

Zusammenfassung

Die GYF-Domäne ist ein neues Mitglied der stetig wachsenden Familie von Adapter-Domänen, die prolinreiche Sequenzen binden (PRD). Prolinreiche Sequenzen kommen häufig in Eukaryoten vor und werden oft von PRD mit nur mittlerer Affinität und vielfach geringer Spezifität gebunden. Daher ist es schwierig die Bindungspartner von PRD und die biologischen Prozesse, an denen sie beteiligt sind, eindeutig zu bestimmen. Deutliche Fortschritte sind auf diesem Gebiet bereits für andere PRD erzielt worden, wie beispielsweise den SH3- und WW-Domänen. Im Gegensatz dazu beschränkten sich die Kenntnisse über GYF-Domänen auf einen einzigen Wechselwirkungspartner, eine systematische Beschreibung der Bindungseigenschaften und Interaktionspartner fehlte.

In der vorliegenden Arbeit habe ich die Erkennungsspezifitäten von fünf GYF-Domänen aus Menschen, der Bäckerhefe (*Saccharomyces cerevisiae*) und der Acker-Schmalwand (*Arabidopsis thaliana*) charakterisiert. Die Auswahl beinhaltet Vertreter der zwei größten Unterfamilien, den so genannten CD2BP2-ähnlichen und den SMY2-ähnlichen GYF-Domänen. Mit Hilfe von Phage-Display wurden möglichen Bindungspartnern aus einer komplexen Peptid-Bibliothek selektiert. SPOT Substitutionsanalysen erlaubten die nähere Bestimmung von Schlüsselresten innerhalb der erhaltenen Peptidliganden. Alle GYF-Domänen erkennen das Motiv PPG Φ , wobei Φ für hydrophobe Aminosäuren steht. Unterschiede bezüglich der Spezifitäten einzelner Domänen sind auf die letzte Position der Bindungssignatur beschränkt. Nur die GYF-Domäne von CD2BP2 erkennt zusätzlich das Bindungsmotiv PPGx(R/K), das neben der prolinreichen Signatur PPG eine positiv geladene Aminosäure enthält.

Eine Suche nach den Erkennungsmotiven der GYF-Domänen in Proteindatenbanken identifizierte potentielle Bindungsstellen in Proteinen. Peptide, die diese potentiellen Bindungsstellen enthalten, wurden auf Membranen synthetisiert und auf ihre Bindungseigenschaften hin getestet. Hefe Two-Hybrid Experimente und zusätzliche biochemische Untersuchungen ergänzten die peptidbasierten Methoden und bestätigten vorhergesagte Bindungspartner. Die möglichen Wechselwirkungspartner weisen auf eine Funktion der GYF-Domänen unter anderem beim Aufbau des Spleißosoms, bei der Regulation des Translationsbeginns und beim mRNA Transport hin. Die Kartierung der Bindungstasche durch NMR-Experimente erlaubte, die Bindungsspezifitäten auf struktureller Ebene zu erklären. Dies führte zur Formulierung dreier Bindungsmodelle, je eines für die zwei Ligandenklassen der CD2BP2 GYF-Domäne, und eines für die Liganden der SMY2-artigen GYF-Domänen. Intramolekulare Ligandenbindung wurde als möglicher Regulationsmechanismus der GYF-Domäne des *Arabidopsis thaliana* Proteins GYN4 identifiziert.

Zusammengefasst beschreibt diese Arbeit die Bindungseigenschaften von GYF-Domänen aus unterschiedlichen Spezies und Unterklassen. Die so gewonnenen Ergebnisse stellen die Grundlage für weitere Untersuchungen der GYF-Domänen enthaltenden Proteine bezüglich ihrer Funktionen beim Spleißen, dem Transport oder der Translation von mRNA dar.

Summary

The GYF domain is a new member of the growing class of proline-rich ligand binding adaptor domains (PRD). Proline-rich sequences (PRS) are frequently found in the eukaryotic kingdom and they often represent protein interaction sites that are recognized by PRD with moderate affinities and limited specificities. Therefore, PRD have the potential to associate with many different PRS and the unambiguous assignment of their binding partners and of the biological processes they mediate is a challenging molecular proteomics task. Significant progress in this area has already been made for other adaptor domains of the PRD superfamily, such as SH3 and WW domains. In contrast, the knowledge about GYF domains was limited to a single prototypic interaction and, therefore, lacked a systematic description of their binding properties and interaction partners.

In the present work, I characterized the recognition specificities of a set of five GYF domains from man, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*. The selection comprised representatives of the two major subfamilies, the so called CD2BP2-like and SMY2-like GYF domains. Phage-display was used to screen a large peptide sequence space for potential binders. SPOT substitution analysis confirmed key residues within the interacting peptides. All GYF domains converge on the binding sequence PPG Φ , with Φ standing for hydrophobic residues. Differences in the specificities of individual domains are limited to the last position of the binding signature. Only the GYF domain of CD2BP2 has a second type of recognition motif, PPGx(R/K), which depends on charge complementarity in addition to the recognition of the proline-rich signature PPG.

Potential natural interaction sites in proteins were identified by database searches. Peptides encompassing these recognition motifs were synthesized on membranes and tested for binding. Yeast two-hybrid screens for two of the five domains and additional biochemical studies complemented the peptide-based methods and confirmed binding of candidate proteins. Likely target proteins for GYF domains hint at a function in the assembly of the spliceosome, regulation of translation initiation or mRNA transport.

A structural understanding of the interaction specificities was gained from NMR epitope mapping experiments. Three binding models are proposed, one for each of the two ligand classes of CD2BP2-GYF and one for the SMY2-type-mediated interactions. Inhibition by intramolecular ligand binding was found to be a possible regulation mechanism of the GYF domain from the *Arabidopsis thaliana* protein GYN4.

Taken together, the work describes the binding properties of GYF domains and sets the basis for functional studies of the respective proteins in the context of mRNA splicing, transport or translation.