects function. Protein folding, allosteric network and binding interactome for a single protein domain and a multi-domain protein are compared. We choose one of the most studied model systems, the PSD-95/Dlg/ZO-1 (PDZ) domain. The choice of model system reflects the association of the PhD thesis with a European Training Network (PDZnet, Marie Curie Fellowship program with 14 individual projects). The specific choice of PDZ3 from postsynaptic density protein 95 (PSD-95) among 268 different PDZ domains is due to the fact that PDZ3 is well characterized from a structural and functional point of view, which makes it a good model system. Thus, the structure and function of PDZ3 and a three domain supramodule, PDZ3-SH3-GK (PSG) from PSD-95 are compared to explore general concepts:

1) How do interdomain interactions in a multi-domain protein affect propagation of allosteric signals in PDZ3? This was elucidated by mapping the allosteric network in PDZ3 and PSG to reveal if they are similar.

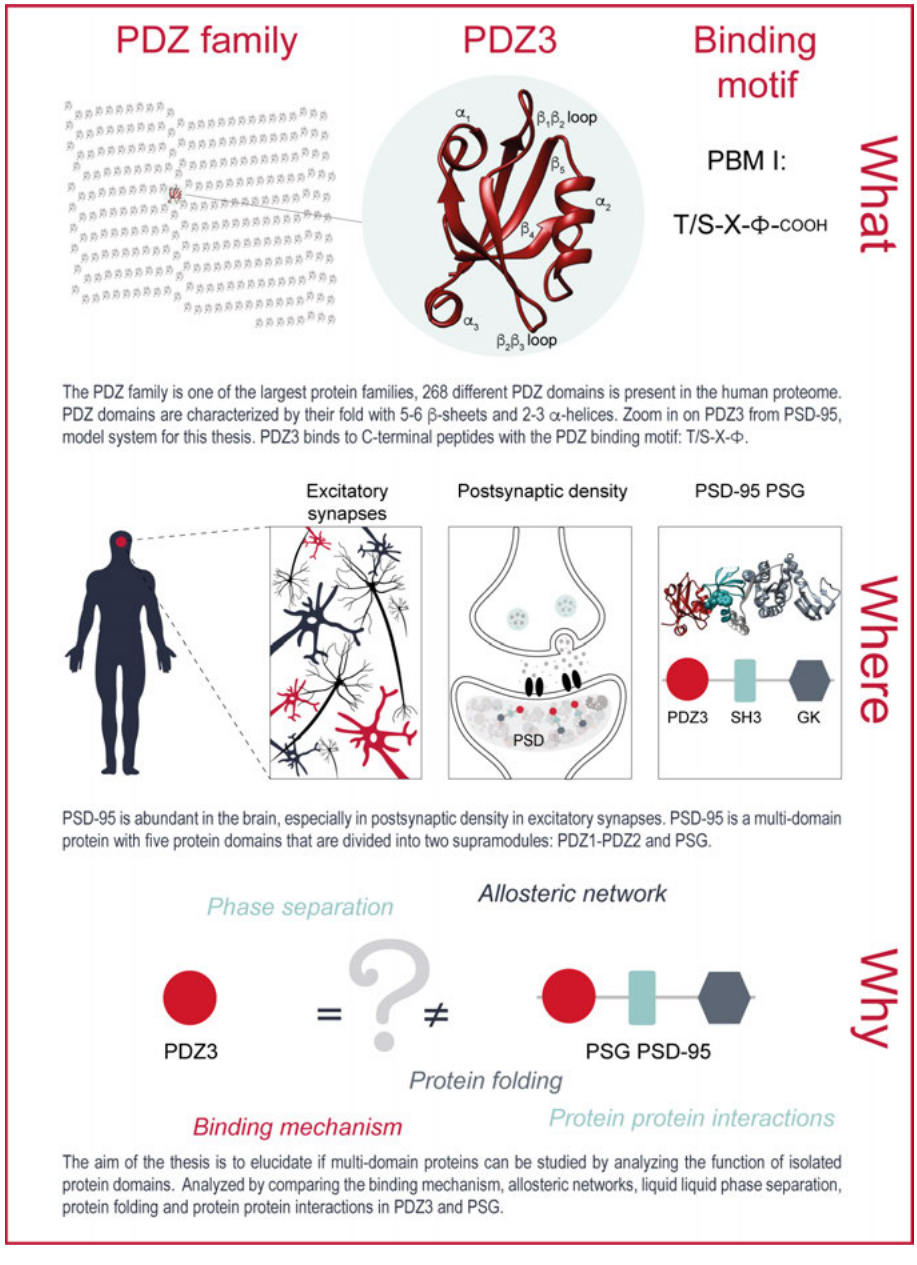
2) Can we use ancestral sequence reconstruction and resurrection to reveal why a binding motif from an essential protein:protein interaction is changed in some species and understand how this affects function?

3) What factors determine if a multi-domain protein acts as a supramodule or as "beads on a string". This was elucidated by comparison of interactome and folding pathway for PDZ3 (single domain) and PSG (multi-domain).

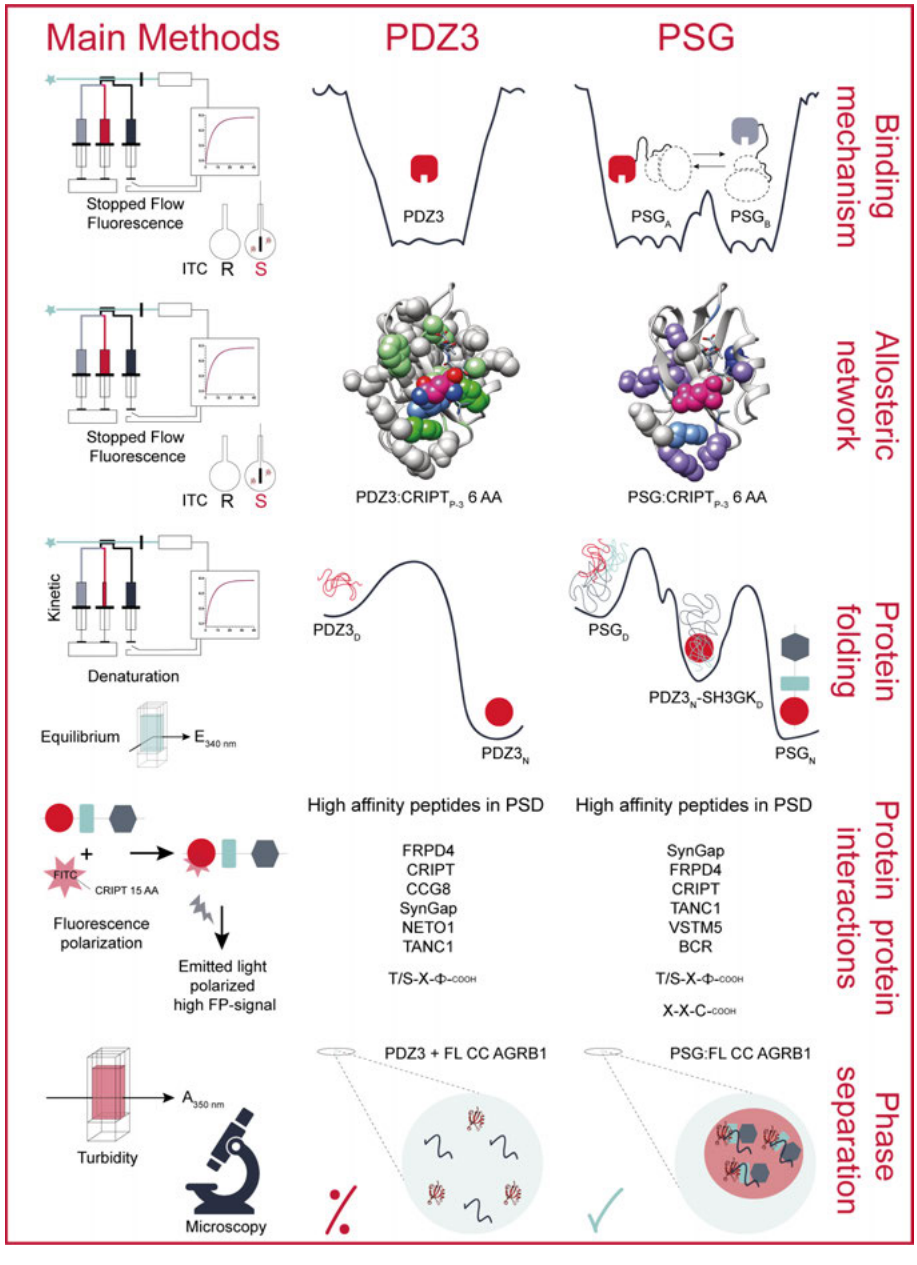
4) Is the specificity profile affected by mutation in the allosteric network? It is postulated that mutation in the allosteric network of PDZ3 can switch its specificity from type I to type II ligand. This was elucidated by screening the potential interactome in the PSD against the PDZ3-SH3-GK supramodule and selected mutants.

I hope the projects of the thesis can plant the seeds to explore multi-domain proteins with supramodule function in a cellular context. A better understanding of supramodular context will expand the druggable proteome by more specific targets. The work in the thesis has contributed to the overall goal of the PDZnet: expand our knowledge of PDZ domains to seek for new potential drug targets in a broad spectrum of diseases such as cancer and neurodegenerative diseases.

Graphical Introduction



Graphical Summary



Brief Summary of Manuscript Contents

Paper I: Functional interplay between protein domains in a supramodular structure involving the postsynaptic density protein PSD-95

Aim: To reveal the binding mechanism upon binding of CRIPT to the PDZ3-SH3-GK (PSG) supramodule.

Results: Kinetic analysis of the PDZ3:CRIPT interaction revealed a simple two state mechanism shown by fitting experimental data to a single exponential function. However, analysis of PSG:CRIPT revealed a more complicated mechanism as shown by fitting experimental data to double exponential function.

Conclusion/Perspective: We suggest that two conformations of PSG are present, and that both bind to CRIPT with similar affinity, but with different rates. This suggests that the interdomain architecture of PSG can shape the energy landscape and modulate protein-protein interactions.

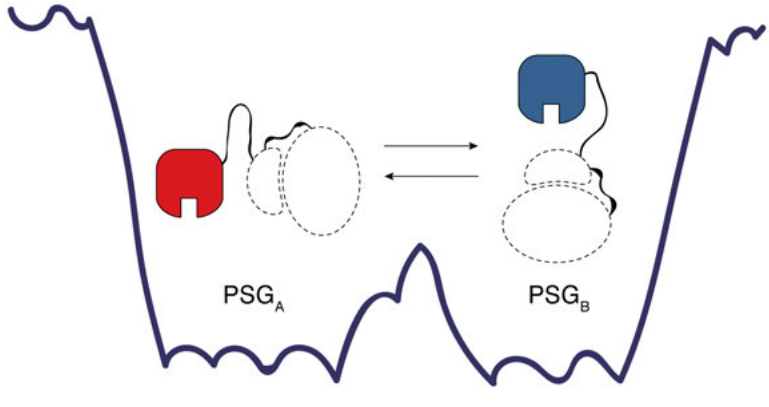


Figure 1: Energy landscape of PSG conformations.
Binding of 6 AA CRIPT to PSG revealed two conformations of PSG named PSG$_A$ and PSG$_B$. Binding studies and modelling of data suggested similar binding constants, but different peptide-binding rates for conformation PSG$_A$ and PSG$_B$.

Paper II: Supertertiary protein structure affects an allosteric network

Aim: To address whether the allosteric network propagation in PDZ3 is affected by interdomain interactions from the adjacent protein domains in PSG supramodule?

Results: Thermodynamic double mutant cycles were applied to map the allosteric network in PDZ3 and PSG upon binding of CRIPT 6 or 15 AA. 15 mutants were analyzed in the background of PDZ3 and PSG. Comparison of allosteric networks mapped onto the structure of PDZ3 and PSG revealed a shift from mainly positive to negative coupling free energies and new coupling patterns in the α_3 helix, $\beta_1\beta_2$ loop and $\beta_2\beta_3$ loop.

Conclusion/Perspective: Allosteric coupling in PDZ3 is affected by the supertertiary structure of the supramodule, which is likely a general feature of the MAGUK family for allosteric propagation. Our findings reveal a striking sensitivity of allosteric networks to perturbation by adjacent domains with general implications for their identification and characterization.

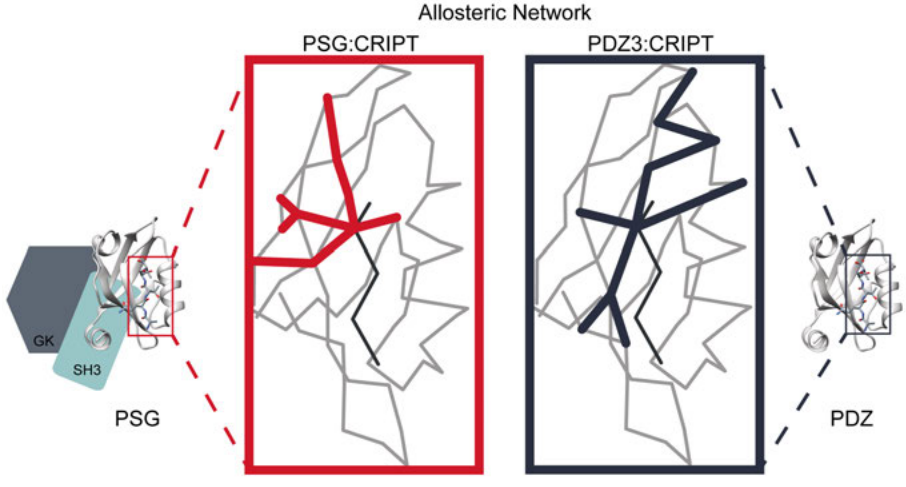


Figure 2: Characterization of allosteric networks in PSG and PDZ3.
Residue by residue in the allosteric networks of PDZ3 and PSG were compared, which revealed that the allosteric network in PDZ3 is affected by the the context of the SH3-GK domain as the allosteric network of PDZ3 and PSG are different.

Paper III: Divergent evolution of a protein-protein interaction revealed through ancestral sequence reconstruction and resurrection

Aim: How is the binding of the PDZ3:CRIPT interaction affected by loss of the type I PDZ binding motif (T/S-X-Φ-COOH) in some species?

Results: Ancestral sequence reconstruction and resurrection were applied on CRIPT and PDZ3. The PDZ3:CRIPT interaction was analyzed horizontally (seven present days species) and vertically (five ancestors). All of the selected extant and resurrected PDZ3 domains bind to the ancestral Eukaryota CRIPT with high affinity. CRIPT from nematodes and insects have lost the type I PBM, therefore a low affinity was reported from these species of the PDZ3:CRIPT complex. An extended α_3 helix can increase the affinity of some weak PDZ3:peptide interactions. The α_3 helix hypothesis was tested for nematodes and insects, but full length α_3 helix could not rescue the weak interaction for species with lost type I PBM in CRIPT.

Conclusion/Perspective: The PDZ3:CRIPT interaction was strong in the ancestors of Bilateria, Deuterostomia and Protostomia, which was revealed by reconstruction and resurrection of ancestral proteins. But a weak PDZ3:CRIPT interaction appeared at the time of the ancestor of Hexapoda animals (including mainly insects). An established protein-protein interaction involved in cellular scaffolding is subject to dynamic evolution at the time point of the common ancestor of Hexapoda, thus raise questions about the physiological function of the PDZ3:CRIPT interaction in present days species from the Hexapoda node.

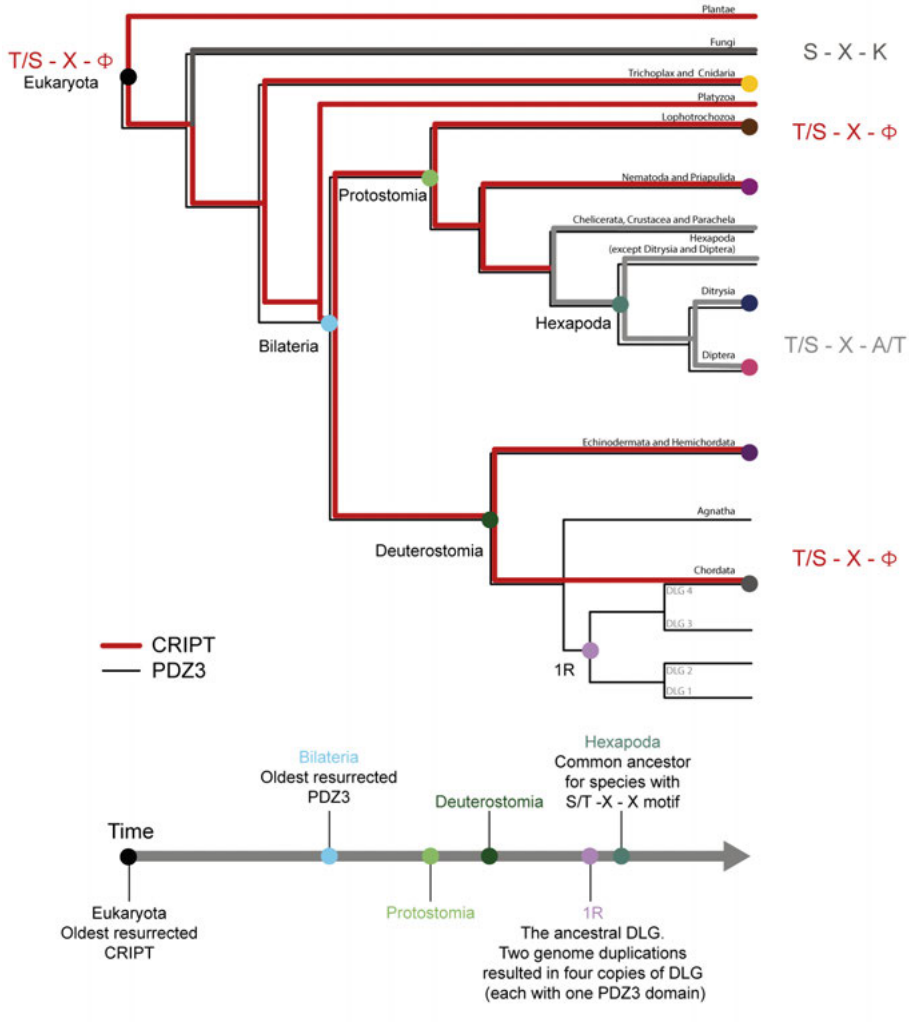


Figure 3: Species tree depicting PDZ3 and CRIPT in different phyla.
Phylogenetic tree showing the presence of PDZ3 (black) and CRIPT (red) at different time points. Proteins were reconstructed, resurrected, expressed and purified to allow horizontal and vertical analysis of the PDZ3:CRIPT interaction. The analysis is performed to reveal how the lost type I PBM in CRIPT from *D. melanogaster*, *D. ponderosae* and Hexapoda affects binding of CRIPT to PDZ3. Colored circles represent the evolutionary nodes where proteins were resurrected, expressed and purified to address the evolution of the PDZ3:CRIPT interaction.

Paper IV: Dissecting inter-domain cooperativity in the folding of a multi-domain protein

Aim: Protein properties are an outcome of biophysical, functional and evolutionary forces, that can clash and form trade-offs. Therefore, we want to understand the protein folding pathway for a supramodule, the three-domain PSG from PSD-95. PSG is a supramodule, as it has properties (binding, dynamic) that are distinct from its single domains, but it is not known how the folding pathway is affected by the supramodular structural organization.

Results: Equilibrium and kinetic folding experiments showed that PDZ3 fold faster and independently from SH3-GK tandem, which folds as one cooperative unit. Folding of PDZ3 domain slows down the folding of SH3-GK by non-native interactions, resulting in an off-pathway folding intermediate. Two kinetic phases were observed for folding of PSG and SH3-GK, therefore we suggested occurrence of a cis-trans proline isomerization in the SH3-GK tandem.

Conclusion/Perspective: Folding of individual domains impair the overall folding of a multi-domain protein. Further experiments need to reveal if the off-pathway folding intermediate is a functional trade-off with a physiological role, as shown for another PDZ multi-domain [21], and if the two observed kinetic phases are due to cis-trans proline isomerization as a mechanism for intrinsic autoinhibition. The main conclusion is that the common SH3 domian folds as a folding unit together with GK domain rather than an independently protein domain.

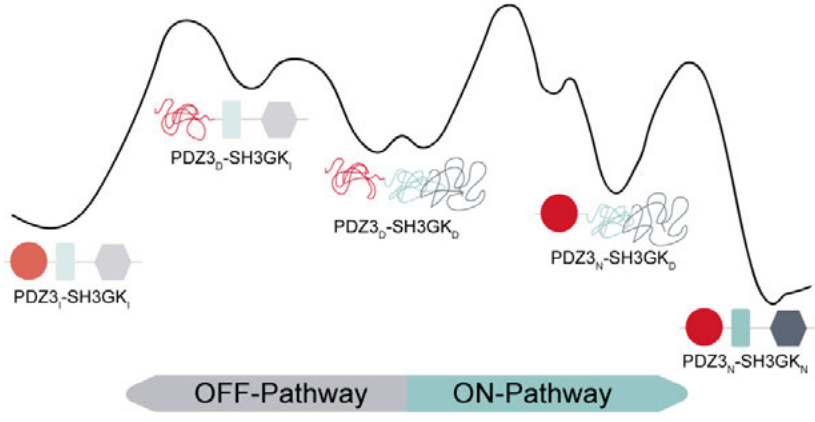


Figure 4: Energy diagram of PSG folding.
The energy diagram shows the main states of PSG in its folding pathway from denatured (D) to native (N) state, illustrated as on-pathway. Due to lower energy barrier from the denatured state of PSG to an intermediate (I) state than to N-state, PSG can be trapped, illustrated as off-pathway.

Paper V: Determinants of affinity, specificity and phase separation in a supramodule from Postsynaptic density protein 95

Aim: To elucidate how the context affects the interactome of PDZ3 in multi-domain protein PSG by comparing specificity of PDZ3 and PSG for C-terminal peptides derived from proteins in PSD. Multimerization is suggested to be essential for liquid liquid phase separation in PSD, which PSG induces upon formation of complex with SynGap by multiple interactions, thus a specific property for PSG in comparison to PDZ3. How will mutants that enhance and disturb affinity of native ligand CRIPT affect ligand specificity and liquid liquid phase separation? Last, can we identify new proteins that induce liquid liquid phase separation in complex with PSG?

Results: PSG and PDZ3 bind with high specificity to ligands with type I PBM. SynGap and AGRB1 bind with significantly higher affinity to PSG than PDZ3 and induce phase separation upon mixing with PSG. PSG_{G322A} and PSG_{G335A} have mutations located outside of the binding pocket in PDZ3 thus increased affinity, specificity and size of liquid droplets, whereas mutations located within the ligand binding pocket relaxed the specificity, and resulted in adaption to a new binding motif, for example INSY2 (Φ-X-Φ) bound with high affinity to PSG_{F325A} and PSG_{I327V}.

Conclusion/Perspective: Our results show how context in terms of supertertiary structure alter affinity and function, and suggest a model for PSD to anchor to the postsynaptic membrane through phase separation with the intracellular C-terminal part of AGRB1.

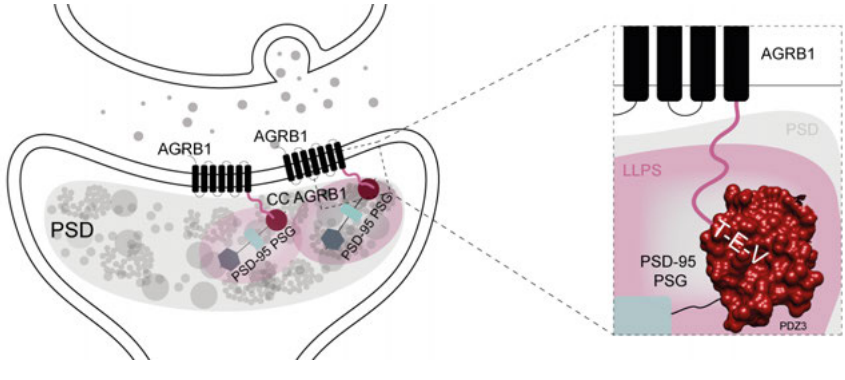


Figure 5: Anchoring of PSD to membrane bilayer through phase separation.
Illustration of an excitatory synapse with PSD located below the membrane bilayer of the synapse. PSG and C-terminal part of AGRB1 are highlighted. The two proteins form phase separation upon mixing. The compartmentalization of PSG and AGRB1 next to the membrane bilayer suggests a mechanism for PSD to anchor to the membrane bilayer.

Review: Seeking allosteric networks in PDZ domains

Aim: To elucidate why we observe different allosteric networks in PDZ domains depending on the use of approach.

Results: Comparison of eight allosteric networks probed by *in silico* and experimental approaches in the PDZ domain showed overlapping patterns around the binding pocket. However, differences were revealed, especially in loops and in α_3 helix of PDZ3.

Conclusion/Perspective: Probing allosteric networks in PDZ domains seems to be method dependent due to the relatively weak energetic connectivity and complexity of amino acid residue interactions involved in the networks. Therefore, we propose that a combination of experimental and *in silico* approaches are required to fully map the allosteric network in single protein domains.

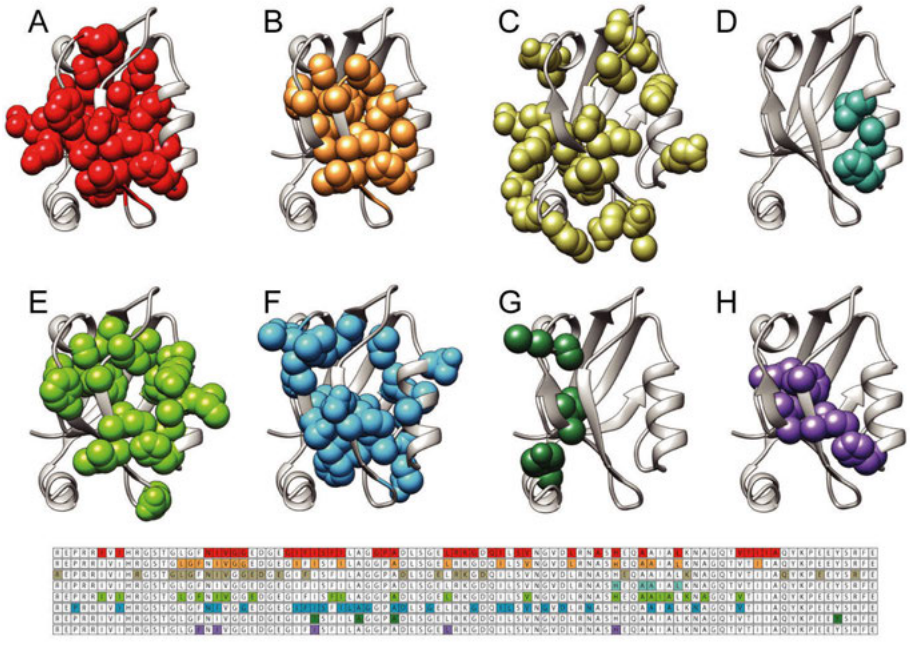


Figure 6: Allosteric networks in PSD-95 PDZ3.
Allosteric networks determined with different methods and shown in form of colored ball and stick in the structure of human PSD-95 PDZ3 (PDB:1BE9): A) Perturbation response scanning (PRS) [129], B) Statistical coupling analysis (SCA) [50], C) Molecular dynamics simulation [39], D) Deep coupling scan (DCS) [40], E) Thermodynamic double mutant cycle (DMC) [128], F) Conservation mutation correlation analysis (CMCA) [127], G) Rigid-residue scan (RRS) [126] and H) Monte Carlo path (MCPath) [125]. Adapted from figure 3 in the review by Gautier et al. [13].

The Model System PDZ3 from PSD-95

PDZ3 from PSD-95 is used as model system in this thesis. PSD-95 is expressed in postsynaptic neurons and is the most abundant protein in the postsynaptic density (Figure 7A). The major functions of PSD-95 are structural organization of the excitatory postsynaptic density and localization of transmembrane proteins to specific sites, that regulate the synaptic strength. Diminished levels of PSD-95 are associated with aging and neurodegenerative diseases [1] such as Huntington, Alzheimer [2], major depressive disorder [3], schizophrenia and autism spectrum disorder [4]. PSD-95 is a scaffold protein characterized by its multiple protein-protein interaction domains: PDZ1, PDZ2, PDZ3, SRC Homology 3 (SH3) and Guanylate kinase like (GK). PDZ1 and PDZ2 form one structural tandem. PDZ3-SH3-GK (PSG) forms a tertiary structure (Figure 7B and C), denoted supramodule, for which the interdomain organization can change upon ligand binding [5]. PSG is the core of the Membrane Associated Guanylate Kinase (MAGUK) protein family (Figure 8), a family of scaffold proteins with essential roles in cell-cell communication, cell polarity and cellular signal transduction [6].

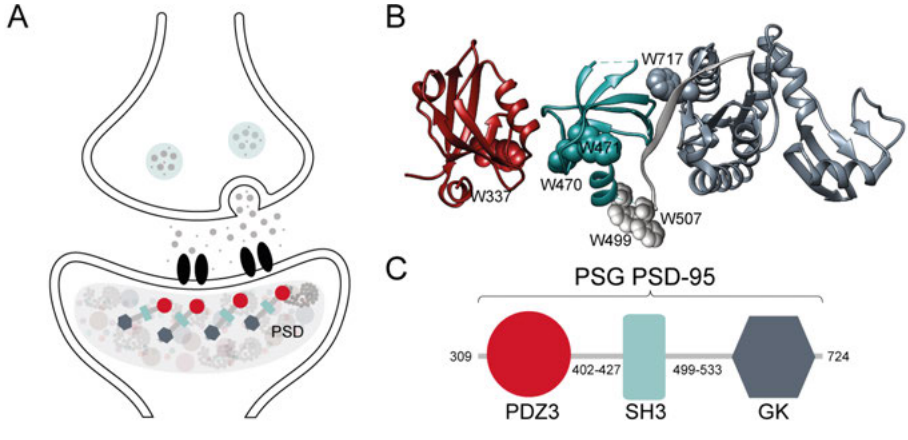


Figure 7: Location and structural organization of PSG supramodule.
A) Illustration of a neuron with the postsynaptic density highlighted in the postsynaptic neuron. PSG from PSD-95 is highlighted in the postsynaptic density due to its high abundance and important function. B) Crystal structures of individual domains in PSD-95 PSG are shown from N- to C-terminal: PDZ3, SH3 and GK. PDB code: PDZ3 (314W) and SH3-GK (5YPR). C) Ribbon diagram of PSG with all native (470, 471, 499, 507, 717) and engineered (337) Trp residues highlighted. Lengths of linkers between PDZ3-SH3 and SH3-GK, respectively, are shown.

PDZ3 from PSD-95 is a small protein domain, that share its compact globular fold with around 268 [7] other PDZ domains in the human proteome. The PDZ domain contains 5-6 β sheets and 2-3 α helices. In the groove, formed between β_2 sheet, α_2 helix and a loop between β_1 and β_2, a C-terminal peptide can bind. The C-terminal peptide will most often arrange as an anti-parallel β strand in connection to the β_2 sheet. This is also the case for cysteine-rich interactor of PDZ3 (CRIPT), a native ligand for PDZ3 (Figure 9A). The

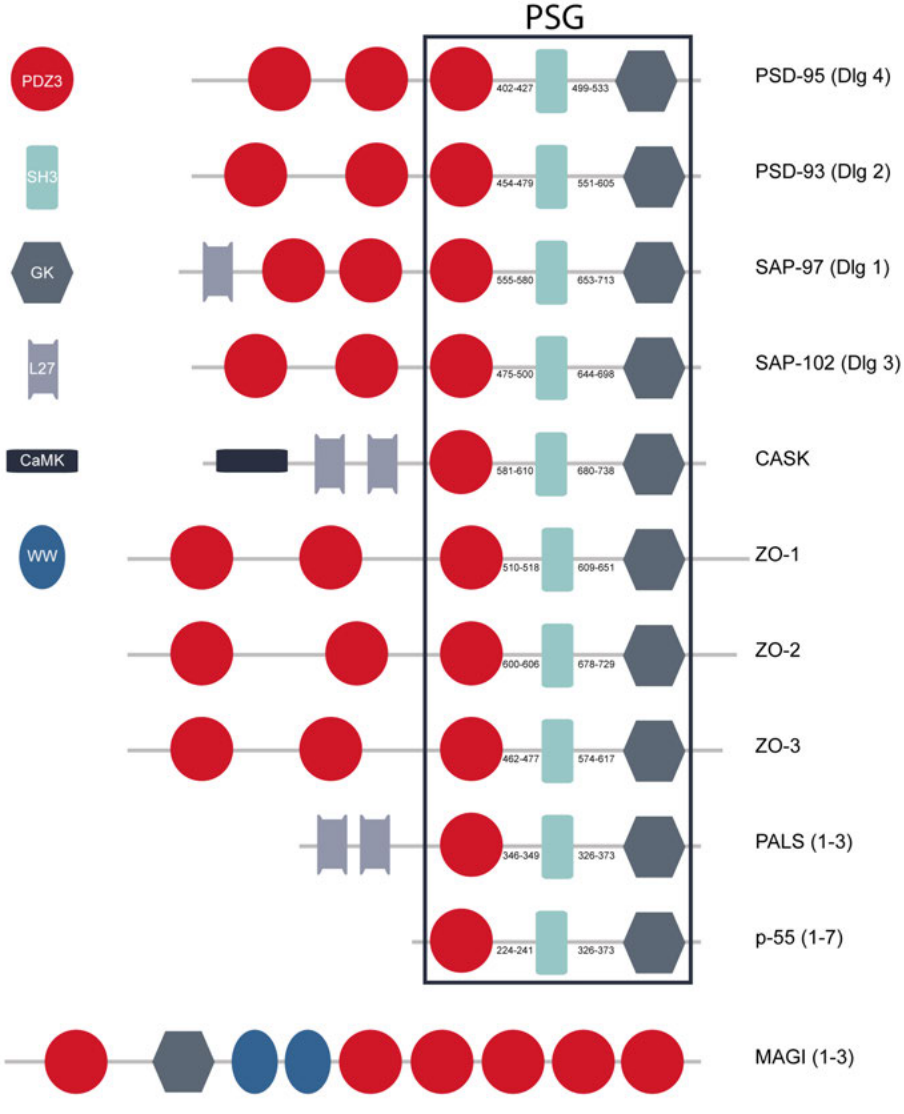


Figure 8: MAGUK family.
Domain structure of selected members from the Membrane associated guanylate kinases (MAGUK) protein family with the basic structural domains: PDZ, SH3 and GK domain. Lengths of PDZ3-SH3 and SH3-GK linkers are highlighted.

interdomain architecture of the supramodule PSG is altered upon binding the flexible C-terminal part of CRIPT to PDZ3. Binding of CRIPT changes the structural network of post synaptic density thus promoting PSG to localize and assemble AMPA receptors and Glutamate A1 subunits [8]. The molecular system is essential for dendritic growth during development. *In vitro* studies have shown that only the last 6 amino acids of CRIPT are required for full affinity (CRIPT binds in the low micro molar range to *Homo sapiens* PDZ3) [9], however it is not known if full length CRIPT (12 kDa) is required to promote the associated functions of the PDZ3:CRIPT interaction. *In vivo* studies show that PDZ3:CRIPT interaction promote localization of AMPA receptors [8] and induce interdomain conformational change in PSG from compact to extended conformation [5].

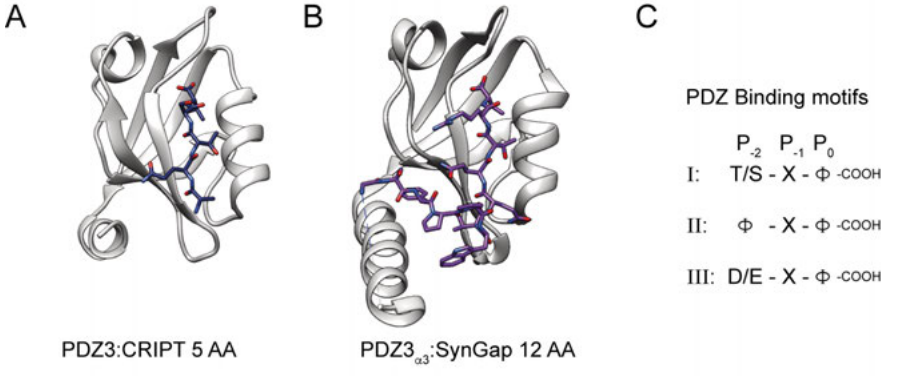


Figure 9: PDZ3 canonical binding of ligands with type I PBM.
A) Structure of PDZ3 with a bound C-terminal peptide of CRIPT with a T-X-V motif. PDB: 1BE9. B) Structure of PDZ3 with a bound C-terminal peptide of SynGap with a T-X-V motif. PDB: 5JXB. C) The three most common PDZ binding motifs: type I (T/S-X-**Φ**), type II (**Φ**-X-**Φ**) and type III (D/E-X-**Φ**), showing the importance of the P_{-2} position for specificity.

PDZ3 has several ligands e.g. Synaptic Ras GTPase-activating protein (SynGap) [10] (Figure 9B), and CRIPT [11] (Figure 9A), but the PDZ3:CRIPT interaction is the most studied interaction from a structural and biophysical point of view. The complex of PDZ3:CRIPT was the first structure of PDZ domain solved by crystallography [12] and has been a conceptual model system to study allostery in single protein domains [13, 14], achieved without conformational change in the domain upon binding of ligand [15]. All the established ligands have one thing in common, they bind to PDZ3 via a T/S-X-**Φ**-COOH motif, also called a type I PDZ binding motif (PBM) (Figure 9C). An early characterization of PDZ domain specificity, based on the observation that amino acid residues P_0 and P_{-2}, counted from the C-terminal, dictate most of the affinity, resulted in a classification system with three types of PDZ binding motifs: type I (T/S-X-**Φ**-COOH), type II (**Φ**-X-**Φ**-COOH) and type III (D/E-X-**Φ**-COOH) [16] (Figure 9C). The classification system is an

oversimplification; thus, specificity depends often on more than the nature of P_{-2} and P_0 [17, 18]. The PDZ3:SynGap interaction is an example of this, since it requires a 10 amino acid SynGap peptide and a full length α_3 helix from PDZ3 [10] (Figure 9B). In figure 10 a list of PDZ3:ligand interactions are reported to illustrate the dependence between primary structure length and associated functions of PDZ3:ligand interactions. Many studies have been conducted with the PDZ3:CRIPT complex as model system, however, very few studies have used the full length α_3 helix, which was reported to be essential for the PDZ3:SynGap interaction and function [10]. This is probably due to the fact that a decent affinity is reported by the use of a short α_3 helix in PDZ3. However, it is not well explored if the function of the PDZ3:CRIPT interaction is affected by the length of the α_3 helix. Therefore, the objectives of the thesis were to explore the importance of the context (e.g. α_3 helix length) regarding affinity, stability, folding and allosteric network propagation. In paper III, a weak PDZ3:CRIPT interaction were reported for nematodes and insects, as CRIPT from these species lack type I PBM in its sequence. To test the hypothesis that extended α_3 helix and longer peptide derived from CRIPT can increase the affinity of weak PDZ:peptide interaction, we measured the affinity between CRIPT 15 AA and $PDZ3_{\alpha3}$. However, the affinity only increased up to two-fold, therefore, not the significant contribution (15-fold) as reported for the PDZ3:SynGap interaction [10]. The function of the α_3 helix is explored indirectly in paper I, II, IV and V, where we compared the allosteric network, phase separation, folding mechanism and ligand specificity for PDZ3 in presence and absence of the SH3-GK tandem. These few cases show how important it can be to consider the primary structure length of a model system to explore the full spectrum of associated functions of protein-protein interactions. In the specific case of PDZ3:CRIPT interaction the dissociation constant is not sensitive to primary structure length, but the associated functions of PDZ3:CRIPT interaction are not well explored to exclude primary structure length dependence. Nevertheless the primary structure length have significant impact on the PDZ3:SynGap and PDZ3:Adhession GPCR receptor B1 (AGRB1) interaction and associated properties of complexes.

Beads on a String or Supramodule

Scaffold proteins in the PSD are multi-domain proteins, with several well folded protein domains without intrinsic enzymatic activity, thus making them ideal for protein network organization [19]. Many scaffold proteins in PSD contain several PDZ domains, and as PSD-95, a highly conserved domain organization through evolution. The conserved multiple PDZ domain organization suggests positive selection for function [19]. Scaffold proteins have been studied as "beads on a string", but a number of studies have demonstrated that the conserved multiple PDZ domain organization have target-specific function [20, 21]. If the properties of a PDZ tandem are distinct from those of its constituent single domains, the structural and functional unit is referred to as a supramodule [19]. A supramodule is a multi-domain functional unit with spatial organization and function distinct from the single domains. In paper I, II, IV and V the PSG supramodule was studied (Figure 10), our studies with the PSG supramodule showed a distinct behavior from isolated PDZ3 in folding (Paper IV), allosteric network (Paper II), peptide interactome and LLPS (Paper V). A recent study reported that a spatial organization of PDZ3, SH3 and GK domain is required for SynGap high affinity interaction and liquid liquid phase separation (LLPS) [10]. Therefore, multi-domain scaffold proteins cannot be studied and described by the simple "beads on a string" model.

PDZ domains are organized in a supramodule to increase stability of single domains [22, 23], expand interactome to dimeric targets [24] and constrain structure [23]. Sequence analysis of multi-domain PDZ proteins reveal that PDZ domains are often connected with a short linker of conserved length. Another example is PDZ1-2 tandem from PSD-95. The linker spacing the two domains is short to decrease conformational space of domains [25]. The same study showed by increasing the length of the linker that the interdomain interaction between PDZ1 and PDZ2 were lost. The obvious question to ask is, if all PDZ tandems act as supramodule or the high prevalence of multiple PDZ domains in scaffold proteins is a consequence of positive selection during evolution for PDZ domain in response to increased signaling in multicellular organisms [26]. Several PDZ tandems have been reported as supramodules, e.g. PDZ1-2 PSD-95 [25], PDZ4-5 GRIP [27], PDZ1-2 Whirlin [20]. The structural organization of all the listed supramodules are

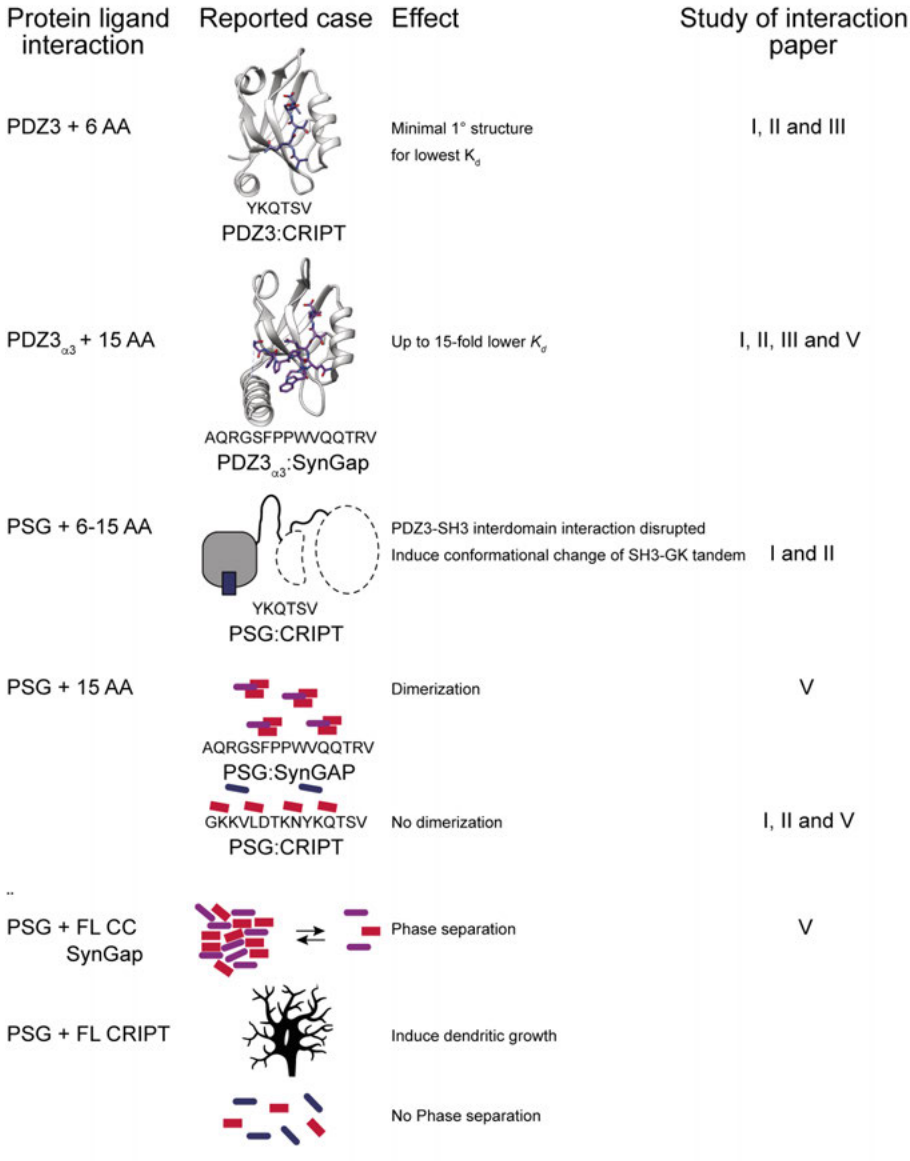


Figure 10: Overview of PDZ3:ligand binding and primary structure length.
Experimental data show that the PDZ3:ligand interaction is dependent of 1° structure length. Six reported cases with use of different 1° structure length for PDZ3 and ligand 1) PDZ3:CRIPT 6 AA (minimum for full affinity), 2) PDZ$_{\alpha3}$:SynGap 15 AA (increases affinity 15 fold), 3) PSG:CRIPT 6-15 AA induces conformational change in interdomain structural organization, 4) PSG:SynGap 15 AA complex induces dimerization, whereas PSG:CRIPT 15 AA complex does not induce dimerization, 5) PSG:FL SynGap induces liquid liquid phase separation 6) PSG:FL CRIPT induces dendrite growth, but no phase separation.

significantly dependent on the length of the linker connecting the domains. Therefore we can assume that the PSG supramodule is present in all DLG proteins (PSD-95, PSD-93, synapse-associated protein (SAP)-97 and SAP-102) [28, 29], whereas further experiments need to reveal if the same is true for the whole MAGUK family [30] (Figure 8). Structural data of the PSG supramodule has been difficult to obtain due to the dynamic and weak interdomain organization of the individual domains in ligand free state [28]. The structure of PALS1 PSG was previously solved reporting an elongated conformation with minimal interdomain interactions in apo conformation, but the structure changes conformation upon ligand binding to PDZ3 to a more compact structure with inter- and intradomain interactions [31]. Experimental data from PSD-95 PSG showed that the interdomain interaction between PDZ3 and SH3 were disrupted upon ligand binding to PDZ3 [28]. Therefore, it can be speculated that PSG supramodule from the MAGUK family have diverse structural organization due to different length of PDZ3-SH3 and SH3-GK linker (Figure 8). Diverse structural organization of the PSG supramodule will increase ligand specificity among proteins in the MAGUK family.

Allostery

How do we define allostery? It is a process where a modulator binds to a distinct allosteric site of a protein and affects the binding activity of another primary site thus enabling regulation of the function of the primary site. This description was formulated over 50 years ago from observations in threonine dehydrogenase [32] and hemoglobin [33] showing a feedback control mechanism for binding of substrate and inhibitor to multiple binding sites. Since then allosteric regulation has been found to play an essential role in many biological systems [34]. As a result it has been suggested that allostery is an intrinsic property of all dynamic proteins [35]. However, there is one major issue with the allosteric concept, even with the potential to discover allostery in hundred or thousand different protein types. The molecular mechanisms associated with allosteric regulation are still not well defined [34, 36]. The allosteric property arises from the perturbation of the native free energy landscape upon perturbation from e.g. modulator binding, protonation, interdomain interactions etc. Therefore, an allosteric behavior can be determined from the following parameters: 1) Relative stabilities of populations of all states in the system. 2) Transition between the states associated with a certain timescale as determined by the energy barriers between them. 3) Modulators (allosteric ligands) or conditions (pH, ionic strength etc.) that can modify the dominant states, thereby remodel the energy landscape of the whole system [34]. However the physical mechanisms that underpin the parameters of allostery are still weakly defined and easy to perturb based on method choice [13]. This is in sharp contrast to protein folding, which is relatively robust to perturbation from choice of method, and synergy between experimental and computational methods is reported [37]. The funneled energy landscape gives rise to a large energy difference between the unfolded and folded state, whereas allostery is mediated by small energetical differences. It has been difficult to report synergy between allosteric data obtained from experimental and computational methods [13] (Figure 11). Therefore Wodak et al. asked the questions, how do we determine allosteric parameters, which roles play thermodynamics by stabilization and destabilization of states and kinetics by time-scales and energy barrier for transition states [34]?

Initially, allostery was only described in dimeric or multi-domain proteins with a clear conformational change. Then allostery could be described on the basis of enthalpy due to the conformational change (Figure 12A). But Cooper et al.

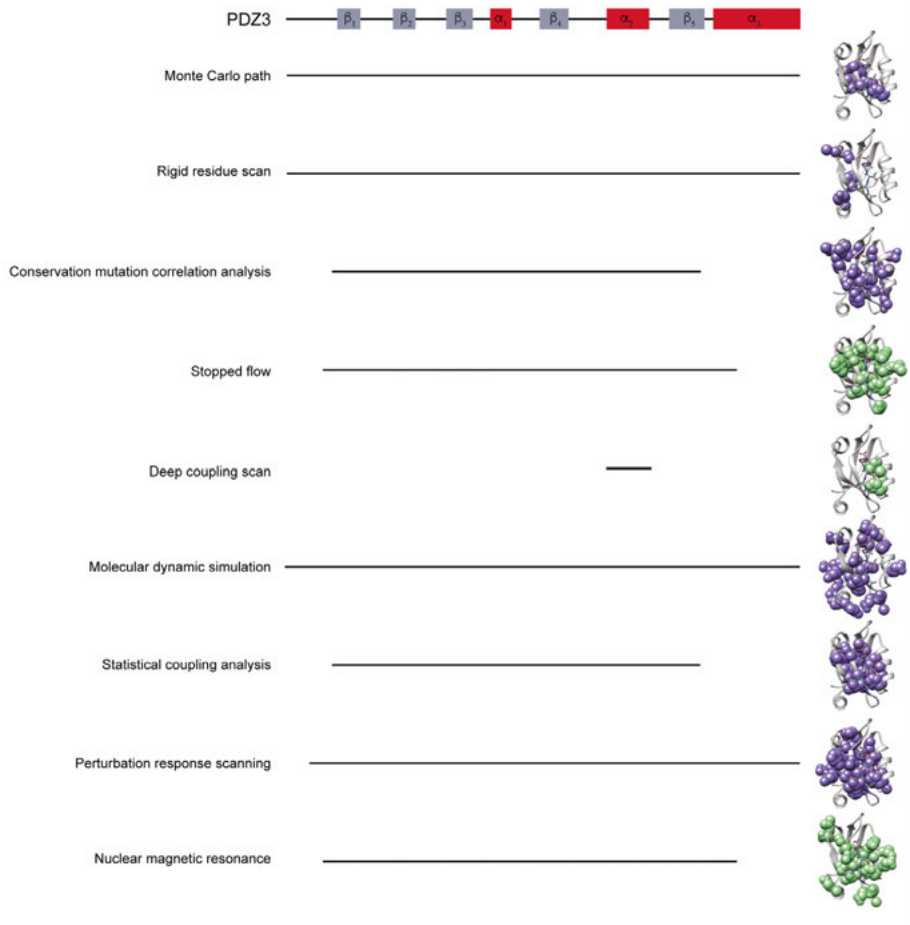


Figure 11: Allosteric networks in PDZ3 determined by different approaches.
2° structure of PDZ3 PSD-95 is illustrated on top. Below is illustrated the 1° structure length used in the different studies. The only exception is the deep coupling scan, which was applied only at α_2 helix, but they used the whole domain for experimental validation. The comparison shows that different N- and C-terminal lengths have been used. Note that the α_3 helix was excluded from the Conservation mutation correlation analysis and Statistical coupling analysis. The following approaches were used: Monte Carlo path (MCPath) [125], Rigid-residue scan (RRS) [126], Conservation mutation correlation analysis (CMCA) [127], Thermodynamic double mutant cycle (DMC) [128], Deep coupling scan (DCS) [40], Molecular dynamics simulation [39], Statistical coupling analysis (SCA) [50], Perturbation response scanning (PRS) [129], Nuclear magnetic resonance [15]. Color code: green (*in vitro*) and purple (*in silico*) approach.

suggested that allostery also can appear in proteins without a conformational change [38] (Figure 12B). This suggests that allostery can be described purely as an entropic effect and was the basis for further studies of dynamic allostery (Figure 13). So what control allosteric regulation? If we look at only one small system, PDZ3 from PSD-95, we will immediately reveal that the physical

principles associated with allostery are difficult to deduce. Initially, PDZ3 was seen as a model system to explore allostery without conformational change, which suggests dynamic allostery caused by an entropic effect (Figure 12B and 13). Evidence for such mechanism was obtained by nuclear magnetic resonance (NMR) experiments [15] since they saw a minimal contribution from enthalpy as the binding affinity dropped 21-fold upon deletion of α_3 helix. Kumawat and Chakrabarty questioned the interpretation and showed that enthalpy plays a role in PDZ3 too. Especially electrostatic interactions are rearranged by the internal energy redistribution caused by enthalpic effects [39]. One thing to notice is that different primary structure length of PDZ3 constructs (Figure 11) were used in the molecular dynamics simulations and NMR experiments. A PDZ3 domain with an extended C-terminus α_3 helix (309-410) was used in the molecular dynamics simulations, whereas the NMR study used a shorter PDZ3 domain that is enough to obtain full affinity with native ligand (Figure 11). The extended α_3 helix (residue 402 to 410) is associated with conformational change upon ligand binding [5]. Further studies (with same primary structure length of PDZ3) will allow a direct comparison and may reveal the entropic and enthalpic contributions to the allosteric network in the PDZ3 domain.

Nine proposed allosteric networks in PDZ3 are illustrated in figure 11 along with the length of the PDZ3 construct used in the respective study. This raises the question how the determined allosteric networks were affected by 1° structure length. Lockless and Ranganathan proposed the first allosteric network in PDZ3 by use of the then novel statistical coupling analysis (SCA) [14]. Allosteric networks reported subsequently in PDZ3 used other methods such as perturbation response scanning (PRS), molecular dynamics simulation, double mutant cycle (Figure 6 and 11). To benchmark the quality of new approaches, most have compared their allosteric networks in PDZ3 with the first identified allosteric network proposed by the SCA approach.

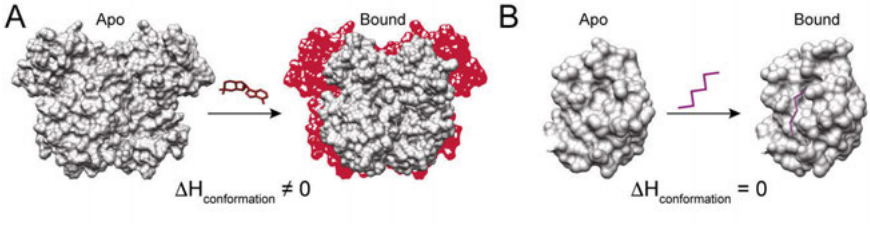


Figure 12: Classic and dynamic allostery.
Illustration of A) Classic allostery is observed by conformational change upon perturbation, apo (2WC2) and bound (1G6N) structure of catabolite gene activator protein. Change in conformation upon perturbation deceases the volume of catabolite gene activator protein structure, the shrink in volume is highlighted with red color. B) Dynamic allostery is present when no conformational change is observed upon perturbation, like in the apo (1BFE) and bound (1BE9) structure of PDZ3.

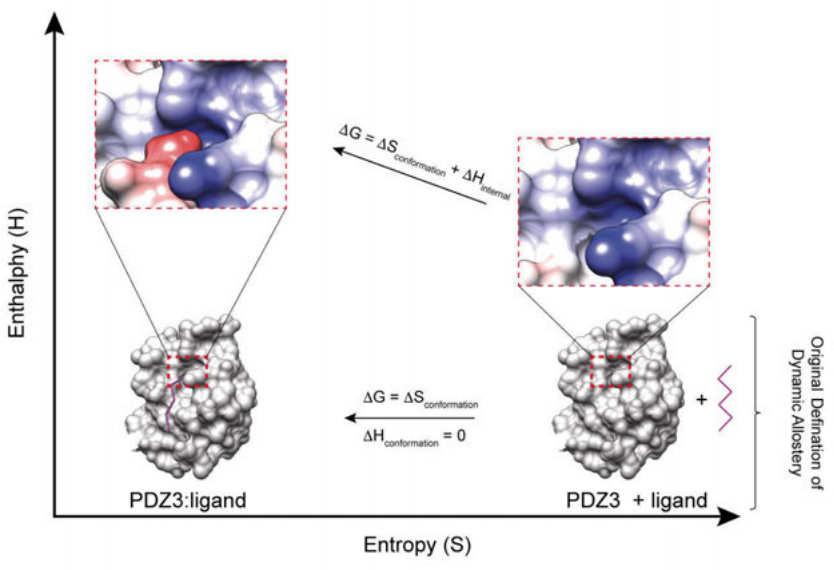


Figure 13: Dynamic allostery.
Illustration of dynamic allostery with and without change in enthalpy. The original view of dynamic allostery was described by an entropic effect with no conformational change (without enthalpy). However a recent study show that no conformational change in secondary structure, does not exclude enthalpy. Enthalphy can arrise from redistribution of internal energies in residues, mainly electrostatic energy from residues (with enthalpy). The figure is adapted from [47].

However, the SCA approach has only recently been experimental validated for five homologous in the PDZ family, but it was only done for the α_2 helix [40]. One thing that rarely has been addressed, why did they not apply the SCA approach only on the subset of PDZ domains that contains an α_3 helix, such as PDZ3 from PSD-95 and compared the result to the analysis done on the whole PDZ family, which may validate if the allosteric network is universal for all PDZ domains. The SCA result was originally experimentally validated by the use of PDZ3 from PSD-95, eventhough the analysis excluded the α_3 helix [14]. In a review article [13] we compared eight allosteric networks reported from experimental and *in silico* approaches and suggested that the discrepancy is caused by choice of method (Figure 6 and 11). Looking in retrospect we did not discuss the fact that the allosteric networks have been probed by using different PDZ constructs, which differ in length of the primary structure even for the same PDZ domain (Figure 6 and 11). This issue is addressed in paper II, where we observe different allosteric networks in PDZ3 and PSG, probed by double mutant cycles and coupling free energy analysis (Figure 14 and 15).

Why can we not define the basic mechanisms that underpin allostery? Gulzar et al. discussed one problem with allosteric research that energy transport is interpreted as allosteric communication even though the allosteric

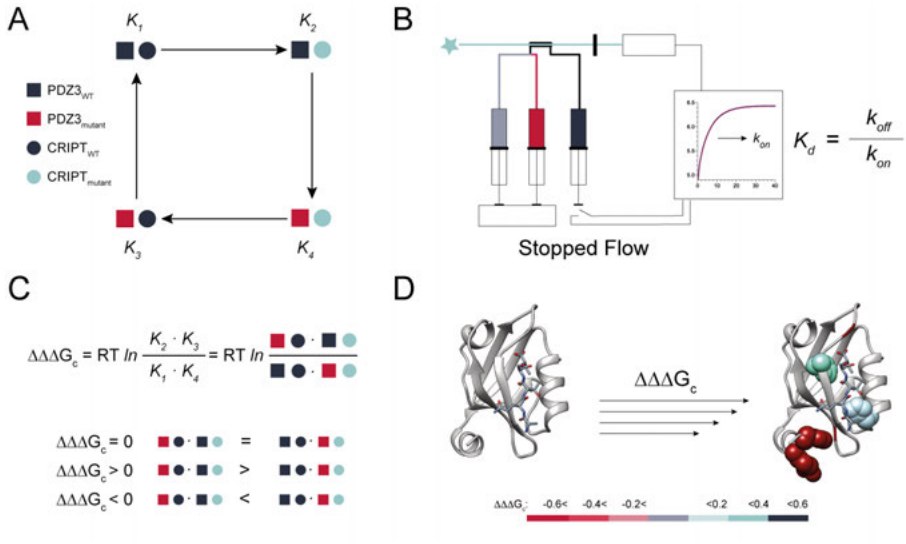


Figure 14: Double mutant cycles and coupling free energy analysis.
Double mutant cycles and coupling free energy analysis is one approach to determine allosteric networks in proteins. Residue pairs are characterized by double mutant cycle to determine their interaction energies, which are used to map allosteric networks. A) The double mutant cycle contains four protein:peptide complexes: 1) WT:WT, 2) WT:mutant, 3) mutant:WT and 4) mutant:mutant. B) The affinity of each protein:peptide complex is measured by stopped flow spectroscopy, that reveals the association and dissociation rate constant, which can be used to obtain the dissociation constant (K_d). C) The coupling free energy of residue pairs is calculated from the product of K_d for complexes with single mutation to the product of K_d for WT complex and double mutant complex. If the products are the same, then the coupling free energy ($\Delta\Delta\Delta G_c$) will equal zero and the residue is experimental displayed as non-allosteric, whereas if the product of single mutant´s K_d is largest (low affinity), then the coupling free energy will be positive and the residue in the protein is allosterically coupled to the residue in the peptide. Oppositely, if the product of K_d for WT and double mutant is largest (low affinity), then the $\Delta\Delta\Delta G_c$ will be negative and the residue in the protein is allosterically coupled to the residue in the peptide. D) Mapping the allosteric network in a protein, residue by residue, by using double mutant cycle for each residue to determine $\Delta\Delta\Delta G_c$. Residues that are analyzed are visualized by spheres and colored by the respective calculated $\Delta\Delta\Delta G_c$.

communication and energy transport arise at different time scales [45]. Do energy transport and allosteric coupling arise from the same tertiary contacts [46]? Gulzar et al. do not support the notion of a well-defined allosteric network as ”dominos on a row” [36]. They suggest that allosteric coupling occurs around the binding pocket. If we look at the allosteric networks shown in figure 6 and 11, a comparison of allosteric networks supports the notion that an allosteric network occurs from the binding pocket and propagates along not well defined paths. All the allosteric networks overlap, but they also show unique patterns. We discuss the difference as consequence of method [13] and construct length (Paper II). In a recent paper it was suggested that all residues are part of the allosteric network in PDZ3, but are stimulated at different time

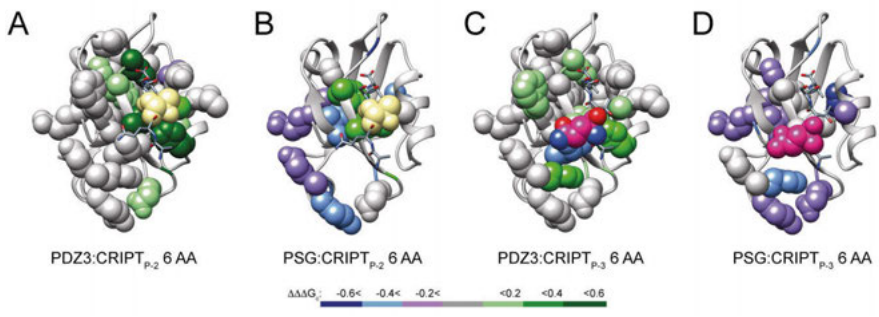


Figure 15: Allosteric networks in PDZ3 and PSG.
Allosteric networks mapped in PDZ3 and PSG by double mutant cycles, using mutational perturbations in the proteins and the peptide ($CRIPT_{P-2}$ 6 AA and $CRIPT_{P-3}$ 6 AA). A) PDZ3:$CRIPT_{P-2}$ 6 AA, B) PSG:$CRIPT_{P-2}$ 6 AA, C) PDZ3:$CRIPT_{P-3}$ 6 AA and D) PSG:$CRIPT_{P-3}$ 6 AA. The structures were made using PDB file 1BE9.

scales [151]. If this observation is true, it will terminate the discussion about the ”truth” of allosteric networks in PDZ3. Since it so far has not been possible to identify two identical allosteric networks in PDZ3, but if all residues are triggered at different time-scales this will explain why. Even though it is difficult to measure allostery in a simple 90 amino acid protein domain and define the physical principles that underpin the mechanism of allostery, there is no doubt about the significance of allosteric applications. Especially drug discovery has shifted paradigm from orthosteric site to allosteric site as drug target [41]. The allosteric site is more specific than the orthosteric site for a given protein family. Allosteric drugs are safer as they possess selectivity and specificity profiles that reduce side effects [42, 43]. However, the paradigm shift has been challenged by slow development of screening methods for allosteric drugs and few successful stories of allosteric blockbuster drugs [44].

What is the origin of allostery in proteins? The main view in the field links allostery with protein function [47, 48]. However, Raman et al. proposed that the origin of allostery is the capacity of proteins to adapt during evolution [49, 50]. The sector in PDZ3 is defined as the group of residues that have coevolved and are associated with allosteric function found by use of SCA approach [14]. In the study it was shown that mutations in the sector of PDZ3 allowed the switch of binding preference from ligand with type I to type II PBM. The mutations were grouped based on their property: class bridging or class switching. A class bridging mutation introduces loss of specificity thus the protein can bind to ligands with both type I and II PBM. On the other hand, a class switching mutation changes the specificity from type I to II PBM. Raman et al. only analyzed the effect of type I to II PBM switch, meaning that they only analyzed the effect of amino acid change at position P_{-2}. However, they interpret their results as a general property of the sector to deduce the

evolution of binding specificity for PDZ3. In retrospect if we ”trusted” the allosteric network proposed by the SCA approach we would not have initiated the study presented in paper III. We could have proposed that *D. melanogaster* and *D. ponderosae* will not bind their native CRIPT ligand without classical type I PBM, where we observe a switch in P_0 residue. When we do multiple sequence alignments of 1° structure from *D. melanogaster* and *D. ponderosae* we do not observe any class switching mutations (Figure 16). Nevertheless, when we look carefully at the data (Figure 16), we can see that four surface residues report class bridging properties, even though they are not part of the allosteric network. Therefore, we choose to apply the ancestral sequence reconstruction (ASR) approach to PDZ3. Hereby, we reconstructed 5 ancestral variants of the PDZ3:CRIPT complex and compared them to seven present-day complexes from animals including *D. melanogaster* and *D. ponderosae.* The aim was to reveal if the change in binding motif is associated with change in function and binding specificity of *D. melanogaster* and *D. ponderosae.* The reasoning for choice of a vertical approach as ASR to study PDZ3:CRIPT interaction is supported by the three major issues with horizontal analysis as SCA belong too. First, the α_3 helix is excluded from the SCA analysis since the approach is based on multiple sequence alignment of all PDZ domains, but only PDZ3 domains contain the α_3 helix. Therefore the α_3 helix was excluded from the original analysis of PDZ family [50]. Second, horizontal analysis best identifies the set of residues important for one function, not the set of residues that can switch function from one to another protein in the family [51].

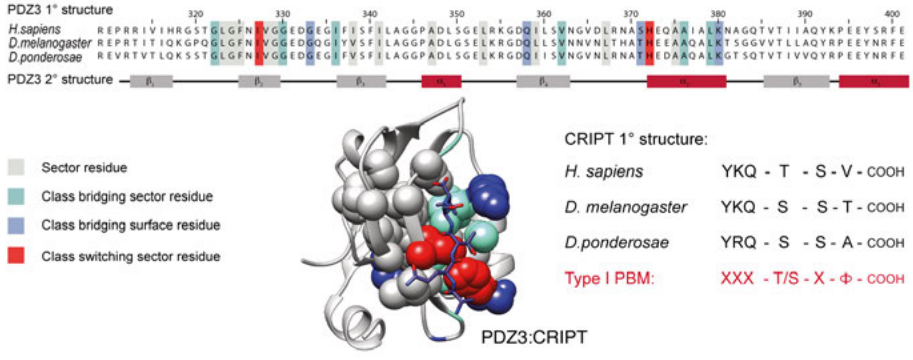


Figure 16: Conserved sector architecture in PDZ3 suggests PBM I preference.
Sequence alignment of CRIPT from *H. sapiens*, *D. melanogaster* and *D. ponderosae* reveals that only *H. sapiens* CRIPT has a type I PBM. Sequence alignment of PDZ3 from *H. sapiens*, *D. melanogaster* and *D.ponderosae* reveals that all class switching residues (red), class bridging residues (light blue), sector residues without experimentally validated function (grey) and class bridging surface residues are conserved. Based on evolution theory by Raman et al. [49] this suggests that *D. melanogaster* and *D. ponderosae* PDZ3 cannot bind their native CRIPT without type I PBM. All the highlighted residues are shown as spheres and with color code in the crystal structure of PDZ3:CRIPT, PDB:1BE9.

Third, specific change in primary structure can cause epistasis. One mutation introduced into five different PDZ domains caused different energetic effects, which reveals high context dependence [52]. Therefore if epistasis is present, the introduction of a mutation in one protein background may not reveal the effect of mutation in other protein family members or ancestral proteins even if a new function evolved from these proteins [51]. It has been shown that a vertical analysis does not have these limitations and can even specifically identify class bridging and class switching mutations [53]. The ASR approach will be discussed in the next section regarding its opportunities and limitations.

The allosteric network in PDZ3 has only been observed *in vitro* or *in silico*, thus set question by its functional application *in vivo* as it may be a concept only observable in the laboratory. In paper V we screened all proteins associated with PSD according to a Uniprot search, and which has a hydrophobic C-terminal residue. The aim was to elucidate if the allostery has implications for PSG specificity. Six mutants were tested, the four single mutants: PSG_{G322A}, PSG_{F325A}, PSG_{I327V} and PSG_{G335A} were previously tested and reported to display interaction energies with the P_{-2} residue in the PSG:CRIPT 6 AA complex. This was probed by mutation of the P_{-2} residue in CRIPT 6 AA and analyzed with a double mutant cycle approach (Figure 14) using stopped flow kinetics to obtain coupling free energies [54] (Figure 15). The SCA approach only identified three of the four residues in PDZ3 and categorized them as listed: $PDZ3_{G322}$ (class bridge), $PDZ3_{F325}$ (sector), $PDZ3_{I327}$ (class switch) and $PDZ3_{G335}$ not identified. This implies that if the SCA network is "true" we will see a significantly different binding profile for the four mutational variants against ligands with type I, II and III PBM. As a class switching residue $PDZ3_{I327}$ is expected to select type II or III over type I PBM. We showed that PSG_{F325A} and PSG_{I327V} display similar binding profile with regard to type I, II and III ligands (Paper V). Therefore further experiments need to be conducted to reveal if SCA approach and associated allosteric networks are valid in PSG supramodule. However, our comparison of the allosteric network in PDZ3 and PSG probed by double mutant cycle, reveals a significant difference in the allosteric networks (Figure 15). All together the significant difference of PDZ3 and PSG allosteric network reported in paper II, suggests that the identified allosteric residues in PDZ3 probed by different approaches [13], will not be the same if SCA or another approach was applied to PSG, thus implicate the importance to study PSG to reveal the context dependent characteristic of PDZ3 as binding mechanism, allostery, binding specificity etc.

Ancestral Sequence Reconstruction

Ancestral sequence reconstruction is a technique used to study structure and function of protein through evolution. Pauling and Zuckerkandl are the pioneers of the research field [55]. They proposed that it would be possible to infer the sequence of an ancestral protein from sequences of recent days proteins. Technology advancements in DNA sequencing, DNA synthesis and better algorithms for reconstruction have made the ASR approach universal [56]. The basic steps of the method are illustrated in figure 17. A brief description of our ASR approach follows: 1) identify sequences in NCBI database by term search and BLAST approach (reference *H. sapiens*), 2) filter and sort sequences by Python programming and manual curation, 3) make a multiple sequence alignment (MSA), 4) extract information about the animal species to obtain a taxonomic tree 5) perform ASR based on a maximum likelihood method thus give posterior probability of all 20 amino acids for every residue in the sequence. Therefore, there are several parameters that determine the quality of ASR, e.g. sequence number, sequence quality and sequence similarity. The conclusions drawn from the first ancestral reconstruction and resurrection studies were therefore limited, due to few available sequences, as input for MSA, which lead to high uncertainty in reconstructed residues [57]. CRIPT and PDZ3 are conserved proteins, and we were therefore able to reconstruct the ancestral sequence of PDZ3 from the common ancestor of all extant bilaterians and the ancestral sequence of CRIPT from the common ancestor of all extant eukaryotes (Figure 16). Fast-evolving proteins, such as intrinsically disordered protein domains, are difficult to reconstruct back to a common ancestor of bilaterian or eukaryotes [58].

Resurrection of reconstructed proteins allow a vertical comparison, which is associated with easier identification of residues that contribute to a shift in protein function. The GK domain from PSG has previously been subjected to ASR analysis. The study showed that GK shares ancestry with the active guanylate kinase enzyme. The ancestor of the GK domain and guanylate kinase enzyme is a protein with guanylate kinase activity, but a single amino acid substitution in the binding pocket allowed evolution of a binding interface and new associated function for the GK domain [59]. The amino acid that shifted the function was already identified from horizontal comparison of GK domains and guanylate kinase enzymes, but a horizontal analysis does

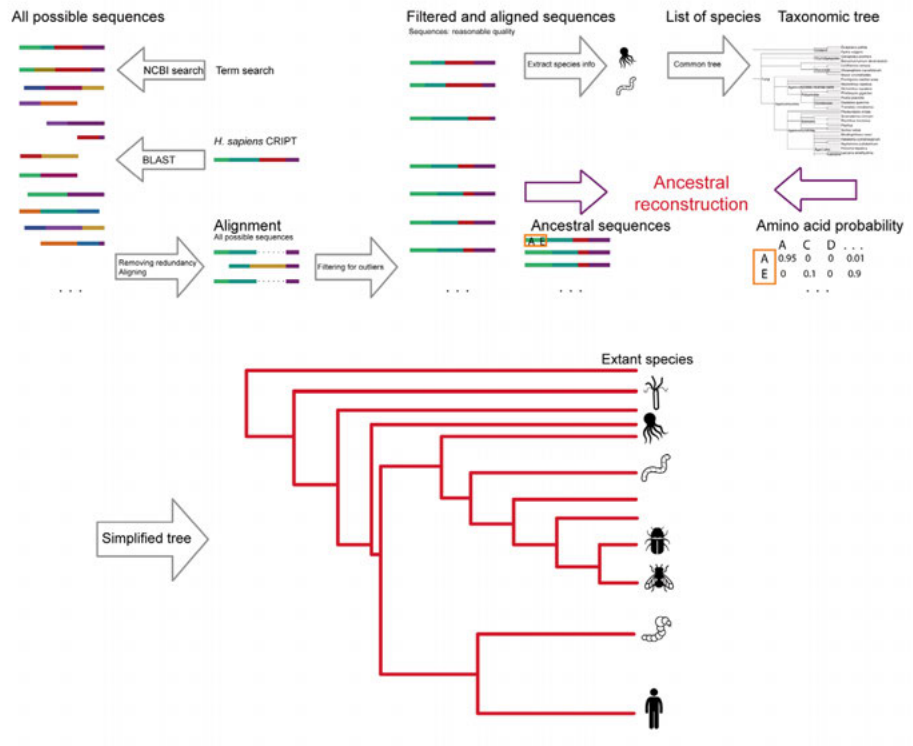


Figure 17: Ancestral sequence reconstruction.
Overview of experimental setup to investigate evolution of the PDZ3:CRIPT interaction using ancestral sequence reconstruction. CRIPT and PDZ3, respectively, were reconstructed based on multiple sequence alignments (MSA) and a species tree. In the reconstruction, 309 unique sequences of CRIPT from 498 different species and 249 unique sequences of PDZ3 from the DLG protein family from 324 different species were included. Sequences were identified in the NCBI database by term search and BLAST approach (reference *H. sapiens*). Sequences were filtered (sequence quality considered) and sorted before MSA by use of Python programming and manual curation. The alignment and the species tree are the input for ancestral sequence reconstruction. Posterior probabilities for the maximum likelihood estimate (ML) of reconstructed ancestral variants are reported for all 20 amino acids at every position in the sequence.

not allow the identification of the biophysical mechanism that mediated the evolution of the new GK domain function. In the present study (Paper III) we reconstruct and resurrect both CRIPT and PDZ3 to deduce if the shift in PBM is associated with shift in the binding interface of PDZ3.

A horizontal comparison of 4196 PDZ domains has previously been reported from 39 eukaryotic species. MSA of 4196 PDZ domains revealed that six amino acids can define the PDZ class. The six identified residues are G330, E334, G335, F340, I341 and A343 and the residues are able to identify PDZ type for 96 percent of all the tested PDZ domains [60]. If we align PDZ3 sequences from the selected seven present days species in paper III we find that *L. loa*

PDZ3 has highest sequence variation. Furthermore *L. loa* only shares one of the six amino acids with the six other species. *D. melanogaster* shares five out of six residues with five of the other species, whereas the remaining five species share all the six classifying positions. This suggests that PDZ3 from *L. loa* has a different binding mode than those from the other species in the alignment. Binding studies with ancestral and present day PDZ3 domains confirmed that *L. loa* has a different binding preference than the rest of the analyzed PDZ3 domains. *L. loa* PDZ3 binds weakly to a CRIPT peptide corresponding to the ancestral sequence of Eukaryota, but *L. loa* $PDZ3_{P335G}$ and $PDZ3_{Q399R}$, with a Gly residue at position 335 or Arg residue at position 399, show 1.7 and 3-fold increased affinity for Eukaryota CRIPT. However a significantly increase in the affinity was observed for *L. loa* $PDZ3_{\alpha 3}$ that increased its affinity with 8-fold for Eukaryota CRIPT in comparisson to *L. loa* PDZ3.

Validation of reconstructed ancient proteins is important to support conclusions made from ASR analysis. Three common approaches are: site directed mutagenesis, Bayesian sampling and the ”AltAll” approach. Site directed mutagenesis was used in the first ASR study, but can be time consuming if the reconstructed proteins have many ambiguously predicted positions [57]. Bayesian sampling is used to reconstruct a set of sequences by sampling amino acids over the posterior probability distribution [61]. The set of the sequences is then resurrected and assayed to show the distribution of function associated with the posterior probability distribution of sequences [62]. Thus, Bayesian sampling integrates phylogenetic uncertainty into ancestral inference [63]. However, it is necessarily not preferable, as shown in a recent study by Eick et al. who compared the different approaches for validation of ancestral sequence reconstruction and found that resurrection of ancestral proteins are generally robust to uncertainty in reconstructed sequences [64]. They found that Bayesian sampling can be biased toward non-functional proteins, because reconstruction is not limited to higher probability residues. We choose to apply the third approach, AltAll, which is the most time-efficient method. The ASR approach proposes the most likely sequence of the ancestral protein based on a maximum likelihood (ML) estimation. A protein conserved through evolution will result in a reconstructed sequence with low uncertainty at every amino acid position. The AltAll sequence represents a “worst case” scenario of the reconstructed sequence. All amino acids that have a posterior probability lower than 0.8 are substituted with the residue with the second highest posterior probability into a sequence denoted AltAll, for alternative sequence with all low probability positions replaced. We chose a 0.8 probability cut-off that is the standard for AltAll approach [64], but lower cut-offs can be applied to introduce more sequence variability between ML and AltAll sequence.

”Ancestral sequence resurrection, the past is the key to the present” [65]. This quotation does well describe the field of enzyme engineering that has been advanced by ASR. Reconstructed ancestral proteins often show higher stability than the present-day protein they were predicted from, which is the general trend in our study too (Paper III). Therefore ASR can be used to engineer functional thermostable proteins [66]. ASR studies suggest that ancient proteins are more thermostable than present day protein, which can be explained by the higher environmental temperature in the past [67]. However, the thermostability trend is not smooth over time [68], especially in more recent time scale, therefore recent days ancestors will not necessary exhibit lower thermostability than ancient ancestors, as thermophilicity can be gained and lost during evolution [69]. Most studies report elevated stabilities for ancestral proteins, which initiated the discussion whether it is a general property of ancient proteins [70] or an artifact by the ASR approach [68, 71]. Future studies need to address this, but from a protein engineering perspective ASR is a reasonable choice of approach to obtain thermostable proteins.

Liquid Liquid Phase Separation

Cellular compartments are numerous in eukaryotic cells to organize biological matter. The best characterized compartments are separated from the surrounding milieu by lipid bilayer membranes. There are also membraneless compartments, that have high concentration of proteins and RNA [72], formed via LLPS. Compartments formed by LLPS are "energy friendly" in comparison to membrane surrounded compartments as they don't require lipid biogenesis and membrane identity maintenance [73, 74]. LLPS is a common phenomenon and often observed during crystallization, however, it was recently discovered that phase separation is essential for formation of membraneless compartments that regulate biological functions and activities [75, 76]. The main functions of membraneless compartments are: I: concentration of biochemical reactions, II: sequestering harmful components, III: storage of biomolecules and IV: signal amplifications [72].

The postsynaptic synapse is a compartment with membrane bilayer, that below its membrane has a thick [77], electron dense, membraneless compartment, known as PSD of excitatory synapse. The PSD contains more than thousand different proteins [78], which have been thoroughly identified and characterized [79], but the assembly and regulation is poorly described. The scaffold protein PSD-95 is one of the most abundant proteins in the PSD [80, 81], and it has an essential function to sustain the molecular organization of PSD [82]. It was recently discovered that synaptic Ras GTPase activating protein (SynGap, isoform α_1) undergoes LLPS by binding to PSG from PSD-95 [10] thus facilitating a dynamic anchoring mechanism for high concentration of SynGap in PSD [83]. Additional to the structural role of SynGap it has enzymatic activity and participate in biochemical signalling pathways [84]. LLPS were observed both *in vitro* and in living cells, therefore Zeng et al. suggested that phase transition is essential for formation of PSD through compartmentalization [10]. It has been a long lasting puzzle [83] why neuronal development is sensitive to SynGap concentration [85], isoform [86] and why the enzyme is present in such high concentration in PSD. Altogether the high SynGap sensitivity, further support the hypothesis that SynGap regulates PSD through LLPS formation. Phase separation of PSG:SynGap complex is protein concentration dependent [10], and LLPS will only occur above a threshold of protein complex that facilitates PSD formation [83].

Biophysical characterization of the PSG:SynGap complex showed that multivalent interactions were essential for LLPS formation [10]. SynGap binds to PDZ3 in PSG PSD-95 by a type I PBM, consequently only one isoform α_1 has the specific binding motif and induces LLPS. Furthermore, SynGap cannot undergo LLPS with PDZ3 alone, as interactions outside of the binding pocket are required for the high affinity of the PSG:SynGap complex. SynGap is present as a trimer in solution, due to intermolecular interactions between the coiled coil part of the protein N-terminal, which binds to two PSG proteins, thus the stoichiometry of the complex is 2:3 [10]. Multivalent interactions are essential for PSD formation through phase separation [87, 88], which is supported by a recent study that showed the possibility to reconstruct PSD from six highly abundant proteins in PSD, including PSG and SynGap [87]. Furthermore, the concentration threshold for PSG and SynGap to form LLPS were significantly lowered upon mixing 6 proteins in comparison to SynGap and PSG alone [10], which underscores the importance of multivalent interactions for compartmentalization of PSD through phase separation. To identify new proteins with potential for LLPS, we examined proteins from PSD with type I PBM, high affinity for PSG, and high predicted coiled coil (CC) content. The C-terminal part of AGRB1 and MTMR2 fulfilled the listed criteria, but only full length (FL) CC AGRB1 formed LLPS with PSG (Paper V). Phase separation of PSG:FL CC AGRB1 complex were observed by light microscopy and quantified by sedimentation and turbidity assay. LLPS is salt dependent [89], phase separation of PSG:FL CC AGRB1 complex was not observable above 250 mM NaCl, which can be due to competitive inhibition of FL CC AGRB1 binding by chloride ions [90]. There are examples of protein complexes that form LLPS without coiled content in the protein. Different intermolecular interactions have been observed in proteins that induce phase separation: oligomerization domain, coiled coil, helix helix, β zipper, pi pi, cation anion, dipole dipole and cation pi contacts [72] (Figure 18).

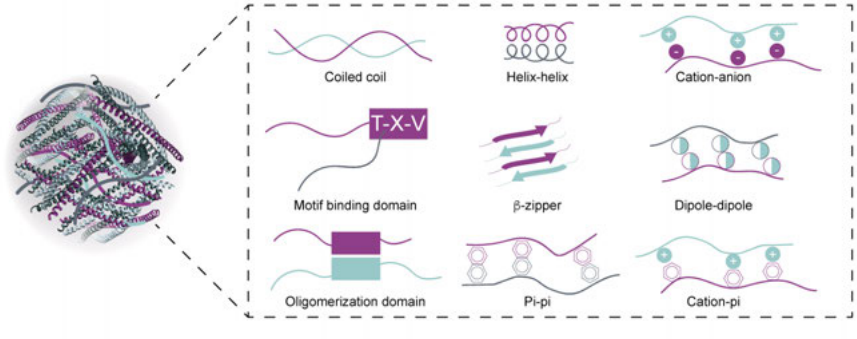


Figure 18: Interactions involved in liquid liquid phase separation.
Overview of different kinds of intermolecular contacts, which have been observed in liquid liquid phase separation. PDB: 5JXC. The figure was adapted from a review in protein phase separation [72].

Oparin et al. proposed that the first step for life was phase separation of macromolecules into liquid compartments [91]. Numerous discoveries of LLPS in cells support the idea. Furthermore it has been proposed that proteins associated with LLPS are remnants of ancient structural organization to make optimal conditions for chemical reactions, as the droplet formation allows high concentration of proteins [92]. A recent study showed that the RNA-binding protein Fused in Sarcoma has preserved its phase separation properties to avoid pathological liquid-to-solid phase separation [93]. Scaffold proteins are highly abundant in LLPS in the PSD, thus making them interesting subjects to study the mechanism of origin of life. We study the evolution of PDZ3 from PSD-95 and show that the protein is highly conserved through evolution, as highly expressed proteins evolve slowly to maintain protein stability and reduce misfolding [94].

Protein Folding

How do proteins fold into their native structure in the crowded cellular environment? A fundamental question, that researchers have spent the past half century to elucidate with a range of methods such as NMR, protein engineering, circular dichroism, kinetics and molecular dynamics simulations [95]. However many of the studies are limited to small single protein domains, thus only present a fraction of the proteome and unlikely to display the true folding mechanism *in vivo* [96]. To address the protein folding problem new methods have to be developed to quantify the folding of protein for a broader set of proteins under native conditions. Recently a large-scale analysis was developed for aggregation-prone proteins, thus quantifying the folding status by co-translational inclusion of protease [97]. Another method is optical tweezers, that allows observation of translation and folding in real time [98]. Expansion of methods to study the protein folding problem will provide the framework to understand the connection between protein folding, disease, cellular function and evolution [96].

PDZ3 is a model protein, which has been used to elucidate general mechanisms of small protein domain folding. Most protein domains fold in an apparently two state mechanism with only native and denatured states significantly populated (Figure 19A and B). But the folding energy landscape is rugged, thus, intermediates accumulate that determine the folding mechanism (Figure 19 C to F). Kinetic and thermodynamic analysis of five different PDZ domains revealed that protein folding is dependent on topology rather than specific sequence [99]. At neutral pH PDZ3 has one transition state, but at low pH and high Urea a second transition state is revealed. To determine if sequence or topology of the PDZ fold drives the folding, the folding mechanism of PTP-BL PDZ2 and PSD-95 PDZ3 were compared. Both PDZ domains have two transition states, where the second transition state is most similar thus emphasizing that protein topology drives protein folding rather than sequence [100]. The energy landscape is rugged and modulated by solution conditions, mutations and topology of PDZ domain. A third transition state was observed under conditions that further stabilized the native state relative to denatured state upon elucidating the folding mechanism of SAP97 PDZ2 and PSD-95 PDZ2 [101]. The mechanism was proposed to be general for all PDZ domains. Notable the folding mechanism of $PDZ3_{\Delta\alpha3}$ (deletion of C-terminal α_3 helix) is

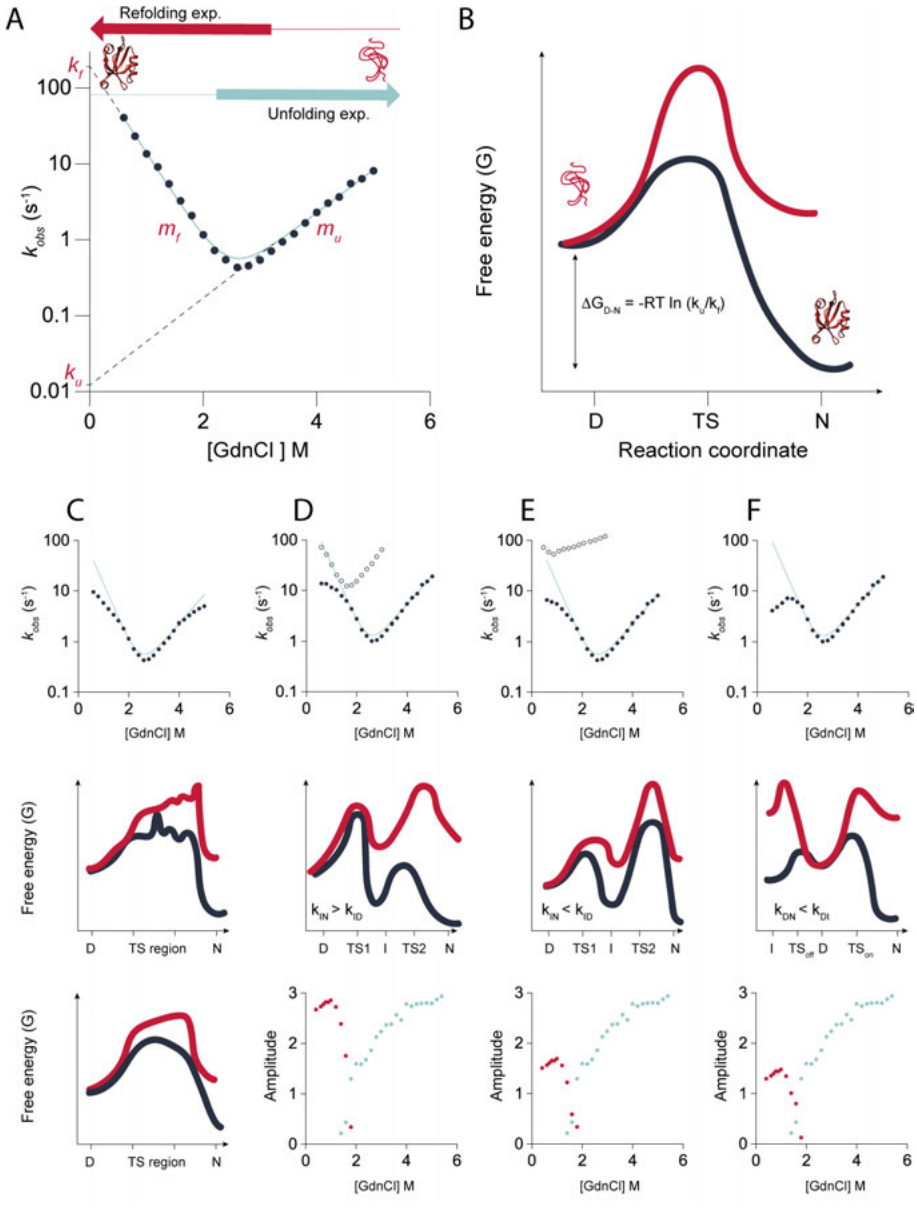


Figure 19: Chevron plots and energy diagrams for different folding mechanisms.
A) Chevron plot for protein with two state folding mechanism. Folding rate constant k_f can be obtained by extrapolating the refolding arm (left) in the chevron plot and unfolding rate constant k_u can be obtained by extrapolating the unfolding arm (right) in the chevron plot. B) Energy diagram at low (blue) and high (red) denaturant corresponding to the simplest case, a two state folding mechanism. The energy differences (ΔG_{D-N}) between native (N) and denatured (D) state determine the stability of the protein domain, larger differences thus more stable protein domain. C-F) different chevron plots and corresponding energy diagrams, that all deviate from the simple two state folding mechanism. C) Chevron plot with rollover at high and low GdnCl, suggests denaturant dependent transition state, thus region with several intermediates.

Corresponding energy diagram at low (blue) and high (red) denaturant, illustrate that transition states are denaturant dependent. An alternative energy diagram for rollover in both re- and unfolding arm is broad transition state. D and E) Forming of "on pathway" intermediate, shown by chevron plot with rollover (positive slope) in the folding arm at low GdnCl. D) $k_{IN} > k_{ID}$ suggest "on-pathway" intermediate, which is confirmed by similar maximal amplitude for refolding and unfolding experiment. However "on pathway" intermediate can even be formed when $k_{IN} < k_{ID}$ (E). The relative "lost" amplitude in refolding experiment can suggest formation of intermediate in the burst phase. F) Forming of "off pathway" intermediate. Chevron plot with rollover (negative slope) in the folding arm at low GdnCl. As $k_{DN} < k_{DI}$ suggest accumulation of stable intermediate "off pathway". Stable intermediate formed in the burst phase, which is supported by the relative "lost" amplitude in refolding (red circles) compared to amplitude for unfolding (turquoise circles).

similar to PDZ3 [102], even though it has been reported that $PDZ3_{\Delta\alpha3}$ is prone to misfolding [103] and displays altered peptide binding [15]. Interestingly Calles et al. observed different intermediates for PSD-95 PDZ3 e.g. dimer under unfolding. However the intermediate states are only observed at high protein concentrations, therefore under dilute experimental conditions these intermediates will not be visible [104, 105]. They suggested that the intermediates are precursors of globular and supramacromolecular complexes. Based on these observations all the studies of protein folding mechanism carried out in dilute samples can be questioned. However, the folding mechanism of PDZ domain family was recently summarized in a review. The author suggested that PDZ domains have the capability to adopt several conformation *in vivo* depending on environment and target [106] since the PDZ domain family (268 proteins with the same topology) show conformational plasticity *in vitro* as the PDZ domains have different binding specificity.

PDZ3 is part of the supramodule PSG, but the folding mechanism has only been studied in isolation as a single domain as described in the previous paragraph. Therefore, we analyzed the folding mechanism of the PSG supramodule to elucidate if the folding of PDZ3 is affected by the additional domains. Few studies have elucidated the folding mechanism of multi-domain proteins due to increased complexity of the analysis, and multi-domain proteins are generally simplified as "beads on and string" for the folding mechanism analysis. A recent study suggested that the folding pathway of a multi-domain protein depends on the topology of domain´s interdomain interactions, a mechanism that should be beneficial to avoid rapid decrease in conformational entropy [107]. Therefore the domains that are cooperative and connected by interdomain interactions will fold first followed by the domains that are independent and detached from interdomain interactions in the multi-domain protein. We considered if that was the case for PSG, thus, SH3 and GK domain should fold first, then followed by folding of PDZ3 domain, as a Rosetta model of PSG suggests interdomain interactions between the SH3 and

GK domain [28]. Last steps in the folding mechanism of PSG supramodule would then be folding of the Hook domain and the PDZ3-SH3 linker, as these subdomains of PSG display large conformational flexibility [28]. However, in paper IV we reported that under neutral pH conditions PDZ3 will fold first followed by SH3-GK tandem along the folding pathway. A putative off-pathway intermediate (Figure 19F) was also detected, as deviation from the classical V-shape chevron plot was observed (Figure 19A). The rollover in the refolding arm can be due to fast folding of the SH3-GK tandem into an intermediate, which is suggested to be formed in the burst phase as the relative maximal amplitude decrease in the refolding experiment. Fast folding and accumulation of the SH3-GK tandem by off-pathway is similar as the above described study about folding mechanism for multi-domain protein folding [107] and initial speculations for PSG´s folding mechanism. Therefore more studies need to be carried out with multi-domain proteins to reveal if topology and interdomain interaction between domains affect the order in which domain fold first and last.

PDZ as Drug Target

PSD-95 is involved in numerous cell signaling pathways, revealed by its large interactome [7] visualized in the STRING database [108]. Hence, PSD-95 is associated with a list of neurological diseases, e.g., neurodevelopmental-, autism spectrum- and depressive disorder [7]. PDZ1, PDZ2 and PDZ3 are involved in numerous protein protein interactions (PPIs), thus making them potential drug targets to modify the associated diseases. PDZ PPI targeting approach has been challenging due to the small binding interface of PDZ domain and the large PDZ domain family with low specificity for different partner proteins. Dimeric peptide-derived inhibitors for PDZ domains in PSD-95 were successfully developed for the PDZ1-2 supramodule as dimeric peptides increase affinity, specificity and plasma stability [7, 109, 110]. In contrast, PDZ3 inhibitors are restricted to a few lead compounds [111, 112] with limited clinical potential. Taken together, the successful targeting of the PDZ1-2 supramodule with a dimeric peptide inhibitor suggests that a similar strategy can be applied for the PSG supramodule to target the PDZ3 domain from PSD-95. The advantage of PSG targeting is increased specificity and selectivity for PSD-95 among the numerous PDZ scaffold proteins in PSD [19]. Revealing the interactome for PSG in PSD among all proteins with hydrophobic C-terminal residue showed that PSG has a more specific interactome in comparison to PDZ3 (Paper V). PSG binds with high affinity to a small set of ligands, whereas PDZ3 reported weaker affinity to several ligands. Our peptide screen showed that context affects the PDZ3:peptide interaction, which is similar to previous findings that SynGap binds with 15-fold higher affinity to PDZ3 in PSG than to PDZ3 alone [10].

PSD-95 belongs to the large MAGUK family of scaffold proteins [6] (Figure 8). Most of the scaffold proteins share the PSG core, but they have various topology due to the length of the PDZ3-SH3 linker. PSG PSD-95 share topology with PSG from PSD-93, SAP-97 and SAP-102 as the four proteins arose as the result of two consecutive whole genome duplications in the vertebrate lineage approximately 440 million years ago [113, 114]. The high sequence similarity among the four proteins make it impossible to selectively inhibit one of the proteins, as knock down studies show that the four proteins compensate for each other [115, 116].

All the above cases modulate protein function by targeting the orthosteric binding site. Competitive binding is a common strategy in drug discovery, but modulation of protein function by direct targeting of the orthosteric binding pocket results in lower specificity and high risk for side-effects. Targeting allosteric binding sites is an alternative strategy to reduce the risk of side-effects [41] as the allosteric binding site is less conserved among proteins in a family [117]. Nevertheless, allosteric drug targeting is not a suitable approach for PDZ family as the site is not well defined and small [118], with low potential for high affinity of possible ligands.

A new paradigm in drug discovery is the modulating of protein function through LLPS [119]. Scaffolding proteins of the PSD (PSD-95, GKAP, Shank and Homer3) form LLPS upon mixing, an enriched condensate of scaffold proteins that exchange rapidly between condensate and aqueous solution. Specific interactions among the scaffold proteins drive the formation of LLPS thus promoting the enrichment of SynGap and NMDA receptors [87]. The scaffold proteins are essential for LLPS, as removal of one of the four scaffold proteins affects the formation of protein condensate dramatically. Our knowledge of LLPS is still limited, but LLPS is a potential drug target in PSD as many neurodegeneration proteins can form LLPS. Mutation in the SynGap gene is associated with mental retardation. One mutation causes an alternative splicing that gives a truncated SynGap protein without binding motif [120]. Lack of type I PBM abolish binding of SynGap to PSD-95 and its association in LLPS [10]. Further studies with LLPS need to be conducted to reveal if it is possible to modulate LLPS in PSD. Common approaches for LLPS modulation are: 1) stabilization or destabilization of proteins in LLPS by small molecules or 2) post-translational modification of proteins in LLPS [121].

The common drug discovery approach is limited to proteins with tertiary structure, even though more than half of the proteome contains proteins with intrinsically disordered regions [122]. Intrinsically disordered proteins (IDPs) are highly abundant in membraneless compartments, due to multiple weak and specific interactions required for phase separation [123]. However important to notice, that compartmentalization by phase separation in PSD is associated with multivalent interactions between protein with tertiary structure [124], thus without IDPs. Increased interest in LLPS as drug target reflect the limited success to target neurodegenerative diseases with common drug strategies. Targeting of LLPS presents new innovative opportunities [119].

Future and New Perspectives on the Projects

In project I we reported that PSG is present in two conformations in solution, that bind CRIPT with similar affinity, but with different rate constants. Conformation B appeared to have a restricted binding pocket, that we suggested to be associated with the dynamics of the SH3-GK tandem. However, we did not characterize the interdomain interaction between PDZ3 and SH3-GK. In the study CRIPT 6 AA and CRIPT 15 AA were characterized upon binding to PSG, however, two kinetic phases were only observed for binding of CRIPT 6 AA and not CRIPT 15 AA. Therefore, it is possible that the conformational equilibrium of PSG, observed upon binding of CRIPT 6 AA, is affected by the sequence of the ligand. SynGap binds with high affinity to PSG, but has significantly different properties, e.g., induces dimerization of PSG, phase separation (Paper V) [10] and intermolecular coupling between domains by stabilization of the linker between PDZ3 and SH3 [130]. Therefore, it will have been interesting if we had performed the same set of experiments with peptide derived from SynGap to analyze if the conformational equilibrium of PSG is affected by sequence of peptide. Elucidating further detail about the PSG conformational equilibrium could have been done by detailed characterization of PSG mutants. In paper II we characterized PSG_{G335A}, that appeared to have only one major conformation of PSG upon binding of CRIPT 6 AA. The mutation is located in the $\beta_2\beta_3$ loop region, which is reported to interact with the SH3 domain [28]. All together our data suggests that the dynamic conformational equilibrium of PSG is affected by interdomain interactions and sequence of ligand, but further studies are needed to reveal the details.

Kinetic heterogeneity was observed upon binding of CRIPT 6 AA to PSG, which we suggested to be due to two conformations of PSG, however the kinetic heterogeneity could also be due to a proline switch. A proline switch will give two conformation, so this is consistent with what we say in paper I. The construct of PSG used for all the kinetic analyses has 15 proline residues: one in the linker to H_5 tag site, three in PDZ3, one in the PDZ3-SH3 linker, one in SH3, seven in GK and two in the C-terminus. In project IV, indications of a cis-trans proline isomerization was observed upon refolding of PSG and SH3-GK. Most proline residues are found in trans configuration in the native conformation of the protein, but under denaturing conditions around twenty percent of the prolines are in cis configuration [131]. We reported that the

two conformations of PSG have similar affinity for CRIPT 6 AA, but the conformations modulate the association and dissociation rate constant (Paper I). In a recent study, the nuclear coactivator binding domain of CBP was characterized upon binding to its binding partner [132]. The kinetic analysis showed that the nuclear coactivator binding domain of CBP is present in two conformations at equilibrium, which modulate association and dissociation rates. The two conformations of the protein were explained by a proline switch; hence 25 percent of the structures were found in the cis-conformation. To support the hypothesis of proline isomerization, a variant of the protein with Pro to Ala substitution was tested. The mutated protein reported only single exponential binding kinetics, suggesting that the heterogeneity in binding of nuclear coactivator binding domain of CBP is due to proline isomerization that induce a conformational switch. To test if the two conformations of PSG observed upon binding of CRIPT 6 AA is due to cis-trans isomerization, we could apply the same approach as described above and substitute every Pro residue with Ala. If the binding of CRIPT 6 AA can be described by single exponential kinetical analysis for a single Pro to Ala substitution it will suggest that the two observed conformations are due to proline isomerization. The same approach can also be applied to test the basis for the slow refolding phase in PSG. Proline cis-trans isomerization have previously been reported in single [134] and multi-domain [135] protein from the SH3 family. A supramodule from Crk contains two SH3 domains that are tethered by a 50 amino acid linker, which has a proline residue in the linker that interconvert between cis and trans conformation [135]. At equilibrium, the cis-conformation of the supramodule dominates, which is autoinhibited by interdomain interactions that prevent ligand binding to N-terminal SH3. A small fraction is found in the elongated trans conformation that allow binding of ligand to N-terminal SH3. Upon ligand binding the equilibrium is shifted to the trans conformation. Autoinhibition of protein activities is common, however, most are reported to be regulated extrinsically, e.g. through post-translational modification, whereas the autoinhibition of the supramodule from Crk is intrinsic through cis-trans isomerization. No full length structure of PSG PSD-95 is available, but it has been suggested from NMR data and crystal structures of isolated domains that PSG has a compact structure with interdomain interactions in the apo state [28], whereas the structure is elongated upon ligand binding to PDZ3 as interdomain interaction between SH3 and PDZ3 is disturbed [28, 130]. Another approach could be to characterize the binding of CRIPT to PSD-95 paralogs, which has been done previously for the SynGap derived peptide. The study used NMR to show that PSD-95 and PSD-93 undergo ligand induced conformational coupling between PDZ3 domain and SH3-GK tandem, noticeable it was not observed for SAP-102. Ligand induced conformational coupling was observed for CASK and PALS1, thus suggests a common mechanism for several MAGUK proteins [130].

Equilibrium unfolding of PSG and SH3-GK domain in acidic conditions showed that the SH3-GK tandem has a midpoint around 3M GdnCl, whereas under basic and neutral conditions the midpoint of SH3-GK tandem was below 2 M GdnCl. However, the kinetic analysis of PSG was not possible under acidic conditions, as the data were too complex for analysis. Kinetic analysis revealed that unfolded PDZ3 traps SH3-GK in an unfolded conformation under neutral conditions. This is reported by the pronounced rollover when both PDZ3 domain and SH3-GK tandem are denatured before refolding, as the productive folding at low denaturant concentration is slowed down. However, in mild denaturant condition PDZ3 retains its native conformation and has a minimal effect on the refolding of SH3-GK tandem, as shown by the similarity to folding of SH3-GK tandem alone. The mild denaturation approach has been used previously for the PDZ tandem supramodule in Whirlin, which showed that the misfolded intermediate was functional and could bind peptides [21]. The authors speculated whether the functional misfolded intermediate is a general feature of PDZ supramodules to increase the number of structures for protein:protein interactions, even in the presence of mechanical unfolding. However it is not known if PSG can bind ligands in its misfolded conformation or how mechanical forces affect the shape and function of PSD [136, 137] as Whirlin is present in hair cells that transform mechanically energy into electrical signal that allow us to hear [138].

Ancestors of PDZ3 and CRIPT were reconstructed, resurrected, expressed and purified to examine if and when the PDZ3:CRIPT interaction was lost, as several present days species lack the type I PBM in CRIPT. However we could also have used the ASR approach to study a protein protein interaction that gain function, as has been carried out previously for the GK domain in Dlg4 (PSD-95 in *H.sapiens*). Anderson et al. reconstructed, resurrected, expressed and purified the GK domain to deduce how the scaffold protein has evolved from an ancient GK enzyme [59]. Amino acid substitutions in the binding surface of GK allowed evolution of a new protein function, which is essential for cell adhesion, and structural organization of PSD in multicellular animals [6]. To study the gain of function as LLPS, SynGap would have been a better choice as the isoform α_1 only seems to be present in mammalian species and are required for LLPS upon binding to PDZ3 [10]. Ancestral sequence reconstruction and resurrection were used to examine the dependence of binding motif for the PDZ3:CRIPT interaction. We showed that PDZ3 has preference for type I PBM (exception *L. loa*). PDZ3 domains from ancestral and present-day species does not bind their native CRIPT if the type I PBM is lost. Interestingly, $PDZ3_{L.loa}$ has novel amino acid substitutions that can explain the reduced binding affinity for CRIPT with type I PBM. Furthermore, we found that one out of five weak PDZ3:CRIPT interactions can increase its affinity by 10-fold upon α_3 helix extension ($PDZ3_{L.Loa}$:$CRIPT_{Eukaryota}$),

which underscores the importance of a full-length α_3 for apparently weak PDZ3:ligand interactions [10].

The PDZ3:CRIPT interaction seems to be dependent of the type I PBM for all present days species and ancestors, as the affinity decreases significantly when the type I PBM is absent, which suggest that the interaction is lost in present-day species evolved from the Hexapoda node. However, it is important to notice that all studies are done *in vitro* in dilute solution. Is the function of a weak interaction affected by the cellular context? In a recent study performed in semi-native environment a PDZ:ligand interaction increased 100-fold in affinity in comparison with *in vitro* experimental data [139]. Therefore, it will be interesting to study the PDZ3:CRIPT interaction in a cellular context to reveal if the lost/weak interaction reported from *L. loa*, *D. melanogaster* and *D. ponderosae* will affect the function of the interaction *in vivo*. The homologous protein Sap97 PDZ3 was recently investigated *in vitro* and *in vivo* upon binding to CRIPT in the worm *C.elegans*. Mutation of P_0 residue in CRIPT removed the type I PBM, which prevented PDZ3 pull down, suggesting that the Sap97 PDZ3 cannot bind to CRIPT without type I PBM [8]. Furthermore, it was shown that CRIPT promotes dendrite growth by triggering Sap97 to traffic AMPA receptors and Glutamate A1 subunits to the extracellular membrane. Nevertheless, it was still reported that CRIPT indeed promotes dendrite growth in *C. elegans*, which has a CRIPT without a type I PBM. Therefore the authors suggested that the biological actions of CRIPT on dendrite growth could be partially independent of the PDZ3:CRIPT interaction [8]. Our findings showed how a protein-protein interaction can undergo dynamic evolution with possibility for loss of interaction, thus supports previous finding that PDZ3:CRIPT interaction through type I PBM recognition is not important for physiological function in these species [8].

PDZ and CRIPT were reconstructed, resurrected, expressed and purified to examine the evolution of the complex, however it will have been even more interesting to reconstruct and resurrect the PDZ3:SynGap interaction due to its associated function in liquid liquid phase separation [10]. Phase separation is linked to PSD formation [10] and organization for cellular compartments without lipid membrane barriers [73], therefore ancestral sequence resurrection and reconstruction could be interesting to do for PSG:SynGap complex, then it may be possible to reveal if the complex is essential for PSD formation at the bilaterian node, the ancestor of all species with nervous system [140]. Ancestral sequence reconstruction has previously been used to study Fused in Sarcoma protein that undergo phase separation. The study showed that natural selection stabilized the liquid forming potential of the protein [93]. The C2 and RASGap domains from SynGap have been resurrected [141], however not the coiled coil domain of SynGap that is required for LLPS

and binding to PSG. Dlg proteins originated in unicellular organism prior to multicellular life [142], therefore before acquisition of neural like cells. Since the time of SynGap's origination is more recent [142], it raises the curiosity, when did the PDZ3:SynGap interaction evolve and is it associated with the formation of neural like cells? Four isoforms of SynGap is present in *H.sapiens* as consequence of genome duplications, but only α_1 isoform with a type I PBM binds to Dlg4 (PSD-95) [10]. SynGap was absent in the metazoan ancestor [142], however, it is not revealed in which isoform, but MSA suggests that isoform α_1 is only present in mammals.

Before the work of this thesis was initiated the allosteric mapping in PDZ3 had only been done in the isolated domain. Therefore we mapped the allosteric network in PDZ3 in its supramodular context together with the SH3-GK tandem and found that interdomain interactions in the supramodule affected the allosteric network in PDZ3. Our findings suggest that the allosteric network is sensitive to perturbation from adjacent domains, which can have general implications for elucidating and characterizing allosteric networks in single domains present in supertertiary structures. However, the study of PDZ3 allosteric network in the presence of the SH3-GK tandem is only one small step closer to the complex and concentrated cellular environment *in vivo*. Does the allosteric network of PDZ3 have any cellular implications or is it just a biophysical concept present in a dilute environment? The allosteric network in a protein domain has not been mapped in a cellular environment due to experimental difficulties. Approaches that have the potential to move the frontiers of allosteric network research are: 1) An *in silico* approach e.g. rigid residue scan, molecular dynamic simulation (Figure 11) studied in simulated crowded cellular environment instead of force field [143]. 2) Supported cell membrane sheet that allows study of protein:protein interactions in semi-native environment, but the approach has its limitations as one protein need to be anchored to the cell-membrane, which neither PSG or CRIPT are [139]. AGRB1 is anchored to the cell membrane, however it is not shown if it induces allosteric response in PDZ3 upon binding. 3) Combine the previously used NMR approach [15] (in solution) for allosteric mapping with in cell NMR [144].

We found that the C-terminal fragment of AGRB1 undergoes LLPS upon mixing with PSG, which is similar as previously observed for SynGap [10]. Therefore we speculated if the PSG:AGRB1 complex undergoes compartmentalization by phase separation to anchor PSD to the membrane bilayer. However the phenomenon phase separation has divided the biology field, in a recent feature article the question was raised, "sloppy science or groundbreaking idea" [145]. Phase separation may answer fundamental questions about cells organization of contents, but the methods are limited to study phase separation. To support

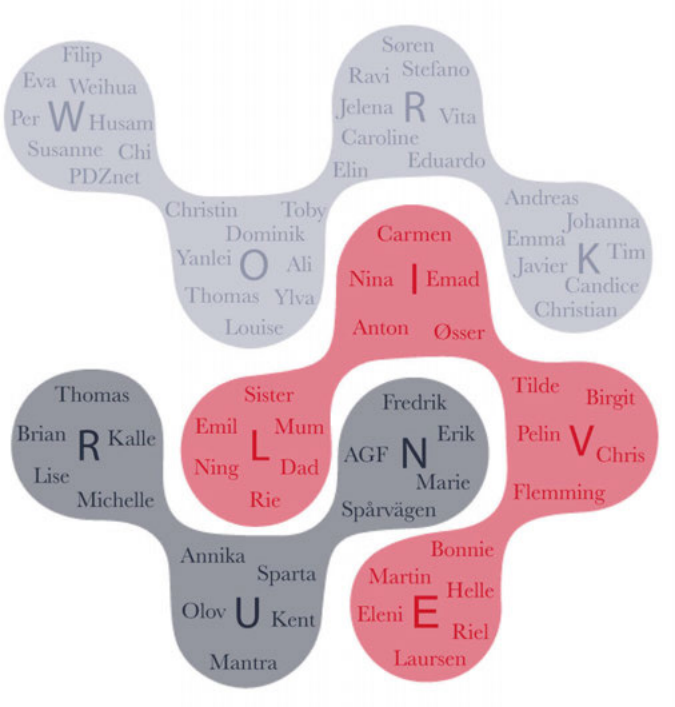
our hypothesis, further experiments need to be performed: 1) reconstitute PSD with several proteins including AGRB1 to test if the concentration threshold can be decreased for phase separation of PSG:AGRB1 complex, as previously reported for PSG:SynGap complex [87] and 2) visualize the clustering of membrane protein (AGRB1) at 2D supported lipid bilayer to characterize the LLPS formation of PSG:AGRB1 complex [73]. If further *in vitro* characterizations of PSG:AGRB1 complex corroborate the initial observations for phase separation of the PSG:AGRB1 complex, then we can speculate about the biological function. Neuronal cells are sensitive to the concentration of SynGap, which has been linked to the formation and compartmentalization of PSD through the concentration dependent phase separation of PSG:SynGap complex. The finding that AGRB1 phase separation is concentration dependent too and maybe associated with PSD assembly and compartmentalization, might help explain why low levels of AGRB1 is associated with cell death and Parkinson's disease [146]. LLPS research has gained increased attention the last couple of years due to the numerous findings of membraneless compartments with significant biological relevance. Nevertheless, we always need to keep in mind, when we observe new protein compartments driven by LLPS, is the *in vitro* observation also biologically relevant [73]? Mechanisms behind LLPS are still scarce and methods for characterization are limited.

All the performed binding and folding studies in this thesis are studied in dilute protein samples, even though PSD-95 is located in liquid droplets with high concentration of proteins. Is the stability, folding and binding mechanism of PSG affected by its association in liquid droplets. Further studies need to reveal that, but we speculate that it might be, as phase separation of SynGap or AGRB1 and PSG is sensitive to choice of buffer. The suggestion that the characteristics of PSG, e.g., folding mechanism, binding, stability etc., will be affected by association of PSG in liquid droplets is supported by recent studies. A model peptide undergoes peptide folding upon LLPS due to intermolecular stacking interactions [147]. PSG can undergo dimerization and multimerization upon SynGap binding, but the dimer conformation is favorable in dilute solutions [130]. Furthermore the major cellular energy metabolite ATP has been shown to induce protein folding, inhibit aggregation, increase stability and dissolve LLPS by specific binding [148]. Altogether the content of the thesis revealed differences in the binding mechanism, allosteric network, phase separation and folding mechanism for single domain PDZ3 and multi-domain PSG. Further studies need to reveal if the characteristics of PSG are similar in a dilute and dense protein context.

Concluding Remarks

The content of the thesis has contributed a functional perspective of the context (multi-domain protein, solution) on the stability, allosteric network, folding and binding profile for a single protein domain in supramodule. Previously, multi-domain proteins such as PSG were considered "beads on a string", a model where the PDZ3, SH3 and GK domains were described as functionally independent domains and therefore the protein domains were studied in isolation [19]. Studies from the last decade have shown that the PSG supramodule has functions that are distinct from functions of isolated domains, for example, SynGap binds with high affinity only to PSG, but not PDZ3 and binding of peptide to PDZ3 induces conformational change in SH3-GK tandem that facilitates ligand binding [5, 10, 149, 150]. Altogether, the findings in the thesis show how binding energy landscape, interactome, allosteric network, folding mechanism and phase separation are dependent on the context, which suggest that we need to be careful in interpretation of data obtained from isolated domains in multi-domain proteins.

Acknowledgements



References

1. Bustos, F.J., et al., *Epigenetic editing of the Dlg4/PSD95 gene improves cognition in aged and Alzheimer's disease mice.* Brain, 2017. **140**(12): p. 3252-3268.
2. Milnerwood, A.J. and L.A. Raymond, *Early synaptic pathophysiology in neurodegeneration: insights from Huntington's disease.* Trends Neurosci, 2010. **33**(11): p. 513-23.
3. Doucet, M.V., A. Harkin, and K.K. Dev, *The PSD-95/nNOS complex: new drugs for depression?* Pharmacol Ther, 2012. **133**(2): p. 218-29.
4. Coley, A.A. and W.J. Gao, *PSD95: A synaptic protein implicated in schizophrenia or autism?* Prog Neuropsychopharmacol Biol Psychiatry, 2018. **82**: p. 187-194.
5. Rademacher, N., et al., *Intramolecular domain dynamics regulate synaptic MAGUK protein interactions.* Elife, 2019. **8**.
6. Funke, L., S. Dakoji, and D.S. Bredt, *Membrane-associated guanylate kinases regulate adhesion and plasticity at cell junctions.* Annu Rev Biochem, 2005. **74**: p. 219-45.
7. Christensen, N.R., et al., *PDZ Domains as Drug Targets.* Adv Ther (Weinh), 2019. **2**(7): p. 1800143.
8. Zhang, L., et al., *SAP97 Binding Partner CRIPT Promotes Dendrite Growth In Vitro and In Vivo.* eNeuro, 2017. **4**(6).
9. Saro, D., et al., *A thermodynamic ligand binding study of the third PDZ domain (PDZ3) from the mammalian neuronal protein PSD-95.* Biochemistry, 2007. **46**(21): p. 6340-52.
10. Zeng, M., et al., *Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity.* Cell, 2016. **166**(5): p. 1163-1175 e12.
11. Niethammer, M., et al., *CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90.* Neuron, 1998. **20**(4): p. 693-707.
12. Doyle, D.A., et al., *Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ.* Cell, 1996. **85**(7): p. 1067-76.
13. Gautier, C., et al., *Seeking allosteric networks in PDZ domains.* Protein Eng Des Sel, 2018. **31**(10): p. 367-373.
14. Lockless, S.W. and R. Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families.* Science, 1999. **286**(5438): p. 295-9.
15. Petit, C.M., et al., *Hidden dynamic allostery in a PDZ domain.* Proc Natl Acad Sci U S A, 2009. **106**(43): p. 18249-54.

16. Songyang, Z., et al., *Recognition of unique carboxyl-terminal motifs by distinct PDZ domains.* Science, 1997. **275**(5296): p. 73-77.
17. Ernst, A., et al., *A structural portrait of the PDZ domain family.* J Mol Biol, 2014. **426**(21): p. 3509-19.
18. Tonikian, R., et al., *A specificity map for the PDZ domain family.* PLoS Biol, 2008. **6**(9): p. e239.
19. Feng, W. and M. Zhang, *Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density.* Nat Rev Neurosci, 2009. **10**(2): p. 87-99.
20. Delhommel, F., et al., *Structural Characterization of Whirlin Reveals an Unexpected and Dynamic Supramodule Conformation of Its PDZ Tandem.* Structure, 2017.
21. Gautier, C., et al., *Hidden kinetic traps in multidomain folding highlight the presence of a misfolded but functionally competent intermediate.* Proc Natl Acad Sci U S A, 2020. **117**(33): p. 19963-19969.
22. Kang, B.S., et al., *PDZ tandem of human syntenin: crystal structure and functional properties.* Structure, 2003. **11**(4): p. 459-68.
23. Cierpicki, T., J.H. Bushweller, and Z.S. Derewenda, *Probing the supramodular architecture of a multidomain protein: the structure of syntenin in solution.* Structure, 2005. **13**(2): p. 319-27.
24. Im, Y.J., et al., *Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization.* J Biol Chem, 2003. **278**(48): p. 48099-104.
25. Long, J.F., et al., *Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95.* J Mol Biol, 2003. **327**(1): p. 203-14.
26. Harris, B.Z. and W.A. Lim, *Mechanism and role of PDZ domains in signaling complex assembly.* J Cell Sci, 2001. **114**(Pt 18): p. 3219-31.
27. Feng, W., et al., *Tandem PDZ repeats in glutamate receptor-interacting proteins have a novel mode of PDZ domain-mediated target binding.* Nat Struct Biol, 2003. **10**(11): p. 972-8.
28. Zhang, J., et al., *Supertertiary structure of the MAGUK core from PSD-95.* Structure, 2013. **21**(3): p. 402-13.
29. McCann, J.J., et al., *Supertertiary structure of the synaptic MAGuK scaffold proteins is conserved.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. 15775-80.
30. de Mendoza, A., H. Suga, and I. Ruiz-Trillo, *Evolution of the MAGUK protein gene family in premetazoan lineages.* BMC Evol Biol, 2010. **10**: p. 93.
31. Li, Y., et al., *Structure of Crumbs tail in complex with the PALS1 PDZ-SH3-GK tandem reveals a highly specific assembly mechanism for the apical Crumbs complex.* Proc Natl Acad Sci U S A, 2014. **111**(49): p. 17444-9.
32. Changeux, J.P., *The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine.* Cold Spring Harb Symp Quant Biol, 1961. **26**: p. 313-8.
33. Monod, J., J. Wyman, and J.P. Changeux, *On the Nature of Allosteric Transitions: A Plausible Model.* J Mol Biol, 1965. **12**: p. 88-118.

34. Wodak, S.J., et al., *Allostery in Its Many Disguises: From Theory to Applications.* Structure, 2019.
35. Gunasekaran, K., B. Ma, and R. Nussinov, *Is allostery an intrinsic property of all dynamic proteins?* Proteins, 2004. **57**(3): p. 433-43.
36. Stock, G. and P. Hamm, *A non-equilibrium approach to allosteric communication.* Philos Trans R Soc Lond B Biol Sci, 2018. **373**(1749).
37. Piana, S., K. Lindorff-Larsen, and D.E. Shaw, *How robust are protein folding simulations with respect to force field parameterization?* Biophys J, 2011. **100**(9): p. L47-9.
38. Cooper, A. and D.T. Dryden, *Allostery without conformational change. A plausible model.* Eur Biophys J, 1984. **11**(2): p. 103-9.
39. Kumawat, A. and S. Chakrabarty, *Hidden electrostatic basis of dynamic allostery in a PDZ domain.* Proc Natl Acad Sci U S A, 2017. **114**(29): p. E5825-E5834.
40. Salinas, V.H. and R. Ranganathan, *Coevolution-based inference of amino acid interactions underlying protein function.* Elife, 2018. **7**.
41. Grover, A.K., *Use of allosteric targets in the discovery of safer drugs.* Med Princ Pract, 2013. **22**(5): p. 418-26.
42. Groebe, D.R., *Screening for positive allosteric modulators of biological targets.* Drug Discov Today, 2006. **11**(13-14): p. 632-9.
43. Zhang, Y., et al., *The binding mode of vilazodone in the human serotonin transporter elucidated by ligand docking and molecular dynamics simulations.* Phys Chem Chem Phys, 2020. **22**(9): p. 5132-5144.
44. Felder, C.C., *GPCR drug discovery-moving beyond the orthosteric to the allosteric domain.* Adv Pharmacol, 2019. **86**: p. 1-20.
45. Gulzar, A., et al., *Energy Transport Pathways in Proteins: A Non-equilibrium Molecular Dynamics Simulation Study.* J Chem Theory Comput, 2019.
46. Ota, N. and D.A. Agard, *Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion.* J Mol Biol, 2005. **351**(2): p. 345-54.
47. Liu, J. and R. Nussinov, *Energetic redistribution in allostery to execute protein function.* Proc Natl Acad Sci U S A, 2017. **114**(29): p. 7480-7482.
48. Hilser, V.J., J.O. Wrabl, and H.N. Motlagh, *Structural and energetic basis of allostery.* Annu Rev Biophys, 2012. **41**: p. 585-609.
49. Raman, A.S., K.I. White, and R. Ranganathan, *Origins of Allostery and Evolvability in Proteins: A Case Study.* Cell, 2016. **166**(2): p. 468-480.
50. McLaughlin, R.N., Jr., et al., *The spatial architecture of protein function and adaptation.* Nature, 2012. **491**(7422): p. 138-42.
51. Harms, M.J. and J.W. Thornton, *Analyzing protein structure and function using ancestral gene reconstruction.* Curr Opin Struct Biol, 2010. **20**(3): p. 360-6.
52. Melero, C., et al., *Quantification of the transferability of a designed protein specificity switch reveals extensive epistasis in molecular recognition.* Proc Natl Acad Sci U S A, 2014. **111**(43): p. 15426-31.
53. Bridgham, J.T., E.A. Ortlund, and J.W. Thornton, *An epistatic ratchet constrains the direction of glucocorticoid receptor evolution.* Nature, 2009. **461**(7263): p. 515-9.
54. Laursen, L., et al., *Supertertiary protein structure affects an allosteric network.* Proc Natl Acad Sci U S A, 2020. **117**(39): p. 24294-24304.

55. Zuckerkandl, E. and L. Pauling, *Molecules as documents of evolutionary history.* J Theor Biol, 1965. **8**(2): p. 357-66.
56. Thornton, J.W., E. Need, and D. Crews, *Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling.* Science, 2003. **301**(5640): p. 1714-7.
57. Jermann, T.M., et al., *Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily.* Nature, 1995. **374**(6517): p. 57-9.
58. Hultqvist, G., et al., *Emergence and evolution of an interaction between intrinsically disordered proteins.* Elife, 2017. **6**.
59. Anderson, D.P., et al., *Evolution of an ancient protein function involved in organized multicellularity in animals.* Elife, 2016. **5**: p. e10147.
60. Sakarya, O., et al., *Evolutionary expansion and specialization of the PDZ domains.* Mol Biol Evol, 2010. **27**(5): p. 1058-69.
61. Yang, Z. and B. Rannala, *Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method.* Mol Biol Evol, 1997. **14**(7): p. 717-24.
62. Williams, P.D., et al., *Assessing the accuracy of ancestral protein reconstruction methods.* PLoS Comput Biol, 2006. **2**(6): p. e69.
63. Hanson-Smith, V., B. Kolaczkowski, and J.W. Thornton, *Robustness of ancestral sequence reconstruction to phylogenetic uncertainty.* Mol Biol Evol, 2010. **27**(9): p. 1988-99.
64. Eick, G.N., et al., *Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty.* Mol Biol Evol, 2017. **34**(2): p. 247-261.
65. Benner, S.A., *The past as the key to the present: resurrection of ancient proteins from eosinophils.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 4760-1.
66. Gumulya, Y., et al., *Engineering highly functional thermostable proteins using ancestral sequence reconstruction.* Nature Catalysis, 2018. **1**(11): p. 878-888.
67. Risso, V.A., J.A. Gavira, and J.M. Sanchez-Ruiz, *Thermostable and promiscuous Precambrian proteins.* Environ Microbiol, 2014. **16**(6): p. 1485-9.
68. Wheeler, L.C., et al., *The thermostability and specificity of ancient proteins.* Curr Opin Struct Biol, 2016. **38**: p. 37-43.
69. Hobbs, J.K., et al., *On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of Bacillus.* Mol Biol Evol, 2012. **29**(2): p. 825-35.
70. Sternke, M., K.W. Tripp, and D. Barrick, *Consensus sequence design as a general strategy to create hyperstable, biologically active proteins.* Proc Natl Acad Sci U S A, 2019. **116**(23): p. 11275-11284.
71. Trudeau, D.L., M. Kaltenbach, and D.S. Tawfik, *On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins.* Mol Biol Evol, 2016. **33**(10): p. 2633-41.
72. Boeynaems, S., et al., *Protein Phase Separation: A New Phase in Cell Biology.* Trends Cell Biol, 2018. **28**(6): p. 420-435.
73. Feng, Z., et al., *Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications.* J Biol Chem, 2019. **294**(40): p. 14823-14835.

74. Rabouille, C. and S. Alberti, *Cell adaptation upon stress: the emerging role of membrane-less compartments.* Current Opinion in Cell Biology, 2017. **47**: p. 34-42.
75. Li, P., et al., *Phase transitions in the assembly of multivalent signalling proteins.* Nature, 2012. **483**(7389): p. 336-40.
76. Brangwynne, C.P., et al., *Germline P granules are liquid droplets that localize by controlled dissolution/condensation.* Science, 2009. **324**(5935): p. 1729-32.
77. Tao, C.L., et al., *Differentiation and Characterization of Excitatory and Inhibitory Synapses by Cryo-electron Tomography and Correlative Microscopy.* J Neurosci, 2018. **38**(6): p. 1493-1510.
78. Kaizuka, T. and T. Takumi, *Postsynaptic density proteins and their involvement in neurodevelopmental disorders.* The Journal of Biochemistry, 2018. **163**(6): p. 447-455.
79. Sheng, M. and E. Kim, *The postsynaptic organization of synapses.* Cold Spring Harb Perspect Biol, 2011. **3**(12).
80. Frank, R.A.W., et al., *Hierarchical organization and genetically separable subfamilies of PSD95 postsynaptic supercomplexes.* J Neurochem, 2017. **142**(4): p. 504-511.
81. Collins, M.O., et al., *Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome.* J Neurochem, 2006. **97 Suppl 1**: p. 16-23.
82. Chen, X., et al., *PSD-95 is required to sustain the molecular organization of the postsynaptic density.* J Neurosci, 2011. **31**(17): p. 6329-38.
83. Zeng, M., G. Bai, and M. Zhang, *Anchoring high concentrations of SynGAP at postsynaptic densities via liquid-liquid phase separation.* Small GTPases, 2019. **10**(4): p. 296-304.
84. Gamache, T.R., Y. Araki, and R.L. Huganir, *Twenty Years of SynGAP Research: From Synapses to Cognition.* J Neurosci, 2020. **40**(8): p. 1596-1605.
85. McMahon, A.C., et al., *SynGAP isoforms exert opposing effects on synaptic strength.* Nat Commun, 2012. **3**: p. 900.
86. Araki, Y., et al., *SynGAP isoforms differentially regulate synaptic plasticity and dendritic development.* Elife, 2020. **9**.
87. Zeng, M., et al., *Reconstituted Postsynaptic Density as a Molecular Platform for Understanding Synapse Formation and Plasticity.* Cell, 2018. **174**(5): p. 1172-1187 e16.
88. Hayashi, Y., et al., *Liquid-Liquid Phase Separation in Physiology and Pathophysiology of the Nervous System.* J Neurosci, 2021. **41**(5): p. 834-844.
89. Bai, G., Y. Wang, and M. Zhang, *Gephyrin-mediated formation of inhibitory postsynaptic density sheet via phase separation.* Cell Res, 2021. **31**(3): p. 312-325.
90. Chi, C.N., et al., *Two conserved residues govern the salt and pH dependencies of the binding reaction of a PDZ domain.* J Biol Chem, 2006. **281**(48): p. 36811-8.
91. Oparin, A.I., *Evolution of the concepts of the origin of life, 1924-1974.* Orig Life, 1976. **7**(1): p. 3-8.

92. Hyman, A.A., C.A. Weber, and F. Julicher, *Liquid-liquid phase separation in biology.* Annu Rev Cell Dev Biol, 2014. **30**: p. 39-58.
93. Dasmeh, P. and A. Wagner, *Natural selection on the phase-separation properties of FUS during 160 million years of mammalian evolution.* Mol Biol Evol, 2020.
94. Zhang, J. and J.R. Yang, *Determinants of the rate of protein sequence evolution.* Nat Rev Genet, 2015. **16**(7): p. 409-20.
95. Dill, K.A. and J.L. MacCallum, *The protein-folding problem, 50 years on.* Science, 2012. **338**(6110): p. 1042-6.
96. Braselmann, E., J.L. Chaney, and P.L. Clark, *Folding the proteome.* Trends Biochem Sci, 2013. **38**(7): p. 337-44.
97. Niwa, T., et al., *Translation-coupled protein folding assay using a protease to monitor the folding status.* Protein Sci, 2019. **28**(7): p. 1252-1261.
98. Wruck, F., et al., *Translation and folding of single proteins in real time.* Proc Natl Acad Sci U S A, 2017. **114**(22): p. E4399-E4407.
99. Chi, C.N., et al., *A conserved folding mechanism for PDZ domains.* FEBS Lett, 2007. **581**(6): p. 1109-13.
100. Calosci, N., et al., *Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins.* Proc Natl Acad Sci U S A, 2008. **105**(49): p. 19241-6.
101. Hultqvist, G., et al., *An expanded view of the protein folding landscape of PDZ domains.* Biochem Biophys Res Commun, 2012. **421**(3): p. 550-3.
102. Gautier, C., et al., *Addressing the role of the alpha-helical extension in the folding of the third PDZ domain from PSD-95.* Sci Rep, 2017. **7**(1): p. 12593.
103. Murciano-Calles, J., et al., *The impact of extra-domain structures and post-translational modifications in the folding/misfolding behaviour of the third PDZ domain of MAGUK neuronal protein PSD-95.* PLoS One, 2014. **9**(5): p. e98124.
104. Murciano-Calles, J., et al., *A thermodynamic study of the third PDZ domain of MAGUK neuronal protein PSD-95 reveals a complex three-state folding behavior.* Biophys Chem, 2014. **185**: p. 1-7.
105. Murciano-Calles, J., et al., *An oligomeric equilibrium intermediate as the precursory nucleus of globular and fibrillar supramacromolecular assemblies in a PDZ domain.* Biophys J, 2010. **99**(1): p. 263-72.
106. Murciano-Calles, J., *The Conformational Plasticity Vista of PDZ Domains.* Life (Basel), 2020. **10**(8).
107. Inanami, T., T.P. Terada, and M. Sasai, *Folding pathway of a multidomain protein depends on its topology of domain connectivity.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 15969-74.
108. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
109. Bach, A., et al., *A high-affinity, dimeric inhibitor of PSD-95 bivalently interacts with PDZ1-2 and protects against ischemic brain damage.* Proc Natl Acad Sci U S A, 2012. **109**(9): p. 3317-22.
110. Sainlos, M., et al., *Biomimetic divalent ligands for the acute disruption of synaptic AMPAR stabilization.* Nat Chem Biol, 2011. **7**(2): p. 81-91.

111. Udugamasooriya, D.G., S.C. Sharma, and M.R. Spaller, *A chemical library approach to organic-modified peptide ligands for PDZ domain proteins: a synthetic, thermodynamic and structural investigation.* Chembiochem, 2008. **9**(10): p. 1587-9.
112. Piserchio, A., et al., *Targeting specific PDZ domains of PSD-95; structural basis for enhanced affinity and enzymatic stability of a cyclic peptide.* Chem Biol, 2004. **11**(4): p. 469-73.
113. McLysaght, A., K. Hokamp, and K.H. Wolfe, *Extensive genomic duplication during early chordate evolution.* Nat Genet, 2002. **31**(2): p. 200-4.
114. Putnam, N.H., et al., *The amphioxus genome and the evolution of the chordate karyotype.* Nature, 2008. **453**(7198): p. 1064-71.
115. Bonnet, S.A., et al., *Synaptic state-dependent functional interplay between postsynaptic density-95 and synapse-associated protein 102.* J Neurosci, 2013. **33**(33): p. 13398-409.
116. Howard, M.A., et al., *The role of SAP97 in synaptic glutamate receptor dynamics.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3805-10.
117. Niello, M., et al., *Allosteric Modulation of Neurotransmitter Transporters as a Therapeutic Strategy.* Trends Pharmacol Sci, 2020. **41**(7): p. 446-463.
118. Zhao, Y., et al., *Cysteine modifiers suggest an allosteric inhibitory site on the CAL PDZ domain.* Biosci Rep, 2018. **38**(4).
119. Wheeler, R.J., *Therapeutics-how to treat phase separation-associated diseases.* Emerg Top Life Sci, 2020. **4**(3): p. 307-318.
120. Hamdan, F.F., et al., *Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation.* N Engl J Med, 2009. **360**(6): p. 599-605.
121. Mullard, A., *Biomolecular condensates pique drug discovery curiosity.* Nat Rev Drug Discov, 2019.
122. Deiana, A., et al., *Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell.* PLoS One, 2019. **14**(8): p. e0217889.
123. Darling, A.L., et al., *Intrinsically Disordered Proteome of Human Membrane-Less Organelles.* Proteomics, 2018. **18**(5-6): p. e1700193.
124. Wu, X., et al., *Liquid-Liquid Phase Separation in Neuronal Development and Synaptic Signaling.* Dev Cell, 2020. **55**(1): p. 18-29.
125. Kaya, C., et al., *MCPath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W249-55.
126. Kalescky, R., J. Liu, and P. Tao, *Identifying key residues for protein allostery through rigid residue scan.* J Phys Chem A, 2015. **119**(9): p. 1689-700.
127. Du, Q.S., et al., *Correlation analysis for protein evolutionary family based on amino acid position mutations and application in PDZ domain.* PLoS One, 2010. **5**(10): p. e13207.
128. Gianni, S., et al., *Sequence-specific long range networks in PSD-95/discs large/ZO-1 (PDZ) domains tune their binding selectivity.* J Biol Chem, 2011. **286**(31): p. 27167-75.
129. Gerek, Z.N. and S.B. Ozkan, *Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning.* PLoS Comput Biol, 2011. **7**(10): p. e1002154.

130. Zeng, M., et al., *PDZ Ligand Binding-Induced Conformational Coupling of the PDZ-SH3-GK Tandems in PSD-95 Family MAGUKs.* J Mol Biol, 2018. **430**(1): p. 69-86.
131. Zimmerman, S.S. and H.A. Scheraga, *Stability of cis, trans, and nonplanar peptide groups.* Macromolecules, 1976. **9**(3): p. 408-16.
132. Zosel, F., et al., *A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction.* Nat Commun, 2018. **9**(1): p. 3332.
133. Craveur, P., et al., *Cis-trans isomerization of omega dihedrals in proteins.* Amino Acids, 2013. **45**(2): p. 279-89.
134. Ogura, K., et al., *Solution structure of N-terminal SH3 domain of Vav and the recognition site for Grb2 C-terminal SH3 domain.* J Biomol NMR, 2002. **22**(1): p. 37-46.
135. Sarkar, P., et al., *Proline cis-trans isomerization controls autoinhibition of a signaling protein.* Mol Cell, 2007. **25**(3): p. 413-26.
136. Kilinc, D., *The Emerging Role of Mechanics in Synapse Formation and Plasticity.* Front Cell Neurosci, 2018. **12**: p. 483.
137. Ahmed, W.W., et al., *Mechanical tension modulates local and global vesicle dynamics in neurons.* Cell Mol Bioeng, 2012. **5**(2): p. 155-164.
138. Gillespie, P.G. and U. Muller, *Mechanotransduction by hair cells: models, molecules, and mechanisms.* Cell, 2009. **139**(1): p. 33-44.
139. Erlendsson, S., et al., *Mechanisms of PDZ domain scaffold assembly illuminated by use of supported cell membrane sheets.* Elife, 2019. **8**.
140. Hartenstein, V. and A. Stollewerk, *The evolution of early neurogenesis.* Dev Cell, 2015. **32**(4): p. 390-407.
141. van Dam, T.J.P., J. Bos, and B. Snel, *Evolution of the Ras-like small GTPases and their regulators.* Small GTPases, 2011. **2**(1): p. 4-16.
142. Alie, A. and M. Manuel, *The backbone of the post-synaptic density originated in a unicellular ancestor of choanoflagellates and metazoans.* BMC Evol Biol, 2010. **10**: p. 34.
143. Feig, M., et al., *Crowding in Cellular Environments at an Atomistic Level from Computer Simulations.* J Phys Chem B, 2017. **121**(34): p. 8009-8025.
144. Serber, Z. and V. Dotsch, *In-cell NMR spectroscopy.* Biochemistry, 2001. **40**(48): p. 14317-23.
145. Leslie, M., *Separation anxiety.* Science, 2021. **371**(6527): p. 336-338.
146. Zhong, J., et al., *Integrated profiling of single cell epigenomic and transcriptomic landscape of Parkinson's disease mouse brain.* bioRxiv, 2020: p. 2020.02.04.933259.
147. Edun, D.N., M.R. Flanagan, and A.L. Serrano, *Does liquid–liquid phase separation drive peptide folding?* Chemical Science, 2021. **12**(7): p. 2474-2479.
148. Kang, J., L. Lim, and J. Song, *ATP induces protein folding, inhibits aggregation and antagonizes destabilization by effectively mediating water-protein-ion interactions, the heart of protein folding and aggregation.* bioRxiv, 2020: p. 2020.06.21.163758.
149. McGee, A.W., et al., *Structure of the SH3-guanylate kinase module from PSD-95 suggests a mechanism for regulated assembly of MAGUK scaffolding proteins.* Mol Cell, 2001. **8**(6): p. 1291-301.

150. Qian, Y. and K.E. Prehoda, *Interdomain interactions in the tumor suppressor discs large regulate binding to the synaptic protein GukHolder.* J Biol Chem, 2006. **281**(47): p. 35757-63.
151. Hayatshahi, H.S., et al., *Probing Protein Allostery as a Residue-Specific Concept via Residue Response Maps.* Journal of Chemical Information and Modeling, 2019. **59**(11): p. 4691-4705.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1741

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-438160

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2021