



COMMUNICATION

Structural Basis for the Bifunctionality of the U5 snRNP 52K Protein (CD2BP2)

Tine K. Nielsen¹, Sunbin Liu², Reinhard Lührmann² and Ralf Ficner^{1*}

¹*Abteilung für Molekulare Strukturbiologie, Institut für Mikrobiologie und Genetik Georg-August-Universität Göttingen, Justus-von-Liebig-Weg 11, 37077 Göttingen Germany*

²*Max-Planck-Institut für Biophysikalische Chemie Abteilung für Zelluläre Biochemie, Am Fassberg 11 37077 Göttingen, Germany*

The bifunctional protein U5-52K is associated with the spliceosomal 20 S U5 snRNP, and it also plays a role in immune response as CD2 receptor binding protein 2 (CD2BP2). U5-52K binds to the CD2 receptor *via* its GYF-domain specifically recognizing a proline-rich motif on the cytoplasmic surface of the receptor. The GYF-domain is also mediating the interaction of the proteins U5-52K and U5-15K within the spliceosomal U5 snRNP. Here we report the crystal structure of the complex of GYF-domain and U5-15K protein revealing the structural basis for the bifunctionality of the U5-52K protein. The complex structure unveils novel interaction sites on both proteins, as neither the polyproline-binding site of the GYF-domain nor the common ligand-binding cleft of thioredoxin-like proteins, to which U5-15K belongs, are involved in the interaction of U5-15K and U5-52K.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: U5 snRNP; pre-mRNA splicing; crystal structure; bifunctionality; CD2-receptor binding protein

*Corresponding author

The “one protein - one structure - one function” paradigm has become outdated due to an increasing number of bi- and multi-functional proteins participating in many diverse cellular processes.^{1,2} Bifunctional proteins are also involved in the splicing of pre-mRNA in the nucleus of eukaryotic cells as recently shown for the U5 snRNP-specific 52K protein, which is identical to the CD2 receptor binding protein 2 (CD2BP2).³ The U5 snRNP is a major component of the spliceosome generally formed by the five snRNPs U1, U2, U4, U5, U6 and several non-snRNP proteins.^{4,5} The human 20 S U5 snRNP consists of the uracil-rich U5 snRNA, seven Sm proteins that are common to all UsnRNPs with the exception of U6, and the U5 particle-specific proteins 15K, 40K, 52K, 100K, 102K, 110K, 116K, 200K and 220K,^{6–8} which mediate important steps during assembly, maturation and dissociation of the spliceosome.⁵ The U5 snRNP assembles with the U4/U6 snRNP and the resulting U4/U6.U5 tri-snRNP joins U1 and U2 snRNPs and other proteins forming the pre-spliceosome leading to the active

spliceosome. During these assembly processes several proteins having fulfilled their function are leaving the complex.⁹ The U5-52K protein belongs to this group, as it is thought to dissociate from the U5 snRNP during assembly of the U4/U6.U5 tri-snRNP. Purified U5 snRNP from *Saccharomyces cerevisiae* also contains the U5-52K protein (SNU40/Lin1p), suggesting a conserved role in pre-mRNA splicing.¹⁰ U5-52K associates with the U5 snRNP by binding to the U5-15K protein,³ an evolutionarily highly conserved component of the U5 and U4/U6.U5 snRNPs.¹¹ This interaction is mediated by the C-terminal domain of U5-52K,³ which is also required for the interaction of U5-52K/CD2BP2 with the CD2 receptor.^{12,13}

The CD2 receptor is located on the surface of T cells, thymocytes and natural killer cells, and interacts with CD2BP2 through two proline-rich motifs located in the cytoplasmic tail of CD2.¹² Overexpression of a CD2-binding U5-52kD fragment significantly enhances the production of interleukin 2, indicating a functional role of U5-52K/CD2BP2 in immune response reactions.¹² The NMR structure of the C-terminal 86 amino acid residues of CD2BP2, which are capable of binding to CD2, revealed a novel polyproline interacting domain, denoted as the GYF-domain.¹³ The structure of the GYF-domain with a bound peptide, that contains the cognate proline-rich PPPPGHR sequence present in CD2, has

Present address: T. K. Nielsen, Ontario Cancer institute, TMdT, 101 College Street, Rm. 4-601, Toronto, Ontario, Canada M5G 1L7.

E-mail address of the corresponding author: rficner@gwdg.de

also been determined by NMR,¹⁴ revealing a novel binding mode for a polyproline motif. In contrast to the SH3 domain, which has two hydrophobic binding pockets, the GYF-domain contains one deep binding pocket.

The GYF-domain of U5-52K/CD2BP2 was shown also to mediate the interaction of the U5-52K and U5-15K proteins within the U5 snRNP. However, U5-15K lacks the cognate proline-rich motif, raising the question about the molecular basis for this specific protein–protein interaction. The previously determined crystal structure of the U5-15K revealed a thioredoxin-like fold that is characterized by a four-stranded β -sheet consisting of pairs of parallel and antiparallel strands flanked by three α -helices.¹¹ Compared to thioredoxin the U5-15K contains additional 37 residues, of which 22 residues extending the C terminus were proposed to form a putative interaction site.^{11,15}

In order to understand the molecular details and specificity of the interaction between the U5-15K and the U5-52K proteins, we have determined the crystal structure of the complex of full-length U5-15K and the GYF-domain of U5-52K revealing novel interaction sites for both proteins.

Complex preparation and structure determination

The human U5-15K protein and the GYF-domain of the human U5-52K protein were prepared separately and the binary protein complex was obtained by *in vitro* reconstitution. Gel-filtration chromatography of this complex revealed an apparent molecular mass of 25 kDa, indicating a 1:1 molar

ratio of both proteins in the complex (data not shown). Crystals of the purified complex were obtained at a temperature of 4 °C. The crystals belong to space group C2 and contain three heterodimeric complexes in the asymmetric unit. The crystallographic phase problem was solved by means of molecular replacement and the crystal structure was refined to a resolution of 2.35 Å with a final *R*-free value of 26.2% (Table 1). The structures of the U5-15K and GYF-domain are well defined in the electron density map except for some flexible regions in the C and/or N termini. The refined model of the U5-15K present in the complex consists of amino acid residues 3–137. The structure of the GYF-domain of U5-52K (residues 256–341) lacks the N-terminal 24 residues due to conformational flexibility. The three copies of each of both proteins present in the asymmetric unit are almost identical. The rmsd values of corresponding superimposed main chain atoms are small (<0.6 Å), and the few

Notes to Table 1:

Recombinant human U5-15K protein and the GYF-domain of U5-52K were purified as described.^{3,11} The His-tag of the GYF-domain was cleaved by thrombin at 4 °C overnight. The proteins were mixed in equimolar ratio and the complex was allowed to form overnight at 4 °C. The complex was purified by gel filtration chromatography (Superdex75-26/60, Amersham Biosciences) using buffer A (20 mM Tris-HCl (pH 7.8), 120 mM NaCl, 2 mM DTT, 1 mM EDTA). The complex crystallized as plate bundles in 14% (w/v) PEG2000 MME, 100 mM Mes (pH 6.5), 100 mM calcium acetate, 4% (v/v) 1,4 butanediol at 4 °C.

X-ray diffraction data were collected at the EMBL beam line X13 (DESY, Hamburg) to a resolution of 2.35 Å. The data were processed with DENZO and SCALEPACK (HKL Research, USA). Crystals contain three U5-52K-GYF-15K complexes in the asymmetric unit, corresponding to a Matthews coefficient of 2.6 Å³/Da²² and a solvent content of 53%.

The structure was determined by molecular replacement using the program MOLREP from the CCP4 program package.²³ The U5-15K protein (PDB code: 1QGV) and the GYF-domain (PDB code: 1GYF) were used as search models. The most representative NMR structure of the GYF-domain (out of the 16 available) was calculated on the OLDERADO server.²⁴ The model was refined with CNS²⁵ using simulated annealing, individual *B*-factor refinement and positional refinement. The model was visualized and manually corrected with the program O.²⁶ The last refinement step was a TLS refinement in REFMAC.²³ The quality of the final model was validated with the program PROCHECK.²³ The total buried surface area upon complex formation was calculated with the program areaimol and the shape complementarity coefficient (sc) was calculated with SC; both programs are from the CCP4 program package.²³ Figures were made in PYMOL (DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, USA). The electrostatic potential was calculated by the Poisson-Boltzmann equation as implemented in the program DelPhi²⁷ and visualized with PYMOL.

^a $R_{\text{sym}} = 100 \cdot \sum |I - \langle I \rangle|^2 / \sum |I|^2$, where *I* is the observed intensity and $\langle I \rangle$ is the average intensity for multiple measurements.

^b $R\text{-factor} = 100 \cdot \sum |F_o - F_c| / \sum |F_o|$, where *F_o* and *F_c* are the observed and calculated structure factors, respectively. *R*-free is the cross-validation *R* factor calculated for 10% of the reflections omitted in the refinement process.

^c Calculated with PROCHECK.

Table 1. Data collection and refinement statistics

<i>Data collection</i>	
Resolution range (Å)	50–2.35
No. of unique reflections	30,397
No. of observed reflections	72,742
Completeness (%)	
Overall	92.4
Last shell (Å)	95.2 (2.41–2.35)
<i>R</i> _{sym} ^a (%)	
Overall	3.1
Last shell (Å)	18.4 (2.41–2.35)
<i>I</i> /σ(<i>I</i>)	
Overall	20.9
Last shell (Å)	2.4 (2.41–2.35)
<i>Refinement statistics</i>	
Resolution range (Å)	50–2.35
<i>R</i> -factor ^b (%)	
Overall	21.6
Last shell (Å)	28.0 (2.39–2.35)
<i>R</i> -free ^b (%)	
Overall	26.2
Last shell (Å)	31.0 (2.39–2.35)
Deviation from ideal geometry	
Bond lengths (Å)	0.008
Bond angles (°)	1.070
Ramachandran plot ^c	
Most favoured, additional, generously allowed (%)	92.8/6.6/0.6

observed differences concern side-chain conformations mostly caused by different crystal contacts.

The overall complex structure

The binary complex comprising U5-15K and the GYF-domain of U5-52K exhibits an elongated overall shape with the dimensions $50 \text{ \AA} \times 25 \text{ \AA} \times 25 \text{ \AA}$ (Figure 1). U5-15K consists of a five-stranded β -sheet surrounded by four α -helices, and the GYF-domain contains an antiparallel β -sheet flanked by three α -helices. The interaction surface of U5-15K resides within the 22 C-terminal residues Glu111–Ser132. These residues interact with residues Lys321–Thr341 at the C terminus of the GYF-domain (Figure 2). The description of residues mediating the interaction of the two proteins is restricted to those, which are consistent for all three complexes present in the asymmetric unit. The GYF-domain and the U5-15K protein are bound mainly by polar interactions (Figure 2(b)) including salt-bridges between Glu111, OE1 (U5-15K) and Gln328, NE2 (U5-52K), hydrogen bonds between side-chains of Asp114, OD2, Glu117, OE2 and Ser132, OG (U5-15K) and side-chains of Tyr330, OH, Lys321, NZ and Arg334, NH2 (U5-52K) and between main chains Ser132, O (U5-15K) and side-chains Arg334, NE (U5-52K). Finally, there is a hydrogen bond between the side-chain of Lys125, NZ (U5-15K) and the C-terminal oxygen of Thr341 (U5-52K). CD1 and CD2 from residue Leu339 (U5-52K) are forming hydrophobic contacts to CG2 of Val130 and CA of Gly122 (U5-15K).

The total surface areas buried upon complex formation for the three complexes in the asymmetric unit are 1328 \AA^2 , 1290 \AA^2 and 1310 \AA^2 , which gives an average value of 1309 \AA^2 . Inspection of the crystal packing reveals that, besides the contact within each complex, both proteins are mainly in contact with their respective symmetry mates, as U5-15K proteins pack against other U5-15K proteins and likewise the GYF-domains are in contact with other GYF-domains. There is only one crystal contact between a U5-15K protein and a GYF-domain molecule belonging to a different complex in the unit cell. The corresponding total buried surface area for this contact between U5-15K and GYF-domain is only 90 \AA^2 . The value of the total buried surface area together with the shape complementarity coefficient of the functional complex (average for the three complex molecules in the asymmetric unit was calculated to be 0.700) correlate well with values suggested by Lawrence *et al.*,¹⁶ which strongly supports that the complex shown in Figure 1 indeed is the biologically functional complex.

The U5-15K and GYF-domain structures

Superimposing the previously determined structures of the individual proteins onto the complex structure reveals that the overall conformation of each protein does not change upon complex formation (Figure 3), nevertheless there are some subtle changes. Residues 96–101 of U5-15K, which are missing in the crystal structure of U5-15K,¹¹ are

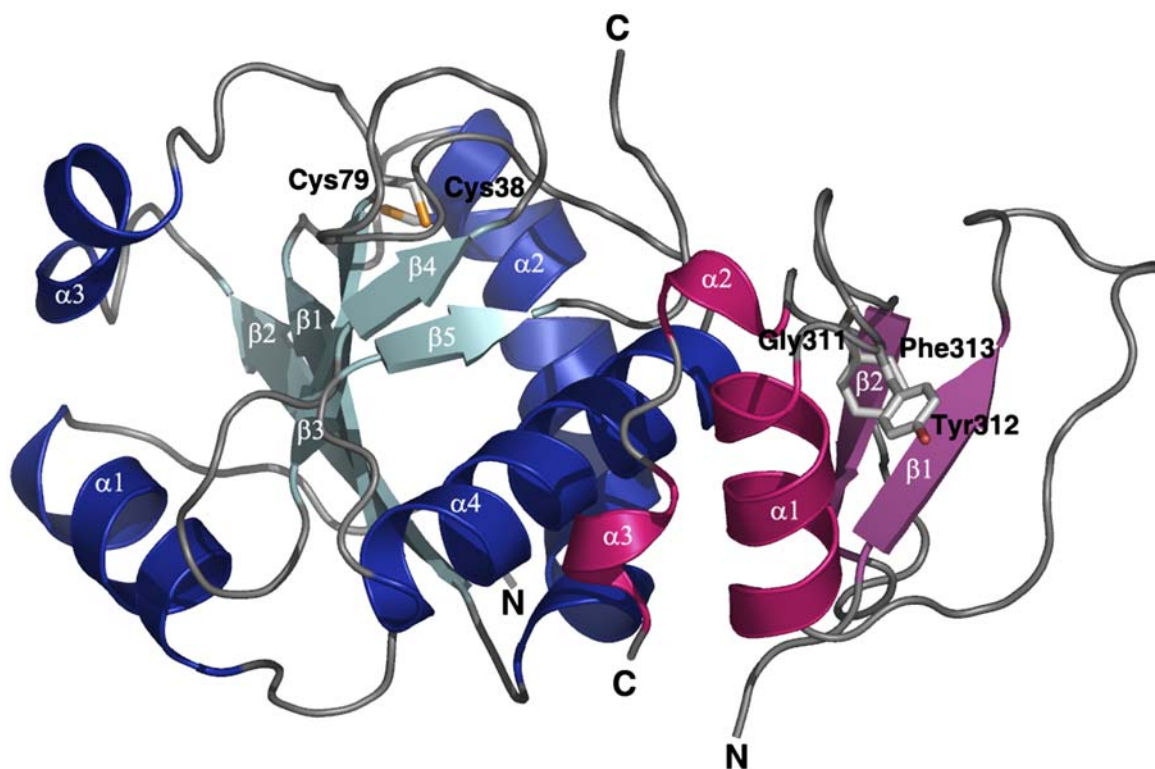


Figure 1. Cartoon representation of the U5-15K:U5-52K complex. U5-15K is shown in cyan and dark blue. Cys38 and Cys79 are shown in ball-and-stick mode. The GYF domain of U5-52K is shown in light magenta and pink. Gly311, Tyr312 and Phe313 are displayed in ball-and-stick mode.

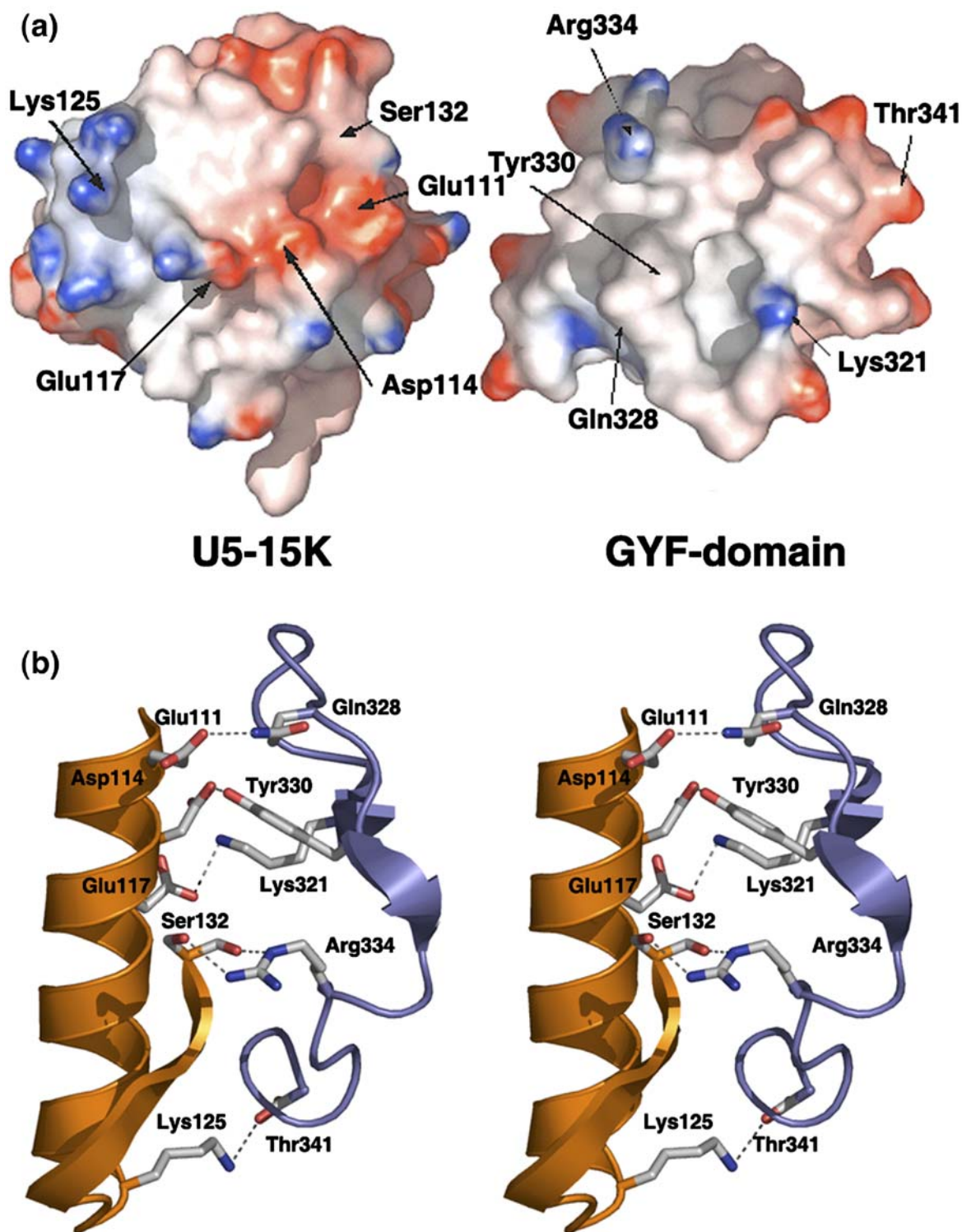


Figure 2. Interactions between the two proteins. (a) Surface representation colored according to the electrostatic potential. The surface complementarity between the two molecules is clearly seen in the open complex. The residues involved in the direct interaction are labeled. (b) A stereo picture of the polar interactions and the distances between the interacting atoms. U5-15K is shown in orange and U5-52K in blue. The interactions are found in all three molecules.

well defined in the electron density of the complex due to stabilization through crystal contacts (Figure 3(a)). Residues 72–77 of U5-15K have a different conformation in the complex, where the location of Tyr73–Leu75 is shifted in the direction

of the C terminus. Cys38 and Cys79, which are capable of forming a disulfide bridge, are in the reduced state. These cysteine residues have earlier been shown to have no functional impact for the U5-15K protein.¹¹

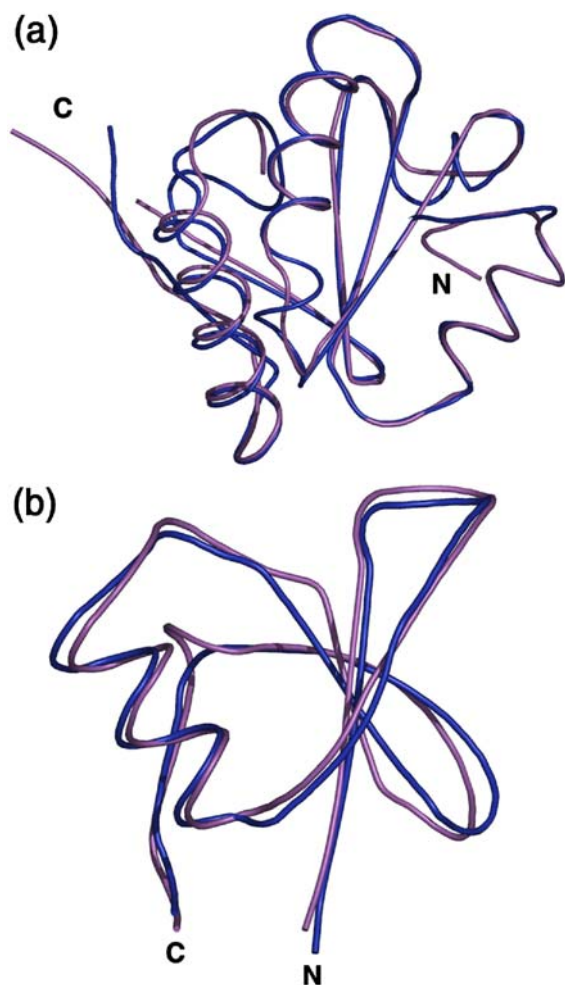


Figure 3. Stereo representations of the superposition of the two components of the complex structure with the uncomplexed structures. The structures of the two proteins in the complex are shown in blue, whereas the uncomplexed structures are depicted in magenta. (a) The superposition of complexed 15K and uncomplexed 15K (PDB code: 1QGV). The main differences are found in the termini of the protein, which is most likely due to flexibility. (b) The superposition of the 52K proteins from the complex and from the NMR structure (PDB code: 1GYF).

The structure of the GYF-domain of U5-52K in the complex is also almost identical to the structure of the uncomplexed GYF-domain (Figure 3(b)) when compared with the most representative NMR structure. It consists of an antiparallel β -sheet followed by an α -helix. As observed in the NMR structure of the GYF-domain alone,¹³ the crystal structure of the complex lacks the N-terminal 24 residues due to conformational flexibility and disorder in the crystal.

A novel protein–protein interaction site for a thioredoxin-like protein

The interaction between the U5-15K protein and the GYF-domain of U5-52K involves regions on

both protein surfaces not expected by the previous structural studies on both proteins. Indeed, the structure of the U5-15K-GYF-domain complex reveals a new interaction site for thioredoxin-like proteins. Among the few known complex structures of thioredoxin or thioredoxin-like proteins the interaction with other proteins mostly involves the cleft close to the redox-active disulfide bridge.^{17–19} The complex between phosphatidylethanolamine transferase (PTE) that contains a C-terminal thioredoxin-like domain, and transducin $\beta\gamma$ ²⁰ shares some similarities with the U5-15K-GYF-domain complex with respect to the thioredoxin part of the complex interface. However, the extension of the phosphatidylethanolamine transferase is more comprehensive than in the U5-15K and the involved residues are also found further towards the C terminus. This example and the U5-15K-GYF complex suggest that different extensions of the thioredoxin fold have been adapted in order to generate novel specific functions of thioredoxin-like proteins that do not involve the redox-active disulfide bridge of thioredoxin.

Structural basis for the bifunctionality of U5-52K

The two known cellular functions of U5-52K comprise the binding to either the CD2 receptor or the spliceosomal U5-15K protein. These two interactions involve clearly separated binding sites on the GYF-domain (Figure 4). The GYF-domain of U5-52K is capable of specifically binding proline-rich sequences (PPPPGHR) of other proteins,¹³ but it does not bind the U5-15K through its polyproline-binding site. This raises the question whether the different cellular functions of U5-52K are strictly related to its two separated protein-binding sites. Recently, it has been shown that the GYF-domain is capable of interacting with the spliceosomal SmB and SmB' proteins by binding a less proline-rich sequence (+/GxxPPGx+).²¹ However, the physiological significance of this binding has to be clarified, as the SmB/B' proteins are common to the snRNPs U1, U2, U4, U5, but the 52K protein is exclusively associated with the U5 snRNP due to its interaction with U5-15K and another U5 specific protein, the U5-102K.³

The sequence alignments (Supplementary Data, Figure 1) of orthologs of the U5-15K protein and the GYF-domain, respectively, further show that the U5-15K is significantly more highly conserved than U5-52K. This indicates that the U5-52K protein has been adjusted to different cellular functions in diverse organisms.

U5-52K is an interesting example of a bifunctional protein, which acts in immune response and also plays a role in the spliceosomal machinery. The basal function of the U5-52K appears to be related to the splicing process, as U5-52K is found in the spliceosomal U5 snRNPs from yeast to human. The other function of U5-52K as CD2-binding protein, however, is restricted to mammals due to the occurrence of T-cell-specific receptors.

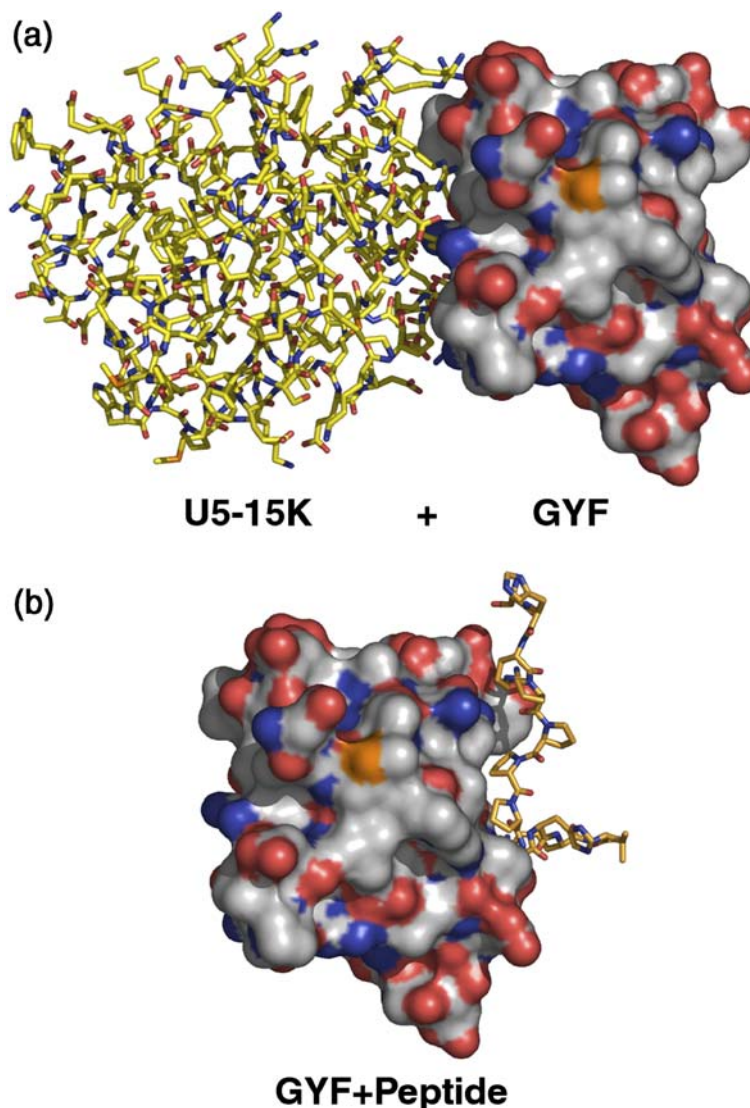


Figure 4. Separated binding sites on the GYF-domain. The GYF-domain is shown in a surface representation either in complex with (a) the U5-15K protein, or (b) a proline-rich peptide. The interaction with the U5-15K protein and the polyproline motif of the CD2 receptor involves two clearly separated binding sites on the GYF-domain.

Protein Data Bank accession code

The coordinates of the complex and the structure factors have been deposited in the RCSB Protein Data Bank and are available under accession code 1SYX.

Acknowledgements

We thank Ehmke Pohl (EMBL, Hamburg) for help during data collection at the EMBL X13 beamline, and Achim Dickmanns for critically reading the manuscript. This work was supported by a grant of the BMBF.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2007.03.077](https://doi.org/10.1016/j.jmb.2007.03.077)

References

1. Moore, B. (2004). Bifunctional and moonlighting enzymes: lighting the way to regulatory control. *Trends Plant. Sci.* **9**, 221–228.
2. Jeffery, C. J. (2003). Moonlighting proteins: old proteins learning new tricks. *Trends Genet.* **19**, 415–417.
3. Lagerbauer, B., Liu, S., Makarov, E., Vornlocher, H. P., Makarova, O., Ingelfinger, D. *et al.* (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.
4. Will, C. L. & Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301.
5. Staley, J. P. & Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92**, 315–326.
6. Bach, M., Winkelmann, G. & Lührmann, R. (1989). 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc. Natl Acad. Sci. USA*, **86**, 6038–6042.

7. Behrens, S. E. & Lührmann, R. (1991). Immunoaffinity purification of a [U4/U6·U5] tri-snRNP from human cells. *Genes Dev.* **5**, 1439–1452.
8. Mougin, A., Gottschalk, A., Fabrizio, P., Lührmann, R. & Branlant, C. (2002). Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6·U5 tri-snRNP particles. *J. Mol. Biol.* **317**, 631–649.
9. Will, C. L. & Lührmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol. Chem.* **386**, 713–724.
10. Stevens, S. W., Barta, I., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D. *et al.* (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6×U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA*, **7**, 1543–1553.
11. Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R. & Ficner, R. (1999). Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kDa protein. *J. Mol. Biol.* **294**, 515–525.
12. Nishizawa, K., Freund, C., Li, J., Wagner, G. & Reinherz, E. L. (1998). Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proc. Natl Acad. Sci. USA*, **95**, 14897–14902.
13. Freund, C., Dotsch, V., Nishizawa, K., Reinherz, E. L. & Wagner, G. (1999). The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nature Struct. Biol.* **6**, 656–660.
14. Freund, C., Kuhne, R., Yang, H., Park, S., Reinherz, E. L. & Wagner, G. (2002). Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *EMBO J.* **21**, 5985–5995.
15. Zhang, Y. Z., Cheng, H., Gould, K. L., Golemis, E. A. & Roder, H. (2003). Structure, stability, and function of hDim1 investigated by NMR, circular dichroism, and mutational analysis. *Biochemistry*, **42**, 9609–9618.
16. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
17. Kumar, J. K., Tabor, S. & Richardson, C. C. (2001). Role of the C-terminal residue of the DNA polymerase of bacteriophage T7. *J. Biol. Chem.* **276**, 34905–34912.
18. Qin, J., Clore, G. M., Kennedy, W. P., Kuszewski, J. & Gronenborn, A. M. (1996). The solution structure of human thioredoxin complexed with its target from Ref-1 reveals peptide chain reversal. *Structure*, **4**, 613–620.
19. James, L. C., Roversi, P. & Tawfik, D. S. (2003). Antibody multispecificity mediated by conformational diversity. *Science*, **299**, 1362–1367.
20. Gaudet, R., Bohm, A. & Sigler, P. B. (1996). Crystal structure at 2.4 angstroms resolution of the complex of transducin betagamma and its regulator, phosducin. *Cell*, **87**, 577–588.
21. 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.
22. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
23. Collaborative Computational Project, No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
24. Kelley, L. A. & Sutcliffe, M. J. (1997). OLDERADO: on-line database of ensemble representatives and domains. *Protein Sci.* **6**, 2628–2630.
25. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
26. Jones, T. A. & Kjeldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208.
27. Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.

Edited by J. Doudna

(Received 20 February 2007; received in revised form 25 March 2007; accepted 29 March 2007)
Available online 4 April 2007