http://dx.doi.org/10.1016/j.jmb.2012.08.027 J. Mol. Biol. (2012) 423, 766–781

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Structural Analysis of the Quaking Homodimerization Interface

Christine Beuck[1], Song Qu[1], W. Samuel Fagg[2], Manuel Ares Jr.[2] and James R. Williamson[1]*

[1]*Department of Molecular Biology, Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA*
[2]*Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, CA 95064, USA*

Received 11 July 2012; received in revised form 30 August 2012; accepted 31 August 2012
Available online 11 September 2012

Edited by C. R. Matthews

Keywords:
STAR/GSG proteins;
Qua1 domain;
crystal structure;
RNA binding;
C2C12 myoblasts

Quaking (QkI) is a prototypical member of the STAR (*s*ignal *t*ransducer and *a*ctivator of *R*NA) protein family, which plays key roles in posttranscriptional gene regulation by controlling mRNA translation, stability and splicing. QkI-5 has been shown to regulate mRNA expression in the central nervous system, but little is known about its roles in other tissues. STAR proteins function as dimers and bind to bipartite RNA sequences; however, the structural and functional roles of homodimerization and heterodimerization are still unclear. Here, we present the crystal structure of the QkI dimerization domain, which adopts a similar stacked helix–turn–helix arrangement as its homologs GLD-1 (*g*erm *l*ine development *d*efective-*1*) and Sam68 (*S*rc-*a*ssociated protein during *m*itosis, *68* kDa) but differs by an additional helix inserted in the dimer interface. Variability of the dimer interface residues likely ensures selective homodimerization by preventing association with non-cognate STAR family proteins in the cell. Mutations that inhibit dimerization also significantly impair RNA binding *in vitro*, alter QkI-5 protein levels and impair QkI function in a splicing assay *in vivo*. Together, our results indicate that a functional Qua1 homodimerization domain is required for QkI-5 function in mammalian cells.

Introduction

Quaking (QkI) is a conserved multifunctional protein in vertebrates that belongs to the STAR (*s*ignal *t*ransducer and *a*ctivator of *R*NA) family of RNA-binding proteins. QkI regulates a multitude of cellular functions, both at different developmental times and in various tissues in mice and humans (reviewed in Ref. 1). STAR family proteins link signaling pathways to various aspects of posttranscriptional regulation of mRNAs. QkI is important for several biological processes including myelination, smooth muscle cell differentiation, vascular development and heart development.[2–4] The cellular processes controlled by QkI are reported to include target mRNA stabilization, translation, subcellular localization and alternative splicing.[5–8] The consensus RNA binding sequence of QkI is A(C/U)UAA(C/U) and is often accompanied by a so-called half-site UAA(C/U) in close proximity.[9–11] The activity of QkI and other STAR family members is regulated via posttranslational modifications including phosphorylation.[12–14] However, the molecular mechanisms linking RNA binding and homodimerization to the diverse regulatory roles of QkI are still poorly understood.

*Corresponding author. E-mail address: jrwill@scripps.edu.

Abbreviations used: CNS, central nervous system; MBP, maltose-binding protein; FP, fluorescence polarization; MR, molecular replacement; PDB, Protein Data Bank; RT, reverse transcription; SeMet, selenomethionine; SSRL, Stanford Synchrotron Radiation Laboratory; wt, wild type.

Several alternatively spliced isoforms of Quaking protein differ in their C-terminal 8- to 30-amino-acid sequences, with the three major isoforms being QkI-5, QkI-6 and QkI-7. These unique tails determine the localization within the cell and confer specific functions. QkI-6 and QkI-7 are mainly cytoplasmic, while QkI-5 contains a noncanonical nuclear localization signal within its tail.[15–17]

The most extensively studied aspect of QkI function is its crucial role in myelin formation in the central nervous system (CNS), where QkI controls the expression of many genes important for myelination through a complex mechanism involving all pathways mentioned above. Defects in QkI expression and/or splicing in the CNS have been implicated in several psychiatric diseases including schizophrenia and ataxia.[18,19] In quaking-viable mutation in mice (qk^v mice), myelin fails to undergo compaction in the CNS due to a large deletion upstream of the quaking locus.[20] The alternative splicing of several myelin-specific genes regulated by QkI has been implicated as an important mechanism in regulation of myelin formation.[5,21,22] Recently, QkI expression was correlated with alternative splicing of the histone variant MacroH2A1 in several different human tissues and cancer cell lines, linking defects in MacroH2A1 splicing to a variety of cancers (including testicular, lung, bladder, cervical, breast, colon, ovarian and endometrial) and suggesting that QkI may be an important tumor suppressor.[23,24]

The conserved STAR domain defines the STAR protein family and consists of a central KH (heterogeneous nuclear ribonucleoprotein *K h*omology) RNA binding domain flanked by two conserved subdomains referred to as Qua1 and Qua2 (Fig. 1).[2] STAR proteins bind RNA via the KH and Qua2 regions, while the Qua1 domain is essential for homodimerization.[25–27] The current model for a STAR protein dimer binding to RNA suggests a bipartite binding mode. One STAR protomer is thought to bind one 6- to 7-nt consensus sequence, while the second protomer binds to an additional full consensus site or shorter (3–4 nt) half-site spaced by 1–20 nt.[9,11,27,28] Alternatively, recognition of two distantly spaced consensus sites could organize mRNA structure for regulation by QkI.[29–31]

The structural basis for the recognition of a single RNA consensus site has been described for the splicing factor SF1,[32] indicating that the Qua2 domain extends the RNA-binding surface of the KH domain and that this KH–Qua2 tandem domain is essential for sequence-specific RNA recognition. Interestingly, SF1 is the only STAR family member that lacks the Qua1 dimerization domain. The solution structure of *Xenopus* Quaking KH–Qua2 domain shows that the Qua2 domain does not contact the KH domain in the absence of RNA.[33] Qua1 homodimerization domain structures have been solved for the GLD-1 (*g*erm *l*ine development *d*efective-*1*) and Sam68 (*S*rc-*a*ssociated protein during *m*itosis, *68* kDa) homologs of QkI, revealing a helix–turn–helix fold.[30,31] The hydrophobic dimer interface is located at the top of the hairpin, and the protomers are stacked at a 90° angle. Although the full STAR domain is conserved within the family, the sequence identity in the Qua1 region is only 30% between QkI and its close homolog GLD-1. Interestingly, the hydrophobic "zipper" residues mediating the contact between the two helices within each monomer are highly conserved, while the hydrophobic patch at the top of the hairpin that forms the dimer interface in the GLD-1 and Sam68 Qua1 structures contains significant variation throughout the STAR family (Fig. 1). Furthermore, the Quaking orthologs feature a 3-residue insertion

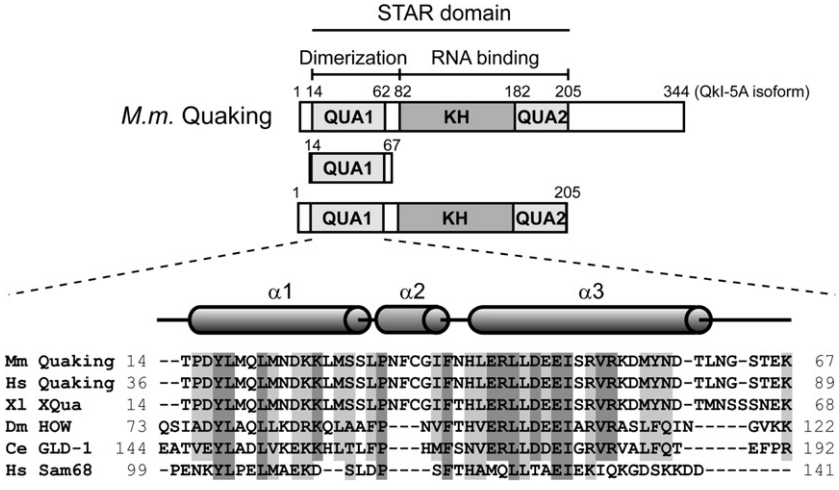

Fig. 1. The STAR family of RNA-binding proteins. Domain structure of QkI and constructs used in this study. The Qua1, KH and Qua2 subdomains are shaded (top). Sequence alignment of the Qua1 domain of representative members of the STAR/GSG protein family (bottom). Identical conserved residues are highlighted in dark gray; similar residues, in light gray. The secondary structure of the Quaking Qua1 domain is shown above. Members of the Quaking subfamily feature a 3-residue insertion between the two large helices that form an additional short helix, instead of a simple turn, which is not present in the Qua1 structures of GLD-1 and Sam68.

in this region that is not present in other STAR proteins.

Some studies suggest that STAR protein dimerization is essential for their function, but little is known about how dimerization translates into biological function on a cellular level. In *Caenorhabditis elegans* GLD-1, the KH–Qua2 domain is sufficient for RNA binding, but dimerization enables the bipartite binding mode and deletion of the Qua1 domain reduces the RNA binding affinity by 1 order of magnitude.[27] QkI mutants that abolish dimerization[26] cause an embryonic lethal phenotype in mice, suggesting that homodimerization serves other important functions in addition to facilitating RNA binding. While the splicing factor SF1 lacks the Qua1 dimerization domain and functions as a monomer, Sam68 mutants that impair homodimerization reduce its activity in an alternative splicing assay,[31] suggesting that proper splicing regulation requires dimerization for at least some STAR family members. The role of QkI dimerization for its function as an alternative splicing regulator is unknown.

Here, we present the crystal structure of the QkI Qua1 domain. A fold similar to a helix–turn–helix is expanded by an additional short helix at the top of the hairpin. Based on thermal melting and RNA binding data, we show that failure to dimerize impairs RNA binding. Furthermore, a cell-based splicing assay in mouse C2C12 myoblasts demonstrates that QkI-5 function in alternative splicing requires its ability to homodimerize.

Results

Structure of the QkI Qua1 homodimer

To select a suitable QkI Qua1 construct for crystallography, we cloned 20 C-terminal truncation constructs based on secondary structure prediction (PredictProtein†) and sequence homology to the GLD-1 and Sam68 Qua1 domains. Residue C35 was changed to Ser for the QkI Qua1 and QkI STAR *in vitro* constructs used in this study to avoid nonspecific aggregation.[34] NMR spectra (^{1}H,^{15}N heteronuclear single quantum coherence; Fig. S1) recorded for six constructs, spanning the region from residue 12 or 14 to between residues 55 and 77, indicated that the majority of the protein is structured and that the structured region is present in all constructs analyzed. The only differences in the spectra between the different C-terminal truncations were observed in the unstructured region (7.8–8.5 ppm) of the spectrum, indicating that the C-terminal region comprising residues 55–77 is in fact unstructured in these constructs.

Native QkI Qua1 (14–67) yielded well-diffracting crystals, but molecular replacement (MR) using the homologous proteins GLD-1 and Sam68 failed to produce a satisfactory model for phasing. The phases were ultimately solved using selenomethionine (SeMet) labeling for multiwavelength anomalous dispersion phasing, and the QkI Qua1 structure was solved at 2.1 Å resolution with one homodimer in the asymmetric unit. A summary of the data collection and refinement statistics is given in Table 1.

The QkI Qua1 monomer consists of two α-helices arranged in a hairpin-like geometry resembling the helix–turn–helix fold observed for GLD-1 and Sam68 Qua1. The two antiparallel helices are held together by a hydrophobic "zipper" and a hydrogen bond between the conserved residues Y17 and E48. The QkI Qua1 structure contains an additional short helix, including the 3-residue insertion, in the linker region between the two main helices (Fig. 2a). This helix is not present in the homologous GLD-1 and Sam68 Qua1 domain structures and lies perpendicular to the two main helices. The C-terminal residues 55–61 show weak electron density and do not form a regular secondary structure. No electron density was observed for residues 62–67, consistent with the NMR data indicating that the C-terminal tail is likely unstructured and flexible.

Structural similarity searches with Dali and PISA[35,36] confirm that the most closely related structures in the Protein Data Bank (PDB) are the Qua1 dimerization domain structures of the homologous STAR proteins GLD-1 (PDB ID 3K6T) and Sam68 (PDB ID 2XA6). The overall hairpin-like structure is conserved in all three Qua1 structures, but QkI Qua1 is unique because of the insertion of an additional helix in the turn region of the hairpin, which is present neither in GLD-1 or Sam68 nor in any other of the many helix–turn–helix structures in the PDB.

The Qua1 dimer is formed by stacking the two monomers at the turn region at an angle of 84°. The dimer interface area covers 815 $Å^2$ per monomer, which corresponds to 19% of the total surface area of the monomer. The hydrophobic dimer interface consists of residues L27, L31, F34, I37 and L41 and the highly conserved residues F38 and L44 (Fig. 2). Several hydrogen bonding interactions along the periphery toward the open end of the hairpin reinforce the hydrophobic interface. Residue D24 forms conserved hydrogen bonds to the backbone amides of residues N39′ and H40′, while intermonomer hydrogen bonds between Q20 and H40′ and Y17–R43′ bridge the dimer interface. These interactions are further facilitated by an intramonomer hydrogen bonding network between H40–R43–E47 and Y17–E48–R51, including the highly conserved Y17–E48 monomer "clamp"

† https://www.predictprotein.org/

Table 1. Data collection and refinement statistics for QkI Qua1 (C35S)

	SeMet			
	SeMet (peak)	SeMet (remote)	SeMet (inflection)	Merged (solve)
Data collection				
Space group		$P2_12_12_1$		
Cell dimensions (Å)				
a, *b*, *c*		33.98, 36.02, 92.81		
Wavelength (Å)	0.9791358	0.9184018	0.979569	
Resolution (Å)	50–2.15 (2.19–2.15)	50–2.10 (2.14–2.10)	50–2.15 (2.19–2.15)	
Total reflections	50,936?	55,390	51,303	
Unique reflections	11,880 (578)	12,905 (659)	11,969 (593)	
R_{sym} (%)	3.6 (28.8)	3.4 (33.9)	3.6 (32.4)	
$I/\sigma(I)$	35.1 (5.2)	40.0 (4.4)	35.4 (4.4)	
Completeness (%)	99.9 (100)	99.8 (100)	99.9 (100)	
Redundancy	4.3 (4.3)	4.3 (4.2)	4.3 (4.2)	
Wilson *B*-factor	41.06	39.83	42.31	
Refinement				
Resolution (Å)				31.91–2.10 (2.65–2.10)
Unique reflections				7067 (3462)
R_{work}				21.63 (22.03)
R_{free}				25.15 (24.66)
Stereochemistry (%)				
Residues in the favored region				98.02
Residues in the allowed region				1.98
Number of atoms				
Protein				856
Water				44
Ca^{2+}				1
B-factors				
Protein				38.84
Water				36.99
Ca^{2+}				119.53
RMSD				
Bond lengths (Å)				0.008
Bond angles (°)				1.102
Residues not modeled				
A				62–67
B				−2–1 (conformer B), 63–67

between the two long helices that has been shown to be important for monomer stability.[30]

Mutational analysis of the dimer interface and monomer zipper

Key dimer interface and zipper residues in QkI Qua1 were mutated to study their contribution to QkI dimerization. The mutation E48G was chosen in addition to E48A because it is biologically relevant, causing an embryonic lethal phenotype in mice. Residue Tyr 17 was mutated to Phe (Y17F) instead of Ala to selectively disrupt the hydrogen bond to residue E48 without disturbing hydrophobic packing of the aromatic side chain. For the GLD-1 Qua1 domain, the zipper, in addition to the dimer interface residues, proved to be crucial for homodimerization by ensuring the structural integrity of the monomer.[30] The homodimer stability of the set of QkI Qua1 mutants was determined by thermal melting monitored by CD spectroscopy (Fig. 3 and Table 2). The curves for most point mutants show a complete melting profile indicating that these constructs are completely folded at low temperatures. Only the QkI Qua1 L21A and E48G point mutants are still significantly unfolded even at temperatures as low as 5 °C. All curves show a single transition suggesting that QkI Qua1, like GLD-1 Qua1, forms obligate dimers in which dimer dissociation and unfolding of the monomer secondary structure occur simultaneously and a folded monomeric Qua1 domain likely does not exist. The melting temperatures for the set of mutants, as a measure of dimer stability, vary over a large range from <15 °C to >85 °C. Mutants in the monomer zipper and the Y17–E48 clamp have a generally more detrimental effect on dimer stability than residues in the dimer interface. However, the dimer interface mutant L21A confers a large destabilization equaling that of the E48G mutant. Residue L21 plays a dual role in the zipper and dimer interface since it is located at the inner edge of the dimer interface close to the dimer symmetry axis (Fig. 3b) and packs against itself (L21′) in the adjacent protomer. Removal of this pair of L21 side chains leads to a large void, while other dimer interface mutants have a more local effect on the hydrophobic

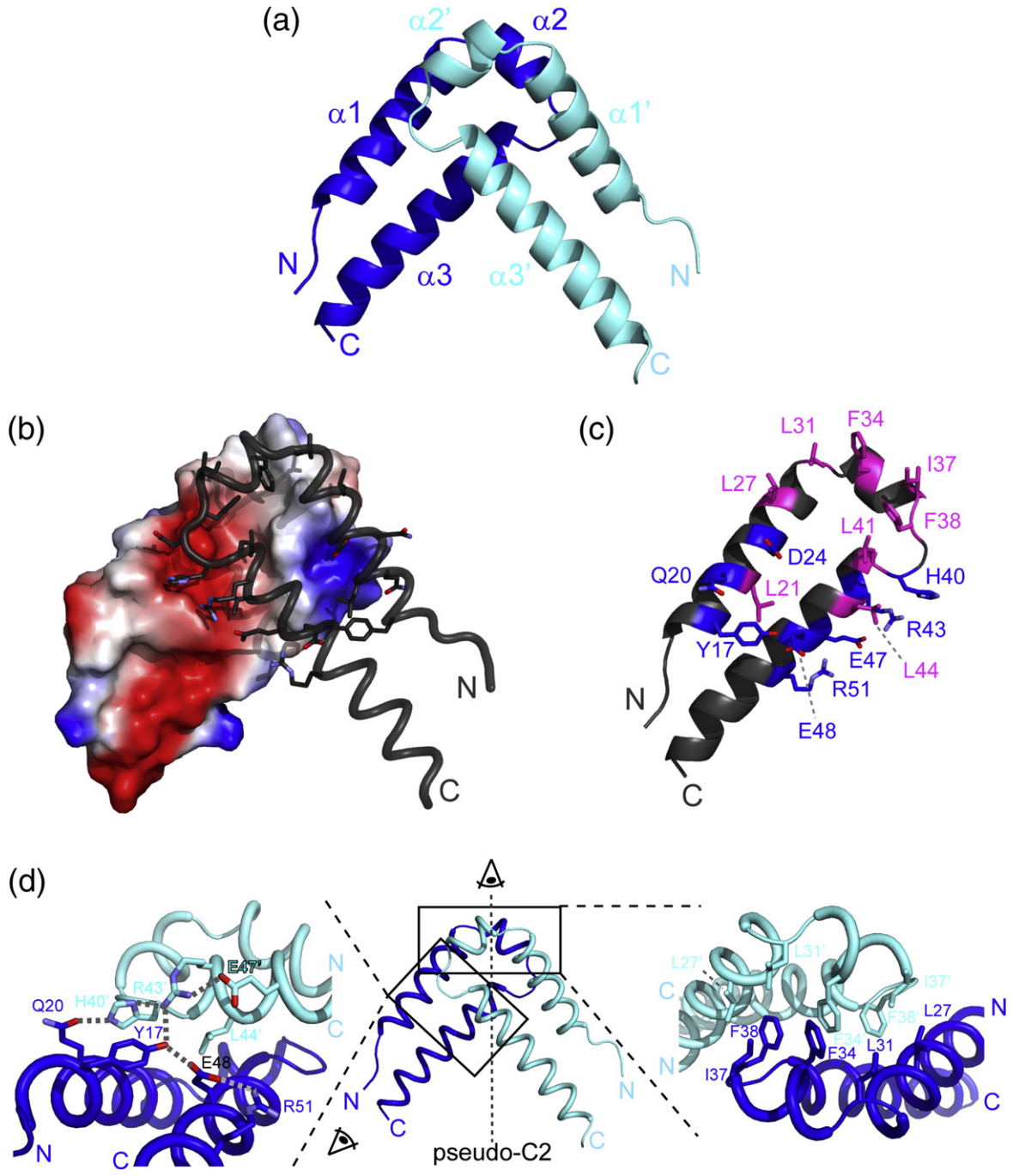

Fig. 2. Overall structure of the QkI Qua1 homodimerization subdomain. (a) Structure of the QkI Qua1 homodimer. Monomers A and B are colored dark blue and light blue, respectively. (b) Dimer interface. Monomer A is shown as electrostatic surface potential, and monomer B is shown as tube with dimer interface residues as sticks. (c) Residues of the QkI Qua1 homodimer interface. The residues that form the hydrophobic core of the interface, containing the conserved Phe34 and Phe38, are highlighted in pink. Residues participating in hydrogen bonds at the edge of the interface are shown in blue. (d) Close-up view on the dimer interface. The prime denotes residues in the other protomer. Side chains of key residues in the dimer interface are shown as sticks, and hydrogen bonds are indicated by gray broken lines.

packing. Furthermore, L21A's bottom face participates in the zipper by packing against I49.

The lack of the Y17–E48 hydrogen bond is associated with a lethal phenotype for mouse Quaking E48G. With the comparison of the E48G and E48A mutants, however, it becomes apparent that the effect of glycine in this position is much greater than simply removing a hydrogen bond.

With the consideration that glycine residues are known to break α-helices, it is likely that loss of secondary structure in helix α3 in the E48G mutant adds significantly to the dimer destabilization.

To investigate the effect of the point mutations within the Qua1 subdomain in the context of the full QkI STAR domain, we cloned the same point mutations into QkI STAR (1–205) fused to maltose-binding protein (MBP). The C35S mutation that prevents aggregation was also introduced into the QkI STAR mutants. This substitution does not affect the RNA binding affinity of QkI STAR (Table 2). The RNA binding affinity of the mutants was measured using a fluorescence polarization (FP) binding assay.[29] The MBP QkI STAR point mutants were titrated to a constant concentration of the previously identified high-affinity RNA

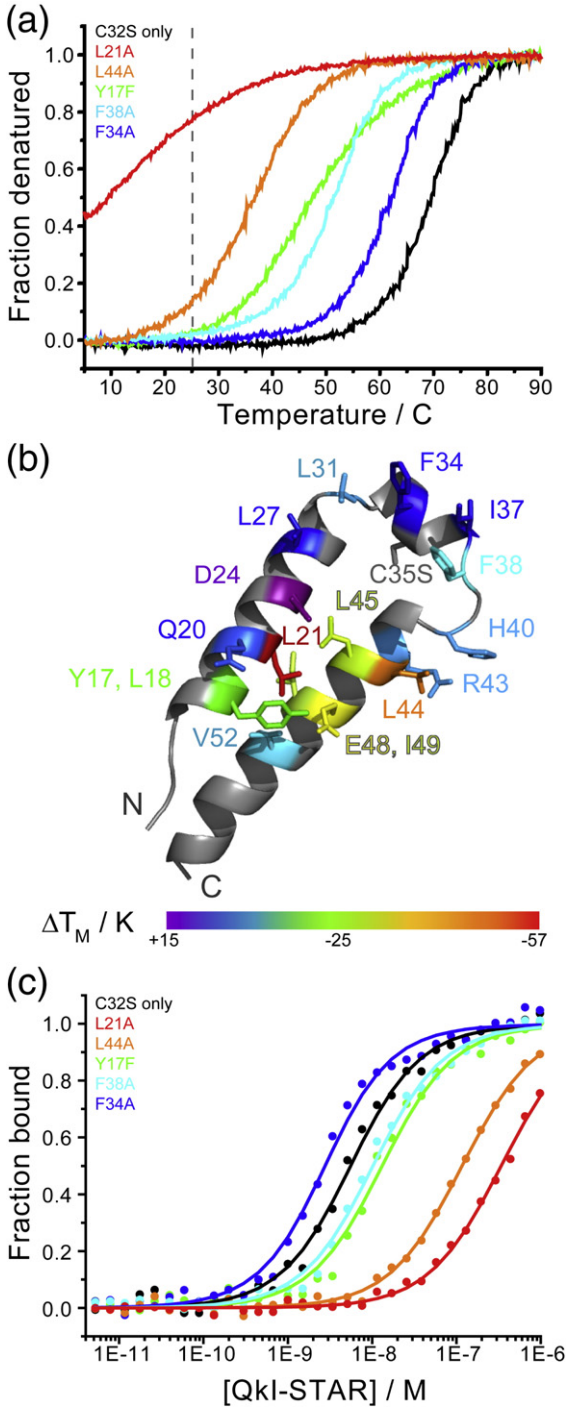

Fig. 3. Qua1 point mutations that destabilize homodimerization also impair RNA binding. All mutants are in the C35S background and are therefore compared to QkI Qua1 C35S. (a) Representative CD melting curves for QkI Qua1 point mutants. Room temperature, at which the RNA binding experiments (c) were performed, is indicated by the gray broken line. (b) ΔT_m mapped onto the monomer structure with residue probes shown as sticks. (c) Representative RNA binding curves for QkI STAR constructs with point mutations in the Qua1 domain. The RNA binding experiments were performed at room temperature [indicated by a gray broken line in (a)]; thus, only those constructs that are significantly destabilized at this temperature are expected to show a significant impact on the RNA binding affinity. The data were fit to the quadratic binding equation for bimolecular binding. The Hill equation does not apply because the RNA concentration cannot be considered "in trace" (probe concentration $\ll 10\ K_d$) for most QkI STAR constructs.

Table 2. Thermal stability of the QkI Qua1 dimer mutants and RNA affinities of QkI STAR mutants

QkI mutant	Qua1 domain (residues 14–67)		Full STAR domain (residues 1–205)	
	T_m (°C)	ΔT_m (K)	K_d (nM)	K_d/K_d (C35S)
wt	n.d.	n.d.	3.4±0.2	0.7
C35S only	69	0	5.1±0.3	1.0
Zipper				
L18A	49	−20	11.6±0.6	2.3
Q20A	65	−4	4.1±0.3	0.8
L44A	37	−32	118±5	23
L45A	40	−29	209±9	41
I49A	38	−31	24±1	4.7
V52A	60	−9	3.8±0.2	0.7
Dimer interface and zipper				
Y17F	45	−24	12.8±0.7	2.5
L21A	<15	<−57	340±30	67
E48A	40	−29	6.7±0.4	1.3
E48G	<15	<−57	22±1	4.3
Dimer interface				
D24A	>85	>+14	4.3±0.2	0.8
L27A	67	−2	3.5±0.2	0.7
L31A	62	−7	n.d.[a]	n.d.[a]
F34A	64	−5	2.3±0.1	0.5
I37A	67	−2	4.4±0.3	0.9
F38A	53	−16	9.8±0.6	1.9
H40A	58	−11	10.8±0.6	2.1
R43A	57	−12	9.0±0.5	1.8

ΔT_m and the K_d ratio are based on the C35S construct since all mutants are within the C35S background.
[a] QkI STAR L31A could not be expressed.

sequence 5′-UAUUUAAUUUCUUAUC**UACUAA**-UAUCUA-3′ labeled with fluorescein as probe (the high-affinity hexamer consensus is in boldface). The dissociation constants (K_d) for both wild type (wt) and C35S-only mutant agree well with the published values.[29] RNA binding is significantly impaired for only those point mutants that drastically decrease the stability of the QkI Qua1 homodimer, reflected by a melting temperature of the QkI Qua1 construct of $T_m \leq 45\,^{\circ}C$, namely, Y17F, L21A, L44A, L45A, E48G and I49A (Table 2). Since the RNA binding assay is performed at room temperature, a drastic effect on RNA binding is only observed for those mutants that are already significantly destabilized at this temperature (Fig. 3c). Interestingly, the mutants E48G and L21A that have similar low melting temperatures in the Qua1 construct show quite different RNA binding affinities. Since different length constructs were used in the different assays, it is possible that destabilization caused by any point mutation could be different in the presence of other parts of the protein compared to an isolated domain, especially since the E48G mutation likely combines at least two destabilizing effects: deletion of a stabilizing hydrogen bond and disruption of secondary structure.

All mutant Qua1 and STAR constructs were analyzed on SDS-PAGE and size-exclusion chromatography (data not shown) to ensure intact protein, equal concentrations and homogeneity. All mutants elute in a single peak, with the exception of the D24A mutant in both Qua1 and STAR constructs. The retention time of destabilized mutants slightly increased compared to wt in accordance with the dimer stability measured by CD spectroscopy, but not as much as it would be expected for a globular monomeric form, which might reflect either a fast equilibrium between monomeric and dimeric forms or an increase of hydrodynamic radius due to unfolding. The D24A mutation causes the appearance of an additional higher-molecular-weight species for both Qua1 and STAR domain constructs, indicating that this mutation might cause aggregation, which could explain the surprisingly high melting temperature of the D24A Qua1 construct.

The mutational analysis shows that QkI homodimerization is correlated to its ability to bind RNA, and mutations that disrupt the homodimer also confer a reduced RNA binding affinity. In addition to the QkI Qua1 dimer interface residues themselves, the conserved zipper serves an important function for both homodimer stability and RNA substrate binding. The zipper is well conserved among STAR proteins. It stabilizes the tertiary structure of the α-helical hairpin and ensures the proper display of the dimer interface residues.

A functional Qua1 dimerization domain is required for robust splicing activation by QkI-5 in muscle cells

Many exons that are activated in muscle development are enriched for a motif similar to the QkI consensus downstream of the activated exon, such as the actin capping protein β-subunit, Capzb, which features two ACUAA sites in the 19–80 region of its exon 9 downstream intron.[37] In mouse C2C12 myoblasts, alternative splicing is regulated during differentiation to myotubes,[38,39] and the STAR motif is enriched near exons that change[38] (M. Hall and M. Ares, unpublished results), including Capzb exon 9. An *in vivo* splicing assay using a Capzb exon 9 minigene construct in C2C12 cells (M. Hall and M. Ares, unpublished results) was used to investigate whether a functional Qua1 domain is required for alternative splicing regulation by QkI-5 (Fig. 4 and Fig. S5). Overexpression of wt Myc-tagged QkI-5 (Myc-QkI-5) increases Capzb exon 9 inclusion up to 10-fold in a dose-dependent manner, saturating at a maximum exon 9 inclusion of 77.5%. Four selected mutants (E48G, E48A, Y17F and L21A) that showed significantly impaired homodimerization in the QkI Qua1 thermal melting assay were investigated in the context of full-length Myc-QkI-5 using the Capzb splicing assay in C2C12 cells. Protein expression was reduced for all mutants, with the lowest protein levels being observed for the mutants E48G and L21A. Capzb exon 9 inclusion is significantly reduced by 20–74% for all four mutants compared to wt, with the E48G and L21A mutants showing the most drastic effect. However, since the protein levels for the QkI-5 mutants vary significantly, these data alone do not reveal whether the decrease of exon inclusion is caused by a reduced splicing efficiency of the point mutants or it is simply by the lower protein levels. Varying amounts of wt Myc-QkI-5 expression plasmid were transfected into C2C12 cells along with a constant amount of the Capzb minigene construct. Capzb exon 9 inclusion values were plotted against Myc-QkI-5 expression levels and fit to a simple saturation model, reflecting the per-molecule splicing efficiency. Adding the QkI-5 mutants to the plot allows comparison of their splicing efficiency to wt Myc-QkI-5 (Fig. 4c and Fig. S5). Points falling on the fitted curve represent a splicing efficiency equal to wt protein, while points falling under the curve imply a reduced splicing efficiency and points above the curve imply an enhanced splicing efficiency. Furthermore, the splicing efficiency for each mutant was calculated as the ratio of observed exon inclusion to the expected wt value for the given mutant expression level that was obtained from the wt titration fit. The data points for all four mutants fall under the wt titration curve, and the splicing efficiency is decreased by 17–73% compared to wt, indicating that the mutations

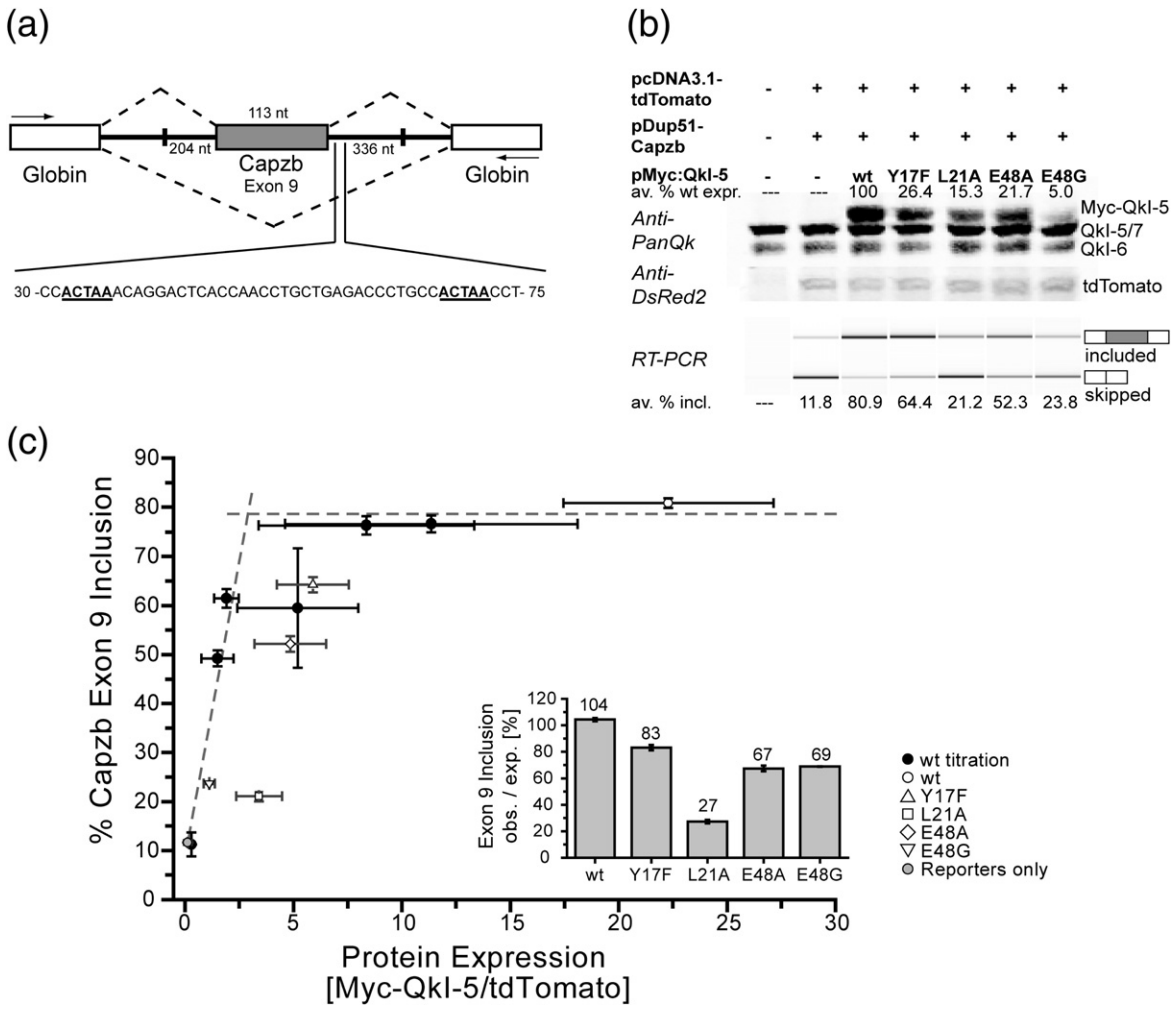

Fig. 4. Dimerization-deficient point mutations in the Qua1 domain reduce QkI-5 splicing activity *in vivo*. (a) Schematic presentation of the Capzb exon9 minigene organization. White boxes represent the globin exons, and the gray box represents the Capzb exon 9. Numbers by the introns flanking the Capzb exon 9 indicate the length of intron sequence that was cloned from the Capzb gene. PCR primer positions are indicated as arrows. The 30- to 75-nt region downstream of the Capzb exon 9 contains two QkI consensus binding sites, shown in boldface and underscored, which are essential for QkI-5 splicing activity in C2C12 myoblast cells. (b) Western blot (top) and RT-PCR analysis of QkI-5 expression levels and minigene splicing patterns from cytoplasmic RNA of C2C12 myoblast cells co-transfected with the Capzb exon 9 minigene construct (pDup51-Capzb), plasmid for expression of wt or mutant QkI-5 protein and a tdTomato expression vector as transfection control. Each QkI-5 construct was assayed in three independent co-transfection experiments. In the Anti-PanQk blot, the two bottom bands represent endogenous QkI, while the top band represents the overexpressed Myc-QkI-5. The Anti-DSRed blot serves as transfection control. Average values for QkI-5 construct expression relative to wt expression are given above the Western blot. The average percentage of Capzb exon 9 inclusion for each construct is given below the RT-PCR gel. Plots of these values including standard deviation can be found in Fig. S5. (c) Capzb exon 9 inclusion plotted *versus* protein expression levels. In order to normalize the splicing efficiencies of the QkI-5 mutants, which show very different protein expression levels, we transfected varying amounts of wt Myc-QkI-5 expression vector in three independent titration experiments. The protein expression level (normalized to tdTomato transfection control) and Capzb minigene splicing efficiency were analyzed as described for the QkI-5 mutants and plotted against each other. Plots of these values including standard deviation can be found in Fig. S5. All individual data points were used to fit a simple saturation model (Fig. S5). This plot shows the wt titration (black circles) and mutant (open symbols) data averaged over the three replicates. Error bars represent one-half standard deviation. The fitted saturation curve is presented as two intersecting broken lines. Points falling below this curve indicate a lower splicing efficiency per QkI-5 molecule compared to wt, while the area above the curve indicates a higher splicing efficiency. The inset shows the observed-to-expected ratio of splicing efficiency (PSI, percent splicing included).

significantly affect the ability of QkI-5 to activate splicing in C2C12 cells.

Discussion

STAR family members show structural differences in their dimer interface

The STAR domain is highly conserved and defines the family of STAR/GSG proteins in metazoans. The Qua1 subdomains across this family show a high degree of sequence similarity, implying that the structures of the Qua1 homodimers are likely similar; however, the fraction of identical residues between homologs is only 30%. Interestingly, the monomer zipper between the two main α-helices shows high sequence identity, while significant variability is observed within the hydrophobic homodimer interface, including the turn region that links the two main helices.

The Quaking subfamily exhibits a characteristic 3-residue insertion within this turn region that is not present in other paralogs. A homology model of the QkI Qua1 domain using the GLD-1 Qua1 structure[30] as template, consistent with secondary structure prediction from the QkI Qua1 sequence,[1] yielded a model where the two helices in the helix–turn–helix motif overlap exactly and the longer linker simply results in a further extending turn (Fig. S4), prompting the initial attempt using MR to solve the QkI Qua1 structure.

The structure of the QkI Qua1 subdomain shows that the two main helices forming the helical hairpin fold are in fact conserved in the QkI Qua1 structure. Interestingly, the top of the hairpin fold stands out because it does not simply consist of a larger turn lacking secondary structure as predicted but forms a short additional α-helix of about 1.5 turns, which lies perpendicular to the two main helices (Figs. 2 and 5). Formation of this third helix is enabled by the Quaking-subfamily-specific 3-residue insertion that is not present in the GLD-1 and Sam68 Qua1 structures,[30,31] which likely explains why MR using the GLD-1 or Sam68 Qua1 structure was not successful.

The QkI Qua1 dimer interface is enlarged by the inserted helix and covers 815 Å^2 per monomer, compared to 700 Å^2 per monomer for GLD-1[30] and 624 Å^2 per monomer for Sam68.[31] The larger interface allows for a more extended hydrogen bonding network at the edge of the dimer interface that, in contrast to GLD-1 and Sam68, also includes the conserved monomer clamp residues Y17 and E48 (Fig. 2d). The differences in surface area and number of interactions are also reflected in the melting temperatures of the QkI (69 °C), GLD-1 (63 °C)[30] and Sam68 (<50 °C)[31] Qua1 constructs. The quaternary dimer structures of GLD-1 and Sam68 Qua1 domains overlay well in the helical segments for both protomers with the main differences in the turn region (backbone RMSD=1.1 Å),[31] and both show a 90° angle between the two protomers. In contrast, the protomers in the QkI Qua1 dimer are stacked at an angle of only 84° (Fig. 5c).

Within the QkI Qua1 dimer interface, the conserved F38 residue and F34 are located close to the hairpin turn region and form the center of the hydrophobic dimer interface. F34 stacks face to face with its own counterpart F34′ in the second protomer. The F34–F34′ interaction lies over the dimer symmetry axis and requires one side chain to face toward the center of the dimer interface while the other one has to rotate outward, which breaks the symmetry between the two protomers. Each of the F34 residues also engages in a face-to-edge stacking interaction with F38 within the same protomer (Fig. 2d). A similar Phe π-sandwich structure is also observed in GLD-1 Qua1, but the position of these side chains within the dimer interface is shifted noticeably. Sam68 with only one Phe (F118) per protomer does not use Phe π-stacking interactions at all but achieves hydrophobic packing with Leu (L114) residues. These differences in the dimer interface between STAR family members might serve to enforce homodimerization as opposed to potential heterodimerization of homologous STAR proteins, for example, QkI and Sam68.

Interestingly, the highest degree of sequence identity in the Qua1 subdomain between QkI, GLD-1 and Sam68, as well as other STAR family members, is found in the zipper residues that form the contact between the two main helices in the helical hairpin fold. The zipper is important for dimerization because it stabilizes the tertiary structure of the monomer and ensures a proper display of the dimer interface residues. This stabilizing effect was also observed for the homologous GLD-1 Qua1 homodimer.[30] The hydrophobic dimer interface residues on the other hand show similarity in that the hydrophobic properties are conserved, but only the Pro residue (P32 in QkI), ending helix α1 and inducing the (first) turn, and one Phe residue (F38 in QkI) are invariant throughout the STAR family. Overlaying the QkI, GLD-1 and Sam68 Qua1 structures (Fig. 5b), it becomes apparent that even the invariant dimer interface residues vary in their spatial arrangement resulting in noticeable variations in the details of interface surface shape, even though the general placement of the surface ridges and grooves and the electrostatic surface potential seems similar (Fig. S3).

Dimer stability and its effect on RNA binding *in vitro*

The dimeric STAR proteins bind to bipartite RNA sequences where each monomer is believed to bind

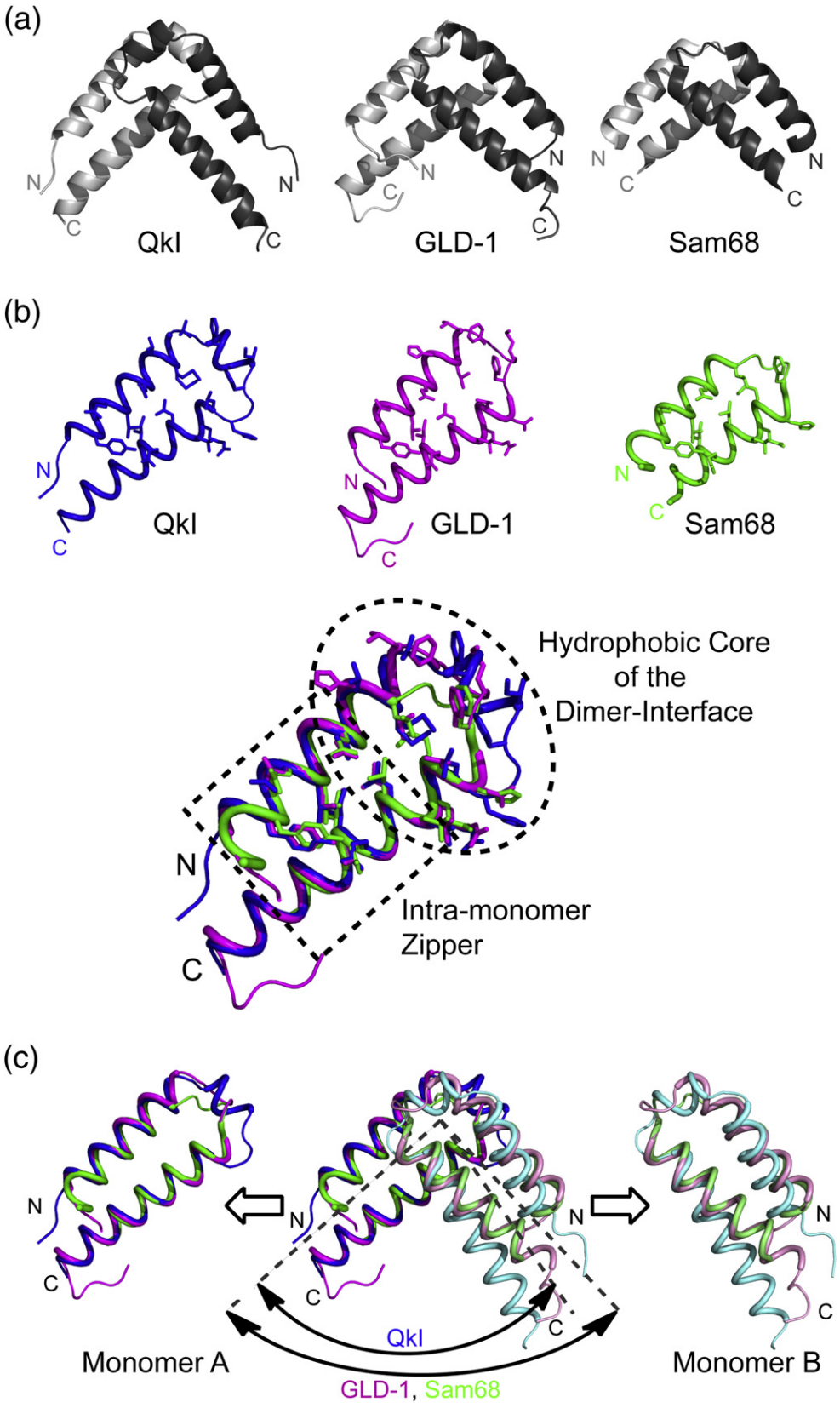

Fig. 5 (*legend on next page*)

to one of the two RNA half-sites.[9–11] Homodimerization is essential for enabling this bipartite binding mode. Close proximity of the half-sites, as well as high local concentration, explains the tighter binding of a STAR dimer to a full site *in vitro* compared to binding of a dimerization-deficient monomeric STAR mutant. For GLD-1, the KH–Qua2 subdomain alone is sufficient to bind RNA, but the absence of the Qua1 domain reduces the RNA binding affinity about 1 order of magnitude.[9] This loss in binding affinity can be explained by the fact that each monomeric GLD-1 protein can only bind to one hexamer consensus site, while the dimer binds a bipartite target site that contains at least one high-affinity consensus element.[9,29] Deletion of the QkI Qua1 domain did not abrogate RNA binding despite the lack of dimerization.[26] However, this experiment was of qualitative nature and did not allow quantitating any changes in QkI RNA binding affinity upon Qua1 deletion.

Mutational analysis of the QkI dimer interface and zipper residues shows that, in addition to the dimer interface residues, the hydrophobic zipper that stabilizes the helix–turn–helix fold of the monomer is essential for homodimerization. Zipper mutations have, on average, a larger effect on the stability of the dimer than mutations within the dimer interface itself as they ensure proper tertiary structure and thus present the dimer interface residues in the correct orientation. This effect was also observed for the GLD-1 Qua1 domain[30] and is consistent with the fact that the zipper residues are more highly conserved within the STAR protein family than the dimer interface residues. Mutational analysis of the Qua1 subdomain within the QkI STAR domain context reveals that RNA binding is indeed significantly reduced in those QkI Qua1 mutants that impair dimerization, whether at the dimer interface or in the zipper, linking QkI dimer stability to its biological functions.

What is the role of homodimerization in QkI function *in vivo*?

To test Qk function *in vivo*, we monitored its role as a direct splicing factor in muscle cells (M. P. Hall *et al.*, unpublished results) using a tranfected splicing reporter carrying Capzb exon 9, which contains two QkI consensus sites in its downstream intron. Exon 9 inclusion is promoted by co-transfection of wt Myc-tagged QkI-5 and requires the presence of the QkI binding sequences near exon 9 in the reporter (Hall and Ares, unpublished results), indicating that QkI-5 binding directly to the intron is required to achieve exon inclusion.

The four QkI mutations (Y17F, L21A, E48A and E48G) tested in the *in vivo* splicing assay significantly impact the stability, as well as the ability to homodimerize, of the QkI Qua1 domain and the RNA binding affinity of the QkI STAR domain *in vitro*. Their significantly reduced splicing activity observed in C2C12 myoblast cells correlates well with the *in vitro* data. Interestingly, the reduction of exon inclusion for the mutants is caused by both a diminished splicing efficiency and lower protein expression levels. The observed lower protein levels suggest that the dimerization-deficient mutants with a likely unfolded Qua1 domain are more prone to protein degradation in C2C12 cells. The impaired splicing efficiencies (Fig. 4) for the mutants correlate well with their *in vitro* RNA binding affinities (Table 2). The L21A mutation confers the lowest RNA binding affinity and splicing efficiency, while both mutants E48A and E48G are being equally impaired in both functions and the least destabilizing Y17F mutant confers the least drastic, albeit still significant, loss of function. Overall, the splicing activity of the QkI point mutants correlates well with their ability to homodimerize, indicating that QkI homodimerization is required for QkI splicing activity *in vivo*.

Homodimerization seems to be important for STAR protein biological function beyond a simple increase in RNA binding affinity. Mutants that fail to dimerize show a loss-of-function phenotype, for example, QkI E48G causes embryonic lethality.[26] It is possible that *in vivo* STAR protein dimers bind to two distant sites looping and arranging the RNA to facilitate binding of other RNA-binding proteins of the spliceosome or transcription regulation complexes.[40] The RNA binding and splicing activity of QkI and other STAR proteins is regulated by phosphorylation,[12–14] and it has been proposed that phosphorylation might also be used to disrupt the

Fig. 5. Comparison of the QkI Qua1 structure (PDB ID 4DNN) to the GLD-1 (PDB ID 3K6T) and Sam68 (PDB ID 2XA6). (a) Ribbon presentation. QkI Qua1 features an additional three residues in the turn region that enable to form an additional short helix and expand the surface area of the dimer interface. (b) Comparison of dimer interface and zipper residues (shown as sticks) of QkI, GLD-1 and Sam68 Qua1 monomers and overlay of the three structures. While the monomer zipper is highly conserved, more variation is found in the hydrophobic dimer interface residues. All three proteins use stacking of conserved Phe residues as the core of the dimer interface, but the exact position of these residues varies to some degree. The additional helix in the QkI structure broadens the dimer interface significantly and allows for more hydrogen bond interactions across the edge of the interface (see also Fig. 2). (c) While the two main helices in each monomer overlay precisely for all three Qua1 domains, the two QkI Qua1 monomers are stacked at a narrower angle compared to the GLD-1 and Sam68 homodimers.

homodimer and switch to a monomeric, inactive form of the protein.[31] Several of the tyrosine residues implicated in phosphorylation are conserved throughout the STAR protein family, suggesting that regulation of dimeric *versus* monomeric forms by phosphorylation might be used as a general switch. It is also possible that other proteins recognize and bind to a STAR protein dimer but not a monomer, or *vice versa*.

QkI exists in multiple alternatively spliced Quaking isoforms, the most prominent ones being QkI-5, QkI-6 and QkI-7. All isoforms are identical in the 1–311 region including the full STAR domain. The C-terminal 8–30 residues differ between isoforms and convey specific functions. QkI-5 (312–341) contains a nuclear localization signal. The 14 C-terminal residues (312–325) of the predominantly cytoplasmic QkI-7 carry an apoptosis-inducing signal. QkI-6 uniquely features an 8-residue tail (residues 312–319) and is found both in the nucleus and in the cytoplasm. Heterodimerization between isoforms can occur,[16] but it is not well understood under what circumstances these isoforms can heterodimerize in cells and how this affects QkI function. When co-transfected and overexpressed in HeLa cells, QkI-5 forms heterodimers with QkI-6 and QkI-7 and self-homodimers. Furthermore, QkI-5 was able to shuttle QkI-7 into the nucleus and suppress the induction of apoptosis by QkI-7. These results are a first indication that heterodimerization between different isoforms might be used to control protein localization and that the balance in the QkI isoforms is critical for the normal function of the QkI proteins and cell viability.

In addition to multiple Quaking isoforms, there are other STAR proteins such as Sam68 present in both the cytoplasm and the nucleus. With the consideration that the topologies of the Qua1 dimer interfaces in the STAR family are similar but not identical, it could be possible that STAR proteins form inter-paralog heterodimers. QkI has been shown to interact with co-transfected GLD-1 but not Sam68 in an immunoprecipitation assay in HeLa cells.[25] The structural differences between STAR family members within the dimer interface likely serve to specify and select dimerization partners. STAR protein heterodimers would be expected to form with a lower affinity, if at all, and thus might be controlled by protein concentrations in the cell, which would add another level of functional regulation. To shed light on these questions, we need further studies addressing the role of STAR protein dimerization *in vivo*, both between isoforms and different family members, and how dimerization affects interaction with RNA and other regulatory proteins. In summary, we have determined the structure of the QkI homodimerization domain and correlated the formation of dimers to RNA binding *in vitro* and to activation of splicing in cells, thus providing insights into how STAR proteins function in regulation of gene expression.

Materials and Methods

Protein expression and purification

MBP-tagged pMAL-QkI-STAR(1–205) was expressed in *Escherichia coli* strain JM109 (New England Biolabs) and purified via affinity chromatography using amylose resin (New England Biolabs), followed by a HiTrap Q anion-exchange column (GE Healthcare), as described previously,[9] but omitting the cation-exchange chromatography step. Purified protein was stored at −80 °C. Cys35 was changed to Ser for consistency with the Qua1 constructs, and point mutants were generated in the C35S background using QuikChange (Agilent).

QkI Qua1 C35S constructs were cloned into pET22b(+) (Novagen), which results in a C-terminal hexahistidine (His) tag, or a pET22b(+)-derived expression vector, introducing an N-terminal His tag and a thrombin cleavage site (MHHHHHHLVPRGS). For a complete list of constructs cloned and tested for expression, see Table S1. All QkI Qua1 constructs, including point mutants, were expressed in *E. coli* BL21-Gold (DE3) cells (Stratagene).

The N-terminally His-tagged QkI Qua1 (14–67) was used for all structural studies and dimer stability experiments. Single point mutations in the QkI Qua1 and QkI STAR context were introduced in the C35S background using QuikChange (Agilent). Unlabeled QkI Qua1 protein was expressed in LB media with 100 mg/l ampicillin for 3 h at 37 °C with 1 mM IPTG, added at a cell density of OD_{600} = 0.6–0.8. Six constructs were labeled with ^{15}N for NMR spectroscopy and were expressed for 6 h at 37 °C in M9 minimal media containing $(^{15}NH_4)_2SO_4$ as the sole nitrogen source, unlabeled glucose and supplemental trace metals and vitamins. SeMet-labeled QkI Qua1 was expressed for 6 h at 37 °C in M9 minimal media containing NH_4Cl as nitrogen source, unlabeled glucose, supplemental trace metals and vitamins and was supplemented with the amino acids SeMet, Lys, Thr, Phe, Leu, Ile, Val and thiamine.

Unlabeled and ^{15}N-labeled QkI Qua1 samples for thermal melting experiments were purified in 20 mM sodium phosphate (pH7.5) buffer, while the SeMet sample for crystallography was purified in 20 mM Tris–HCl (pH7.5) buffer. Cells were lysed by sonication in buffers containing 15 mM imidazole and Complete ethylenediaminetetraacetic acid-free protease inhibitor cocktail (Roche). Soluble protein was purified in two steps by Ni-affinity chromatography followed by anion-exchange chromatography. The lysate was passed over a 5-ml HisTrap HP column (GE Healthcare), washed with 25 ml of purification buffer with 15 mM imidazole, followed by 15 ml of purification buffer with 20 mM imidazole and finally eluted in a single step with 25 ml of purification buffer with 200 mM imidazole. Pure unlabeled protein fractions were dialyzed against 20 mM sodium phosphate buffer (pH7.5) at 4 °C overnight. For SeMet-labeled protein fractions, thrombin (5 U/mg of protein) was added, and the protein was dialyzed against 20 mM

Tris–HCl (pH7.5) with 0.02% NaN_3 at room temperature for 24 h. Complete thrombin cleavage was confirmed by SDS-*N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine-PAGE. The protein was further purified by anion-exchange chromatography on a 5-ml HiTrap Q HP column (GE Healthcare), where SeMet-labeled QkI Qua1 eluted with a 120-ml gradient from 0 to 1 M NaCl, while unlabeled, still His-tagged, QkI Qua1 constructs were found in the flow through fraction. Unlabeled samples for thermal melting experiments were dialyzed against 20 mM sodium phosphate (pH7.5) and 0.02% NaN_3, while SeMet-labeled protein for crystallization was dialyzed against 20 mM Tris–HCl (pH7.5) and 0.02% NaN_3. All protein samples were concentrated in Centriprep concentrators (3-kDa cutoff; Millipore), and purified protein was stored at 4 °C. Protein concentration was determined by measuring the UV absorption at 280 nm. All protein samples (50 μM for QkI Qua1 and 30 μM for QkI STAR) were run on an SDS-PAGE gel to ensure sample integrity and equal concentrations in the assays.

NMR spectroscopy

The following constructs were labeled with ^{15}N to study by NMR spectroscopy: QkI Qua1 (14–77)-C-His_6, QkI Qua1 (14–77)-N-His_6, QkI Qua1 (12–77)-N-His_6, QkI Qua1 (14–63)-C-His_6, QkI Qua1 (12–60)-C-His_6 and QkI Qua1 (14–55)-C-His_6. NMR samples were prepared in 20 mM sodium phosphate (pH7.5) and 0.02% NaN_3 containing 10% D_2O/90% H_2O with a final protein concentration of 0.5 mM. NMR spectra were recorded at 35 °C on a 750-MHz Bruker Avance spectrometer with 5 mm TXI/HCN triple resonance probes.

Analytical size-exclusion chromatography

All QkI Qua1 and STAR mutants were subjected to analytical size-exclusion chromatography at room temperature to ensure sample integrity and multimerization status. QkI Qua1 proteins were loaded onto a Superdex 75 10/300 gel-filtration column (GE Healthcare) and eluted in 20 mM sodium phosphate buffer (pH7.5). QkI STAR proteins were loaded onto a Superdex 200 10/300 gel-filtration column (GE Healthcare) and eluted in 50 mM Tris–HCl (pH7.9), 20 mM NaCl and 2 mM DTT. Gel-filtration standard (Bio-Rad) with thyroglobulin (bovine, 670,000 Da), γ-globulin (bovine, 158,000 Da), ovalbumin (chicken, 44,000 Da), myoglobin (horse, 17,000 Da) and vitamin B12 (1350 Da) was used to calibrate molecular weight.

Protein crystallography

Protein crystallization

SeMet-QkI-Qua1(14–67) was crystallized by vapor diffusion in 24-well sitting-drop plates, 2 μl drop volume, at 22 °C. Crystals of SeMet-QkI-Qua1 were obtained by mixing 1 μl of protein (5 mg/ml) in 20 mM Tris–HCl (pH7.5), 0.02% NaN_3 with 1 μl of 0.1 M sodium cacodylate (pH6.5), 0.2 M calcium acetate and 29% (v/v) polyethylene glycol 600. Crystals were flash-frozen in liquid nitrogen without additional cryoprotectant.

Data collection and processing

Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) BL 11-1 processed with HKL2000.[41] MR was attempted using the PHENIX[42] AutoSol. With the use of the Sam68 Qua1 (PDB ID 2XA6) monomer or dimer, no solution was obtained. With the use of GLD-1 Qua1 monomer or dimer, solutions in space group $P2_12_12_1$ were found. However, the electron density maps were poorly defined, especially in the turn region. The log-likelihood gains of LLG=34 for GLD-1 Qua1 dimer and LLG=79 for GLD-1 Qua1 monomer as model were both lower than the acceptable value of 100‡, indicating a poor quality of these solutions. The PHENIX package[42] was also used for multiwavelength anomalous dispersion phasing of SeMet-QkI-Qua1. Initial model building after density modification was carried out in PHENIX AutoBuild, followed by alternating rounds of refinement with PHENIX Refine and manual model building and evaluation in Coot.[43] In later rounds of refinement, TLS refinement using each monomer as an independent TLS group was performed in addition to coordinate and atomic *B*-factor refinement. Adding NCS restrains during refinement did not improve the statistic and was therefore not used. Iterative rebuilding and refinement converged on a final model with R=21.6% and R_{free}=25.2% and good stereochemistry (98% of residues in the most favored region/2% of residues in the additionally allowed region). For complete crystallographic statistics, see Table 1.

Thermal melting experiments by CD spectroscopy

Thermal melting experiments were performed using a Jasco J-815 CD spectrometer with Jasco PFD-425S/15 temperature unit and a 1-mm-quartz cuvette. The samples consist of 50 μM protein in 20 mM sodium phosphate (pH7.5). Melting curves were recorded between 5 °C and 90 °C in 0.2 °C steps with a heating rate of 1 °C/min monitoring the signal at 222 nm. The raw CD data were normalized to fraction denatured and corrected for the slopes for folded and unfolded signals according to:

$$F_u(T) = \frac{\text{mdeg}-\text{min}-s_f\cdot(T-T_{min})}{\text{max}-\text{min}+s_u\cdot(T-T_{max})-s_f\cdot(T-T_{min})}$$

with F_u=fraction unfolded, T=temperature in degrees Celsius, mdeg=CD signal in millidegrees, min=minimum CD signal (at 5 °C), max=maximum CD signal (at 90 °C), T_{min}=5 °C, T_{max}=90 °C, s_f=slope for folded state and s_u=slope for unfolded state.

T_m was determined as the inflection point of the normalized melting curves.

Samples were run on an SDS-PAGE gel after the melting experiment to ensure sample integrity and equal concentrations in the assay.

Quantitative RNA binding assay using FP

Quantitative analysis of QkI STAR–RNA interaction was performed using an FP assay as described previously.[9] A

‡ http://www.phenix-online.org/documentation/

constant concentration of 1 nM fluorescein (Fluo)-labeled RNA (5′-Fluo-AUUUAAUUUCUUAUCUACUAAUAU-CUA-3′) was equilibrated with varying concentrations of QkI STAR in a total volume of 100 μl in 96-well opaque 200-μl Fluotrak plates (Greiner) for 3 h at room temperature. The blank-corrected polarization (mP) was measured using an EnVision 2104 Multilabel Reader (Perkin Elmer). Each experiment was performed in triplicate. The apparent equilibrium dissociation constant K_d was determined by fitting the FP data from all repeats globally using the quadratic binding equation for a bimolecular complex, keeping the RNA concentration R_0 constant:

$$F_B(P_0) = \frac{mP - mP_{max}}{mP_{max} - mP_{min}}$$

$$= \frac{P_0 + R_0 + K_d - \sqrt{(P_0 + R_0 + K_d)^2 - 4 \cdot P_0 \cdot R_0}}{2 \cdot R_0}$$

with F_B=RNA fraction bound, mP=polarization value in millipolarization units, mP_{max}=maximum mP value at saturation, mP_{min}=minimum mP value in the absence of protein, P_0=protein concentration, R_0=labeled RNA concentration and K_d=apparent equilibrium dissociation constant.

The Hill equation does not apply because the RNA concentration cannot be considered "in trace" (probe concentration $\ll 10\ K_d$) for most QkI STAR constructs. The data points shown in Fig. 3 are the average of the normalized F_B values.

QkI splicing assay

Plasmid construction

The pDup51 vector was used as a splicing reporter as previously described.[44] Exon 9 of Capzb, including 204 nt 5′ of the exon and 336 nt 3′ of the exon, was PCR-amplified and cloned into the ApaI and BglII sites in pDup51. The pMyc:Qk5 vector was a kind gift from Sean Ryder. Mutations were made by site-directed mutagenesis and verified by sequencing. The pcDNA3.1-tdTomato vector was a kind gift from Rohinton Kamakaka.

Cell culture and transfections

Mouse C2C12 myoblasts (American Type Culture Collection) were cultured in 1× Dulbecco's modified Eagle's media (Gibco/Invitrogen) supplemented with 10% fetal bovine serum (Gibco/Invitrogen) under sub-confluent conditions as recommended by the distributor. Cells were co-transfected in 6-well plates with 1 μg pcDNA3.1-tdTomato, 0.5 μg pDup51-CapzbEx9 and either 2 μg of mutant and wt pMyc:Qk5 (for mutant splicing assays) or varying amounts of wt pMyc:Qk5 (for Qk5 titrations). Transfections were performed using Lipofectamine2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested and RNA was extracted using TriZOL Reagent (Invitrogen), while protein was extracted by lysis in Laemmli buffer.

RT-PCR and bioanalyzer analysis

Reverse transcription (RT) was carried out on 2 μg total RNA using Superscript II (Invitrogen) according to the manufacturer's instructions. Equal volumes of resulting cDNA were used as PCR template for each sample. Reactions were carried out for 30 cycles with primers directed toward the flanking Dup51 reporter exons and visualized by ethidium bromide staining of agarose gel analysis. To measure percent inclusion for splicing, we purified PCR products then separated samples on the Agilent Bioanalyzer. This reported the size and concentration of both skipped and Capzb exon 9 included PCR products.

Western blotting

Equal volumes of cell lysate were loaded onto SDS-polyacrylamide gels and electrophoresed. Gels were transferred to pure nitrocellulose then blocked for 1 h at room temperature. Primary antibodies directed toward Pan-Quaking (Antibodies Incorporated) and DsRed2 (Clontech) were used at respective dilutions of 1:2000 and 1:500 and incubated with membranes overnight at 4 °C. Both IRDye 800CW conjugated goat anti-mouse IgG2b (Li-Cor) and IRDye 680LT conjugated goat anti-mouse IgG1 secondary (Li-Cor) antibodies were diluted to 1:20,000 and simultaneously used to probe the membrane for 60 min at room temperature. The resulting signals were visualized on a Li-Cor Odyssey Infrared Imager. Bands were quantified by corresponding infrared counts in each signal wavelength, and ectopic Myc-tagged Qk5 expression was normalized to the expression of the tdTomato transfection control.

Calculation of normalized splicing efficiency

wt titration data were plotted as Capzb exon 9 inclusion *versus* protein expression and fitted to a simple saturation model keeping the y intercept fixed to the value determined from the average of the reporter only data: $y=(21\pm3)\cdot x+11.8$ for $x\leq3.14$ and $y=(77\pm1)$ for $x>3.14$.

Mutant splicing efficiency relative to wt was calculated as the ratio of the averaged exon inclusion for the mutants divided by the expected exon inclusion for wt protein at the given mutant protein level according to the fit above.

Accession numbers

Coordinates and structure factors for QkI Qua1 (14–67) C35S have been deposited in the PDB with the accession code 4DNN.

Acknowledgements

We thank Ian Wilson for the use of X-ray equipment, Kurt Wüthrich for the use of CD spectrometer and Steven Brown for the use of the fluorescence plate reader. X-ray data were collected at SSRL, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the U.S. Department of Energy, Office of Basic

Energy Sciences, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. This work was supported by the National Institutes of Health grant GM053320 (J.R.W.), a Training Grant 5T32GM008646-14 (W.S.F.) and a postdoctoral fellowship by the Deutsche Forschungsgemeinschaft (C.B.).

Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2012.08.027

References

1. Artzt, K. & Wu, J. I. (2010). STAR trek: an introduction to STAR family proteins and review of quaking (QKI). *Adv. Exp. Med. Biol.* **693**, 1–24.
2. Vernet, C. & Artzt, K. (1997). STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**, 479–484.
3. Lobbardi, R., Lambert, G., Zhao, J., Geisler, R., Kim, H. R. & Rosa, F. M. (2011). Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development. *Development*, **138**, 1783–1794.
4. Noveroske, J. K., Lai, L., Gaussin, V., Northrop, J. L., Nakamura, H., Hirschi, K. K. & Justice, M. J. (2002). Quaking is essential for blood vessel development. *Genesis*, **32**, 218–230.
5. Wu, J. I., Reed, R. B., Grabowski, P. J. & Artzt, K. (2002). Function of quaking in myelination: regulation of alternative splicing. *Proc. Natl Acad. Sci. USA*, **99**, 4233–4238.
6. Saccomanno, L., Loushin, C., Jan, E., Punkay, E., Artzt, K. & Goodwin, E. B. (1999). The STAR protein QKI-6 is a translational repressor. *Proc. Natl Acad. Sci. USA*, **96**, 12605–12610.
7. Larocque, D., Galarneau, A., Liu, H. N., Scott, M., Almazan, G. & Richard, S. (2005). Protection of p27(Kip1) mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat. Neurosci.* **8**, 27–33.
8. Larocque, D., Pilotte, J., Chen, T., Cloutier, F., Massie, B., Pedraza, L. *et al.* (2002). Nuclear retention of MBP mRNAs in the quaking viable mice. *Neuron*, **36**, 815–829.
9. Ryder, S. P. & Williamson, J. R. (2004). Specificity of the STAR/GSG domain protein Qk1: implications for the regulation of myelination. *RNA*, **10**, 1449–1458.
10. Galarneau, A. & Richard, S. (2005). Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat. Struct. Mol. Biol.* **12**, 691–698.
11. 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.
12. 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.
13. Matter, N., Herrlich, P. & Konig, H. (2002). Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature*, **420**, 691–695.
14. Nir, R., Grossman, R., Paroush, Z. & Volk, T. (2012). Phosphorylation of the *Drosophila melanogaster* RNA-binding protein HOW by MAPK/ERK enhances its dimerization and activity. *PLoS Genet.* **8**, e1002632.
15. Ebersole, T. A., Chen, Q., Justice, M. J. & 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.
16. Pilotte, J., Larocque, D. & Richard, S. (2001). Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes Dev.* **15**, 845–858.
17. Wu, J., Zhou, L., Tonissen, K., Tee, R. & Artzt, K. (1999). The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm. *J. Biol. Chem.* **274**, 29202–29210.
18. Haroutunian, V., Katsel, P., Dracheva, S. & Davis, K. L. (2006). The human homolog of the QKI gene affected in the severe dysmyelination "quaking" mouse phenotype: downregulated in multiple brain regions in schizophrenia. *Am. J. Psychiatry*, **163**, 1834–1837.
19. Noveroske, J. K., Hardy, R., Dapper, J. D., Vogel, H. & Justice, M. J. (2005). A new ENU-induced allele of mouse quaking causes severe CNS dysmyelination. *Mamm. Genome*, **16**, 672–682.
20. Ebersole, T., Rho, O. & Artzt, K. (1992). The proximal end of mouse chromosome 17: new molecular markers identify a deletion associated with quaking-viable. *Genetics*, **131**, 183–190.
21. Zhao, L., Mandler, M. D., Yi, H. & Feng, Y. (2010). Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein. *Proc. Natl Acad. Sci. USA*, **107**, 19061–19066.
22. Zearfoss, N. R., Clingman, C. C., Farley, B. M., McCoig, L. M. & Ryder, S. P. (2011). Quaking regulates Hnrnpa1 expression through its 3′ UTR in oligodendrocyte precursor cells. *PLoS Genet.* **7**, e1001269.
23. Novikov, L., Park, J. W., Chen, H., Klerman, H., Jalloh, A. S. & Gamble, M. J. (2011). QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Mol. Cell. Biol.* **31**, 4244–4255.
24. Yang, G., Fu, H., Zhang, J., Lu, X., Yu, F., Jin, L. *et al.* (2010). RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology*, **138**, 231–240; e1–e5.
25. Chen, T., Damaj, B. B., Herrera, C., Lasko, P. & Richard, S. (1997). Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. *Mol. Cell. Biol.* **17**, 5707–5718.
26. Chen, T. & Richard, S. (1998). Structure-function analysis of Qk1: a lethal point mutation in mouse

quaking prevents homodimerization. *Mol. Cell. Biol.* **18**, 4863–4871.
27. Ryder, S. P., Frater, L. A., Abramovitz, D. L., Goodwin, E. B. & Williamson, J. R. (2004). RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. *Nat. Struct. Mol. Biol.* **11**, 20–28.
28. 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.
29. Carmel, A. B., Wu, J., Lehmann-Blount, K. A. & Williamson, J. R. (2010). High-affinity consensus binding of target RNAs by the STAR/GSG proteins GLD-1, STAR-2 and Quaking. *BMC Mol. Biol.* **11**, 48.
30. Beuck, C., Szymczyna, B. R., Kerkow, D. E., Carmel, A. B., Columbus, L., Stanfield, R. L. & Williamson, J. R. (2010). Structure of the GLD-1 homodimerization domain: insights into STAR protein-mediated translational regulation. *Structure*, **18**, 377–389.
31. Meyer, N. H., Tripsianes, K., Vincendeau, M., Madl, T., Kateb, F., Brack-Werner, R. & Sattler, M. (2010). Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain. *J. Biol. Chem.* **285**, 28893–28901.
32. Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., 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.
33. Maguire, M. L., Guler-Gane, G., Nietlispach, D., Raine, A. R., Zorn, A. M., Standart, N. & Broadhurst, R. W. (2005). Solution structure and backbone dynamics of the KH-QUA2 region of the *Xenopus* STAR/GSG quaking protein. *J. Mol. Biol.* **348**, 265–279.
34. Ali, M. & Broadhurst, R. W. (2008). Structural studies of Quaking protein Qua1 domain. In International Conference on Magnetic Resonance in Biological Systems, vol. XXIII, pp. 193, San Diego, CA.
35. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). Searching protein structure databases with DaliLite v. 3. *Bioinformatics*, **24**, 2780–2781.
36. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
37. Sugnet, C. W., Srinivasan, K., Clark, T. A., O'Brien, G., Cline, M. S., Wang, H. *et al.* (2006). Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2**, e4.
38. Bland, C. S., Wang, E. T., Vu, A., David, M. P., Castle, J. C., Johnson, J. M. *et al.* (2010). Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res.* **38**, 7651–7664.
39. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J. *et al.* (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.
40. Jungkamp, A. C., Stoeckius, M., Mecenas, D., Grun, D., Mastrobuoni, G., Kempa, S. & Rajewsky, N. (2011). *In vivo* and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell*, **44**, 828–840.
41. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr., Part A*, **276**, 307–326.
42. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W. *et al.* (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1948–1954.
43. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
44. Dominski, Z. & Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.* **11**, 6075–6083.