the FEBS Journal

REVIEW ARTICLE

The GYF domain

Michael M. Kofler and Christian Freund

Protein Engineering Group, Free University and FMP Berlin, Germany

Keywords
adaptor domains; function; GYF domains; interaction partners; proline-rich sequences; proline-rich sequence-recognition domains

Correspondence
M. M. Kofler, C. Freund, Protein Engineering Group, Free University and FMP Berlin, Robert-Rössle-Str. 10, 13125 Berlin, Germany
Fax: +49 30 94793 189
Tel: +49 30 94793 181
E-mail: freund@fmp-berlin.de, kofler@fmp-berlin.de

(Received 6 October 2005, accepted 24 November 2005)

doi:10.1111/j.1742-4658.2005.05078.x

GYF domains are small, versatile adaptor domains that recognize proline-rich sequences (PRS). They are present in most eukaryotic species sequenced so far, but in contrast to other PRS-recognition domains (PRD), GYF domains have not experienced the same amplification in metazoa during evolution. Mutational and structural analysis has shown the conserved signature W-X-Y-X_{6-11}-GPF-X_4-M-X_2-W-X_3-GYF to be the site of interaction with proline-rich peptides. In contrast, composition and length of the C-terminal half of GYF domains are not conserved. Similar to other PRD, GYF domains bind to many different PRS that converge on a minimal consensus sequence. All GYF domains analyzed so far selected for the core motif PPG, whereas amino-acid preferences adjacent to this motif vary. As a result of this analysis, two subfamilies have been identified: CD2BP2-type and SMY2-type GYF domains. The latter subfamily comprises most GYF domains and is characterized by a shorter β_1–β_2 loop and an aspartate instead of the tryptophan found at position 8 in CD2BP2-type GYF domains. Recent analysis of binding specificities for GYF domains allowed identification of novel interaction partners. Thereby proteomics has contributed to a functional understanding of GYF domain-containing proteins and sets the stage for a more systematic investigation of their functions *in vivo*.

Structural elucidation of a GYF–ligand complex

Originally identified in the CD2BP2 protein as the domain responsible for binding to the PPPGHR amino-acid repeats of the T cell adhesion molecule CD2, databank searches soon revealed a set of proteins that contain GYF domains [1]. Conserved amino acids of the fold, including the GYF tripeptide, are contained within the signature W-X-Y-X_{6-11}-GPF-X_4-M-X_2-W-X_3-GYF as exemplified by the alignment of a set of GYF domains from various species (Fig. 1). BLAST searches [2] revealed the presence of GYF domains in most, if not all, eukaryotic genomes, although the number of GYF domain-containing proteins is small in a given species. NMR chemical shift mapping experiments indicated this conserved signature to comprise residues important for binding of the CD2 ligand. Subsequent determination of the structure of both the isolated GYF domain and the domain in complex with the CD2-derived peptide SHRPPPPGHRV allowed a molecular understanding of the observed binding behaviour [3,4] (Fig. 2). The GYF domain of CD2BP2 forms a compact fold with a single α-helix packed against a small sheet. The helix is tilted away from the sheet, and Fig. 2A shows that this inclination allows several bulky aromatic amino-acid side chains to be placed between the helix and the sheet. Thereby an array of stacked aromatic amino acids is created which forms a hydrophobic surface depression that accommodates the PPG motif of the CD2 ligand (Fig. 2B). The four proline residues of the ligand adopt a PPII

Abbreviations
eIF, eukaryotic initiation factor; GYN4, GYF domain-containing protein binding to Not4; PRS, proline-rich sequence(s); PRD, proline-rich sequence recognition domain(s); snRNP, small nuclear ribonucelotide particle; SYH1, SMY2 homolog 1.

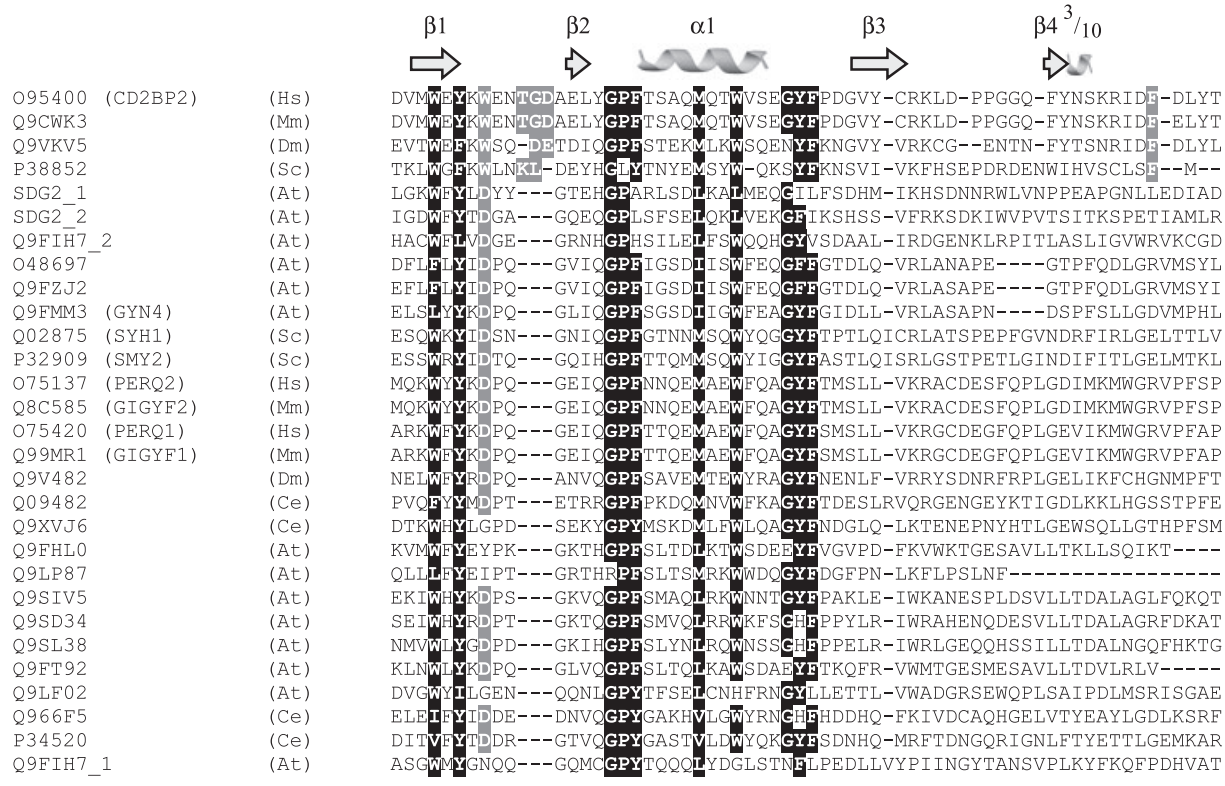

Fig. 1. Sequence alignment of GYF domains from various species. GYF domains of the CD2BP2 subfamily (first four sequences) and the SMY2 subfamily (all other sequences) were identified by BLAST search. Swiss-Prot, TrEMBL and EnsEMBL entry/protein names and the origins of the proteins are indicated. At, Ce, Dm, Hs, Mm and Sc stand for the following species: *A. thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *H. sapiens*, *Mus muculus*, and *S. cerevisiae*, respectively. Residues that are characteristic of GYF domains and substitutions for amino acids with similar physicochemical properties are depicted as white bold letters on black background. White bold letters on grey background represent residues conserved in either the CD2BP2 or SMY2 subfamily.

helical conformation, an extended left-handed helix with a periodicity of three. The glycine in the PPG motif induces a kink in the peptide backbone, which possibly helps to prevent a steric clash with the amino acids W8, E15 and Y17 of the GYF domain. Flanking the hydrophobic interaction sites, there are a number of complementary charges between the domain (E9, E15, E31, D36) and the peptide (R3, R10), which add to the specificity of the interaction. Results of peptide substitution analyses were largely in agreement with the structural data [5]. However, a mutant bearing a $G \rightarrow W$ substitution (leading to the peptide SHRPPPPWHRV) bound with similar affinity to that of the wild-type peptide to the CDBP2-GYF domain. Assuming the same binding mode, the mutant peptide would face the problem of steric hindrance introduced by the bulky tryptophan side chain. Yet molecular-dynamics simulations and peptide-binding studies suggested that the mutant peptide is able to exert a register shift, placing the last proline at the position of the glycine in the wild-type peptide [6].

Comparison of the GYF domain with other proline-rich ligand-binding domains

Several fold families that recognize proline-rich sequences (PRS) have emerged during the course of evolution. SH3 [7,8], WW [9,10], EVH1 [11], GYF [1,3], and UEV domains [12] as well as profilin [13] share the use of stacked aromatic amino acids to recognize PxxP or xPPx proline motifs contained in different ligands [14,15]. One or two hydrophobic pockets thereby contact at least two of the prolines in the canonical ligands of the PRS-recognition domain (PRD), as exemplified for SH3, GYF and WW domains in Fig. 2. The ligand forms a PPII helical conformation in all cases, and this conserved binding mode helps to explain the observation that overlap of sequence space *in vitro* is commonly found for PRD-mediated interactions. Loops in spatial proximity to the proline-binding pockets may contribute to selectivity by interacting with ligand residues neighbouring the proline-rich core. The specificity regions within the

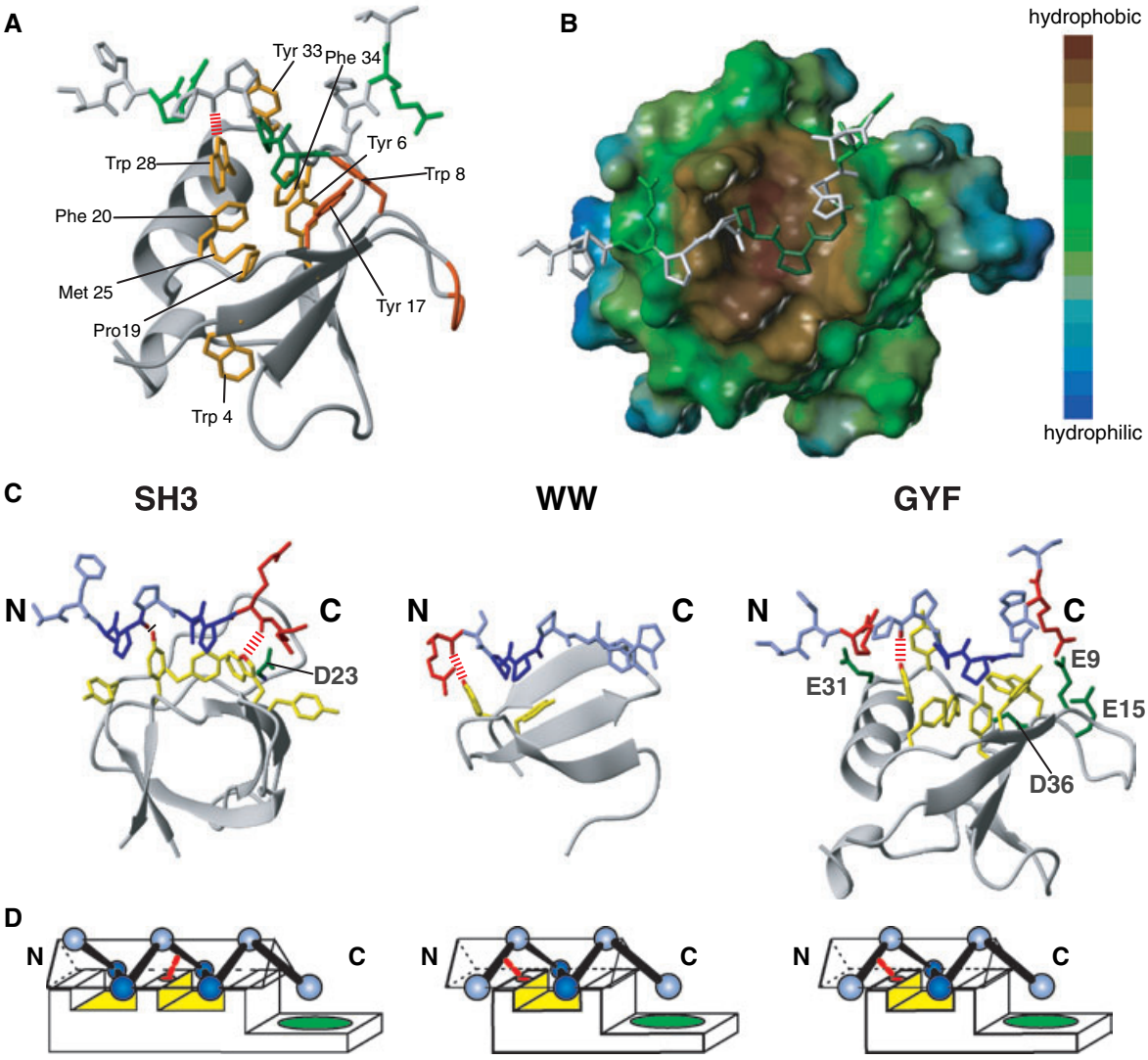

Fig. 2. Structure of GYF, SH3 and WW domains in complex with proline-rich ligands. (A) Conserved residues and topology of the GYF domain of CD2BP2. Side chains of conserved residues are depicted in orange, the CD2BP2 GYF-specific residue W8 and the extended loop between strand β_1 and β_2 as well as residue Y17 are highlighted in red–orange. Residue Y17 is included because it is conserved in mammalian CD2BP2-like GYF domains and also forms part of the ligand binding site. The two prolines of the ligand that directly interact with the conserved aromatic residues of the domains are shown in dark green, and flanking arginines are visualized in light green. (B) Lipophilic surface potential of CD2BP2-GYF as calculated with the program Sybyl (Tripos Inc., St Louis, MO, USA). The ligand-binding surface of CD2BP2-GYF in complex with the CD2 peptide is shown. The surface is colour coded according to the lipophilic potential, ranging from brown for hydrophobic to blue for hydrophilic surface areas. The ligand is coloured as in (A). (C) Structures and binding mechanisms of PRD. The structures of SH3 and GYF domains are shown in complex with ligands that use charge complementarity for selective binding. For WW domains, no complex structure of a domain recognizing positively charged ligands is yet available. The structure of a WW domain in complex with a PPxY ligand is shown instead. Aromatic residues that form part of the hydrophobic binding site (aromatic cradle) are depicted in yellow. Proline residues of the ligands that directly interact with the hydrophobic pockets are coloured dark blue, positively charged arginines are depicted in red, and complementarily charged residues in the domains are shown in green. The hydrogen bond between carbonyl oxygens of the ligands and the conserved tryptophan of the domains are represented as dashed red lines. Protein Data Base (PDB) accession codes for presented structures are: 1prm (SH3), 1eg4 (WW) and 1l2z (GYF). (D) Cartoons of the PRD–ligand interactions shown in (C) [14]. The hydrophobic grooves accommodating conserved proline residues of the ligand and the specificity patches are coloured yellow and green, respectively, similar to the corresponding residues shown in (C). The positions of ligand residues are depicted as blue spheres. The binding mechanism of SH3 domains is delineated in the first model. SH3 domains comprise a specificity patch and two shallow hydrophobic grooves which recognize two xP dipeptides organised as xP-x-xP core. The second and third models represent the binding mechanism of WW and GYF domains, respectively, with a single xP binding groove binding the xPPx core, in addition to a specificity patch.

domain families are less well conserved than the residues forming the proline-recognition pockets [16,17]. Therefore it is not surprising that the actual amino-acid composition of the flexible loops largely defines the ligand specificity of a given PRD. Two of the CD2BP2-GYF interaction partners, namely CD2 and SmB/B′, are also recognized by SH3 and WW domain-containing proteins, respectively. CD2 contains the motif PPPGHR, which binds with similar affinity to CD2BP2-GYF and Fyn-SH3 [4]. The SmB/B′ motif PPPGMR is recognized by the WW domain of FBP21 [18] in addition to CD2BP2-GYF [5], and both FBP21-WW and Fyn-SH3 interact with the PRS in Sam68 [19]. These observations highlight the overlapping ligand-binding properties of CD2BP2-GYF, FBP21-WW and Fyn-SH3. Convergence on similar binding mechanisms is particularly illustrated by WW and GYF domains. Both folds recognize two consecutive proline residues in the ligand (Fig. 2C). The PPGΦ core motif is the recognition code for most GYF domains and it is comprised in ligands of the large class of PPxY-recognizing WW domains. Conversely, two domains belonging to the same fold family may recognize quite different sets of ligands. This is exemplified by the inability of the SH3 domain of Lck to bind to the PPPGHR sequences of CD2, in contrast to the highly homologous Fyn-SH3 domain. Homology modelling suggested that a single T → H mutation within the flexible RT loop of the SH3 domains accounts for the observed difference in binding behaviour between Fyn-SH3 and Lck SH3 [4]. The histidine side chain forms an intramolecular hydrogen bond with an RT-loop aspartate. This aspartate is critical for the recognition of the CD2 peptide by Fyn-SH3, as it provides a donor group for an intermolecular salt bridge. More generally, charge complementation of arginine/lysine-containing PRS by aspartate and glutamate side chains of the PRD seems to be important for several ligand subclasses of SH3, WW and GYF domains [4,20–22] (Fig. 2).

Interaction of CD2BP2-GYF with CD2

The GYF domain-binding sites in the cytoplasmic tail of CD2 have been shown to be important for intracellular signalling events mediated by CD2 ectodomain cross-linking [23]. However, the CD2–CD2BP2 interaction is of low affinity, when a single PRS from CD2 is analysed. Indeed, NMR and fluorescence titration experiments revealed the K_d of the GYF domain–SHRPPPPGHRV peptide interaction to be only ≈ 190 μM [3]. The presence of a second, neighbouring PPPGHR motif increases the overall affinity for the GYF–CD2 interaction severalfold and confirms the importance of local concentrations of binding sites for GYF domain-mediated interactions [4]. Whereas the isolated GYF domain is monomeric in solution, the full-length CD2BP2 protein is predicted to contain coil regions that may contribute to avidity enhancement by formation of protein dimers or oligomers. Interestingly, the CD2 interaction partner Fyn mainly localizes to the plasma membrane, whereas CD2BP2 is present in the soluble fraction of the cell (especially the nucleus, see below). A model has been proposed in which the translocation of CD2 into the raft fraction of the membrane allows the interaction with Fyn to take place, thereby enabling the phosphorylation of critical downstream targets [4]. The predominant nuclear localization of CD2BP2 in resting Jurkat and HeLa cells suggests that the CD2BP2–CD2 interaction may only be observed under certain conditions within primary cells, as has been observed for other nuclear proteins in the context of adhesion [24].

Interaction of CD2BP2-GYF with spliceosomal proteins

The striking similarity between the PPPGHR motif present in CD2 and the PPPGMR motifs contained in the C-terminal tail of the core-splicing protein SmB/B′ immediately suggested its interaction with CD2BP2-GYF [5]. SmB/B′ is part of a ring of seven Sm proteins that constitutes the core in most of the small nuclear ribonucelotide particles (snRNPs) [25]. Within this heptameric structure, the N-terminal Sm domain of SmB/B′ is suggested to contact the Sm site in snRNA [25], whereas the C-terminus contains five PPPGMR motifs and is probably accessible to other proteins. Peptide substitution analysis confirmed the structure-based assumption that the H → M mutation would not affect the binding affinity [5]. As expected, peptides corresponding to the PRS of SmB/B′ bind to the same epitope on the GYF domains as CD2-derived peptides. Pull-down experiments confirmed the direct interaction between the GYF domain and full-length SmB/B′ protein and revealed the PRS in SmB/B′ to be accessible for binding under these conditions. Furthermore, confocal laser scanning microscopy showed that both proteins localize to the nucleus of HeLa and Jurkat cells, indicating a likely *in vivo* interaction between the two proteins [5].

A recent finding identified CD2BP2 as a component of the U5 snRNP [26]. After the binding of U1 and U2 snRNPs to the 5′ splice site and branch site of the pre-mRNA, respectively, the U5 snRNP, as part of the U4/U6.U5 tri-snRNP completes the formation of

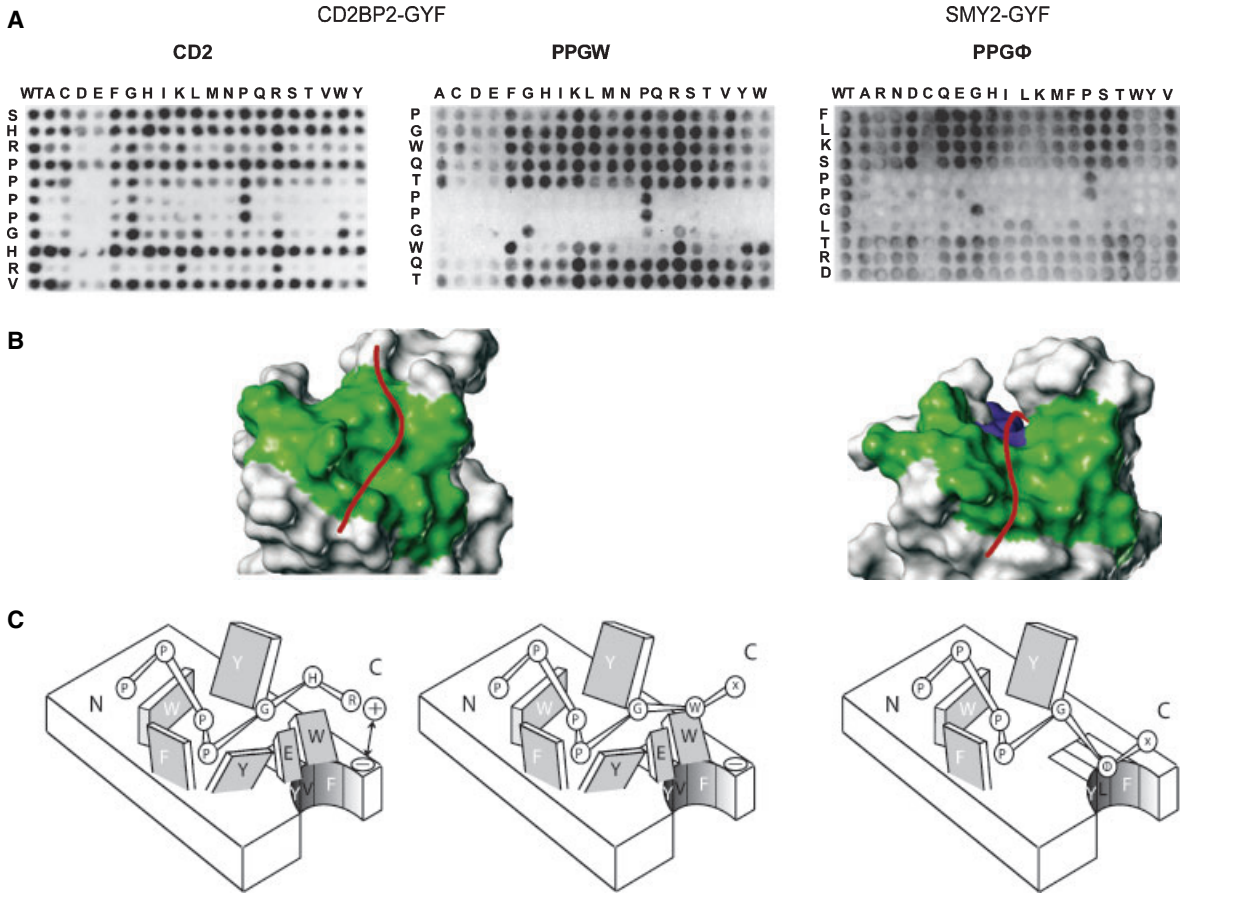

Fig. 3. Ligand classes and binding models for CD2BP2 and SMY2 GYF domains. (A) Substitution analysis of the two CD2BP2-GYF ligand classes, namely the CD2 and the PPGW class, and of the single PPGΦ ligand class for SMY2 GYF. All possible single-substitution analogs of the peptide SHRPPPPGHRV (from CD2), PGWQTPPGWQT (from MAGD1) and the phage display-derived ligand FLKPPGLTRD were synthesized on a membrane. The single-letter code above each column indicates the amino acid that replaces the corresponding wild-type residue. The rows define the position of the substitution within the peptides. WT columns comprise the unsubstituted peptides [5,20,27]. (B) Binding surface of CD2BP2-GYF (left) and the suggested binding epitope for GYF domains of the SMY2 subfamily, deduced from NMR titration experiments with SMY2-GYF and plotted on to the structure of the SMY2-like GYF domain of the *A. thaliana* protein Q9FT92 (right, PDB code 1wh2) [27]. The surface areas in green correspond to residues comprising the common binding epitope of CD2BP2- and SMY2-GYF. The blue patch represents the extension of the binding epitope in SMY2-GYF, namely residues L289, Q290 and I291, which show significant chemical-shift changes upon ligand binding. Only small chemical-shift changes were observed for the corresponding residues in CD2BP2-GYF (residues V38, Y39 and C40). The known orientation of a CD2 class ligand bound to CD2BP2-GYF and the suggested orientation of a ligand (PPGΦ) for the SMY2-GYF domain are depicted as red lines. (C) Proposed models for the binding mode of the CD2BP2 (left and centre) and SMY2 (right) subfamilies of GYF domains. Three models are shown; the first two account for the different ligand specificities of the CD2BP2-GYF domain, which recognizes ligands comprising the motif PPGx(R/K) (CD2 class ligands, left) or PPGW class ligands (centre). The third model (right) depicts the suggested binding mode of SMY2-type GYF domains. Side chains of the GYF domains that create the 'walls' of the pocket are shown as rhombic structures. Side chains that form part of the floor of the binding pocket are largely buried by W8 in CD2BP2-GYF, but are more accessible in SMY2-like GYF domains. These side chains (Y6, F34 and V38 in CD2BP2-GYF) are indicated as rectangular structures, embedded in the core of the domains. Conserved residues Y6, F20, W28 Y33 and F34 (CD2 numbering) are marked in white, the CD2BP2 GYF-specific residues W8 and E15 as well as residue Y17 and the most solvent exposed residue of the extended SMY2 epitope (L289 in SMY2 which corresponds to V38 in CD2BP2) are depicted as black letters on grey background. Positions of ligand residues are shown as spheres connected by sticks. The charge complementarity between the arginine of the CD2 class peptides and the negatively charged surface of the CD2BP2-GYF domain is indicated by an arrow (left panel). In the case of the PPGW class, the W8 of the CD2BP2-GYF domain and the tryptophan of the peptide are suggested to make hydrophobic interactions. Substitution of residues W8, E15 and Y17 in SMY2 partly exposes Y33, F34 (CD2BP2 numbering) and L289 (the V38 analogue in SMY2). Thereby, the contiguous hydrophobic patch on the domain surface is extended and possibly interacts with the hydrophobic side chain of the residue in the Φ position of the ligand.

spliceosomal complex B. This complex has to undergo several rearrangements to give rise to the catalytically active complex B*. CD2BP2 was shown to dissociate from the U5 snRNP before formation of the U4/U6.U5 tri-snRNP, suggesting a role in the assembly of this subcomplex. Within the U5 snRNP, CD2BP2 interacts with the essential splicing protein U5-15K. The interaction is mediated by the GYF domain of CD2BP2, but is independent of PRS.

GYF domain-binding specificities

Peptide substitution analysis was used to evaluate the contribution of individual amino acids to the interaction between the GYF domain of CD2BP2 and the CD2 peptide SHRPPPPGHRV [5]. This analysis revealed the PPGx+ motif to be required for efficient binding, in agreement with the NMR structural data that showed several van der Waals contacts between the PPG core and the aromatic side chains of the GYF domain. Positively charged residues flanking the PPG core also contribute to binding, and negatively charged amino acids at any position in the peptide are disfavoured. Phage display, using an X_9 (X = any amino acid) peptide library fused to the major capside protein g8p, was used for a comprehensive search of new recognition motifs [20,27]. The predominant motif contained in CD2BP2-GYF domain-binding peptides was PPGW, but other, mostly aromatic, residues were also tolerated. The CD2-derived motif PPGx+ was found rarely, showing that charge dependence is only observed for a subset of CD2BP2-GYF domain ligands. The search for interaction sites of GYF domains from *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Homo sapiens* identified the signature PPGΦ as the common sequence recognition motif for SMY2-type GYF domains (Φ = hydrophobic amino acid) [27]. The strong similarity to the recognition motif PPGW of CD2BP2-GYF indicated an evolutionarily conserved binding mode for all GYF domains. The identity of the hydrophobic amino acid following the PPG motif was found to be domain specific. PERQ2, SMY2 and SMY2 homolog 1 (SYH1) GYF domains preferentially bind to peptides with a hydrophobic amino acid displaying either an aliphatic side chain or a phenyl moiety, whereas tryptophan is not favoured. In contrast, CD2BP2-GYF has a strong preference for tryptophan in the last position of the motif. The GYF domain of GYN4 (GYF domain-containing protein binding to Not4) predominantly binds to the motif PPGF. For the SMY2-type GYF domains, hydrophobic interactions with the central four amino acids of the ligand seem to be solely responsible for peptide recognition. To corroborate these findings, peptide substitution analyses were performed with phage display-derived peptides containing the consensus sequences (Fig. 3A). The four-amino-acid code defined by the phage display consensus is quite a strict requirement for binding in each case, and only certain substitutions for the hydrophobic amino acids adjacent to the PPG motif are allowed, albeit with distinct specificities for each domain (Fig. 3).

Novel interaction partners for GYF domains

The defined binding propensities of the GYF domains allowed proteome wide searches of novel interaction partners. Relaxed signatures (PPG[AEFHILMSTWY] for CD2BP2 and PERQ2, PPG[ACFGILMVWY] for SMY2 and SYH1, and PPG[FILMVWY] for GYN4), deduced from the phage display results to include most binding sites, were identified in the proteomes of the respective organisms by searching the Swiss-Prot and TrEMBL databases [20,27]. As the presence of several neighbouring consensus motifs was shown to enhance the avidity of the GYF domain-mediated interactions [4], human and *A. thaliana* proteins with a minimum number of two such consensus motifs were regarded as more likely interaction partners. The yeast proteome is small enough to allow experimental evaluation of all sequences containing the relaxed binding motif. PRS obtained by this search were spotted on to a nitrocellulose membrane and tested for direct binding to glutathione S-transferase fusions of the different GYF domains. In principle, this strategy is similar to the WISE strategy as performed by Landgraf *et al.* [28], with the additional feature of motif repetition as an important hallmark of natural targets. Interacting peptides were classified according to signal intensities of their peptide spots, and the respective proteins containing several high-affinity binding sites were further investigated by yeast two-hybrid analysis and pull-down experiments. The functions of potential interaction partners allowed the GYF domain-containing proteins to be placed into biological context (summarized in Table 1). In addition, the domain–ligand interface for the new class of CD2BP2-GYF ligands was mapped on to the domain and revealed a very similar binding epitope to that for the CD2 peptide [20].

Several proteins containing high-affinity sites for the GYF domain of CD2BP2 were subjected to yeast two-hybrid analysis. This allowed the verification of peptide-mediated interactions within larger protein fragments under more physiological conditions. In addition to CD2 and SmB/B′, the protein PI31 showed GYF domain-dependent binding to CD2BP2. PI31 has

Table 1. Biological functions of GYF domains. The functions of GYF domain-containing proteins were deduced from the biological implications of identified interaction partners and associated complexes. The first three columns refer to the origin and recognition motif of different GYF domains. Interaction partners and their biological functions are shown in the fourth and fifth columns. References for the biological functions of the interaction partners and for the GYF domain-containing proteins themselves are listed in the last column. The interaction motif for LIN1 has not yet been determined (n.d.), and SMY2 was found to genetically interact with MYO D, but a direct encounter remains to be established. Overlapping interaction partners for the GYF domains of CD2BP2 and PERQ2 or SMY2 and SYH1 are shown in bold in the column 'Candidate'.

Organism	GYF domain	Binding motif	Candidate	Biological implications of candidate	References
H. sapiens	CD2BP2	PPGW PPGx+	CD2	T cell adhesion and signalling	[55,56]
			PI31	Regulation of immunoproteasome	[29,30]
			SmB	Splicing	[57]
			NpwBP	Splicing	[36]
	PERQ2	PPGΦ	hSF1 SWAN	Splicing	[58,59]
A. thaliana	GYN4	PPGF	At3g45630.1	Proposed homologe to human CNOT4; inhibition of transcription, ubiquitin ligase	[27,60,61]
S. cerevisiae	LIN1	n.d.	PRP8	Splicing	[62,63]
	SMY2	Genetic interaction	MYO D	Transport functions	[39]
		PPGΦ	**MSL5**	Splicing, nuclear export of mRNA	[38,64]
			EAP1	Regulation of translation initiation	[40,42,65–67]
	SYH1	Yeast two-hybrid	p33 and p39	Translation initiation	[49,68,69]

been shown to inhibit the 20S proteasome *in vitro* [29], and overexpression of PI31 abrogates major histo compatibility complex class I peptide presentation [30].

A further interesting candidate derived from the proteomic approach was the nuclear protein NpwBP (alternatively named SIPP1, WBP11 or SNP70). This protein was initially described as a nuclear protein with RNA-binding capability which associates with the protein Npw38 [31]. In a subsequent study that more extensively investigated its localization, the protein was found in the cytoplasm of dividing cells, and partial colocalization with perinuclear intermediate filaments in HEK293 cells could be observed [32]. Recently, NpwBP was shown to shuttle between cytoplasm and nucleus [33]. The PRS of NpwBP that bind to CD2BP2-GYF were also shown to interact with the WW domain of Npw38 [31]. This raises interesting questions about whether binding of NpwBP by CD2BP2 and Npw38 is competitive, independent or co-operative, and to what extent cross-talk between the three proteins contributes to spliceosome assembly. Npw38 and CD2BP2 also bind to U5-15K, an essential splicing protein associated with the U5 snRNP [26,34,35]. The N-terminal portion of CD2BP2, excluding the GYF domain, was found to interact with U5-102K, another component of the U5 snRNP [26]. It will be important to decipher the relative contribution of PRS recognition by GYF and WW domains with regard to the additional interactions of the multiprotein complexes that form during the dynamic assembly of the spliceosomal complex B. CD2BP2 has been suggested to be dispensable for the catalytic steps of splicing [26], NpwBP has recently been shown to stimulate splicing [33,36]. Therefore, CD2BP2 is more likely to be involved in the early events of spliceosomal complex formation, possibly linking them to splicing-related processes, such as transcription, mRNA transport, mRNA decay, and translation. Such a peripheral function in splicing could conceivably be mediated by regions of CD2BP2 that have not been characterized so far.

The human GYF domain-containing protein PERQ2 (and probably the close homologue PERQ1) was found to interact with two candidate proteins that also interact with the CD2BP2-GYF domain, namely SmB/B′ and NpwBP [27]. However, the SH3/WW domain anchor protein in the nucleus (SWAN) interacted only with the PERQ2-GYF, but not the CD2BP2-GYF, domain. SWAN (Swiss-Prot accession number Q9NTZ6) is a hnRNP-like adaptor protein with several RNA-recognition motifs. In mouse, the proteins homologous to PERQ1 and PERQ2, named GIGYF1 and GIGYF2, have been shown to interact with the Grb10 adaptor protein [37]. Consequently, a possible functional role

in Grb10-mediated insulin-receptor signalling has been suggested. However, PRS from human Grb10 do not obey the sequence requirements for GYF domain binding, and it remains to be investigated whether Grb10 association can be observed for the PERQ1 and PERQ2 proteins in human cells.

For the two yeast GYF domains SMY2 and SYH1, the MSL5 protein has been confirmed as an interaction partner. Pull-down experiments in the presence of purified peptides representing one of the proposed binding sites inhibited the recognition of MSL5 by the GYF domains of these proteins. This further strengthens the assumption that these sequences represent the actual target sites *in vivo*. As MSL5 is a branch point binding protein involved in the formation of the commitment complex 2 of splicing, a potential role for SMY2 and its paralogue SYH1 in splicing-associated processes is likely. However, as MSL5 is also important for the nuclear export of mRNA [38], a potential role for these two GYF domain-containing proteins in transport functions is also conceivable. The original finding that SMY2 is a suppressor of the myo2-66 mutation in type V myosin MYO D [39] is in line with such an argument. In addition to the known target sites within MSL5, several sequence motifs in the yeast protein EAP1 have been found to strongly interact with the SMY2 and SYH1 GYF domains. EAP1 has been identified as an inhibitor of cap-dependent translation, similar to p20, competing with eukaryotic initiation factor (eIF) 4G for binding to the cap-binding protein eIF4E [40]. The involvement of EAP1 in the regulation of translation initiation is further substantiated by the findings that EAP1 attenuated the rapamycin-induced translation of GCN4 [41] and caused general attenuation of translation in cells with mutations in the secretory pathway [42]. Furthermore, EAP1 has been found in messenger ribonucleoprotein complexes together with PABP1 [43,44], the poly(A)-binding protein that has been shown to increase the efficiency of translation [45]. These results place SMY2 and SYH1 in the functional context of regulation of translation initiation. Additional support for the involvement of SYH1 in processes that regulate initiation of translation comes from yeast two-hybrid experiments. SYH1 binds to p33 and p39 [46], two subunits of the eIF3 complex [47–49]. eIF3 is part of a multifactor complex [50], which assists in the assembly of the 48S preinitiation complex that scans the mRNA for an AUG start codon [49,51]. Besides its role in translation initiation, eIF4E has also been implicated in cap-mediated transport and decay of mRNA [52]. As a lethal phenotype, as would be expected from a general defect in splicing or translation, is not observed in SMY2 or SYH1 yeast deletion strains [53], it is likely that the GYF domain-containing proteins in yeast are functionally redundant or involved in the spliceosomal and/or translational control of specific genes.

GYF domain-binding models

On the basis of NMR experiments, different binding models for CD2BP2-like and SMY2-like GYF domains and their respective ligands have emerged (Fig. 3). Mapping of the binding epitope of PPGW ligands on to CD2BP2-GYF suggested a binding mode, similar to that of the CD2 peptide [4,5]. However, the aromatic residue in PPGW-type ligands is assumed to be oriented towards the domain and undergoes hydrophobic interactions with W8 [20]. This is in contrast to the side chain of the corresponding residue H9 in the CD2 peptide, which is oriented towards the solvent. In SMY2-type GYF domains, tryptophan at position 8 (CD2BP2-GYF numbering) is replaced by an aspartate (Fig. 1). Therefore, comparable interactions with the hydrophobic residues that follow the PPG core could not account for the observed ligand specificity of SMY-like GYF domains. NMR studies confirmed the conserved aromatic residues of SMY2-GYF to be involved in ligand binding, but the domain–ligand interface is extended. The structure of a SMY2-type GYF domain (PDB code 1wh2) suggested that the extended epitope falls into the region of the W8D replacement, which partially exposes hydrophobic side chains, including those of the conserved residues Y6 and F34 (CD2BP2-GYF numbering). Accordingly, the additionally exposed surface may preferentially accommodate the hydrophobic amino acid found in the Φ position of the ligands. In the CD2BP2-GYF domain subfamily this hydrophobic patch is shielded by W8, favouring inter-molecular stacking with ligand trytophan (Fig. 3).

Intramolecular recognition of proline-rich motifs by GYF domains

Some of the proteins that contain GYF domains also display sequence motifs matching the respective consensus of the domains. To test the hypothesis that intramolecular recognition can inhibit intermolecular interactions, phage display was performed with protein fragments containing the GYF domain and the intramolecular target site of the proteins SYH1 (yeast) and GYN4 (*A. thaliana*). The PPGI motif of SYH1 is adjacent to the N-terminal domain border as predicted by sequence homology, whereas in the *A. thaliana* protein GYN4, an amino acid spacer of 5–8 residues at the

C-terminus, links the assumed domain border with the PRS. Interestingly, for the SYH1 GYF domain construct, comprising the N-terminal PRS, the same recognition motif as for the shorter SYH1 GYF construct without PRS was selected from the randomized peptide library. Similarly, the PRS did not interfere with the binding of the SYH1 GYF constructs in glutathione S-transferase pull-down experiments. In contrast, the GYN4 GYF domain comprising the intramolecular PRS at the C-terminus failed to show interaction with any peptide or protein fragments tested by phage display and yeast two-hybrid analysis. These findings suggest that intramolecular interaction between the GYF domain and a C-terminal PRS masks the binding properties of the GYN4 GYF domain, whereas for the SYH1 protein, the intramolecular sequence at the N-terminus of the GYF domain does not influence its interaction competence. The assumption that a C-terminal linker sequence of 5–8 amino acids is sufficient to allow intramolecular interactions to occur is supported by investigations of an artificial GYF–linker–peptide construct used to obtain NMR restraints for a detailed structural description of the CD2BP2–CD2 interaction [54]. In this example, only the C-terminal linker resulted in the formation of an intramolecular interaction and argued for the preferred orientation of the CD2 peptide SHRPPPPGHRV as displayed in Fig. 2. Similarly, the linker within the GYN4 protein is probably of sufficient length to enable the autoinhibited conformation, whereas the attachment of a PRS directly N-terminal to the GYF domain, as in the SYH1 protein, excludes such an encounter. The observation of intramolecular masking is probably of functional relevance and may prevent unwanted low-affinity interactions to take place. Co-compartmentalization of a protein containing a high-affinity binding site can be envisaged to compete with the intramolecular interaction under appropriate conditions. In addition, post-translational modification of residues close to the binding site, such as phosphorylation or methylation, may act as regulatory mechanisms for the GYF domain-mediated interactions.

Conclusions

Detailed analysis of the structure–function relationship of the CD2BP2-GYF domain has laid the foundation for an understanding of GYF domains in general. The use of a conserved set of aromatic residues forming a central hydrophobic pocket for ligand binding seems to hold true for most, if not all, GYF domains. Studies of the binding propensities for the CD2BP2-GYF domain and four members of the large SMY2 subfamily have generated the framework for comprehensive mapping of the GYF domain ligand space. An interesting twist comes from the observation that the GYF domain of CD2BP2 can bind to a protein devoid of PRS. Such an adaptor function may be a general feature of GYF domains and would allow the simultaneous binding of a second interaction partner. So far, only a few physiological binding partners for GYF domains have been identified, and the exact role of GYF domain-containing proteins is poorly understood. However, defining putative interaction partners hints at possible functions of these proteins. Involvement in splicing or splicing-associated processes is a recurrent functional theme for proteins containing GYF domains. In particular, GYF domain-containing proteins may mediate mRNA export from the nucleus or couple splicing with transcription or translation. To what extent GYF domain-mediated interactions with proteins involved in immune cell function or antigen presentation are related to splicing-associated functions is another exciting aspect of future research. Moreover, masking by intramolecular target sites and the interplay of such autoinhibited conformations with intermolecular binding or post-translational modifications is just emerging as an important topic in the understanding of GYF domains. We are thus at the beginning of an era where a true functional understanding of this small, yet versatile, domain can be envisaged.

Acknowledgements

This work was supported in part by a grant from the *Deutsche Forschungsgemeinschaft* (DFG 1325/2-1) to C.F.

References

1 Nishizawa K, Freund C, Li J, Wagner G & Reinherz EL (1998) Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proc Natl Acad Sci USA* **95**, 14897–14902.
2 Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
3 Freund C, Dotsch V, Nishizawa K, Reinherz EL & Wagner G (1999) The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nat Struct Biol* **6**, 656–660.
4 Freund C, Kuhne R, Yang H, Park S, Reinherz EL & Wagner G (2002) Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *EMBO J* **21**, 5985–5995.
5 Kofler M, Heuer K, Zech T & Freund C (2004) Recognition sequences for the GYF domain reveal a possible

spliceosomal function of CD2BP2. *J Biol Chem* **279**, 28292–28297.
6 Gu W, Kofler M, Antes I, Freund C & Helms V (2005) Alternative binding modes of proline-rich peptides binding to the GYF domain. *Biochemistry* **44**, 6404–6415.
7 Mayer BJ, Hamaguchi M & Hanafusa H (1988) A novel viral oncogene with structural similarity to phospholipase C. *Nature* **332**, 272–275.
8 Varmus H, Hirai H, Morgan D, Kaplan J & Bishop JM (1989) Function, location, and regulation of the src protein-tyrosine kinase. *Princess Takamatsu Symp* **20**, 63–70.
9 Bork P & Sudol M (1994) The WW domain: a signalling site in dystrophin? *Trends Biochem Sci* **19**, 531–533.
10 Sudol M, Chen HI, Bougeret C, Einbond A & Bork P (1995) Characterization of a novel protein-binding module: the WW domain. *FEBS Lett* **369**, 67–71.
11 Niebuhr K, Ebel F, Frank R, Reinhard M, Domann E, Carl UD, Walter U, Gertler FB, Wehland J & Chakraborty T (1997) A novel proline-rich motif present in ActA of *Listeria monocytogenes* and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. *EMBO J* **16**, 5433–5444.
12 Pornillos O, Alam SL, Rich RL, Myszka DG, Davis DR & Sundquist WI (2002) Structure and functional interactions of the Tsg101 UEV domain. *EMBO J* **21**, 2397–2406.
13 Carlsson L, Nystrom LE, Sundkvist I, Markey F & Lindberg U (1977) Actin polymerizability is influenced by profilin, a low molecular weight protein in non-muscle cells. *J Mol Biol* **115**, 465–483.
14 Zarrinpar A, Bhattacharyya RP & Lim WA (2003) The structure and function of proline recognition domains. *Sci STKE* RE8.
15 Ball LJ, Kuhne R, Schneider-Mergener J & Oschkinat H (2005) Recognition of proline-rich motifs by protein–protein interaction domains. *Angew Chem Int Ed Engl* **44**, 2852–2869.
16 Mayer BJ (2001) SH3 domains: complexity in moderation. *J Cell Sci* **114**, 1253–1263.
17 Kay BK, Williamson MP & Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14**, 231–241.
18 Bedford MT, Reed R & Leder P (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. *Proc Natl Acad Sci USA* **95**, 10602–10607.
19 Espejo A, Cote J, Bednarek A, Richard S & Bedford MT (2002) A protein-domain microarray identifies novel protein–protein interactions. *Biochem J* **367**, 697–702.
20 Kofler M, Motzny K, Beyermann M & Freund C (2005) Novel interaction partners of the CD2BP2-GYF domain. *J Biol Chem* **280**, 33397–33402.
21 Cesareni G, Panni S, Nardelli G & Castagnoli L (2002) Can we infer peptide recognition specificity mediated by SH3 domains? *FEBS Lett* **513**, 38–44.
22 Otte L, Wiedemann U, Schlegel B, Pires JR, Beyermann M, Schmieder P, Krause G, Volkmer-Engert R, Schneider-Mergener J & Oschkinat H (2003) WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. *Protein Sci* **12**, 491–500.
23 Chang HC, Moingeon P, Pedersen R, Lucich J, Stebbins C & Reinherz EL (1990) Involvement of the PPPGHR motif in T cell activation via CD2. *J Exp Med* **172**, 351–355.
24 de Hoog CL, Foster LJ & Mann M (2004) RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. *Cell* **117**, 649–662.
25 Stark H, Dube P, Luhrmann R & Kastner B (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539–542.
26 Laggerbauer B, Liu S, Makarov E, Vornlocher HP, Makarova O, Ingelfinger D, Achsel T & Luhrmann R (2005) The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation. *RNA* **11**, 598–608.
27 Kofler M, Motzny K & Freund C (2005) GYF domain proteomics reveals interaction sites in known and novel target proteins. *Mol Cell Proteomics* **4**, 1797–1811.
28 Landgraf C, Panni S, Montecchi-Palazzi L, Castagnoli L, Schneider-Mergener J, Volkmer-Engert R & Cesareni G (2004) Protein interaction networks by proteome peptide scanning. *Plos Biol* **2**, E14.
29 McCutchen-Maloney SL, Matsuda K, Shimbara N, Binns DD, Tanaka K, Slaughter CA & DeMartino GN (2000) cDNA cloning, expression, and functional characterization of PI31, a proline-rich inhibitor of the proteasome. *J Biol Chem* **275**, 18557–18565.
30 Zaiss DM, Standera S, Kloetzel PM & Sijts AJ (2002) PI31 is a modulator of proteasome formation and antigen processing. *Proc Natl Acad Sci USA* **99**, 14344–14349.
31 Komuro A, Saeki M & Kato S (1999) Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-rich motif containing glycine and arginine. *J Biol Chem* **274**, 36513–36519.
32 Craggs G, Finan PM, Lawson D, Wingfield J, Perera T, Gadher S, Totty NF & Kellie S (2001) A nuclear SH3 domain-binding protein that colocalizes with

mRNA splicing factors and intermediate filament-containing perinuclear networks. *J Biol Chem* **276**, 30552–30560.
33 Llorian M, Beullens M, Lesage B, Nicolaescu E, Beke L, Landuyt W, Ortiz JM & Bollen M (2005) Nucleocytoplasmic shuttling of the splicing factor SIPP1. *J Biol Chem* **280**, 38862–38869.
34 Zhang Y, Lindblom T, Chang A, Sudol M, Sluder AE & Golemis EA (2000) Evidence that dim1 associates with proteins involved in pre-mRNA splicing, and delineation of residues essential for dim1 interactions with hnRNP F and Npw38/PQBP-1. *Gene* **257**, 33–43.
35 Waragai M, Junn E, Kajikawa M, Takeuchi S, Kanazawa I, Shibata M, Mouradian MM & Okazawa H (2000) PQBP-1/Npw38, a nuclear protein binding to the polyglutamine tract, interacts with U5-15kD/dim1p via the carboxyl-terminal domain. *Biochem Biophys Res Commun* **273**, 592–595.
36 Llorian M, Beullens M, Andres I, Ortiz JM & Bollen M (2004) SIPP1, a novel pre-mRNA splicing factor and interactor of protein phosphatase-1. *Biochem J* **378**, 229–238.
37 Giovannone B, Lee E, Laviola L, Giorgino F, Cleveland KA & Smith RJ (2003) Two novel proteins that are linked to insulin-like growth factor (IGF-I) receptors by the Grb10 adapter and modulate IGF-I signaling. *J Biol Chem* **278**, 31564–31573.
38 Rutz B & Seraphin B (2000) A dual role for BBP/ScSF1 in nuclear pre-mRNA retention and splicing. *EMBO J* **19**, 1873–1886.
39 Lillie SH & Brown SS (1992) Suppression of a myosin defect by a kinesin-related gene. *Nature* **356**, 358–361.
40 Cosentino GP, Schmelzle T, Haghighat A, Helliwell SB, Hall MN & Sonenberg N (2000) Eap1p, a novel eukaryotic translation initiation factor 4E-associated protein in *Saccharomyces cerevisiae*. *Mol Cell Biol* **20**, 4604–4613.
41 Matsuo R, Kubota H, Obata T, Kito K, Ota K, Kitazono T, Ibayashi S, Sasaki T, Iida M & Ito T (2005) The yeast eIF4E-associated protein Eap1p attenuates GCN4 translation upon TOR-inactivation. *FEBS Lett* **579**, 2433–2438.
42 Deloche OCJ, Kressler D, Doere M & Linder P (2004) A membrane transport defect leads to a rapid attenuation of translation initiation in *Saccharomyces cerevisiae*. *Mol Cell* **13**, 357–366.
43 Lang BD & Fridovich-Keil JL (2000) Scp160p, a multiple KH-domain protein, is a component of mRNP complexes in yeast. *Nucleic Acids Res* **28**, 1576–1584.
44 Mendelsohn BA, Li AM, Vargas CA, Riehman K, Watson A & Fridovich-Keil JL (2003) Genetic and biochemical interactions between SCP160 and EAP1 in yeast. *Nucleic Acids Res* **31**, 5838–5847.
45 von Der HT, Ball PD & McCarthy JE (2000) Stabilization of eukaryotic initiation factor 4E binding to the mRNA 5′-Cap by domains of eIF4G. *J Biol Chem* **275**, 30551–30555.
46 Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M & Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* **98**, 4569–4574.
47 Verlhac MH, Chen RH, Hanachi P, Hershey JW & Derynck R (1997) Identification of partners of TIF34, a component of the yeast eIF3 complex, required for cell proliferation and translation initiation. *EMBO J* **16**, 6812–6822.
48 Naranda T, Kainuma M, MacMillan SE & Hershey JW (1997) The 39-kilodalton subunit of eukaryotic translation initiation factor 3 is essential for the complex's integrity and for cell viability in Saccharomyces cerevisiae. *Mol Cell Biol* **17**, 145–153.
49 Hershey JW & Merrick WC (2000) The pathway and mechanism of initiation of protein synthesis. In *Translational Control of Gene Expression* (Sonenberg N, Hershey JW & Mathews MB, eds), pp. 33–88. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
50 Valasek L, Nielsen KH & Hinnebusch AG (2002) Direct eIF2-eIF3 contact in the multifactor complex is important for translation initiation in vivo. *EMBO J* **21**, 5886–5898.
51 Hinnebusch AG (2000) Mechanism and regulation of initiator methionyl-tRNA binding to ribosomes. In *Translational Control of Gene Expression* (Sonenberg N, Hershey JW & Mathews MB, eds), pp. 185–243. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
52 Andrei MA, Ingelfinger D, Heintzmann R, Achsel T, Rivera-Pomar R & Luhrmann R (2005) A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *RNA* **11**, 717–727.
53 Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, Dunn B, Dwight S, Riles L, Mortimer RK & Botstein D (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387**, 67–73.
54 Freund C, Kuhne R, Park S, Thiemke K, Reinherz EL & Wagner G (2003) Structural investigations of a GYF domain covalently linked to a proline-rich peptide. *J Biomol NMR* **27**, 143–149.
55 Siliciano RF, Pratt JC, Schmidt RE, Ritz J & Reinherz EL (1985) Activation of cytolytic T lymphocyte and natural killer cell function through the T11 sheep erythrocyte binding protein. *Nature* **317**, 428–430.
56 Hahn WC & Bierer BE (1993) Separable portions of the CD2 cytoplasmic domain involved in signaling and ligand avidity regulation. *J Exp Med* **178**, 1831–1836.
57 Stark H, Dube P, Luehrmann R & Kastner B (2001) Arrangement of RNA and proteins in the spliceosomal

U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539–542.
58 Kramer A (1992) Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Mol Cell Biol* **12**, 4545–4552.
59 Arning S, Gruter P, Bilbe G & Kramer A (1996) Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* **2**, 794–810.
60 Albert TK, Hanzawa H, Legtenberg YI, de Ruwe MJ, van den Heuvel FA, Collart MA, Boelens R & Timmers HT (2002) Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex. *EMBO J* **21**, 355–364.
61 Collart MA (2003) Global control of gene expression in yeast by the Ccr4-Not complex. *Gene* **313**, 1–16.
62 Bialkowska A & Kurlandzka A (2002) Proteins interacting with Lin 1p, a putative link between chromosome segregation, mRNA splicing and DNA replication in *Saccharomyces cerevisiae*. *Yeast* **19**, 1323–1333.
63 Stevens SW, Barta I, Ge HY, Moore RE, Young MK, Lee TD & Abelson J (2001) Biochemical and genetic analyses of the U5, U65 and U4/U6 center dot U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7**, 1543–1553.
64 Abovich N & Rosbash M (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403–412.
65 Matsuo R, Kubota H, Obata T, Kito K, Ota K, Kitazono T, Ibayashi S, Sasaki T, Iida M & Ito T (2005) The yeast eIF4E-associated protein Eap1p attenuates GCN4 translation upon TOR-inactivation. *FEBS Lett* **579**, 2433–2438.
66 Lang BD & Fridovich-Keil JL (2000) Scp160p, a multiple KH-domain protein, is a component of mRNP complexes in yeast. *Nucleic Acids Res* **28**, 1576–1584.
67 Mendelsohn BA, Li AM, Vargas CA, Riehman K, Watson A & Fridovich-Keil JL (2003) Genetic and biochemical interactions between SCP160 and EAP1 in yeast. *Nucleic Acids Res* **31**, 5838–5847.
68 Verlhac MH, Chen RH, Hanachi P, Hershey JW & Derynck R (1997) Identification of partners of TIF34, a component of the yeast eIF3 complex, required for cell proliferation and translation initiation. *EMBO J* **16**, 6812–6822.
69 Naranda T, Kainuma M, MacMillan SE & Hershey JW (1997) The 39-kilodalton subunit of eukaryotic translation initiation factor 3 is essential for the complex's integrity and for cell viability in *Saccharomyces cerevisiae*. *Mol Cell Biol* **17**, 145–153.