CHAPTER 11

Discussion

Identification and characterization of CD2BP2 defined GYF domains as a new family of PRD. In comparison with the SH3, WW, and EVH1 families, the biological implications and recognition motifs of GYF domains were largely elusive. Understanding of the functions of GYF domain comprising proteins can be gained from analysis of their interaction partners, since the formation of protein complexes is central to most biological processes. In this work, the binding properties of different GYF domains were characterized and potential interaction partners identified. For several selected candidates, encounter of the GYF domains was verified by complementary methods, putting the GYF domain containing proteins into diverse functional contexts.

11.1 Recognition Codes of GYF Domains

The strategy to define the recognition signatures of individual GYF domains comprised phage display and SPOT substitution analysis, for a global and local analysis of the sequence space of binders. Interaction sites within natural proteins were subsequently detected by additional SPOT experiments with the implementation of motif repetition as an important feature[319]. Seven GYF domains were selected for analysis to include members of both the CD2BP2 and SMY2 subfamilies of GYF domains from the three eukaryotic kingdoms plantae, fungi, and animalia. The first subfamily, absent from plants, was represented by human CD2BP2-, *Drosophila melanogaster* Q9VKV5-, and yeast LIN1-GYF. SMY2 subfamily representatives were human PERQ2-GYF, the GYN4-GYF domain from *Arabidopsis thaliana* and the GYF domains of the two yeast paralogs SMY2 and SYH1. SYH1 and GYN4 comprise proline-rich sites N- and C-terminal to the GYF domain, respectively, calling for an investigation with respect to possible intramolecular interactions. Binding to sites within the same molecule has been shown to be employed as means of regulation by certain SH3 domain containing proteins[74,80].

The GYF domain of LIN1 could not be expressed as stable protein in amounts sufficient for binding experiments and therefore had to be excluded from analysis. The remaining GYF domains expressed well and were soluble, at least when fused to GST. Correct folding of SMY2-type GYF domains was exemplified for SMY2- and PERQ2-GYF (Chapter 5). The Q9VKV5-

GYF construct turned out to lack peptide binding competence, presumably due to misfolding or partial degradation of the domain attached to GST. This limited analysis of the CD2BP2 subfamily to its prototypic GYF domain.

11.1.1 Definition of the Recognition Motifs

Initial substitution analysis of the CD2 peptide ligand, containing the motif PPPGHR[23,42], delineated the signature (R/K/G)xxPPGx(R/K) to be important for the interaction with CD2BP2-GYF (Chapter 6). This finding is in agreement with the NMR structural data that showed several van der Waals contacts between the PPG motif and the aromatic side-chains of the GYF domain. Double substitution analysis (Fig. 8.2) clearly demonstrated the positively charged amino acid following the PPG motif to be the more discriminatory position of the CD2 class of ligands than the N-terminal position defined by (R/K/G). Further support came from binding analysis of peptides from various proteins (Fig. 6.3, Fig. 6.7, and Fig. 8.3). Thereby, the consensus for the CD2 class of ligands was redefined as the more relaxed signature PPGx(R/K) (Chapter 8).

Phage display was employed to scan a large peptide sequence space for potential interaction signatures. Subsequent substitution analysis of binding peptides confirmed and refined the obtained consensus motifs. All GYF domains converged on the recognition of the signature PPGΦ, with Φ standing for hydrophobic residues.

PERQ2-, SMY2-, and SYH1-GYF domains bound peptides displaying either a phenylic or aliphatic side-chain at the Φ position, with a subtle preference for the latter. In contrast, CD2BP2-GYF strongly selected for aromatic residues at the last position of the motif, particularly favoring tryptophan. The short GYF domain construct of GYN4 without internal PRS, revealed an even tighter ligand spectrum, preferentially binding to the motif PPGF. Interestingly, the PPGx(R/K) binding motif was rarely found by phage display with CD2BP2-GYF and six rounds of panning were required to detect it (3 incidences in 20 sequenced, phage displayed peptides, Table 8.1). In contrast, the predominating PPGW binding motif was already apparent after three rounds. These observations are in line with initial phage display experiments (Fig. 5.3), showing significant enrichment of PPGL- and PPGW- but not CD2-motif displaying phages on the respective GYF domains from yeast and CD2BP2. Peptides MSL5S1 (PPGL motif) and PD1 (PPGW motif) both have a more than 3 times higher affinity for their GYF domains compared to the CD2 peptide (Fig. 8.4 and Table 9.4, only NMR-derived values compared). Therefore, the low binding affinity of the CD2 motif for CD2BP2-GYF is possibly the reason for poor enrichment. Consequently, more rounds of panning were required to detect

the motif in the small number of analyzed phage inserts. Other reasons for the differential behavior of the ligand classes, such as bias for/against a ligand motif in the original phage libraries or general deleterious effects of the PPGx(R/K) motif on phage stability, infectivity or cell survival, could be excluded. No significant bias against charged residues in the unselected focused library was observed (data not shown) and the X_9 library has been used extensively for selection of SH3 ligands with similarly charged recognition signatures[46,47].

11.1.2 Evaluation of the Recognition Motifs

In conclusion, the combination of phage display and substitution analysis allowed the definition of the recognition motif PPGΦ for all GYF domains analyzed. The rare selection for CD2 ligands marked clearly the limitations of phage display. Similar to other screening methods based on short peptide libraries, phage display selects mostly for binding sites with highest affinity present in a library (e.g. PPGW ligands for CD2BP2). However, nature frequently utilizes several suboptimal binding sites to establish productive encounter between two proteins[319], as observed for the PPGx(R/K) motifs in CD2[23,42] and SmB/B' (Chapter 6). To acknowledge this fact, databases were searched to uncover multiple relaxed consensus sites in proteins of the respective species. Supportive interactions of flanking residues, however, could compensate for deviations from the relaxed recognition motif, as was indicated by the substitution analysis of the MSL5S1 peptide (Fig. 9.10). Consequently, lower affinity binding sites that do not comprise a PPG core motif are conceivable. Yeast two-hybrid screens were conducted to investigate if natural proteins which bind to GYF domains do indeed tend to contain the consensus motifs, as they were defined by the recognition motif-driven approach. Since the GYF domain of CD2BP2, other CD2BP2 fragments (K4 bait and K6 bait, see Table 4.2), and the full-length protein displayed auto-activation in the yeast two-hybrid system (unpublished observations), and since proteome wide two-hybrid screens for yeast proteins had been performed[46,248,320], analysis was restricted to GYF domains of PERQ2 and GYN4 (Chapter 9). The screens were expected to pick up all sites with superior binding properties, regardless of the affinity of individual, embedded recognition motifs. All proteins obtained were among the most likely interaction partners according to the consensus-based strategy. Their binding could be explained solely on the basis of the recognition motifs identified before, demonstrating the comprehensiveness of the interaction analysis.

Candidates were largely enriched for spliceosomal proteins but hints at other functional implications emerged as well (Chapters 6, 8, and 9). Interestingly, the PPGW motif is the preferred binding site of CD2BP2-GYF in phage display and titration experiments; however, binding to CD2BP2 could not be confirmed for any of the selected candidates with this

recognition signature. Analysis of other candidates is necessary to reveal the physiological importance of the PPGW consensus motif as CD2BP2-GYF docking site.

11.2 Functional Implications of GYF Domains

11.2.1 Involvement in CD2 Signaling

CD2BP2 was identified as interaction partner of the transmembrane adhesion receptor CD2, and hence its implication in signal transduction of T lymphocytes was assumed[42,43]. Preferential nuclear localization observed for CD2BP2 in Jurkat J77 and HeLa S3 cells and its homolog LIN1 in yeast[209,217], however, disfavors such a function. Moreover, CD2 overexpression in HeLa S3 cells did not redirect CD2BP2 to the membrane (Fig. 6.8). Mann and coworkers found SmB/B' in spreading initiation centers which are potential precursors to focal adhesions[321]. These ribonucleoprotein complexes were exclusively detected in primary or non-tumor-derived cells and only during very early stages of spreading. The results highlight the limitation of cell lines to reflect the different physiological aspects of cellular behavior and protein functions. In this regard, a cytoplasmic localization of CD2BP2 under certain conditions and/or in primary cells is not disproved and conclusions about its implication in CD2 signaling await further analysis.

11.2.2 Spliceosomal Functions

Several lines of evidence suggest CD2BP2 and the yeast homolog LIN1 to be predominantly involved in splicing or splicing associated processes. Firstly, they are primarily nuclear[209,217] (see 11.2.1 above) and CD2BP2 colocalizes with the splicing factor SmB/B' (Fig. 6.6). Secondly, direct interactions with splicing factors were shown for the two orthologs, defining them as components of the U5 snRNP. CD2BP2 bound SmB/B', NPWBP (Chapters 6 and 8), U5-15K, and U5-102K[217], while LIN1 recognized PRP8[209,322]. In addition, SPOT analysis data hint at interactions of several other splicing factors with CD2BP2 (Fig. 6.7 and Fig. 8.3). The contribution of a 'proline front' in spliceosomal proteins to binding of adaptor domains has already been proposed for the FBP21-WW domains[207] and could be extended to include GYF domains. CD2BP2 and LIN1 have been isolated from spliceosomal complexes[216,217,322], providing a third source of evidence that supports a relevant function in splicing.

The current model associates CD2BP2 with tri-snRNP formation. Interactions of U5-102K and U5-15K with CD2BP2 locate it to the U5 snRNP and potentially mask protein-binding sites in the premature snRNP. Phosphorylation of CD2BP2 coincides with its release from the

spliceosomal component during formation of the U4/U6•U5 tri-snRNP[217,323]. The dissociation could unmask relevant interfaces to initiate productive tri-snRNP assembly. Alternative spliceosomal maturation pathways, involvement in the splicing of specific mRNAs or partial functional redundancy are the likely reasons for the lack of an obvious phenotype upon depletion of CD2BP2[217] or LIN1[209,322,324,N]. Future work will have to address this issue to allow for a detailed understanding of the spliceosomal functions of CD2BP2.

A similar function to that of CD2BP2 is anticipated for PERQ2. It was found to interact with the splicing factors SmB/B', SmN, and NPWBP. SPOT analysis results proposed additional splicing factors to be recognized by PERQ2-GYF, including different isoforms of BBP/hSF1 (Fig. 9.3). Conservation of the PERQ2–BBP/hSF1 interaction in yeast, namely between the PERQ2 homologs SMY2 or SYH1 and MSL5[210] (see Chapter 9), corroborates spliceosomal functions of PERQ2, SMY2, and SYH1. Cytoplasmic localization of the two GYF domain containing proteins in yeast argues against a physiological interaction with nuclear MSL5[325]. However, as pointed out previously, compartmentalization of individual proteins can drastically change under distinct conditions. Localization studies will help to clarify the functional implications of the human counterparts. As for CD2BP2, both nuclear and cytoplasmic compartmentalization is conceivable for PERQ2. The interaction partners presented herein are predominantly found in the nucleus and therefore favor a similar localization of PERQ2. In the case that PERQ2 is restricted to the cytoplasm, it could associate with NPWBP and SmB/B' under certain conditions[321,326] or throughout snRNP maturation[177]. Due to the high similarity of their GYF domains (86 %; also see Fig. 1.3), the functional predictions for PERQ2, based on the GYF domain interaction partners, can be extended to PERQ1. Such an extension can be legitimized since it proved correct for SMY2 and SYH1; these two paralogs share less similarity in their GYF domains (67 %), yet reveal an almost identical binding behavior. Sequence diversity outside the GYF domain, however, indicates overlapping, rather than redundant functions of the protein pairs.

11.2.3 Regulation of Translation Initiation

EAP1 was identified as inhibitor of cap-dependent translation, similar to p20, competing with eukaryotic initiation factor 4G (eIF4G) for binding to the cap binding protein eIF4E[308]. The involvement of EAP1 in the regulation of translation initiation was further substantiated by the findings that EAP1 attenuated the rapamycin induced translation of GCN4[327] and it exhibited a general attenuation of translation in cells with mutations in the secretory pathway[328]. Moreover, EAP1 was found in messenger ribonucleoprotein complexes together with the poly (A) binding

protein 1[329,330], a protein which increases the efficiency of translation[331]. These results suggest that the EAP1 interaction partners SMY2 and SYH1 may function in regulation of translation initiation. Support for the implication of SYH1 in processes regulating initiation of translation arose also from yeast two-hybrid experiments. SYH1 was identified as an interaction partner of p33 and p39[248], two subunits of eIF3[332-334]. As part of a multifactor complex[335], eIF3 assists in the assembly of the 48 S preinitiation complex which scans mRNA for the AUG start codon[334,336]. Global localization studies of yeast proteins[325] revealed a cytoplasmic localization of SMY2, SYH1, and EAP1, supporting EAP1 as physiological interaction partner.

11.2.4 Other Possible Functions

In addition to the proposed implications in CD2 signaling, mRNA splicing, and regulation of translation initiation, other putative functions for GYF domain containing proteins emerged. Interaction with the regulator of immunoproteasomes, PI31, could potentially link splicing and presentation of peptides by MHC I[337].

The *Arabidopsis thaliana* homolog of CNOT4 was identified as interaction partner for GYN4, and both the yeast and the human counterparts comprise consensus motifs recognized by the respective GYF domains *in vitro* (Fig. 9.3). CNOT4 forms part of the transcriptional repressor complex CCR4-NOT and possesses E3 ubiquitin ligase activity[338,339]. Encounter of CNOT4 could either control the degradation of GYF domain-containing proteins via the ubiquitin-mediated degradation pathway or link these proteins to transcriptional regulation.

Proteins with a GYF domain could also exert transport functions (see Chapter 1.6.3). SMY2 and its *Schizosaccharomyces pombe* homolog, MPD2, are multicopy suppressors of mutations in MYO2 and PTR1, respectively, proteins with ascribed transport functionality. MYO2 is a motor protein, important for actin-mediated transport phenomena[128], while a mutation in PTR1 results in an mRNA export defect[225]. SYH1 is synthetically lethal with ric1, which encodes a protein with vesicular transport function[222-224]. The SMY2 and SYH1 binding partner EAP1 genetically interacts with mutations in proteins of the secretory pathway (see above); MSL5 is involved in the nuclear retention of mRNA[340]; and MUD2 functionally interacts with SUB2[226], also important for mRNA export[227]. The eIF4E protein, known to associate with EAP1[308], also regulates the nuclear export of mRNA[341] in addition to its major function in translation initiation. Furthermore, the homolog of the CD2BP2-GYF interaction partner U5-15K in *Schizosaccharomyces pombe*, DIM1, has been implicated in the export of specific mRNAs[342].

The identification of binding sites in viral proteins uncovers another exciting twist in the theme of GYF domain functions. It is known that intracellular pathogens exploit host proteins for their

own benefit[312,343]. Interactions between phosphotyrosine 112 in the vaccina virus protein A36R and the SH2 domain of the adaptor protein Nck, for example, are implicated in the recruitment of the cellular actin assembly machinery for viral motility[312]. The same principle utilizes the intracellular parasite *Listeria monocytogenes* to propel through the host cell[54,312]. Here, the bacterial surface protein ActA comprises four proline-rich recognition sites for the EVH1 domains of the host proteins Mena and VASP. Moreover, many viruses such as HIV and Ebola usurp the vacuolar protein sorting machinery to facilitate viral release. Normally, the vacuolar protein sorting complex is involved in vesicle budding into the late endosome/multivesicular bodies[343-346]. The late domain of viral proteins is proposed to mimic hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs) in interacting with tumor susceptibility gene 101 (Tsg101) protein or NEDD4 ubiquitin ligase. The association of Hrs with these proteins is believed to be a crucial step for vesicular budding into late endosomes[315,343,347]. HIV gag late domain comprises the conserved PTAP motif to recruit the UEV domain of Tsg101[24,60,313,343] and the Ebola protein VP40 additionally contains a PPxY motif recognized by the third WW domain of NEDD4[314-316].
A search for GYF recognition motifs in viral proteomes lead to several candidates for direct encounter. One of them is the protein Vpx found in HIV-2 and different SIV strains[317,318]. The protein is important for the viral life cycle[348], playing a crucial role in the nuclear import of the preintegration complex[349]. Especially the C-terminal PRS of Vpx was associated with the nuclear import machinery[350] and is the proposed binding site for α-actinin 1[351]. The latter protein is implicated in cytoskeletal reorganization[352], and it could be important for the nuclear import of Vpx and the associated complex. The PRS is also recognized by CD2BP2 *in vitro* (Fig. 8.6), but the binding site PPGL is a typical PERQ2-GYF consensus motif, making the PERQ proteins the likely interaction partners. Future work will have to confirm the predicted encounter of Vpx by GYF domain containing proteins *in vivo*. Furthermore, potential cooperativity or competition with binding of α-actinin 1 and a possible contribution of GYF domains to the HIV life cycle will have to be addressed.

11.3 Binding Models for the GYF Domain Subfamilies

The identified recognition motifs not only defined putative interaction partners and consequently novel functions for the GYF domain containing proteins, but also led to the proposal of binding models to explain the observed specificities on an atomic scale. The NMR structure of CD2BP2-GYF in complex with the CD2 peptide revealed a binding mechanism, highly analogous to those employed by the evolutionary distinct WW and SH3 domains (see Chapter 10). The characteristic

residues conserved in all GYF domains (Trp 4, Tyr 6, Gly 18, Pro 19, Phe 20, Met 25, Trp 28, Gly 32, Tyr 33, and Phe 34) adopt similar conformations in the highly divergent proteins CD2BP2 and Q9FT92 (Fig. 1.3), suggesting comparable hydrophobic surface patches as a communality of GYF domains. The aromatic cradle of CD2BP2-GYF intimately contacts the PPG tripeptide of the CD2 ligand in the complex structure. Presence of the PPG signature in all GYF domain ligands, identified herein, abet a common ligand-domain interface.

11.3.1 Binding Modes for CD2BP2-type GYF Domains

The mode of interaction of CD2BP2-GYF with the CD2 class of ligands was exhibited by the NMR complex structure. Complementary charges in the domain and ligand are flanking the common binding epitope. To account for the charge independence of the PPGW recognition motif, a different binding mechanism was postulated. According to this model, the PPG tripeptide adopts a similar conformation as in CD2 ligands, while the side-chain of the ligand tryptophan residue faces the domain and undergoes hydrophobic interactions with Trp 8 (Fig. 8.1 and Fig. 10.3). Such hydrophobic interactions are conceivable, regarding the strong selection for tryptophan in the PPGW recognition motif of CD2BP2-GYF (Table 8.1). Strict conservation of the aromatic surface residues (Fig. 1.3) suggests other CD2BP2-like GYF domains to recognize PPGW ligands as well.

The proposed binding models were challenged by the substitution of tryptophan for glycine (G8W) in the CD2 ligand. This exchange had a minimal effect on the binding affinity to CD2BP2-GYF (Fig. 6.2 and Fig. 7.4), although a clash of the two tryptophan side-chains in the ligand and the domain was anticipated to abolish binding. Molecular dynamics simulations predicted a register shift, a rotation and translation of the G8W mutant peptide with respect to the peptide in the known structure. This 'screw-like' transition puts the ligand tryptophan (W8) into the favorable Φ position (in reference to the PPGΦ motif), while P7 falls into the region normally occupied by G8. P5 and P6, instead of P6 and P7, interact with the aromatic binding groove (Chapter 7). However, the model orients the tryptophan side-chain of the G8W mutant peptide towards the solvent, a conformation similar to that seen for the H9 side-chain of the CD2 peptide in the NMR structure. Taking into account the model for PPGW-ligand binding, it is conceivable that the G8W substitution induces a register shift and concurrently repositions the tryptophan side-chain to face the domain.

11.3.2 Binding Mode for SMY2-type GYF Domains

The binding mode of SMY2-type GYF domains is expected to be closely related to the predicted recognition mechanisms of CD2BP2-GYF for PPGW ligands. Exclusive selection of hydrophobic amino acids at the fourth position of the recognition pattern PPGΦ argues for a direct interaction of all motif residues with the domain—in analogy to the postulated PPGW binding mechanism of CD2BP2-GYF. The structure of the Q9FT92-GYF domain illustrates the possible reason for the differential specificities observed for SMY2-type GYF domains and CD2BP2-GYF. The subtype-characteristic substitution of aspartate for Trp 8 in the SMY2 subfamily increases solvent exposure of Tyr 6, Phe 34, and the mostly aliphatic residue 38 (CD2BP2-GYF numbering; Fig. 1.3), thereby extending the major hydrophobic groove of the domain (Fig. 9.7 and Fig. 10.3). The aliphatic and aromatic residues which are part of the typical binding motif for this GYF domain subfamily, could well be accommodated in this extended groove. Tryptophan, on the other hand, would probably be too bulky for an efficient interaction. In line with this hypothesis is the extended binding epitope of SMY2-GYF, observed by NMR mapping experiments (Fig. 9.7). Residues Leu 38, Gln 39, and Ile 40 in SMY2 (CD2BP2-GYF numbering) are part of or in close vicinity to the extended hydrophobic groove. These residues showed significant chemical shift perturbations upon addition of a ligand, indicating a direct interaction.

Similar to the charge dependent CD2 class of ligands for CD2BP2-GYF, residues flanking the hydrophobic binding site in other GYF domains might introduce additional specificities. These alternative motifs would remain largely elusive in phage-display experiments because of their weaker binding affinities. More generalized conclusions demand binding analysis of additional, possibly more divergent GYF domains.

In summary, two binding mechanisms are suggested for GYF domains: a charge dependent interaction, so far exclusively observed for CD2BP2-GYF with respective peptides termed CD2 class ligands, and interactions, proposed to be entirely hydrophobic/aromatic in nature. The latter are postulated to be universal for all GYF domains, with CD2BP2-GYF preferring the motif PPGW, whereas GYF domains of the SMY2 subfamily select for aliphatic or aromatic residues, excluding tryptophan. Both mechanisms are expected to occur without major rearrangements of the domains, similar to ligand encounter by other PRD[124,125].

11.4 Perspectives

11.4.1 Evolution of GYF Domains

In animals, the increase in complexity (measured here as size of the genome in mega base pairs) is orchestrated, in part, by an augmentation of PDZ, SH2, SH3, and WW domains[19] (Table 1.1). In contrast, the number of GYF domains and profilins has remained relatively constant. Comparison of *Arabidopsis thaliana* with *Drosophila melanogaster*, which share similar-sized genomes, clearly indicated PDZ, SH2, SH3, and, to a lesser extent, WW domains to be under-represented in the plant species. GYF domains and profilins, however, experienced a significant amplification in *Arabidopsis thaliana* and *Oriza sativa* (not shown) although a general absence of the CD2BP2 subfamily from plants has been observed (Chapter 1.4). The frequencies of GYF domains and profilins relative to the other adaptor domains (presented in Table 1.1), are 0.8 % and 0.3 % in *Drosophila melanogaster*, but 4.7 % and 3.6 % in *Arabidopsis thaliana*, respectively. The reason for the different frequencies is not known. Comprehensive evaluation of the general distribution of adaptor domains in plants, fungi, and animals of different complexity will have to be conducted to give these observations statistical significance. Such a study would reveal whether the trends seen for *Arabidopsis thaliana* are real and reflect those of other plants as well. It will be interesting to see how these results relate to our current understanding of plant and animal organization principles and their evolutionary history. Absence of the CD2BP2 subfamily in plants could be due to loss in the plant ancestor. Alternatively, the SMY2 subfamily could be ancestral and therefore shared by plants, fungi, and animals, while the CD2BP2 subfamily appeared early after separation of plants, but before the differentiation into fungi and animals.

Despite its amplification in plants, the number of GYF domains in all organisms is rather small. A reason for the low number, compared to other PRD, could be the strong contribution of conserved residues to both structural integrity and ligand binding capacity. The highly conserved aromatic residues are the major determinants for ligand binding and they are also responsible for the packing of the helix against the sheet. Consequently, these residues probably experienced strong evolutionary constrains with most mutations having detrimental effects on the fold stability. The CD2BP2 subfamily characteristic features (Trp in position 8 and an extended loop between strands β_1 and β_2) are not crucial for the domain fold (Fig. 1.4 and Fig. 1.5) and confer only small changes of the ligand spectrum (shift from PPGΦ recognition by SMY-type GYF domains to the PPGW and CD2 ligand classes of CD2BP2-GYF). This is in line with a limitation of the GYF family of adaptor domains in exploring new binding specificities.

In contrast, in SH3 and WW domains, the residues contributing to structural integrity are clearly distinct from those involved in ligand recognition. In the first instance, the binding sites are mainly formed by the RT- and N-Src loop and the 3_{10} helix. Heterogeneity in the loops defines different recognition classes of SH3 domains[18]. Residues important for the domain fold are localized on the β-strands[124]. The three stranded β-sheet of WW domains is robust and amino acid exchanges on the ligand binding site are not anticipated to corrupt the packing of the hydrophobic core on the opposite side of the sheet[353,354]. The separation of positions, preserving the integrity of the domain from those forming the ligand binding site could explain the utility of SH3 and WW domains during evolution. New functions could easily be adopted as seen by the different ligand classes of these two PRD families (see Table 1.2).

Strikingly, conservation of residues in GYF domains is limited to the N-terminal half and there is high sequential flexibility in the C-terminal portion of the domains. This region defines the surface, opposing the PRS recognition pocket and comprises the binding site for U5-15K in CD2BP2-GYF. It is tempting to speculate that the organization of GYF domains on one hand restricts the variability of the N-terminal half of the domain and consequently the evolution of new PRS recognition motifs due to structural requirements. On the other hand, however, the C-terminal half of the domain may be evolutionary unconstrained and ultimately could have evolved to fulfill diverse functions. In this respect, the low number of GYF domains in the genome would originate from the obligate connection of these newly evolved functions to the conserved binding properties of the N-terminus of the domain.

A definite conclusions about the relationship between domain stability and binding competence for the highly conserved residues and their influence on domain evolution awaits structural and functional investigations of less conserved GYF domains and mutational analysis.

11.4.2 Regulation of GYF Domain Functions

Many different mechanisms evolved to regulate ligand binding of adaptor domains for a precise tempo-spatial control of protein-protein interaction networks. Regulation strategies include changes in the effective concentrations of interacting proteins and modulation of their binding affinities (although the latter could well be interpreted as a change in the concentration of binding competent species). Control over the effective concentrations could be established by co-compartmentalization, cooperative assembly in multiprotein complexes[3,38] (Chapter 1.1.1), competitive interactions with other (intramolecular) ligands and regulation of protein expression or stability. The tyrosine kinase Fyn is paradigmatic for some of these mechanisms. It colocalizes with diverse interaction partners, such as CD3ζ and CD2, to the plasma membrane, thereby

increasing the relative local concentrations of adaptor domains and ligands. Concurrently, intramolecular interactions limit the amount of free Fyn-SH2 and -SH3 domains in a competitive manner. Binding of its SH2 domain to phosphotyrosines directs Fyn into larger molecular complexes and promotes SH3–ligand encounter[75,355].

Modulation of the interaction affinities is frequently accomplished by posttranslational modifications of the ligand. For WW domains, both supporting and diminishing effects on binding have been reported upon ligand phosphorylation[25,236]. Similarly, arginine methylation appears to affect the association of PRS with SH3 domains[356]. Assuming an important contribution of the dynamic interactions, established by GYF domains, to diverse aspects of cellular life (see Chapter 11.2 above), corresponding regulation mechanisms for their fine-tuning are plausible. Several hints support this idea.

The expression level of CD2BP2 could be regulated by a silencing mechanism. CD2BP2 is expressed in different tissues[135,L]. A pseudogene in the mouse and human genomes has been predicted by the program Genescan[357,R], consisting of an isolated GYF domain with high similarity to CD2BP2-GYF. However, no data indicate the expression of the proposed gene products. The human pseudogene localizes to the non-coding strand in the twelfth intron of the human staphylococcal nuclease domain containing protein 1 (Swiss-Prot accession number Q7KZF4). In the introduction to this work, it was pointed out that introns could comprise non-coding RNAs such as snoRNAs[158] or miRNAs[159]. The pseudogene in the excised intron twelve is complementary to the GYF domain encoding region in the CD2BP2 mRNA and, similar to small interfering RNAs, could reduce the expression of CD2BP2.

The integration into multiprotein complexes possibly limits the number of GYF domain interaction partners *in vivo*. CD2BP2 is a component of the U5 snRNP and binds to U5-15K and U5-102K[217] (Chapter 11.2.2). Its GYF domain associates with both PRS and U5-15K, and the location of the two interfaces on the domain surface allows for simultaneous encounter[120] (Kofler, unpublished results, Fig. 1.4, and Fig. 11.1a). Similarly, U5-15K can possibly bind at the same time to the carboxyl-terminal domain (CTD) of PQBP-1 (polyglutamine tract-binding protein 1)[358,359] and to CD2BP2-GYF, as revealed by the U5-15K–CD2BP2-GYF complex structure[120] and alanine substitution analysis of the U5-15K interaction with PQBP-1[358]. A similar scenario was found for the protein SAP. It consists almost exclusively of an SH2 domain with unusual binding site for the SH3 domain of Fyn adjacent to its canonical phosphotyrosine interaction pocket[360]. SAP serves to bridge Fyn to the receptor SLAM by concomitant recognition of Fyn-SH3 and phosphotyrosines in SLAM.

[R] Genscan data implemented in www.ensembl.org

CD2BP2-GYF could serve to bridge PRS and U5-15K, while the latter could link CD2BP2 to PQBP-1 and its WW domain. The presence of GYF and WW domains in such a complex could restrict the binding spectrum of CD2BP2-GYF. Since CD2BP2-GYF as well as the WW domain of PQBP-1 bind NPWBP[361] (Chapter 9) and since NPWBP comprises several PRS, recognition of NPWBP by both PRD could stabilize the weak CD2BP2-GYF–NPWBP association within a tetrameric complex (CD2BP2-GYF and PQBP-1 linked via NPWBP and U5-15K; Fig. 11.1b). However, formation of this complex remains elusive and PQBP-1 has not yet been described as U5 snRNP component. Nevertheless, stable integration of CD2BP2 into U5 snRNPs plausibly favors the encounter of PRS associated with the U5 snRNP, over those found in the other snRNPs.

For the PERQ proteins, association with PQBP-1 could have a similar stabilizing effect on ligand binding. Here, PERQs and PQBP-1 could directly interact with each other, rather than binding indirectly via U5-15K. The central polar amino acid-rich domain of PQBP-1 was found to bind polyglutamine tracts[362,363] and both PERQ proteins are rich in glutamine. In particular, PERQ2

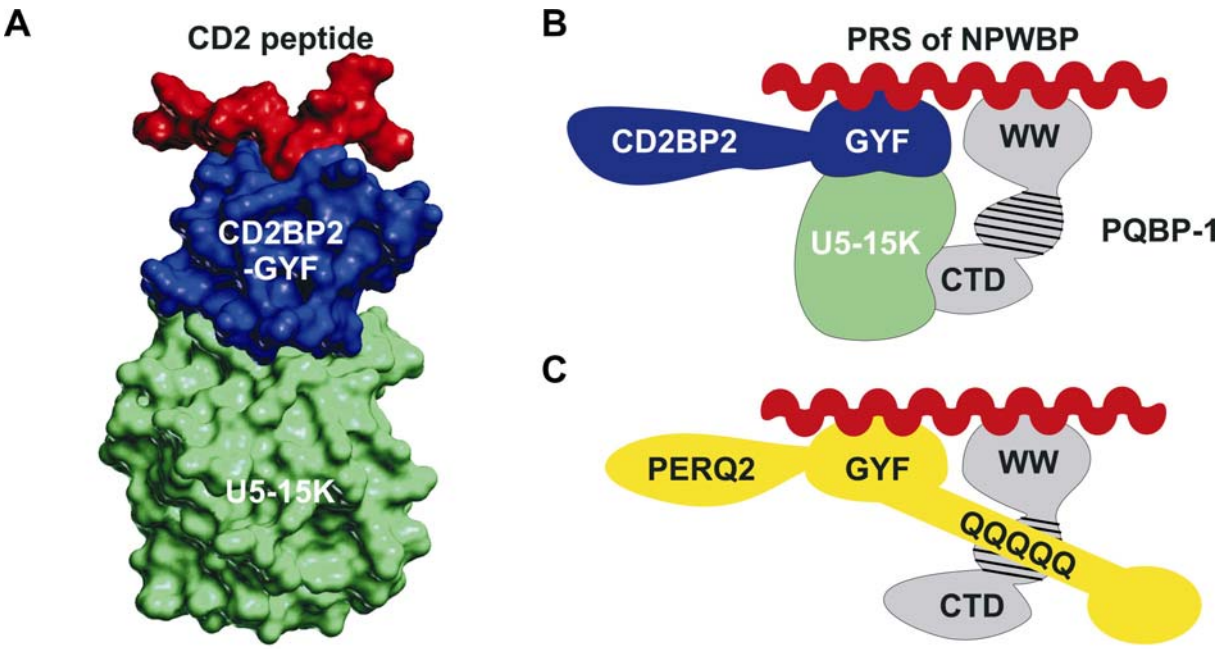

Fig. 11.1: Models of CD2BP2 and PERQ2 containing complexes
(A) The structures of the binary complexes of CD2BP2-GYF and the peptide SHRPPPPGHRV (1L2Z) or the protein U5-15K (1SYX) were superimposed. U5-15K, GYF domain, and proline-rich peptide are depicted as surface model in green, blue, and red, respectively. (B) Model of the proposed tetrameric complex of CD2BP2 (blue), the PRS of NPWBP (red), U5-15K (green), and PQBP-1 (grey). The WW domain of PQBP-1 binds to the PRS in NPWBP, its carboxyl-terminal domain (CTD) mediates association with U5-15K. (C) Model of the proposed trimeric complex of PERQ2 (yellow), NPWBP, and PQBP-1. The PERQ2–PQBP-1 interaction is suggested to occur via the central polar amino acid-rich domain (dashed region in PQBP-1) and the short polyglutamine tracts in PERQ2 (QQQQQ).

comprises several regions of up to nine consecutive Gln residues, giving rise to longer, disrupted polyglutamine tracts. Weak association with PQBP-1 could support binding of PERQ2-GYF to PRS in NPWBP and result in a stable trimeric complex (Fig. 11.1c).

Competitive inhibition is another possible means to regulate GYF domain binding. Given the promiscuity of PRD, several potential competitive inhibitors for GYF domains with overlapping binding specificities exist such as FBP21 and PQBP-1[50,207,361]. On the one hand, competitive inhibition is expected only under conditions of limiting binding sites. The multitude of recognition motifs in many of the GYF interaction partners renders such a regulation mechanism unlikely. On the other hand, the weak interactions could be sensitive to only subtle changes of available binding sites which could translate into a significant shift of local concentrations. Therefore, a quantitative description of such a system is required to predict its behavior correctly.

Competitive inhibition by intramolecular ligands, however, is an appreciated mechanism to regulate binding of PRD to ligands. For GYN4 such an intramolecular association has been suggested to regulate the binding competence of its GYF domain (Chapter 9). The assumption is founded on two observations: firstly, the binding to ligands in the presence of the intramolecular recognition site is diminished; secondly, the GYN4-GYF domain and its intramolecular PRS display an organization, analogous to the single-chain constructs of the CD2BP2-GYF–CD2 ligand, with demonstrated intramolecular encounter.

Often the intramolecular blockade is overcome by posttranslational modifications. In the Fyn tyrosine kinase, intramolecular inhibition is alleviated by dephosphorylation, disrupting the association of the SH2 domain with an internal phosphotyrosine. Similarly, in GYN4, serine phosphorylation could reduce the affinity for the intramolecular ligand. NetPhos[364,M], a web based program, predicts a serine in the linker region (S3 according to CD2 ligand numbering) to be a potential phosphorylation site in GYN4. Although negative charges were not found to inhibit binding, if introduced into the internal interaction site, represented as membrane bound peptide (Fig. 9.2), the situation could be different for an intramolecular encounter. In the context of a single polypeptide, residues linking the C-terminus of the domain to the ligand might adopt a restricted conformation. Assuming the structure and the domain borders of GYN4-GYF to be similar to those of CD2BP2- or Q9FT92-GYF, the linker would be 5–6 residues shorter than the one used in the single-chain construct of CD2BP2-GYF–CD2 ligand, supporting limited conformational freedom in the cis-bound state. Introduction of the bulky, negatively charged phosphate group could possibly interfere with a linker conformation required for binding and it would be in close proximity to residue Asp 24 (CD2BP2-GYF numbering).

The same prediction program also identified Tyr 61 of CD2BP2-GYF as a potential phosphorylation site. This residue localizes to the U5-15K binding patch, on the side opposite to the PRS recognition pocket[120]. Intriguingly, phosphorylation of CD2BP2 coincides with its release from the U5 snRNP during formation of the U4/U6•U5 tri-snRNP[217,323] and dissociation of CD2BP2 was suggested to initiate tri-snRNP assembly[217]. The proposed phosphorylation of Tyr 61 could well disrupt the GYF-U5-15K interaction and trigger tri-snRNP formation. This hypothesis draws a detailed picture of the molecular events that could lead to the dissociation of CD2BP2 during assembly of the spliceosome and explain the implication of phosphorylation in this process. However, phosphorylation of Tyr 61 and its proposed effect on U5-15K binding have to be proven experimentally to substantiate the suggested hypothesis.

In summary, many aspects of GYF domain functions remain to be clarified, including the regulation of their binding functions, their evolution, and their precise role within the cell. The elucidation of these issues requires ongoing investigations. The recently identified adaptor family of GYF domains, in each case and in all respects, remains an interesting field for future studies.