GYF Domain Proteomics Reveals Interaction Sites in Known and Novel Target Proteins*

Michael Kofler, Kathrin Motzny, and Christian Freund‡

GYF domains are conserved eukaryotic adaptor domains that recognize proline-rich sequences. Although the structure and function of the prototypic GYF domain from the human CD2BP2 protein have been characterized in detail, very little is known about GYF domains from other proteins and species. Here we describe the binding properties of four GYF domains of various origins. Phage display in combination with SPOT analysis revealed the PPG(F/I/L/M/V) motif as a general recognition signature. Based on these results, the proteomes of human, yeast, and Arabidopsis thaliana were searched for potential interaction sites. Binding of several candidate proteins was confirmed by pull-down experiments or yeast two-hybrid analysis. The binding epitope of the GYF domain from the yeast SMY2 protein was mapped by NMR spectroscopy and led to a structural model that accounts for the different binding properties of SMY2-type GYF domains and the CD2BP2-GYF domain. Molecular & Cellular Proteomics 4:1797–1811, 2005.

The GYF domain is a protein interaction domain ubiquitously expressed in eukaryotic species (1, 2). It belongs to the functional class of proline-rich sequence (PRS)¹ recognition domains such as SH3 (3, 4), WW (5), EVH1 (6), and UEV domains (7, 8) as well as profilin (9). Although the sequence requirements have been investigated in detail for most of the PRS-binding domains, only little is known about the recognition code of GYF domains, and the biological role of GYF domain-containing proteins remains largely unknown. So far, an in-depth analysis has been performed solely for the GYF domain of CD2BP2. For this GYF domain a role in CD2 receptor-dependent T cell signaling has been observed (1, 10), and the identification of CD2BP2 as part of the U5 snRNP (11, 12) and as interaction partner of the core splicing protein SmB/B’ (13) suggests an independent role in splicing or splicing-associated processes. The structure of the complex of the CD2BP2-GYF domain with a CD2-derived proline-rich peptide defined a set of mostly conserved aromatic amino acids of the domain to act as the primary contact site for the interaction (2, 10). Analysis of the binding properties revealed two classes of ligands for CD2BP2-GYF. The CD2 class shows a charge dependence of binding and is characterized by the recognition motif PPGX(R/K), whereas the so-called PPGW class requires a hydrophobic residue directly C-terminal to the PPG core, and its recognition motif was identified as PPG(W/F/Y/M/L) (39). Sequence alignment suggests that the GYF domain of CD2BP2 belongs to a GYF domain subfamily that is characterized by a tryptophan at position 8 and an extended loop between β-strands 1 and 2 (10) (see Fig. 1). Most GYF domains contain an aspartate at position 8 and a shorter β1-β2 loop (Fig. 1) forming a second subfamily. Here we describe the analysis of four GYF domains from yeast, plant, and man that belong to the second subfamily, which we name the SMY2 subfamily of GYF domains. SMY2 was originally cloned as a suppressor of the myo2-66 mutation in the motor protein MYO2 in Saccharomyces cerevisiae (14), and our investigation of the GYF domains of SMY2 and three other proteins shows that all sequences recognized by the different GYF domains converge on the PPG motif followed by one of the hydrophobic residues Phe, Ile, Leu, Met, and Val (Φ). Except for this PPGΦ motif there is little dependence on additional flanking residues within the respective ligands. The consensus motifs that were derived by a combination of phage display and substitution analysis allowed us to identify natural target sites. Peptides comprising these sites were tested for binding by membrane SPOT analysis. Several interactions were further investigated by fluorescence and NMR titration experiments revealing binding epitopes and affinities for the novel ligands. Yeast two-hybrid analysis and pull-down experiments confirmed the association of selected proteins with GYF domains under more physiological conditions and allowed the placement of GYF domain-containing proteins into known functional contexts. In S. cerevisiae, the GYF domains of SMY2 and its paralog YPL105C (named SYH1 for SMY2 homolog 1) interact with proteins that have been im-
plicated in pre-mRNA branch point binding (MSL5) or regulation of translation (EAP1). For the human O75137 protein (also indicated as PERQ2 because of its homology to human PERQ amino acid rich with GYF domain protein 1 (PERQ1)), our work identified the core splicing proteins SmB/B' and snRNP-N as the most likely interaction partners, whereas the Arabidopsis thaliana Q9FMM3-GYF domain selects for sequences in the homolog of the human transcription regulator CNOT4 (accordingly termed GYN4, GYF domain-containing protein binding to Nof4 homolog). GYN4 and SYH1 contain internal binding sites for their respective GYF domains, and we show that intramolecular association precludes intermolecular binding in the case of GYN4.

MATERIALS AND METHODS

Constructs—The estimated domain borders of the GYF domains of proteins PERQ2 (Swiss-Prot accession number O75137; residues 531–596), GYN4 (Swiss-Prot accession number Q9FMM3; residues 546–604), SYM2 (Swiss-Prot accession number P32909; residues 243–340), and SYH1 (Swiss-Prot accession number Q02875; residues 150–226) were obtained from sequence alignments, and the corresponding fragments were subcloned into expression vectors. For GYN4 and SYH1, longer constructs containing an internal PRS (GYN4 residues 546–619, termed GYN4-PR, and SYH1 residues 141–226, termed PR-SYH1; see Fig. 1) were also cloned into protein expression vectors. All constructs were amplified from the following DNA clones. PERQ2 was derived from the cDNA clone HU03496 and GYN4 was derived from genomic P1 clone MBD2 (kind gifts from Takahiro Nagase and Satoshi Tabata, Kazusa DNA Research Institute, respectively), whereas SMY2 and SYH1 were amplified from genomic DNA of yeast strain S288C. For GST fusion protein expression, the fragments of PERQ2, SYH1, and SMY2 were cloned into pGEX4T-1 (Amersham Biosciences) via BamHI and Xhol restriction sites, and the fragments of GYN4 were cloned via BamHI and NotI restriction sites. For yeast two-hybrid screens the GYF domains of PERQ2 and GYN4 were cloned into the bait vector pGADT7 (Clontech) via Ncol and BamHI and with Ncol and NotI restriction sites, respectively. For the yeast two-hybrid analysis of selected candidates, the same fragment of PERQ2 was also introduced into the prey vector pGADT7 (Clontech). Fragments of the candidates NPWB (Swiss-Prot accession number Q9Y2W2; residues 388–551), SmB (Swiss-Prot accession number P14678-2; residues 148–231), and SWAN (Swiss-Prot accession number Q8NTZ6; residues 700–869) were amplified from the following I.M.A.G.E. Consortium (Lawrence Livermore National Laboratory) cDNA clones (15) obtained from the Deutsches Ressourcenzentrum für Genomforschung GmbH; IRAIP962/0114Q2 (NPWB), IMAGp998/1841503 (SmB), and IRAIP962/4172502 (SWAN), the respective I.M.A.G.E. Consortium clone identification numbers are 3829990, 3445210, and 3956772. These fragments, comprising the proline-rich regions of the respective proteins, and the cytoplasmic tail of CD2 (Swiss-Prot accession number P60729; residues 245–351) were inserted into pGBK7 via Ncol-NotI or EcoRI and BamHI restriction sites in the latter case. The Hisα-tagged GYF domain of SMY2 was expressed from a modified PET28 vector, containing an N-terminal Hisα tag, followed by a thrombin cleavage site and BamHI and Xhol restriction sites, which allowed cloning similar to that for the GST fusion construct; for Hisα-tagged PERQ2, the corresponding fragment was cloned into the pTFT74 vector (16) via Ncol and HindIII sites with an N-terminal Hisα tag introduced via PCR.

Cloning of the focused library RKSHRXKPPXXVQ into PC89 was similar to the procedure described elsewhere (17). PC89 and the PC89 nonapeptide library (Xj) were a gift from Gianni Cesareni (Dipartimento di Biologia, Università di Roma).

Protein Preparation—Proteins were expressed in Escherichia coli BL21 (DE3-pLys S) and purified from the soluble fraction after sonication. GST-GYF and Hisα-tagged GYF domains were purified by affinity chromatography using glutathione-Sepharose and Ni2+-nitrilotriacetic acid-agarose according to the manufacturer’s manual (Amersham Biosciences), respectively, and dialyzed against PBS. To obtain NMR samples for titration experiments and backbone resonance assignments of the SMY-GYF domain, cells were grown on defined media supplemented with [15N]NH₄Cl and/or [13C]glucose. The Hisα tag of SMY2-GYF and GST tag of SYH1-GYF were removed by thrombin cleavage (Calbiochem, 10 units/mg of protein, at 4 °C and 16 °C overnight in PBS, respectively), and the domains were purified by subsequent gel filtration (Superdex® 75, Amersham Biosciences) in PBS.

Phage Display—Phage displaying the nonapeptide (Xj) or the focused peptide library fused to the major capsid protein were produced by transforming E. coli XL-1 Blue cells with PC89 constructs followed by superinfection with the VCS-M13 helper phage (Stratagene). After overnight incubation in 2× YT medium in the presence of ampicillin and kanamycin (30 °C at 270 rpm) phage particles were purified by three successive polyethylene glycol/NaCl precipitations (18). Library screening was performed as follows. 30–50 μl of GST-GYF-loaded glutathione-Sepharose 4B beads (Amersham Biosciences) were incubated with 5 × 10−5–5 × 10−11 infectious particles at 4 °C overnight in PBS. After washing three times with PBS, bound phages were eluted with 100 mM glycine HCl, pH 2.2, and neutralized with 2 mM Tris. For phage amplification, E. coli XL-1 Blue cells were infected with eluted phage followed by superinfection with helper phage and subsequent incubation as described above. After six rounds of panning, the inserts of selected phage were sequenced to identify their displayed peptides.

SPOT Analysis—Peptides were synthesized on Whatman 50 paper using an Auto-Spot Robot (ASP 222, Invertas AG, Köln, Germany). Peptide synthesis using Fmoc (N-α-(9-fluorenylmethyloxycarbonyl)) chemistry on β-alanine functionalized cellulose membranes was performed according to standard protocols (19). Membranes were probed with GST-GYF protein as described elsewhere (20). Briefly the membranes were incubated with GST-GYF (40 μg/ml) overnight. After washing with PBS, bound GST fusion protein was detected with rabbit polyclonal anti-GST antibody (Z-5, Santa Cruz Biotechnology) and horseradish peroxidase-coupled anti-rabbit IgG antibodies (Roche). An enhanced chemiluminescence substrate (SuperSignal West Pico, Pierce) was used for detection by a Lumilimegger™ (Roche Applied Science).

NMR Spectroscopy—The NMR experiments were performed at 297 or 299 K on either a Bruker DRX600 or DMX750 instrument equipped with standard triple resonance probes. Data processing and analysis were carried out with the XWINNMR (Bruker) and SPARKY (40) software packages. Backbone assignment of the SMY2-GYF domain was based on the CBCA(CO)NH (21), the CB-SPARKY (40) software packages. Backbone assignment of the SMY2-GYF domain was based on the CBCA(CO)NH (21), the CB-SPARKY (40) software packages. Backbone assignment of the SMY2-GYF domain was based on the CBCA(CO)NH (21), the CB-SPARKY (40) software packages. Backbone assignment of the SMY2-GYF domain was based on the CBCA(CO)NH (21), the CB-SPARKY (40) software packages.
RESULTS

Phage Display—GYF domains from evolutionary distant species of the eukaryotic kingdom were chosen for analysis. The sequence alignment of these domains is shown in Fig. 1 and highlights the conserved amino acids of the N-terminal part of the domains. In the case of GYN4 and SYH1, two GYF domain constructs of varying length were used to account for the presence of an internal PRS that is localized directly adjacent to the anticipated domain borders (Fig. 1). The phage libraries applied for the screen were of the format X₉PPPₓ₈ or X₉, and the peptides were expressed as gene VIII fusion proteins (17). Individual clones were sequenced after six rounds of panning, and the results are summarized in Table I. As can be seen, the obtained sequence motifs for both libraries are similar, and the preference for the PPG core motif is observed for all of the GYF domains. A strict requirement for a third consecutive proline is not seen in the case of the X₉ library and is in agreement with previous structural data of the CD2BP2-GYF domain-CD2 peptide complex, showing that the PPG motif contributes most of the van der Waals interactions (10). The general requirement for the PPG motif suggests that the binding mode for the four GYF domains and CD2BP2-GYF is similar. The requirement for only two prolines...
gyf domain proteomics

raises the question, however, whether formation of the polyproline II helix is a prerequisite for binding of ligands to GYF domains in general.

For the position directly C-terminal to the PPG motif, the GYF domains of SMY2, SYH1, and PERQ2 seem to accept most hydrophobic amino acids with a preference for leucine or isoleucine. For the GYN4-GYF domain (containing amino acids 546–604 of the full-length protein), phenylalanine is the preferred amino acid at this position, whereas a longer construct (GYN4-PR; amino acids 546–619) that includes the proline-rich sequence KSGPPPPGFTG failed to select for peptides defining a consensus (Table I and data not shown). Interestingly the internal PRS comprises the consensus motif PPGF for this GYF domain as can be seen from Table I. Because previous experiments from our group have shown that a C-terminally linked proline-rich ligand can bind intramolecularly to the CD2BP2-GYF domain (Fig. 1) (27), we suggest that the intramolecular binding of the ligand in GYN4-PR masks the GYF domain binding site accordingly. In contrast, the presence of the internal PRS directly N-terminal to the

| Libraries | PERQ2 | SMY2 | SYH1 | PR-SYH1 | GYN4 |
|-----------|-------|------|------|---------|------|
| **Focused Library** | | | | | |
| RRKSHRXXXPFFXXXVQ | | | | | |
| **X9-Library** | | | | | |
| XXXXXXXX | | | | | |
| **Strict Consensus** | PPG[FLMW] | PPG[ILMY] | PPGF | |
| **Relaxed Consensus** | PPG[AEFHILMSTWY] | PPG[ACFGILMVWY] | PPG[FLMVWY] | |
GYF domain in SYH1 does not interfere with phage selection (Table I). This is in agreement with experiments of CD2BP2-GYF domain single chain constructs, showing that a short linker sequence is not sufficient to allow the intramolecular interaction between an N-terminal binding motif and the GYF domain to take place (27).

**Analysis of Phage Display Peptides**—Substitution analysis of individual peptides selected by phage display were conducted to identify residues within the peptide that are important for binding. Therefore, peptides were synthesized on a cellulose membrane, and the binding to GST-GYF domains was analyzed by anti-GST antibody detection (Fig. 2). The substitution analysis highlights the requirement for the PPG(F/I/L/V) motif for the GYF domains of PERQ2, SMY2, and SYH1. A large hydrophobic amino acid is clearly preferred as the most C-terminal amino acid of the motif in these cases, whereas GYN4-GYF exclusively binds to the PPGF motif. For all four domains, amino acids outside of the central PPG motif of the phage display-derived peptides are not critical for binding. This suggests that hydrophobic interactions almost entirely account for the observed binding behavior. In summary, these results identify the general recognition code for the SMY2 subfamily of GYF domains to be PPG(F/I/L/M/V) with subtle domain-dependent variations in specificity (Table I and Fig. 2).

**Proteome Analysis for GYF Domain Binding Sites**—The phage display results allowed us to define the recognition signature for each individual domain (Table I, strict consensus). Strict consensus sequences included the most frequently found PPGX motifs or motifs with amino acids at position X that show physicochemical properties similar to the frequently found amino acid. The relaxed consensus motifs comprised the essential binding motifs derived from the substitution analysis. Additional positions were not taken into account for both strict and relaxed consensus because the substitution analysis results showed that the PPGX motif is the major determinant for binding. We performed database searches in the Swiss-Prot/TrEMBL, The Arabidopsis Information Resource (TAIR), and SGD databases for the different GYF domains with relaxed and strict consensus motifs (Table I) following a strategy that has also been used by others (28). For the yeast GYF domains, the search was performed with the relaxed consensus PPG(A/C/F/G/I/L/M/V/W/Y). For the human PERQ2-GYF domain and the *A. thaliana* GYF domain of GYN4, the relaxed consensus PPG(A/E/F/H/I/L/M/S/T/W/Y) and PPG(F/I/L/M/V/W/Y) were applied in the database search, respectively. In the latter two cases, motifs were only selected if present twice in the protein with a maximal linker distance of 40 amino acids. This accounts for avidity effects of tandem binding motifs and reduces the number of database hits below 1000. Based on our phage display results, database searches were also performed with the strict motifs PPGQ and PPGF for the GYF domains of PERQ2 and GYN4, respectively, to identify optimal individual binding motifs. All human proteins (except for procollagens, collagens, and hypothetical proteins) were considered in the search for PERQ2. 349 non-redundant motifs in 152 proteins for the relaxed and 166 non-redundant motifs in 157 proteins for the strict consensus were found. Correspondingly for *A. thaliana* the relaxed motif was found 34 times in 15 proteins, and the strict
consensus was found 227 times in 221 proteins of the *A. thaliana* proteome. For *S. cerevisiae*, 168 motifs in 153 sequences containing the relaxed consensus were analyzed. To test binding, the peptides obtained from the database searches were synthesized on a membrane and incubated with the GST-GYF fusion constructs of the indicated proteins. The consensus sequences that were used as profiles in the respective searches are depicted above the membranes. Peptide spots comprising the internal recognition motif for GYN4 and SYH1, respectively, are circled.

![Search for GYF domain binding sites in natural proteins](image)

**Fig. 3.** Search for GYF domain binding sites in natural proteins. Swiss-Prot/TrEMBL, TAIR, or the SGD databases were searched for sequences meeting the relaxed or strict consensus motifs of the GYF domains of PERQ2, GYN4, SYH1, and SMY2. Selected candidates were synthesized on a membrane and incubated with the GST-GYF fusion constructs of the indicated proteins. The consensus sequences that were used as profiles in the respective searches are depicted above the membranes. Peptide spots comprising the internal recognition motif for GYN4 and SYH1, respectively, are circled.

Further analysis according to signal strength and number of consensus motifs present within the proteins (Table II). In addition, fluorescence and NMR titrations of GYF domains with selected peptides were performed. The dissociation constants for the GYF domain peptide interactions are in the range of 4–300 μM (see Table IV and Supplemental Fig. 1) when a simple two-state binding model was assumed.

**Pull-down Experiments: Novel Interaction Partners of the Yeast GYF Domains—** From the proteome analysis a large set of interactions was derived for the individual GYF domains. For the two yeast GYF domains of SMY2 and SHY1, the most
likely interaction candidates are MSL5, EAP1, and PRP8 with seven, four, and three motifs fitting to the relaxed consensus, respectively (Table II), whereas three other proteins contain the motif twice. SPOT analysis confirmed that several of the PRSs in the yeast proteins MSL5 and EAP1 present bona fide binding sites for SMY2 and SYH1 in vitro, and avidity effects likely contribute to enhanced binding in vivo. To validate MSL5 and EAP1 as interaction partners, we performed pull-down analysis with SYH1 and SMY2.

| GYF Domain strict consensus | Candidate | Proline-rich Motifs within Candidate Protein |
|-------------------------------|-----------|-------------------------------------------|
| PERQ2 PPGL(FWM) PPGL(AEHILMTWY) | NPWB (358-351) | SQQQAPQQSVPSSQIQAPMPGPPPLPGPPAPRPG | |
| SmB (148-231) | TAAAASISACAPTQPQPPGRGPPPMGRGAPP | |
| SWAN (700-859) | AGGBHAFILTGVSEKANNGGNPAGPFGSPAP | |

| GYN4 PPGL PPGL(FILMVY) | At2g28540.1 | SSYYKPTSVPDFVSAPPAPGGVTPSSRPPGG | |
| At3g45630.1 | LSYKSPVARTQVSGAPGAPPANRPPGGSSHQRGLSS | |
| At5g06170.1 | FSFKNPVSPQVSAAPGSGAPRLLPPFFSAHVTIQLPPGPTQFTVQ | |

| Controls | CD2 (245-351) | NDEEEETRAHRVTAEERGRKPGQPIASTPQNPATSPQHPP | |
| | | | |
| | lamin C | devoid of PRS | |

**Table II**

Candidates from the proteome analysis

Selected candidates obtained from the proteome SPOT analysis of proline-rich peptides comprising the identified recognition motifs are shown. For the different GYF domains, candidates were selected from the respective proteome according to binding strength and number of motifs present. Portions of the proteins that comprise the putative binding sites are shown, and motifs fitting to the indicated relaxed or strict consensus motifs are depicted as underlined or bold and underlined letters, respectively. The indicated fragments of candidates for PERQ2 were tested for binding in the yeast two-hybrid system with CD2 and lamin C as controls. Numbers in parentheses below the protein names refer to the position of the cloned region within the full-length protein. The two candidates MSL5 and EAP1 were subjected to pull-down analysis with SYH1 and SMY2.
and 5 proline-rich peptide MSL5L1 (Fig. 4, yeast GYF domains and both MSL5 and EAP1 (Fig. 4, lane 3–6 contain GST-GYF-loaded beads. Lane 1 contains lysate, lane 2 contains GST-loaded beads, and lanes 3–6 contain GST-GYF-loaded beads.

Pull-down experiments with the two GST-tagged yeast GYF domains and lysates of cells overexpressing HA-tagged MSL5 or EAP1. The experiment confirmed the interaction between the yeast GYF domains and both MSL5 and EAP1 (Fig. 4, lanes 3 and 5), and the interactions could be competed out with the proline-rich peptide MSL5L1 (Fig. 4, lanes 4 and 6, and see Table IV). This establishes that the PRSs of MSL5 and EAP1 mediate the direct interaction with SYM2 and SYH1. Interestingly MSL5 has been reported as an interaction partner for both SMY2 and SYH1 in a yeast two-hybrid screen (see the SGD at www.yeastgenome.org), further supporting our findings. EAP1 was described as a protein binding to the cap-binding protein eIF4E (25), and subsequently a role in translational attenuation in response to vesicular transport defects has been reported (29). Because SMY2 was cloned as a suppressor of the myo2-66 mutation of type V myosin (14), a motor protein playing a decisive role in vesicular transport, this interaction establishes a particularly interesting link between transport processes and splicing or translational control. From a biochemical point of view, the binding of identical target proteins by SYM2- and SYH1-GYF is not surprising because the two GYF domains are very homologous. The biological implications of this redundancy are not clear, but the identified interaction partners hint to a role of SYM2 and/or SYH1 in processes that are coupled to the transport of splicing factors and translational attenuators.

Yeast Two-hybrid Analysis: Novel Interaction Partners of the PERQ2- and GYN4-GYF Domains—For the GYF domain of the human PERQ2 protein we chose the three nuclear proteins SmB, NPWBp, and SWAN for further investigations by yeast two-hybrid analysis. As can be seen from Fig. 5, an interaction with PERQ2-GYF could be established for all three proteins, and the results show that the proposed interaction can take place under more physiological conditions. In addition to this knowledge-driven identification of binding partners, we performed a yeast two-hybrid screen for the human PERQ2-GYF and the A. thaliana GYN4-GYF domain. Unbiased screening of a human lung cDNA library for interaction partners of the PERQ2-GYF domain resulted in the identification of a number of putative interactors, including the small nuclear ribonucleoproteins SmB/B’, the closely related snRNP-associated protein N (snRNP-N, SmN), U1 snRNP protein C, and the SWAN protein (Fig. 6A and Table III). These findings are in good agreement with the interaction partners that were proposed by the knowledge-based strategy (Figs. 3 and 5) described elsewhere (28) because all of the above mentioned proteins contain the phage display-derived recognition motif (Table I). SmB/B’ was previously identified as a binding partner of CD2BP2 (13), and it will be interesting to further investigate whether SmB/B’ and SmN also attract the PERQ2-GYF domain in vivo. Such an interaction would imply that the two GYF domain-containing proteins converge in the recognition of proteins of the SmB/SmN family. Because U1 snRNP protein C and SWAN also represent nuclear proteins and because NPWB colocalizes with splicing factors (30), a function of PERQ2-GYF in splicing or splicing-associated processes is well conceivable. Initial localization experiments with GFP-PERQ2-GYF support this notion, showing a predominant nuclear localization (data not shown). However, the mouse homolog of PERQ2, GIGYF2, was identified as a binding partner of the cytosolic adaptor protein Grb10 by a yeast two-hybrid screen. Because human Grb10, in contrast to mouse Grb10, is devoid of a PPGΦ recognition motif, the function of the human PERQ2 and the mouse GIGYF2 protein may well be different.

For the GYN4-GYF domain of A. thaliana, the yeast two-hybrid screen resulted in the selection of At2g28540.1, At3g45630.1, At5g60170.1, and At5g65410.1 (Fig. 6B and Table III). All of these proteins contained the PPGF motif that was also selected by the phage display screen (Table I). Interestingly the first three proteins obtained from the yeast two-hybrid anal-
Fig. 6. Analysis of yeast two-hybrid screening results. Clones of interaction partners for PERQ2-GYF (A) or GYN4-GYF (B) selected by yeast two-hybrid screens were retransformed with different bait constructs and spread on non-selective (−Leu, −Trp) or high stringency medium (−Ade, −His, −Leu, −Trp). Plates were analyzed after 4 days of incubation at 30 °C. The bait constructs containing the GYF domains of PERQ2, GYN4, or GYN4 containing the internal PRS are indicated by PERQ2, GYN4, and GYN4-PR, respectively. Empty bait and prey vectors as well as the bait vector containing the unrelated protein lamin C (Lam) serve as negative control. Names or accession numbers of interaction partners are shown on the left side of the panels.

NMR Analysis of the SMY2/MSL5 Interaction—To obtain a detailed molecular picture of the interaction between a GYF domain of the SMY2 subfamily of GYF domains and a PRS, backbone assignments of NMR heteronuclear triple resonance spectra of the 13C/15N isotope-labeled SMY2-GYF domain were performed. This allowed us to map the binding site for the MSL5S1 peptide (Table IV). The overlay of spectra of isolated SMY2-GYF (red) with SMY2-GYF domain in the presence of equimolar amounts (green) and a 10-fold excess (blue) of MSL5S1 peptide is shown in Fig. 7A. The deduced NH chemical shift changes of residues of the SMY2- and CD2BP2-GYF domain upon addition of MSL5S1 and CD2S peptides, respectively, are depicted in Fig. 7B. Most of the residues that show large chemical shift changes are part of the characteristic bulge-helix-bulge motif within the N-terminal half of the domains and are indicated as green bars. The three residues Leu-289, Gln-290, and Ile-291 in the C-terminal half of SMY2-GYF display significant chemical shift changes, but the homologous residues in CD2BP2-GYF only show small changes (Fig. 7B, blue bars). Because large chemical shift changes are indicative of residues that are close in space to the ligand, the data can be qualitatively used to compare the binding epitopes of SMY2 (or homologous SMY2-type GYF domains) and CD2B2-GYF. The structure of one SMY2-type GYF domains was recently deposited in the Protein Data Bank (GYF domain from the A. thaliana protein Q9FT92, Protein Data Bank code 1WH2), and based on sequence comparisons (Fig. 1) we plotted the results for the SMY2-GYF domain (Fig. 7B) onto the surface of the GYF domain of Q9FT92 (Fig. 7C). Residues forming the novel binding epitope, as depicted in blue in Fig. 7, B and C, are partially solvent-exposed in the case of SMY2-like GYF domains but not in CD2BP2-GYF, which might explain the differences seen in the binding preferences for the two types of GYF domains (see “Discussion”).

Intramolecular Recognition of Proline-rich Motifs by GYF Domains—The two proteins GYN4 and SYH1 display internal sequence motifs that match the consensus of the respective GYF domain. Fig. 3 shows that the two GYF domains recognize these motifs when synthesized on a membrane. The presence of the internal PRS in the GYF domain construct PR-SYH1 had no influence on phage display (Table I) or pull-down experiments (Fig. 4), whereas the KSGPPPGFTG sequence of GYN4 interfered with binding in phage display (data not shown) and yeast two-hybrid experiments (Fig. 6). In contrast, the shorter GYN4-PR construct allowed us to define a recognition motif (Table I) and to identify interaction partners (Fig. 6). Misfolding of the domain due to the proline-rich extension could be ruled out because the GST fusion protein of GYN4-PR was able to bind to peptides synthesized on a membrane where concentrations are high enough to compete out the intramolecular interaction (data not shown).
Taken together, these findings suggest that interaction in cis between the GYN4-GYF domain and the internal PRS masks its binding site, whereas for the SYH1 protein the intramolecular sequence does not influence the interaction competence of its GYF domain. Further support for intramolecular encounter between GYN4-GYF and its internal PRS is derived from fluorescence titration experiments (Fig. 8). For the GYF domain-only construct (residues 546–604 of GYN4), a $K_{D}$ value of $1101$ nM for binding to the GYN4 peptide was obtained (Table IV). The presence of the internal recognition motif within the construct GYN4-PR decreased the affinity to the titrated GYN4 peptide by a factor of 8. In addition, the spectrum of the GYN4-PR and the shorter construct in its bound form display similar centroid positions (Fig. 8). The most likely explanation for the different behavior of GYN4-PR is the intramolecular binding of a large fraction of the GYF domain to the internal PRS. Additional binding is only observed at a large excess of externally added peptide. The assumption of an intramolecular interaction is substantiated by studies on artificial CD2BP2-GYF single chain constructs with N- or C-terminally linked interaction sites (Fig. 1) that were used to obtain NMR restraints for structural investigations of the CD2BP2/CD2 interaction (27). Only C-terminal linkage resulted in the formation of a soluble, intramolecular interaction and supported the view that the CD2 peptide SHRPPPPGHRV binds exclusively in one orientation to the CD2BP2-GYF domain (27). Similarly the linker between the GYF domain and the C-terminal interaction site within the GYN4 protein is of sufficient length to enable the autoinhibited conformation, whereas the attachment of a PRS directly N-terminal to the GYF domain, as it is
Different GYF domains were tested with various ligands. \( K_D \) values were either determined by using NMR chemical shift changes (NMR) or the shift of the fluorescence emission spectrum (Fluorescence), defined by the change in the position of the centroid, of the respective GYF domain upon ligand addition. \( K_D \) values were calculated assuming a two-state binding model (see also Supplemental Fig. 1).

| Table IV | Titration data of different GYF domains |
|----------|----------------------------------------|
| GYF domain | Ligand | Sequence | \( K_D \) (\( \mu M \)) |
| PERQ2    | CD2S   | SHRPPPGLGKV             | 302 |
|          | CD2L   | HPPPPPGRHSAQPSPHRPPPPGHRQH | 162 |
|          | SmB2   | GTPMGPPPGRPPPPMGLL     | -20 |
| snRNP    | AKNA   | VSNKPPGQAS              | -33 |
| SMY2     | MSL5S1 | SSIAPPPLGLSG            | 43  |
|          | MSL5L1 | SIAPPLGPLSGPGFSN        | 25  |
|          | MSL5L2 | DINKPTIPGPLQIPPGAL      | 47  |
| SYH1     | MSL5S1 | SSIAPPPLGLSG            | 12  |
|          | MSL5L1 | SIAPPLGPLSGPGFSN        | 5   |
|          | MSL5L2 | DINKPTIPGPLQIPPGAL      | 27  |
| GYN4     | Internal | AKSGPPPGFTGLAKQN       | 4   |
| GYN4-PR  | Internal | AKSGPPPGFTAKQN         | 31  |

present within the SYH1 protein, excludes such an encounter. The observation of intramolecular masking is likely to be of functional relevance and might prevent unwanted low affinity interactions from taking place. Co-compartmentalization of a protein containing a high affinity binding site can be envisaged to outcompete the intramolecular interaction under appropriate conditions. In addition, post-translational modification of residues close to the binding site, as for example phosphorylation or methylation, might act as regulatory switches for the GYF domain-mediated interactions.

**DISCUSSION**

In this first comprehensive study of GYF domains of the SMY2 subfamily, it is shown that PRS recognition of these domains converges on the four-amino acid recognition motif PPG\( \Phi \). Because all of the analyzed GYF domains contain the conserved signature \((Y/F/W)(Y/F)\times_{6-11}PFX_{3}(M/V/I)\times_{2-8}W_{3-4}^{\text{GYF}}\), it is very likely that the hydrophobic cavity formed mainly by the aromatic side chains of this signature constitutes the main side of interaction with the PPG\( \Phi \) moiety. Albeit there is no three-dimensional complex structure available for a GYF domain of the large SMY2 subfamily, the high conservation of the peptide-binding signature of GYF domains allows us to put our new results into a structural context and to compare the recognition code defined in this study to the well characterized CD2BP2-GYF-CD2 peptide complex (see below). Defining the recognition rules sets the stage for further functional investigations of GYF domain-containing proteins. While our results reveal a considerable number of possible protein targets for GYF domains, yeast two-hybrid analysis and pull-down experiments indicate that the selected interactions could take place under more physiological conditions.

**PERQ2**—It was of some surprise to find the core splicing protein SmB/B’ as the most prominent interaction partner obtained from the yeast two-hybrid analysis of the PERQ2-GYF domain. SmB/B’ contains numerous proline-rich motifs in its C-terminal tail that represent several binding sites for GYF domains. It was previously shown that CD2BP2 and SmB/B’ interact and colocalize within the same nuclear compartment and that the PPPGMR motifs within the SmB/B’ tail represent the target sites for CD2BP2-GYF (2, 13). According to our data, PERQ2-GYF can also bind to these PPPGMR motifs. It is therefore plausible that CD2BP2 and PERQ2 can simultaneously interact with SmB/B’ under conditions where SmB/B’ is not limiting. Alternatively CD2BP2 could be the major interaction partner for SmB/B’ in most cells, whereas PERQ2 could preferentially bind to SmN, the brain-specific variant of SmB/B’. Because the two human proteins CD2BP2 and PERQ2 display no sequence homology in regions outside the GYF domain, functional redundancy of the two proteins is unlikely. It is rather possible that the two proteins utilize the SmB/B’ recruitment in a different functional context. The third human protein containing a GYF domain is PERQ1. Its GYF domain shows a high degree of sequence homology to PERQ2-GYF and is therefore expected to exhibit very similar binding specificities. Competition for the same interaction partners, as for example SmB/B’, is a possible scenario for PERQ1 and PERQ2 proteins, and the presence of homologous regions in the two proteins suggests similar but not necessarily identical functions. The role of SmB/B’ as part of the seven-protein core of all snRNP complexes is firmly established and supports its role as an important structural component of the spliceosomal machinery (32). A recent report identified SmB/B’ at the sites of initial adhesion of certain primary cells, so-called spreading initiation centers (33). Strikingly no mRNA was found at these sites, opening the possibility that SmB/B’ and also other Sm proteins play an important role in processes other than splicing. It is also known that the snRNPs undergo a complex maturation process in Homo...
**Fig. 7.** SMY2-GYF epitope mapping and comparison with the CD2BP2-GYF binding epitope. 

A, overlay of $^{15}$N-$^1$H correlation spectra of the isolated SMY2-GYF domain (red), the SMY2-GYF domain upon equimolar addition of the MSL5S1 peptide (green), and the SMY2-GYF domain in the presence of a 10-fold excess of MSL5S1 peptide (blue). NH resonances of residues with weighted geometrical differences of
that extend the binding epitope in SMY2-GYF. The NMR titration experiments were performed at 297 K with 0.2 mM sample of the 15N-labeled residues comprising the common binding epitope of CD2BP2-GYF and SMY2-GYF are depicted as domains (Fig. 1). The known orientation of a CD2 class ligand bound to CD2BP2-GYF (Fig. 2) and CD2BP2- or SMY2-GYF domain in 50 mM sodium phosphate buffer, pH 6.3, and PBS, pH 7.3, respectively. The SMY2- and CD2BP2-GYF domains represent two subfamilies of GYF domains. All the GYF domains investigated here belong to the SMY2 type. This large subfamily is characterized by a short \( \beta_1-\beta_2 \) loop and an aspartate as the last residue of the \( \beta_1 \) strand (position 8), whereas the CD2BP2-GYF domain contains a cryptophan (Trp-8) at this position. Trp-8 contributes to binding of the CD2 ligand SHRPPPPGHRV by directly contacting glycine 8 and arginine 10 of this peptide. At the same time, the bulky Trp-8 side chain of the domain shields the conserved Tyr-6 and Phe 34 residues from solvent. A recent NMR structure of the A. thaliana Q9FT92-GYF domain (Protein Data Bank code 1W2H) shows that Asp-8 allows a partial solvent exposure of the conserved aromatic residues Tyr-6 and Phe-34, which are now able to contribute to the hydrophobic surface that most likely represents the interaction site for the proline-rich ligand. Fig. 7C shows a comparison of the molecular surface of the CD2BP2-GYF domain and the Q9FT92-GYF domain from A. thaliana. The binding epitope for Q9FT92-GYF/ligand interactions is not known; however, alignment of the SMY2- and Q9FT92-GYF domain (Fig. 1) was used to color the suggested Q9FT92 epitope based on the binding of the CD2 class ligand bound to CD2BP2-GYF (left) and the suggested orientation of a ligand for the SMY2 domain (right) is depicted as a red line above the indicated epitopes. The binding epitopes are color coded as in B. Residues Lys-7, Pro-19, and Pro-33 in CD2BP2-GYF and the corresponding residues in Q9FT92 are also shown. EAP1 contains MSL5 C-terminal tail. MSL5 is the yeast branch point binding protein that is involved in spliceosome assembly (36). Global analysis of yeast protein localization suggests a predominant nuclear occurrence of MSL5, whereas SMY2 and SYH1 are mostly found in the cytoplasm (37). A candidate for a cytoplasmic interaction partner of SMY2 and SYH1 is EAP1 that could also be detected by pull-down experiments of cellular lysates (Fig. 4). EAP1 was originally found as a translational inhibitor protein that competes with eIF4G and p20 for interaction with the eIF4E cap-binding protein. In addition, EAP1 has a separate function in mediating genetic stability (38), and a recent report highlights the involvement of EAP1 in the attenuation of translation in cells with mutations in the secretory pathway (29). Further investigations should address the question whether the two GYF domain-containing proteins are involved in transport processes that depend on the cap structure of mRNA.

Comparing Different Classes of GYF Domains—The SMY2- and CD2BP2-GYF domains represent two subfamilies of GYF domains. All the GYF domains investigated here belong to the SMY2 type. This large subfamily is characterized by a short \( \beta_1-\beta_2 \) loop and an aspartate as the last residue of the \( \beta_1 \) strand (position 8), whereas the CD2BP2-GYF domain contains a cryptophan (Trp-8) at this position. Trp-8 contributes to binding of the CD2 ligand SHRPPPPGHRV by directly contacting glycine 8 and arginine 10 of this peptide. At the same time, the bulky Trp-8 side chain of the domain shields the conserved Tyr-6 and Phe 34 residues from solvent. A recent NMR structure of the A. thaliana Q9FT92-GYF domain (Protein Data Bank code 1W2H) shows that Asp-8 allows a partial solvent exposure of the conserved aromatic residues Tyr-6 and Phe-34, which are now able to contribute to the hydrophobic surface that most likely represents the interaction site for the proline-rich ligand. Fig. 7C shows a comparison of the molecular surface of the CD2BP2-GYF domain and the Q9FT92-GYF domain from A. thaliana. The binding epitope for Q9FT92-GYF/ligand interactions is not known; however, alignment of the SMY2- and Q9FT92-GYF domain (Fig. 1) was used to color the suggested Q9FT92 epitope based on the chemical shifts colored in green or blue in B are labeled according to residue type and number. B, combined chemical shift changes of CD2BP2-GYF and SYM2-GYF for their respective ligands, CD2S and MSL5S1. The weighted geometrical differences of the chemical shifts for each residue of the domains upon addition of a 10-fold excess of peptide are plotted against the corresponding sequence. Residues of both domains (black for CD2BP2-GYF and red for SYM2-GYF) have been aligned for a better comparison. The key residues that define the conserved GYF domain signature are depicted as white letters on a black background. Weighted geometrical differences of the chemical shifts of residues comprising the common binding epitope of CD2BP2-GYF and SYM2-GYF are depicted as green bars. Blue bars belong to residues that extend the binding epitope in SMY2-GYF. The NMR titration experiments were performed at 297 K with 0.2 mM sample of the 15N-labeled CD2BP2- or SMY2-GYF domain in 50 mM sodium phosphate buffer, pH 6.3, and PBS, pH 7.3, respectively. C, proposed model of ligand binding for the SYM2 subfamily in comparison to the known orientation of the CD2BP2-GYF ligand CD2. Because the structure of the SMY2-GYF domain is not known so far, the binding epitope of SMY2-GYF-MSL5S1 was plotted onto the structure of Q9FT92-GYF based on sequence homology (Fig. 1). The known orientation of a CD2 class ligand bound to CD2BP2-GYF (left) and the suggested orientation of a ligand for the SMY2 domain (right) is depicted as a red line above the indicated epitopes. The binding epitopes are color coded as in B. Residues Lys-7, Pro-19, and Pro-33 in CD2BP2-GYF and the corresponding residues in Q9FT92 are also depicted in green because they are part of the binding epitope. Prolines cannot be observed in 15N heteronuclear single quantum coherence spectra, whereas the backbone NH resonance of Lys-7 displayed line broadening and could therefore not be followed during the titration.
results for SMY2 (Fig. 7B). The substitution W8D in Q9FT92 extends the hydrophobic cleft in the A. thaliana GYF domain, probably allowing the ligand to bind along its main axis as indicated by the red line (Fig. 7C). This mode of binding is supported by the observation of significant chemical shift changes for residues 38–40 of the SMY2-GYF domain (indicated as a blue surface) when interacting with the MSL5S1 peptide. A reorganization of the binding pocket by the W8D replacement could explain the preference for the large hydrophobic residues Phe, Leu, and Ile that was found in the consensus motifs of all GYF domains analyzed in this study. On the other hand, our results with CD2BP2-GYF show that the PPGW motif represents an optimal binding motif (39) that is rarely found in the phage display sequences of the SMY2 subfamily of GYF domains. We propose that the tryptophan of the ligand forms a “stacked” aromatic interaction with the Trp-8 of CD2BP2-GYF at the surface of the domain, whereas it is too bulky to be optimally placed in the deep binding pocket of the SMY2-GYF subfamily. In conclusion, the data suggest a common binding mode for PPGΦ-containing ligands of the SMY2 subfamily of GYF domains. For the CD2BP2-GYF domain, two classes of ligands were identified, the PPGW class and the PPGX(R/K) class (CD2 class), that require a positively charged amino acid within the flanking regions of the PPG core. The PPGW class for the CD2BP2-GYF domain and the PPGΦ class of ligands are charged-independent and dominated by hydrophobic interactions. However, although stacked aromatic interactions at the surface of the domain probably characterize the CD2BP2-GYF/ligand interaction, the hydrophobic residue of the PPGΦ motif is expected to insert into the extended hydrophobic cleft of the SMY2-GYF domain family (Fig. 7C).

Acknowledgments — We are thankful to Gianni Cesareni for providing the pC89 vector and the randomized X8 phage library. We are also grateful to Angelikia Ehrlich for peptide membrane synthesis, M. Beyermann for peptide synthesis, and Sebastian Modersohn for sharing the SpecWin software. Furthermore we thank Nahum Sonenberg and Michael Rosbash for providing EAP1 and MSL5 clones, respectively. We acknowledge Benjamin Horwitz, Hong Ma, and the Arabidopsis Biological Resource Center for providing A. thaliana Horwitz and Ma two-hybrid cdNA library and Takahiro Nagase and Satoshi Tabata from the Kazusa DNA Research Institute for providing the cdNA clone HU03496 and the genomic P1 clone MBD2, respectively. We also thank Christine Lang for the yeast strain BY4741.

* This work was supported by Deutsche Forschungsgemeinschaft Grant FR 1325/2-1 (to C. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked ‘advertisement’ in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

‡ To whom correspondence should be addressed. Tel.: 49-030-94793-181; Fax: 49-030-94793-189; E-mail: freund@fmp-berlin.de.
22. Grzesiek, S., and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. J. Magn. Reson. 99, 201–207
23. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) 3-Dimensional triple-resonance NMR-spectroscopy of isotopically enriched proteins. J. Magn. Reson. 89, 496–514
24. Fan, H. Y., Hu, Y., Tudor, M., and Ma, H. (1997) Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J. 12, 999–1010
25. Cosentino, G. P., Schmelzle, T., Haghighat, A., Helliwell, S. B., Hall, M. N., and Sonenberg, N. (2000) Eap1p, a novel eukaryotic translation initiation factor 4E-associated protein in Saccharomyces cerevisiae. Mol. Cell. Biol. 20, 4604–4613
26. Abovich, N., and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89, 403–412
27. Freund, C., Kuhne, R., Park, S., Thiemke, K., Reinherz, E. L., and Wagner, G. (2003) Structural investigations of a GYF domain covalently linked to a proline-rich peptide. J. Biomol. NMR 27, 143–149
28. Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R., and Cesareni, G. (2004) Protein interaction networks by proteome peptide scanning. PLoS Biol. 2, 94–103
29. Deloche, O., de la Cruz, J., Kressler, D., Doere, M., and Linder, P. (2004) A membrane transport defect leads to a rapid attenuation of translation initiation in Saccharomyces cerevisiae. Mol. Cell 13, 357–366
30. Craggs, G., Finan, P. M., Lawson, D., Wingfield, J., Perera, T., Gadher, S., Totty, N. F., and Kellett, D. (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
31. Albert, T. K., Hanzawa, H., Legtenberg, Y. I., de Ruwe, M. J., van den Heuvel, F. A., Collart, M. A., Boelens, R., and Timmers, H. T. (2002) Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex. EMBO J. 21, 355–364
32. Will, C. L., and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. Curr. Opin. Cell Biol. 13, 290–301
33. de Hoog, C. L., Foster, L. J., and Mann, M. (2004) RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. Cell 117, 649–662
34. Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P., and Richard, S. (2000) Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. J. Biol. Chem. 275, 16030–16036
35. Bedfor, M. T., Reed, R., and 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. U. S. A. 95, 10602–10607
36. Rutz, B., and Seraphin, B. (2000) A dual role for BBP/ScSF1 in nuclear pre-mRNA retention and splicing. EMBO J. 19, 1873–1886
37. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O'Shea, E. K. (2003) Global analysis of protein localization in budding yeast. Nature 425, 686–691
38. Chial, H. J., Stemm-Wolf, A. J., McBratney, S., and Winey, M. (2000) Yeast Eap1p, an eIF4E-associated protein, has a separate function involving genetic stability. Curr. Biol. 10, 1519–1522
39. Kofler, M. M., Motzny, K., Beyermann, M., and Freund, C. (2005) Novel interaction partners of the CD2BP2-GYF domain. J. Biol. Chem., in press
40. Goddard, T. D., and Kneller, D. G. (2000) SPARKY 3, Version 3.1, University of California, San Francisco, CA