

## **Structural investigations of the RNA binding properties of STAR proteins**

Mikael Feracci, Jaelle Foot, and Cyril Dominguez\*

Department of Biochemistry, Henry Wellcome Laboratories of Structural Biology, University of Leicester, Lancaster Road, LE1 9HN, Leicester, UK

\*: To whom correspondence should be addressed:

Henry Wellcome building

Lancaster Road

Leicester, LE1 9HN

Phone: +44 116 2297073

Email: [cyril.dominguez@le.ac.uk](mailto:cyril.dominguez@le.ac.uk)

**Keywords:** STAR proteins, protein-RNA complexes, alternative splicing, cell signalling, structural biology.

**Abbreviations:** GLD-1, defective in Germ Line Development protein 1; HOW, Held Out of Wings; KH, hnRNP K homology; Sam68, Src-associated in mitosis of 68kDa; SLM1 and SLM2, Sam68-like mammalian proteins 1 and 2; SF1, Splicing Factor 1; STAR, Signal Transduction and Activation of RNA;

**Abstract**

STAR (Signal Transduction and Activation of RNA) proteins are a family of RNA binding proteins that regulate post-transcriptional gene regulation events at various levels, such as pre-mRNA alternative splicing, RNA export, translation and stability. Most of these proteins are regulated by signalling pathways through post-translational modifications, such as phosphorylation and arginine methylation. These proteins share a highly conserved RNA binding domain, denoted STAR domain. Structural investigations of this STAR domain in complex with RNA have highlighted how a subset of STAR proteins specifically recognizes its RNA targets. The present review focuses on the structural basis of RNA recognition by this family of proteins.

### **The STAR family of Protein**

Throughout the 1990's, a novel family of proteins that functions in RNA metabolism and is regulated through signal transduction pathways was discovered and named the STAR (Signal Transduction and Activation of RNA) family [1]. First, Sam68 (Src-associated in mitosis of 68kDa) was initially identified as a 62kDa protein binding to Ras-GTP and was later found to be a downstream target of Src and Fyn [2,3]. Sam68 is now the best-characterised member of the STAR family, and two mammalian orthologs have also been identified, SLM-1 and SLM-2/T-STAR (Sam68-like mammalian proteins 1 and 2 / testis STAR) [4,5]. Shortly after the discovery of Sam68, a novel subfamily of evolutionarily conserved proteins that share sequence homology with Sam68 was described. These proteins are *c. elegans* GLD-1 (defective in Germ Line Development protein 1) [6], mouse Quaking (QK1) [7], and drosophila HOW (Held Out of Wings) [8,9] and play roles in post-transcriptional gene regulation and developmental processes including muscle development, oogenesis and spermatogenesis. Finally, another member, SF1 (splicing factor 1) was identified in 1996 [10].

### **Domain organization of STAR proteins**

Together, the STAR family of proteins are related by an evolutionarily conserved domain of approximately 200 amino acids, which is responsible for RNA binding. This domain, denoted STAR domain, can be divided into 3 distinct subdomains: the central well-known KH (hnRNP K homology) RNA binding domain surrounded by two flanking regions, the QUA1 and QUA2, which are involved in dimerization and RNA binding, respectively [1] (Figure 1). In addition to this STAR domain, most STAR proteins contain additional domains or regions important for their proper function (Figure 1). Sam68 possess a nuclear localisation signal and a Tyrosine-rich domain at the C-terminus, which is phosphorylated by Tyrosine kinases such as Src and Brk [2,11]. Sam68 also contain six Proline-rich motifs, three to the N-terminal and three to the C-terminal of the STAR domain, which bind SH3 and WW domains to facilitate tyrosine phosphorylation [2], and two RG-rich (Arginine-Glycine rich) motifs that are target sites for Arginine methylation by PRMT1 (Protein methyltransferase 1) [12]. Finally, Sam68 is also subject to Lysine Acetylation [13], Sumoylation [14] and Serine/Threonine phosphorylation by kinases such as Cdc2 [15], Erk1/2 [16], and Nek2 [17]. The domain architectures of SLM1 and SLM2/T-STAR are very similar to that of Sam68, with a deletion of the first 100 amino acids that form the N-terminus of Sam68 [4,5]. In addition to its STAR domain, QKI also contains a nuclear localization signal, a tyrosine-rich and a proline-rich regions. QKI has also been shown to be tyrosine phosphorylated by Src and Fyn [18]. In contrast, GLD-1 lacks such domains but has been recently shown to be phosphorylated by the Serine/Threonine kinases Cyclin E and Cdk2 [19]. SF1 is more the most divergent member of the STAR family. Its STAR domain contains the KH and the QUA2 regions but lacks the QUA1 dimerization domain. In addition, an N-terminal nuclear localization signal, and C-terminal zinc-knuckle and Pro-rich regions are present in SF1.

### **Function of STAR proteins in RNA metabolism**

All identified STAR proteins have been shown to bind RNA and to be involved in post-transcriptional gene regulation. Sam68 affect various cellular processes such as differentiation, cell cycle progression and apoptosis, through its direct involvement in alternative splicing [20]. Indeed, many pre-mRNA targets of Sam68 have been identified that encode for proteins spanning a variety of cellular functions: mTOR in adipogenesis [21], Neurexin 1 in synapse

function [22], Cyclin D1 in cell cycle progression [23], ASF1/SF2 in splicing [24], CD44 in cell migration [16], and Bcl-x in apoptosis [25]. In addition, Sam68 is involved in the nuclear export of the unspliced HIV (human immunodeficiency virus) RNA through binding directly the viral Rev protein and the viral RNA [26]. While Sam68 is ubiquitously expressed in all tissues, SLM1 and SLM2/T-STAR exhibit preferential expression in the brain where they specifically act as alternative splicing regulators of the Neurexin 1-3 genes [27,28]. QKI is involved in the regulation of alternative splicing, RNA export, and mRNA stability of myelin-associated pre-mRNAs [29,30]. GLD-1 is a cytoplasmic protein that function as a translational repressor [31,32]. SF1 is an essential splicing factor that binds the pre-mRNA branchpoint sequence during the early stage of spliceosome formation [33].

### **Specific RNA recognition by STAR proteins**

Several studies have investigated the RNA binding specificity of STAR proteins. First, SELEX (Systematic evolution of ligands by Exponential enrichment) experiments have identified purine-rich RNA sequences specifically bound by Sam68 [34]. Most of these sequences contained a conserved UAAA motif. This consensus sequence has been later confirmed and extended to SLM2/TSTAR that was found, like Sam68, to bind specifically a U(A/U)AA motif [35]. The RNA binding motifs specifically bound by QKI and GLD-1 have also been identified [36,37]. Both proteins recognize very similar motifs, with the consensus sequences being UACU(C/A)A for GLD-1 [36] and NA(A/C)UAA for QKI [37]. These motifs have been confirmed in vitro by SELEX [35] and in vivo by RIP-CHIP (RNP immunoprecipitation microarray) for GLD-1 [38] and PAR-CLIP (photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation) for QKI [39]. The RNA motif recognized by QKI and GLD-1 is very similar to the highly conserved branchpoint sequence (UACUAAAC) specifically bound by SF1 [33]. STAR proteins can therefore be divided into two subfamilies based on their RNA binding specificities: the QKI/GLD-1/SF1 subfamily that recognizes a six-nucleotide UACUAA motif and the Sam68/SLM1/SLM2/TSTAR subfamily that recognizes specifically AU-rich RNAs containing the four-nucleotide UAAA motif.

### **Structural investigations of RNA binding by STAR proteins.**

The RNA binding STAR domain is very well conserved within the STAR family (Figure 2) and this domain is crucial for the proper function of these proteins in RNA metabolism. Structural studies of STAR proteins have mainly so far focused on the STAR domain free and in complex with RNA. These structures, as detailed below, explain the specificity of RNA recognition by certain members of the STAR family.

The first structure of a KH domain from the human FMR1 protein was solved by NMR (nuclear magnetic resonance) in 1996 [40]. This structure showed that the KH domain adopts a  $\beta\alpha\alpha\beta\alpha$  fold and demonstrated that a single point mutation on this protein (I304N) leading to the phenotype of fragile X Syndrome [41] was responsible for the unfolding of the KH domain. Later the structure of NOVA-2 KH3 was solved in complex with a RNA hairpin [42]. This structure revealed how KH domains specifically bind RNA molecules and highlighted the importance of a highly conserved GXXG motif in RNA binding.

In 2001, the first structure of a KH domain from a STAR protein, SF1, was solved by NMR in complex with the canonical branchpoint sequence, UAUACUAACAA [43]. In contrast to the KH domains of FMR1 and Nova, the KH domain of SF1 contains an additional flexible loop

located between the second and third  $\beta$ -strands. Interestingly, this additional loop is conserved among the STAR family of proteins (Figure 2). Although this loop is not involved in RNA binding by SF1, deletion of this loop in Sam68 disrupted its RNA binding ability [34]. The RNA recognition by SF1 is driven by a mixture of hydrophobic, electrostatic and hydrogen bond contacts. The KH domain of SF1 specifically recognizes the <sup>6</sup>UAAC<sup>9</sup> sequence of the <sup>1</sup>UAUACUAACAA<sup>11</sup> RNA. The GXXG motif (GPRG in the case of SF1) recognizes specifically the base of U6 and contacts the backbone of the RNA. Specific hydrogen bonds to the bases of the RNA are formed between E149 of SF1 and A7, I177 and A8 and S182 and A11. In addition, N151 contacts specifically the base of C5 via a hydrogen bond. Interestingly, this structure also revealed that the QUA2 region that is conserved among the STAR family (Figure 2) adopts a helical conformation, packs against the KH domain and is also involved in RNA binding. Indeed, the QUA2 domain recognizes 4 nucleotides, <sup>2</sup>AUAC<sup>5</sup>, of the RNA. Specifically, T253 contacts A2 through specific hydrogen bonds, and R255, A248, L244 and L247 make hydrophobic contacts to A4 and C5. This study also demonstrated that, although the QUA2 is flexible in absence of RNA, the  $\alpha$ -helix is already preformed in solution and does not occur upon RNA binding. This has been later confirmed through an NMR dynamic study of the KH-QUA2 domain of the Xenopus QKI protein showing that the QUA2 region is helical but highly flexible in absence of RNA and become rigid in presence of RNA [44].

More recently, the structure of the QUA1 domains of Sam68, GLD-1 and QKI have been solved [45-47]. These structures showed that the QUA1 region of these proteins consists of a helix-turn-helix motif stabilized by a network of hydrophobic contacts. The dimer is formed by an almost perpendicular stacking of the two monomers and is stabilized by a network of hydrophobic contacts and hydrogen bonds involving mainly residues of the turn between the two helices. The structures demonstrated that the QUA1 domain alone is sufficient for STAR protein dimerization. In the case of Sam68, it has been proposed that the phosphorylation of one tyrosine (Y103), localized in the QUA1 domain, could be involved in the dissociation of the homodimer and mutation of this residue leads to a loss of splicing activity [45]. This suggests that the dimerization of the protein is compulsory for his function in alternative splicing and that tyrosine phosphorylation could regulate Sam68 function by disrupting its ability to dimerize.

In 2013, the structures of the full STAR domains of QKI and GLD-1 in complex with their target RNAs have been solved by X-ray crystallography [48]. These structures revealed that the RNA is only bound by the KH and the QUA2 regions while the QUA1 region is involved in dimerization and protein-protein interaction with the QUA2 region. The RNA sequences bound by QKI and GLD-1 are very similar to the one bound by SF1 (CUAAC) and consequently the RNA recognition by GLD1 and QKI proteins involves the same conserved residues of the KH and the QUA2 as SF1. Specifically, the base of C1 is specifically recognized by K190 and K313 of QKI and GLD-1, respectively. The UAA motif that overlaps perfectly on all the structures contacts Q193 (QKI) and Q316 (GLD1) for the U2, and N97 (QKI) for A3. The last adenine forms two hydrogen bonds with the main chain of the protein, a valine in both QKI and GLD-1. The last cytosine (C5) is recognized by an arginine (R124 and R247 for QKI and GLD-1 respectively). As for the structure of SF1, the QUA2 domain of QKI and GLD1 is strongly involved in the interaction with the RNA. These structures also show that the QUA1 domain interacts with the QUA2 region, forming a three-helix bundle. This interaction induces a precise orientation of the KH-QUA2 domains in the dimer and positions the two RNA binding surfaces at opposite ends of the dimer. This orientation suggests that one dimer can bind a single RNA molecule only if the two binding motifs (CUAAC) are separated by more than 10 nucleotides. In

addition to the QUA1-QUA2 interaction, the structure of GLD-1 STAR domain shows that the QUA1-KH linker contacts the KH domain through hydrogen bonds and van der Waals contacts. It was proposed that these contacts facilitate the orientation of the QUA1, KH and QUA2 regions within the STAR domain. It should be noted, however, that this linker is clearly visible in the GLD1 structure, but not in the QKI structure suggesting that the linker is more flexible in the case of QKI. This flexibility could be important for STAR proteins to bind two RNA motifs that are relatively close (less than 10 nucleotides on the same RNA) [49]. Further structural studies will be needed to address this issue.

In contrast to SF1, QKI and GLD-1 that specifically bind a CUAAC RNA motif, Sam68, SLM1 and SLM2/TSTAR bind specifically AU-rich RNA sequences (see above). The molecular details of the specific AU-rich RNA recognition by these proteins remain unknown. We have therefore initiated a structural study of Sam68 and STAR in complex with AU-rich RNA [50]. Using NMR spectroscopy, we have identified the optimal protein constructs of Sam68 and SLM2/TSTAR, and, through an AU-rich RNA NMR screen, the optimal RNA sequence for the structure determination of these complexes.

**Funding:**

This work was supported by a Medical Research Council Career Development Award to C.D. (G1000526) and by a College of Medicine, Biological Sciences and Psychology, University of Leicester, studentship to J.F.

## References

- [1] C Vernet, Artzt K. (1997) STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**, 479–484.
- [2] Taylor SJ, Shalloway D. (1994) An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature.* **368**, 867–871.
- [3] Fumagalli S, Totty NF, Hsuan JJ, Courtneidge SA. (1994) A target for Src in mitosis. *Nature.* **368**, 871–874.
- [4] Di Fruscio M, Chen T, Richard S. (1999) Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. *Proc Natl Acad Sci USA.* **96**, 2710–2715.
- [5] JP Venables, C Vernet, SL Chew, DJ Elliot, RB Cowmeadow, J Wu, et al. (1999) T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis. *Hum. Mol. Genet.* **8**, 959–969.
- [6] Jones AR, Schedl T. (1995) Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* **9**, 1491–1504.
- [7] Ebersole TA, Chen Q, Justice MJ, Artzt K. (1996) The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nat Genet.* **12**, 260–265.
- [8] Zaffran S, Astier M, Gratecos D, Semeriva M. (1997) The held out wings (how) *Drosophila* gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. *Development.* **124**, 2087–2098.
- [9] Baehrecke EH. (1997) who encodes a KH RNA binding protein that functions in muscle development. *Development.* **124**, 1323–1332.
- [10] Arning S, Gruter P, Graeme B, Kramer A. (1996) Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA.* **2**, 794–810.
- [11] Derry JJ, Richard S, Valderrama Carvajal H, Ye X, Vasioukhin V, Cochrane AW, et al. (2000) Sik (BRK) Phosphorylates Sam68 in the Nucleus and Negatively Regulates Its RNA Binding Ability. *Mol Cell Biol.* **20**, 6114–6126.
- [12] Côté J, Boisvert F-M, Boulanger M-C, Bedford MT, Richard S. (2003) Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. *Mol Biol Cell.* **14**, 274–287.
- [13] Babic I, Jakymiw A, Fujita DJ. (2004) The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity. *Oncogene.* **23**, 3781–3789.

- [14] Babic I, Cherry E, Fujita DJ. (2006) SUMO modification of Sam68 enhances its ability to repress cyclin D1 expression and inhibits its ability to induce apoptosis. *Oncogene*. **25**, 4955–4964.
- [15] Resnick RJ, Taylor SJ, Lin Q, Shalloway D. (1997) Phosphorylation of the Src substrate Sam68 by Cdc2 during mitosis. *Oncogene*. **15**, 1247–1253.
- [16] Matter N, Herrlich P, König H. (2002) Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature*. **420**, 691–695.
- [17] Naro C, Barbagallo F, Chieffi P, Bourgeois CF, Paronetto MP, Sette C. (2014) The centrosomal kinase NEK2 is a novel splicing factor kinase involved in cell survival. *Nucleic Acids Res.* **42**, 3218–3227.
- [18] Zhang Y, Lu Z, Ku L, Chen Y, Wang H, Feng Y. (2003) Tyrosine phosphorylation of QKI mediates developmental signals to regulate mRNA metabolism. *EMBO J.* **22**, 1801–1810.
- [19] Jeong J, Verheyden JM, Kimble J. (2011) Cyclin E and Cdk2 Control GLD-1, the Mitosis/Meiosis Decision, and Germline Stem Cells in *Caenorhabditis elegans*. *PLoS Genet.* **7**, e1001348.
- [20] Sánchez-Jiménez F, Sánchez-Margalet V. (2013) Role of Sam68 in Post-Transcriptional Gene Regulation. *Int. J. Mol. Sci.* **14**, 23402–23419.
- [21] Huot M-É, Vogel G, Zabarauskas A, Ngo CT-A, Coulombe-Huntington J, Majewski J, et al. (2012) The Sam68 STAR RNA-Binding Protein Regulates mTOR Alternative Splicing during Adipogenesis. *Mol. Cell.* **46**, 187–199.
- [22] Iijima T, Wu K, Witte H, Hanno-Iijima Y, Glatter T, Richard S, et al. (2011) SAM68 Regulates Neuronal Activity-Dependent Alternative Splicing of Neurexin-1. *Cell.* **147**, 1601–1614.
- [23] Paronetto MP, Cappellari M, Busà R, Pedrotti S, Vitali R, Comstock C, et al. (2010) Alternative splicing of the cyclin D1 proto-oncogene is regulated by the RNA-binding protein Sam68. *Cancer Res.* **70**, 229–239.
- [24] Valacca C, Bonomi S, Buratti E, Pedrotti S, Baralle FE, Sette C, et al. (2010) Sam68 regulates EMT through alternative splicing-activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. *J. Cell Biol.* **191**, 87–99.
- [25] Paronetto MP, Achsel T, Massiello A, Chalfant CE, Sette C. (2007) The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* **176**, 929–939.
- [26] Reddy TR, Xu W, Mau JK, Goodwin CD, Suhasini M, Tang H, et al. (1999) Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev. *Nat Med.* **5**, 635–642.



- [27] Ehrmann I, Dalglish C, Liu Y, Danilenko M, Crosier M, Overman L, et al. (2013) The Tissue-Specific RNA Binding Protein T-STAR Controls Regional Splicing Patterns of Neurexin Pre-mRNAs in the Brain. *PLoS Genet.* **9**, e1003474.
- [28] Iijima T, Iijima Y, Witte H, Scheiffele P. (2014) Neuronal cell type-specific alternative splicing is regulated by the KH domain protein SLM1. *J. Cell Biol.* **204**, 331–342.
- [29] Wu JI, Reed RB, Grabowski PJ, Artzt K. (2002) Function of quaking in myelination: Regulation of alternative splicing. *Proc Natl Acad Sci USA.* **99**, 4233–4238.
- [30] Li Z, Zhang Y, Li D, Feng Y. (2000) Destabilization and Mislocalization of Myelin Basic Protein mRNAs in *quaking* Dysmyelination Lacking the QKI RNA-Binding Proteins. *J. Neurosci.* **20**, 4944–4953.
- [31] Jan E, Motzny CK, Graves LE, Goodwin EB. (1999) The STAR protein, GLD-1, is a translational regulator of sexual identity in *Caenorhabditis elegans*. *EMBO J.* **18**, 258–269.
- [32] Schumacher B, Hanazawa M, Lee M-H, Nayak S, Volkmann K, Hofmann R, et al. (2005) Translational Repression of *C. elegans* p53 by GLD-1 Regulates DNA Damage-Induced Apoptosis. *Cell.* **120**, 357–368.
- [33] Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. (1997) The Splicing Factor BBP Interacts Specifically with the Pre-mRNA Branchpoint Sequence UACUAAC. *Cell.* **89**, 781–787.
- [34] Lin Q, Taylor SJ, Shalloway D. (1997) Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains. *J Biol Chem.* **272**, 27274–27280.
- [35] Galarneau A, Richard S. (2009) The STAR RNA binding proteins GLD-1, QKI, SAM68 and SLM-2 bind bipartite RNA motifs. *BMC Mol Biol.* **10**, 47.
- [36] Ryder SP, Frater LA, Abramovitz DL, Goodwin EB, Williamson JR. (2004) RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. *Nat Struct Mol Biol.* **11**, 20–28.
- [37] Ryder SP. (2004) Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination. *RNA.* **10**, 1449–1458.
- [38] Wright JE, Gaidatzis D, Senften M, Farley BM, Westhof E, Ryder SP, et al. (2011) A quantitative RNA code for mRNA target selection by the germline fate determinant GLD-1. *EMBO J.* **30**, 533–545.
- [39] Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. (2010) Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell.* **141**, 129–141.

- [40] Musco G, Stier G, Joseph C, Morelli MAC, Nilges M, Gibson TJ, et al. (1996) Three-Dimensional Structure and Stability of the KH Domain: Molecular Insights into the Fragile X Syndrome. *Cell*. **85**, 237–245.
- [41] De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, et al. (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet*. **3**, 31–35.
- [42] Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB, et al. (2000) Sequence-Specific RNA Binding by a Nova KH Domain: Implications for Paraneoplastic Disease and the Fragile X Syndrome. *Cell*. **100**, 323–332.
- [43] Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, et al. (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*. **294**, 1098–1102.
- [44] Maguire ML, Guler-Gane G, Nietlispach D, Raine ARC, Zorn AM, Standart N, et al. (2005) Solution structure and backbone dynamics of the KH-QUA2 region of the *Xenopus* STAR/GSG quaking protein. *J Mol Biol*. **348**, 265–279.
- [45] Meyer NH, Tripsianes K, Vincendeau M, Madl T, Kateb F, Brack-Werner R, et al. (2010) Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain. *J Biol Chem*. **285**, 28893–28901.
- [46] Beuck C, Szymczyna BR, Kerkow DE, Carmel AB, Columbus L, Stanfield RL, et al. (2010) Structure of the GLD-1 homodimerization domain: insights into STAR protein-mediated translational regulation. *Structure*. **18**, 377–389.
- [47] Beuck C, Qu S, Fagg WS, Jr MA, Williamson JR. (2012) Structural Analysis of the Quaking Homodimerization Interface. *J Mol Biol*. **423**, 766–781.
- [48] Teplova M, Hafner M, Teplov D, Essig K, Tuschl T, Patel DJ. (2013) Structure-function studies of STAR family Quaking proteins bound to their in vivo RNA target sites. *Genes Dev*. **27**, 928–940.
- [49] Cukier CD, Ramos A. (2010) Creating a twin STAR. *Structure*. **18**, 279–80.
- [50] Foot JN, Feracci M, Dominguez C. (2014) Screening protein - Single stranded RNA complexes by NMR spectroscopy for structure determination. *Methods*. **65**, 288–301.

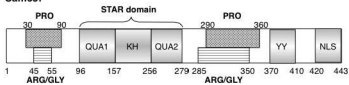
### **Figure legends:**

Figure 1: Domain organization of STAR proteins. All STAR protein possess a STAR domain responsible for RNA binding and composed of a central KH domain flanked by the QUA1 and QUA2 regions. In addition, other domains or motifs are present in STAR proteins, such as nuclear localization signals (NLS), tyrosine-rich (YY) region, proline-rich (PRO) regions, and arginine/glycine-rich regions (ARG/GLY).

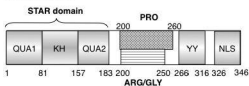
Figure 2: Amino acid sequence alignment of the STAR domain of Sam68, SLM2/TSTAR, QKI and GLD-1. The KH domain is highlighted in yellow. Secondary structures elements are shown above the sequences.

Figure 3: Structures of members of the STAR family protein with their RNA. (a) SF1, (b) QKI and (c) GLD-1. The QUA1 domain is colored in blue, the KH domain in grey, the QUA2 domain in red, and the RNA in orange and green. The structures of GLD-1 and QKI show the dimerization of the QUA1 domain. Specific contacts between the protein and the RNA are enlarged and labeled.

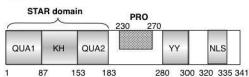
**Sam68:**



**SLM2/TSTAR:**



**Quaking:**



**GLD-1:**

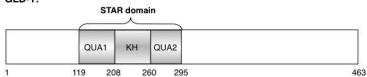


Figure 1

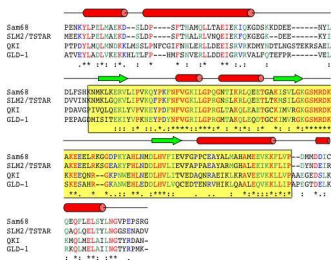


Figure 2

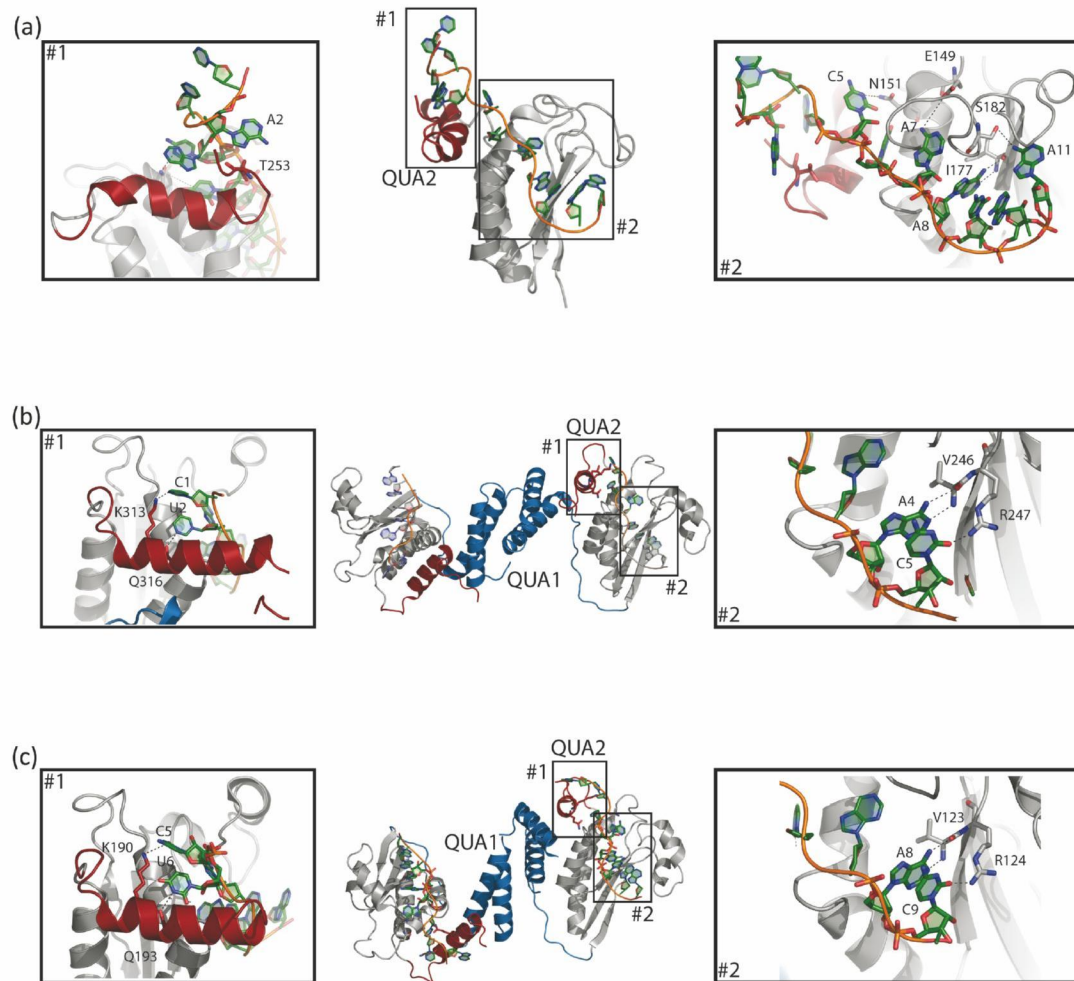


FIGURE 3