nature structural & molecular biology

RBFOX and SUP-12 sandwich a G base to cooperatively regulate tissue-specific splicing

Kanako Kuwasako[1–3,14], Mari Takahashi[1,2,14], Satoru Unzai[4], Kengo Tsuda[1,2], Seiko Yoshikawa[1], Fahu He[1], Naohiro Kobayashi[1,5], Peter Güntert[6–9], Mikako Shirouzu[1,2], Takuhiro Ito[1,2], Akiko Tanaka[1,2], Shigeyuki Yokoyama[1,10], Masatoshi Hagiwara[11,12], Hidehito Kuroyanagi[11,13] & Yutaka Muto[1–3]

Tissue-specific alternative pre-mRNA splicing is often cooperatively regulated by multiple splicing factors, but the structural basis of cooperative RNA recognition is poorly understood. In *Caenorhabditis elegans*, ligand binding specificity of fibroblast growth factor receptors (FGFRs) is determined by mutually exclusive alternative splicing of the sole FGFR gene, *egl-15*. Here we determined the solution structure of a ternary complex of the RNA-recognition motif (RRM) domains from the RBFOX protein ASD-1, SUP-12 and their target RNA from *egl-15*. The two RRM domains cooperatively interact with the RNA by sandwiching a G base to form the stable complex. Multichromatic fluorescence splicing reporters confirmed the requirement of the G and the juxtaposition of the respective *cis* elements for effective splicing regulation *in vivo*. Moreover, we identified a new target for the heterologous complex through an element search, confirming the functional significance of the intermolecular coordination.

Alternative splicing plays a substantial part in the production of appropriate protein isoforms for various developmental stages and tissue types[1], and a number of diseases have been reported to be associated with dysfunction of alternative splicing[2,3]. A variety of splicing-regulatory factors prerequisitely bind to specific sites on precursor mRNAs (pre-mRNAs) to control exon inclusion or exclusion[4–9]. However, the binding sequences for many of the RNA-binding proteins lack tight consensus, and therefore cooperative recognition of the target pre-mRNAs by multiple splicing-regulatory factors is expected for accurate splicing regulation *in vivo*.

RNA-binding protein, fox-1 homolog (*C. elegans*) (RBFOX) family proteins are well-documented and evolutionarily conserved splicing regulators with important roles in tissue-specific alternative splicing in metazoans[10]. The RBFOX family proteins are characterized by their single and nearly invariant RRM domain, which specifically binds to a (U)GCAUG element with extraordinarily high affinity[11]. The (U)GCAUG element is enriched in flanking introns of alternative exons regulated in neuron-specific or muscle-specific manners in mammals[7,12–15]. Indeed, RBFOX1 and RBFOX2 are predominantly expressed in the brain, skeletal muscle and heart, although RBFOX2 is present in many cell types[16,17], and RBFOX3 is expressed exclusively in neurons[18]. Comparative genomic analysis of evolutionarily conserved intronic UGCAUG elements[19] as well as genome-wide identification of binding sites for RBFOX2 in human embryonic stem cells[20] has revealed many alternative-splicing events regulated by the RBFOX family and has elucidated a position-dependent trend of splicing regulation via the (U)GCAUG elements. Nevertheless, some of the alternative exons with the UGCAUG element(s) in the flanking introns are oppositely regulated in brain and muscle tissues or are unresponsive to expression or knockdown of the RBFOX proteins[19], thus suggesting that other tissue-specific splicing regulators may be required for proper regulation of the brain- or muscle-specific splicing of the RBFOX-target exons.

In *C. elegans*, ligand binding specificity of FGFRs is determined by mutually exclusive selection of exon 5A or exon 5B of the *egl-15* gene encoding the sole homolog of the FGFRs[21]. The EGL-15(5A) isoform is required for directed migration of hermaphrodite sex myoblasts toward the center of the body, where its specific ligand, EGL-17, is expressed[21]. However, the EGL-15(5B) isoform and its specific ligand, LET-756, exert essential functions in embryonic and larval development[22–25]. We previously visualized the tissue-specific selection patterns of *egl-15* exon 5A and exon 5B in living worms with bichromatic fluorescence reporter minigenes[26]. Through genetic and biochemical analyses, we demonstrated that the *C. elegans* RBFOX family proteins ASD-1 and FOX-1 cooperate with SUP-12, a member of the SUP-12–RBM24–RBM38 family proteins for muscle-specific production of EGL-15(5A) *in vivo*[27]. Both ASD-1 and FOX-1 are expressed in a wide variety of tissues[27], like RBFOX2 in mammals,

[1]RIKEN Systems and Structural Biology Center, Yokohama, Japan. [2]RIKEN Center for Life Science Technologies, Yokohama, Japan. [3]Faculty of Pharmacy and Research Institute of Pharmaceutical Sciences, Musashino University, Nishitokyo, Japan. [4]Graduate School of Nanobioscience, Yokohama City University, Yokohama, Japan. [5]Institute for Protein Research, Osaka University, Suita, Japan. [6]Tatsuo Miyazawa Memorial Program, RIKEN Genomic Sciences Center, Yokohama, Japan. [7]Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt, Frankfurt am Main, Germany. [8]Frankfurt Institute of Advanced Studies, Goethe University Frankfurt, Frankfurt am Main, Germany. [9]Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, Hachioji, Japan. [10]RIKEN Structural Biology Laboratory, Yokohama, Japan. [11]Graduate School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan. [12]Department of Anatomy and Developmental Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [13]Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan. [14]These authors contributed equally to this work. Correspondence should be addressed to M.H. (hagiwara.masatoshi.8c@kyoto-u.ac.jp), H.K. (kuroyana.end@tmd.ac.jp) or Y.M. (ymuto@musashino-u.ac.jp).

Received 26 May; accepted 14 July; published online 17 August 2014; doi:10.1038/nsmb.2870

whereas SUP-12 is muscle specific[28], thus indicating that functional coordination between these factors is crucial for the muscle-specific selection of *egl-15* exon 5A. Furthermore, *asd-1; fox-1* and *asd-1; sup-12* double-mutant worms are defective in the directed migration of the sex myoblasts and in egg laying (Egl-d)[26,27], the same phenotypes as in the *egl-15(5A)* and *egl-17* mutants, thus emphasizing that *egl-15* exon 5 is the essential physiological target for the RBFOX family in nematodes.

We previously identified a UGCAUGGUGUGC stretch located just upstream of the splice-acceptor site for exon 5B as a *cis* element allowing the RBFOX family and SUP-12 to cooperatively repress exon 5B[26,27] (**Fig. 1a**). The RBFOX family proteins specifically recognize the first 6-nt UGCAUG segment, whereas SUP-12, which also has a single RRM domain, is expected to bind to the latter GUGUGC segment[27]. However, the precise RNA sequence specificity of SUP-12 is not yet clear, and it remains obscure why two RRM proteins are necessary for the effective regulation of the *egl-15* pre-mRNA.

We set out to elucidate the structural basis for the cooperative RNA recognition by the RBFOX family proteins and SUP-12. We thus determined solution structures of a binary complex consisting of SUP-12 RRM and 5′-GUGUGC-3′ and a ternary complex consisting of ASD-1 RRM, SUP-12 RRM and 5′-UGCAUGGUGUGC-3′. We found that muscle-specific regulation of *egl-15* alternative splicing requires appropriate positioning of the two RRM domains that sandwich the G base at the seventh position in the RNA sequence. We confirmed the rigidity and the flexibility in their cooperativity predicted from the structural analysis with our *egl-15* splicing-reporter system *in vivo*.

RESULTS

Cooperative RNA binding by RRM domains of ASD-1 and SUP-12

We previously reported cooperative binding of full-length ASD-1 and full-length SUP-12 to the *egl-15* pre-mRNA *in vitro*[27]. Here, we first created ASD-1 and SUP-12 fragments and then determined which of them were capable of RNA binding. For ASD-1, we used a fragment spanning residues 97–190 (hereafter termed ASD-1 RRM), according to a previous report on the identical RRM domains of mammalian RBFOX1 (ref. 11), which share high sequence identity with that of ASD-1 (**Supplementary Fig. 1**). For SUP-12, NMR chemical-shift-perturbation analyses revealed that a fragment spanning residues 20–121 (SUP-12 RRM) was sufficient for binding to the RNA oligonucleotide 5′-GUGUGC-3′ (RNA_6) (**Supplementary Fig. 2a**,**b**). Next, we tested whether the RRM domains are sufficient for cooperative RNA binding by performing an electrophoretic mobility shift assay (EMSA) experiment with 10 nM of the RNA oligonucleotide 5′-UGCAUGGUGUGC-3′ (RNA_{12}) and ASD-1 RRM and/or SUP-12 RRM. In the case of ASD-1 RRM, a 1:1 protein–RNA binary complex (ASD-1–RNA_{12} complex) was mainly formed (**Fig. 1b**). When we added only SUP-12 RRM to RNA_{12}, we observed a smeared signal, and the mobility of the RNA gradually shifted according to the increase in the amount of SUP-12 RRM (**Fig. 1c**), thus implying that SUP-12 RRM alone lacked strict RNA sequence specificity and that multiple SUP-12 RRM molecules can bind to multiple sites in RNA_{12}. Then we added SUP-12 RRM to RNA_{12} in the presence of 25.6 μM ASD-1 RRM (**Fig. 1d**). In this case, SUP-12 RRM supershifted a band corresponding to the ASD-1–RNA_{12} complex to form another clear band (**Fig. 1d**), indicating the formation of a ternary complex. The ternary

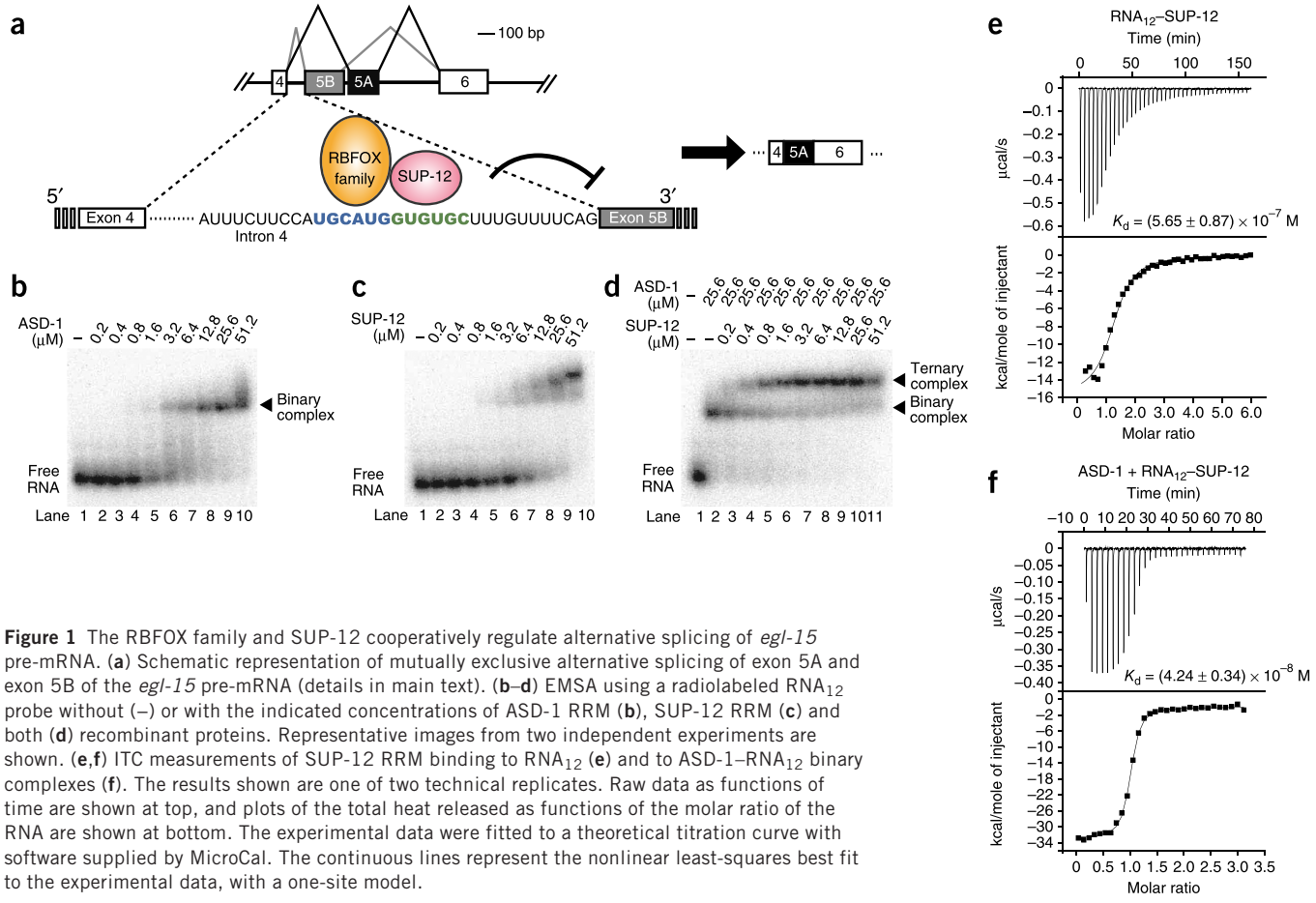

Figure 1 The RBFOX family and SUP-12 cooperatively regulate alternative splicing of *egl-15* pre-mRNA. (**a**) Schematic representation of mutually exclusive alternative splicing of exon 5A and exon 5B of the *egl-15* pre-mRNA (details in main text). (**b**–**d**) EMSA using a radiolabeled RNA_{12} probe without (–) or with the indicated concentrations of ASD-1 RRM (**b**), SUP-12 RRM (**c**) and both (**d**) recombinant proteins. Representative images from two independent experiments are shown. (**e**,**f**) ITC measurements of SUP-12 RRM binding to RNA_{12} (**e**) and to ASD-1–RNA_{12} binary complexes (**f**). The results shown are one of two technical replicates. Raw data as functions of time are shown at top, and plots of the total heat released as functions of the molar ratio of the RNA are shown at bottom. The experimental data were fitted to a theoretical titration curve with software supplied by MicroCal. The continuous lines represent the nonlinear least-squares best fit to the experimental data, with a one-site model.

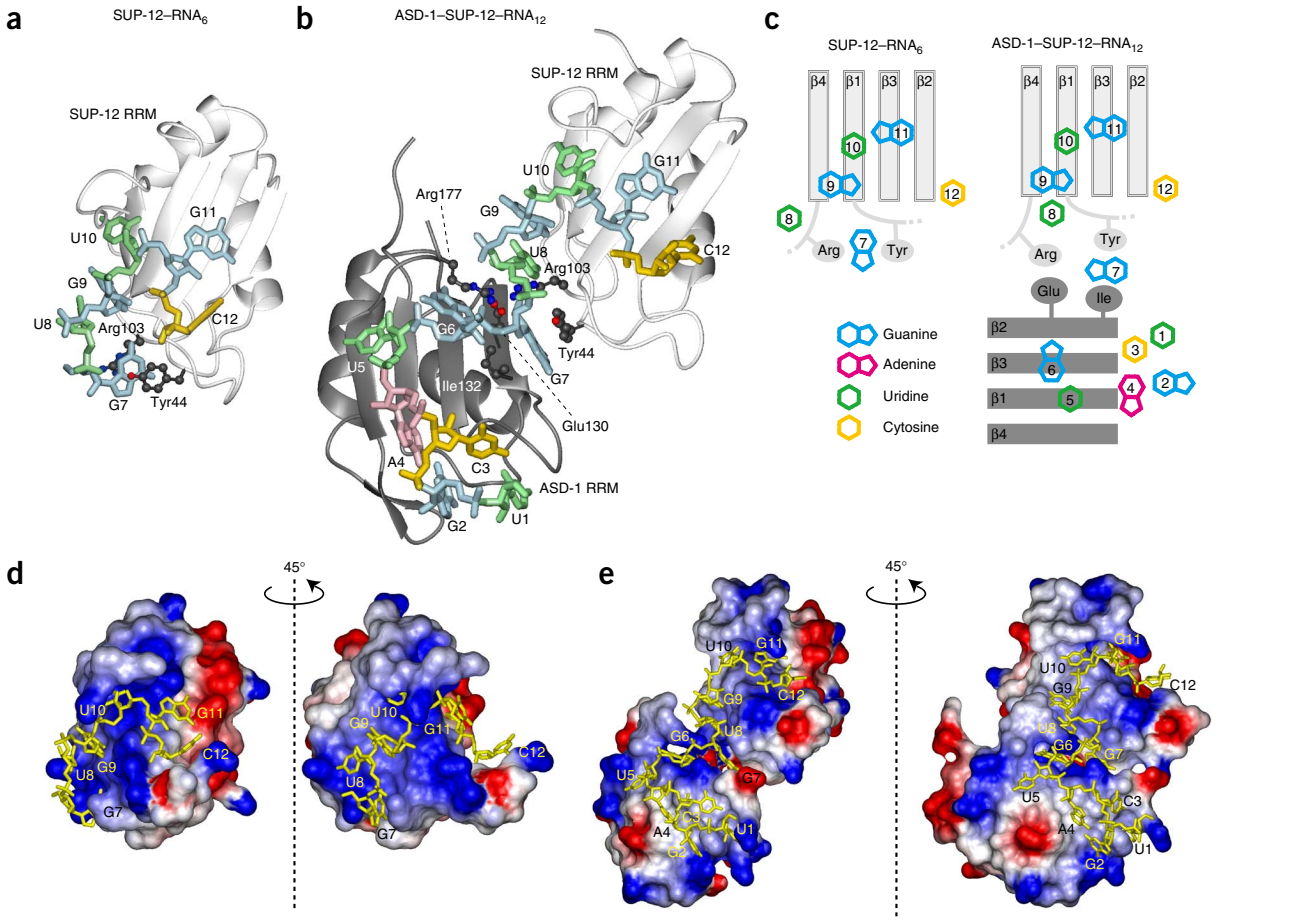

Figure 2 Solution structures of SUP-12–RNA_6 and ASD-1–SUP-12–RNA_{12} complexes. (**a**,**b**) Ribbon representations of the SUP-12–RNA_6 complex (**a**) and the ASD-1–SUP-12–RNA_{12} complex (**b**). SUP-12 RRM and ASD-1 RRM are colored white and gray, respectively. The G, A, U and C nucleotides are shown in light blue, pink, light green and yellow, respectively. The side chains in the RRM domains involved in the G7-base recognition are represented as follows: black, carbon; red, oxygen; blue, nitrogen. (**c**) Schematic diagrams of the SUP-12–RNA_6 complex (left) and the ASD-1–SUP-12–RNA_{12} complex (right). (**d**,**e**) Electrostatic-potential surfaces of the RRM domains in the SUP-12–RNA_6 complex (**d**) and the ASD-1–SUP-12–RNA_{12} complex (**e**). Red and blue indicate negative and positive charges, respectively. The RNA molecule is colored yellow.

complex was formed at a lower concentration of SUP-12 RRM than was required for the formation of the SUP-12–RNA_{12} binary complex (**Fig. 1c**). These results suggested that ASD-1 RRM anchored SUP-12 RRM to the appropriate position on the target RNA to form a stable ternary complex at a molar ratio of 1:1:1.

In order to assess the cooperativity between ASD-1 RRM and SUP-12 RRM, we also performed isothermal titration calorimetry (ITC) experiments by using SUP-12 RRM, the RNA oligonucleotides and a preformed ASD-1–RNA_{12} binary complex (**Fig. 1e**,**f**). SUP-12 RRM exhibited an apparent K_d value of 565 nM for binding RNA_{12} (**Fig. 1e**). Consistently with the EMSA experiments, SUP-12 RRM bound more effectively to the preformed ASD-1–RNA_{12} binary complex, with a K_d of 42.4 nM (**Fig. 1f**). The ITC experiments also revealed that SUP-12 RRM by itself could bind to some extent to other RNA stretches with GU and/or UG element(s) flanking the canonical binding site (**Supplementary Fig. 3a–e**), whereas ASD-1 RRM specifically bound to the UGCAUGG stretch (**Supplementary Fig. 3a**,**f–h**). The wobble RNA recognition of SUP-12 RRM may explain the EMSA results in which SUP-12 RRM alone shifted the mobility of RNA_{12} in a non-uniform manner and hypershifted the mobility more effectively than ASD-1 RRM alone (**Fig. 1b–d**). Thus, the EMSA and ITC experiments indicated that ASD-1 RRM guides SUP-12 RRM to the proper site to effectively form the stable ASD-1–SUP-12–RNA_{12} ternary complex.

Structural basis for cooperative RNA recognition

In order to elucidate the structural basis for the cooperation between ASD-1 and SUP-12, we solved a solution structure of a binary complex composed of SUP-12 RRM and 5′-GUGUGC-3′ (SUP-12–RNA_6 complex) (**Fig. 2a**,**c**,**d**, **Supplementary Fig. 2c**,**d**, **Table 1** and **Supplementary Tables 1** and **2**) and that of the ternary complex composed of ASD-1 RRM, SUP-12 RRM and 5′-UGCAUGGUGUGC-3′ (ASD-1–SUP-12–RNA_{12} complex) (**Fig. 2b**,**c**,**e**, **Supplementary Fig. 4**, **Table 1** and **Supplementary Tables 3** and **4**).

In the ternary complex, ASD-1 RRM and SUP-12 RRM were arranged side by side in fixed positions and formed positively charged surfaces for RNA recognition (**Fig. 2b**,**c**,**e** and **Supplementary Fig. 4c**,**d**). All of the nucleotide residues (U1–C12) interacted with either or both of the RRM domains: U1–G6 were located on ASD-1 RRM, G7 was sandwiched between the two RRM domains, and U8–C12 were on SUP-12 RRM (**Fig. 2b** and **Supplementary Fig. 4c–e**).

The RNA-recognition mode of ASD-1 RRM in the ternary complex was essentially the same as that of the single RRM domain of mammalian RBFOX1 in a binary complex with 5′-UGCAUGU-3′ except for the base at the seventh position; the U base is not fixed on the body of the RRM domain of RBFOX1 (ref. 11). The mode of recognition of G9–C12 by SUP-12 RRM was common between the SUP-12–RNA_6 and the ASD-1–SUP-12–RNA_{12} complexes (**Fig. 2a–c**). The G9 base

Table 1 Data collection and refinement statistics

	SUP-12–RNA_6	ASD-1–SUP-12–RNA_{12}
NMR distance and dihedral constraints		
Distance restraints		
Total NOE	1,640	1,839
Intra-residue	428	602
Inter-residue		
Sequential ($\|i-j\|=1$)	400	519
Non-sequential ($\|i-j\|>1$)	812	718
Hydrogen bonds[a]	40	112
Protein–nucleic acid intermolecular	67	122
Total dihedral angle restraints		
Protein		
ϕ	2	5
ψ	2	5
Nucleic acid		
Base pair	0	4
Sugar pucker	6	12
Backbone	6	12
Total RDCs (HN-N)[b]	–	69
Average correlation	–	0.963 ± 0.007
Average *Q*	–	0.272 ± 0.021
Structure statistics		
Violations (mean and s.d.)		
Distance constraints (Å) > 0.1	0.145 ± 0.028	0.134 ± 0.036
Dihedral angle constraints (°)	0	0
Max. dihedral angle violation (°)	0	0
Max. distance constraint violation (Å)	0.186 ± 0.024	0.188 ± 0.047
Deviations from idealized geometry		
Bond lengths (Å)	0.014	0.014
Bond angles (°)	1.8	1.9
Average pairwise r.m.s. deviation (Å)		
Protein[c]		
Heavy	1.19 ± 0.14	2.52 ± 0.69
Backbone	0.45 ± 0.09	2.09 ± 0.69
RNA[d]		
All RNA heavy	1.27 ± 0.47	2.13 ± 0.68
Complex[e]		
Protein and nucleic acid heavy	1.24 ± 0.15	2.59 ± 1.00

[a]Used only in CYANA calculation. [b]Used only in AMBER calculation. [c]For the calculated residues, the proteins were Ile37–Ala114 of SUP-12 RRM and Leu102–Ala174 of ASD-1 RRM. [d]For the calculated residues, the RNAs were G7–C12 in the SUP-12–RNA_6 complex and U1–C12 in the ASD-1–SUP-12–RNA_{12}. [e]For the calculated residues, the proteins and RNAs were Ile37–Ala114 and G7–C12 in the SUP-12–RNA_6 complex and Leu102–Ala174 and U1–C12 of ASD-1–SUP-12–RNA_{12}.

was located in a pocket composed of Phe38, Tyr78 and Asn106 of SUP-12 (**Fig. 3a**). The O6 atom of G9 was recognized by the H^{δ} proton of Asn106 in the pocket (**Fig. 3a**). The U10 base was sandwiched between the conserved Phe38 residue of the RNP2 motif and a C-terminal extension of the RRM domain spanning residues Leu112 to Lys115 (**Fig. 3b**). Its functional moieties, the O4 atom and the imino proton, were recognized by hydrogen-bonding with the H^{δ} amide proton of Asn108 and the main chain carbonyl oxygen atom of Ala110, respectively (**Fig. 3b**). The G11 base was stacked with the side chain of Phe80 on the RNP1 motif and that of Lys115 (**Fig. 3c**). Two hydrogen-bonding interactions were formed between the O6 atom of G11 and the amide proton of Lys115 and between the OP2 atom of G11 and the $H^{\zeta 3}$ atom of Lys115 (**Fig. 3c**). The C12 base stuck into the space composed of Ile67 and Arg70; the O2 atom of C12 could be hydrogen-bonded to a proton of the guanidyl group of Arg70 (**Fig. 3d**).

ASD-1 RRM altered the mode of RNA recognition by SUP-12 RRM

In the ASD-1–SUP-12–RNA_{12} complex, the side chain of Arg177 of ASD-1 RRM intruded into the space between the G6 nucleotide on ASD-1 RRM and the U8 nucleotide on SUP-12 RRM to interact with the phosphate backbones of the G6, G7 and G9 nucleotides (**Figs. 2b** and **3e**). Accordingly, the recognition modes for the G7 and U8 bases by SUP-12 RRM were quite different between the SUP-12–RNA_6 and the ASD-1–SUP-12–RNA_{12} complexes.

The G7 base was buried in a groove formed by the β1–α1 and the α2–β4 loops and was sandwiched between Tyr44 and Arg103 in the SUP-12–RNA_6 complex (**Fig. 2a**,**c**,**d**). The imino proton of the G7 base formed a hydrogen bond with the main chain carboxyl oxygen of Leu42 and exhibited several nuclear Overhauser effects (NOEs) with SUP-12 RRM (**Supplementary Fig. 5a**). Thus, the G7 base was fairly strictly recognized in the SUP-12–RNA_6 binary complex. Correspondingly, ITC experiments revealed that SUP-12 RRM exhibited the lower affinity for the G7-substituted RNA oligonucleotides AUGUGC, CUGUGU and UUGUGC (with substituted oligonucleotides underlined here and subsequently) than for RNA_6 (**Supplementary Fig. 5c–f**).

In the ASD-1–SUP-12–RNA_{12} complex, the G7 base was sandwiched between ASD-1 Ile132 and SUP-12 Tyr44 and was surrounded by a wall formed by ASD-1 Glu130 and SUP-12 Arg103 to fix the spatial relationship between ASD-1 RRM and SUP-12 RRM (**Fig. 3e**). The χ_2 angle of SUP-12 Tyr44 was altered to maintain the stacking interaction with the G7 base as described above (**Figs. 2b** and **3e**). Correspondingly, the imino-proton resonance from the G7 base was missing in the ASD-1–SUP-12–RNA_{12} complex (**Supplementary Fig. 5b**).

The U8 base was barely recognized in the SUP-12–RNA_6 complex (**Fig. 3f**). In the ASD-1–SUP-12–RNA_{12} complex, the alternation of the χ_2 angle of SUP-12 Tyr44 induced re-formation of the pocket that accommodated G9 (**Figs. 2b** and **3f**), and the U8 base intruded into the pocket to stack with the G9 base. The sugar moiety of U8 was also stacked with ASD-1 Arg177, and the O4 atom of U8 was recognized by the H^{ε} and $H^{\eta 21}$ amide protons of SUP-12 Arg103 (**Fig. 3f**). Consistently with the relocation of U8 to the pocket, five intermolecular NOEs were present between the U8 base and the two RRM domains in the ASD-1–SUP-12–RNA_{12} complex, whereas only one NOE was present between the U8 base and SUP-12 RRM in the SUP-12–RNA_6 complex (**Supplementary Tables 1** and **3**). The solution structure thus revealed that the stacking of the U8 base with the G9 base and ASD-1 Arg177 was a dominant interaction for U8 recognition in the ASD-1–SUP-12–RNA_{12} complex. However, the space beside the U8 base was large enough for the other bases (**Supplementary Fig. 6a**).

The role of the G7 base in ternary-complex formation

The G7-recognition mode in the ASD-1–SUP-12–RNA_{12} ternary complex was distinct from that in the SUP-12–RNA_6 binary complex, as described above. Analyses of the solution structures described above predicted that G7 was the key nucleotide for the cooperation between ASD-1 RRM and SUP-12 RRM through the sandwich-like hydrophobic interactions, because a G base is preferable for such hydrophobic interactions with hydrophobic amino acid residues. To test whether an A base could substitute for the G at this position, we obtained an NMR spectrum of a mutant ternary complex containing a G7A-mutant oligonucleotide 5′-UGCAUGAUGUGC-3′ (RNA_{12}(G7A)). The spectrum indicated that both of the RRM domains bound to RNA_{12}(G7A) and that the A base resided at the position corresponding to the G7 base (**Fig. 4a**). In contrast to the wild-type ASD-1–SUP-12–RNA_{12} complex (**Fig. 4b**), however, the resonances of the following UG dinucleotide (U8 and G9) were missing in the G7A-mutant complex (**Fig. 4a**),

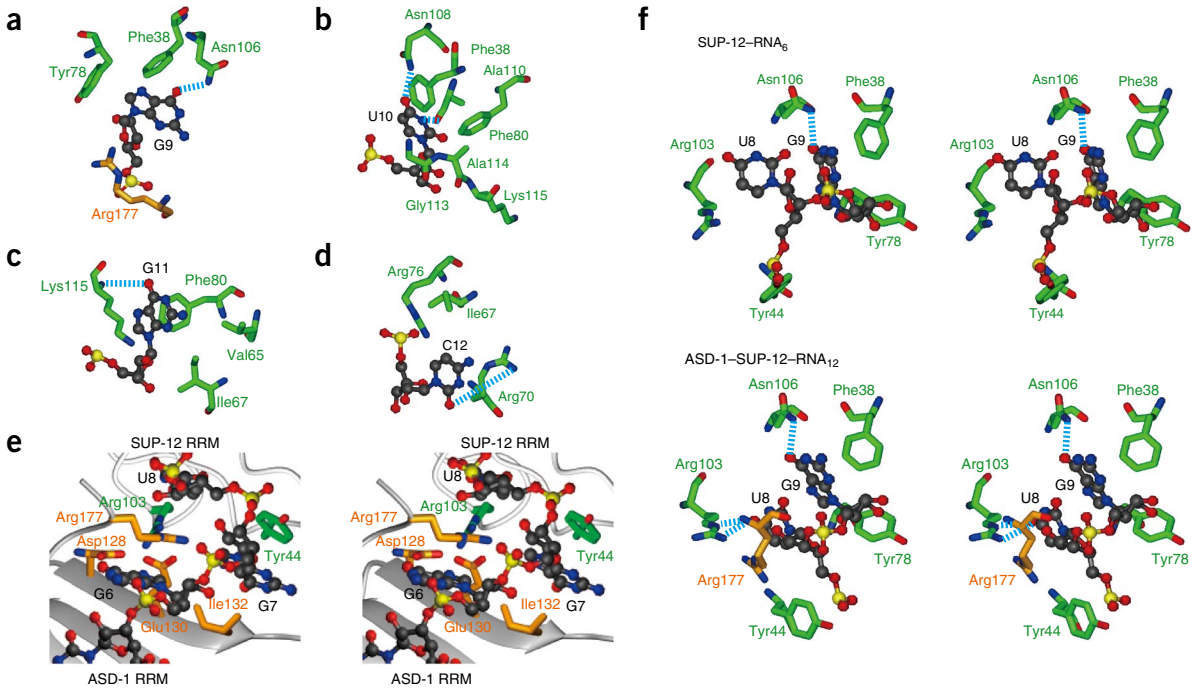

Figure 3 Close-up views of the ASD-1–SUP-12–RNA_{12} and SUP-12–RNA_6 complexes. (**a**–**d**) The recognition contacts for G9 (**a**), U10 (**b**), G11 (**c**) and C12 (**d**) in the ASD-1–SUP-12–RNA_{12} complex. The side chains of the RRM domains involved in RNA recognition are represented as follows: green, carbons in SUP-12 RRM; orange, carbons in ASD-1 RRM; red, oxygens; blue, nitrogens. The RNA molecule is represented by a ball-and-stick model with atoms colored as follows: dark gray, carbon; red, oxygen; blue, nitrogen; yellow, phosphorus. The base-specific hydrogen bonds, calculated by AMBER, are represented by dashed cyan lines (also **Supplementary Tables 3** and **4**). (**e**) Stereo view of cooperative G7 recognition by ASD-1 RRM and SUP-12 RRM in the ASD-1–SUP-12–RNA_{12} complex. (**f**) Comparison of U8- and G9-recognition contacts between the SUP-12–RNA_6 complex (top) and the ASD-1–SUP-12–RNA_{12} complex (bottom) in stereo views.

thus indicating that the recognition of these two nucleotides by SUP-12 RRM was weakened, and the cooperativity between the two RRM domains was probably lost.

Then we analyzed NMR dynamics parameters for the wild-type and the mutant ternary complexes by measuring backbone-nitrogen relaxation times T_1 and T_2 and proton-nitrogen heteronuclear NOEs. The correlation time (τ_c = 10.2 ns) and relatively large deviations of the T_1 and T_1/T_2 values for each of the RRM domains indicated that the ASD-1–SUP-12–RNA_{12} complex was shaped like a prolate spheroid and behaved in one united body (**Supplementary Fig. 6b**). In the G7A-mutant complex, in contrast, the deviations were in a small range, and the correlation time did not correspond to the appropriate molecular weight for the rigid ternary complex (**Supplementary Fig. 6c**). These data suggested that the low binding activity of SUP-12 RRM to the mutant RNA and/or the poor fitting of A7–G9 on SUP-12 RRM caused the lower rigidity of the G7A-mutant complex.

We also performed ITC analysis with a preformed ASD-1–RNA_{12}(G7A) binary complex and found that binding of SUP-12 RRM to the ASD-1–RNA_{12}(G7A) complex (**Fig. 5a**) was reduced as

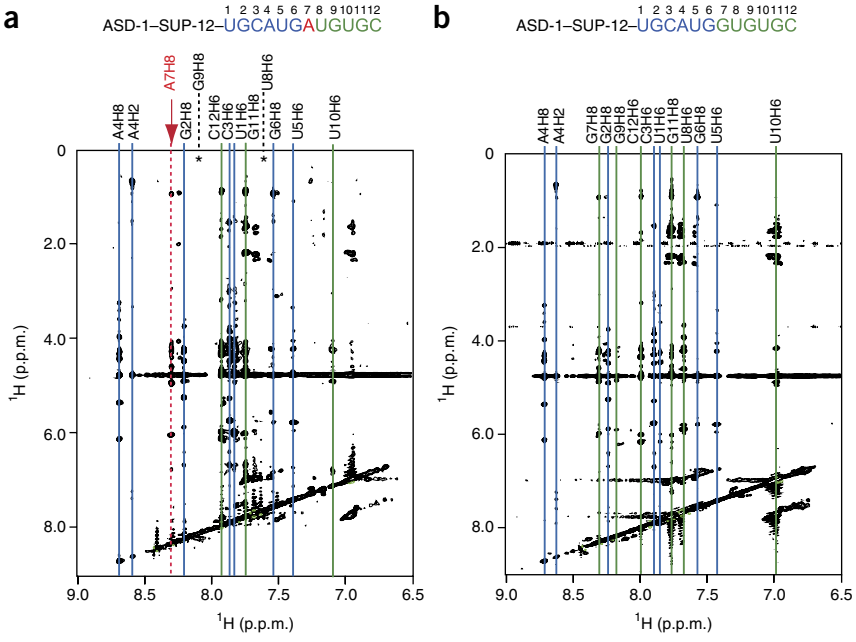

Figure 4 Recognition of the U8 and G9 bases is weakened in the G7A-mutant complex. (**a**,**b**) The H2, H6 and H8 regions of 2D ^{1}H-^{1}H NOESY spectra of the G7A mutant (**a**) and the wild-type (**b**) ASD-1–SUP-12–RNA_{12} complexes at 298 K. The assignments of the H2, H6 and H8 atoms in the RNA molecules are indicated with blue and green lines. The assignment of the H8 atom of A7 is indicated with a dashed red line in **a**. Asterisks in **a** denote unsuccessful assignment of the H6 atom of U8 and the H8 atom of G9 for the mutant complex, indicating weakened recognition of the U8 and G9 bases by SUP-12 RRM.

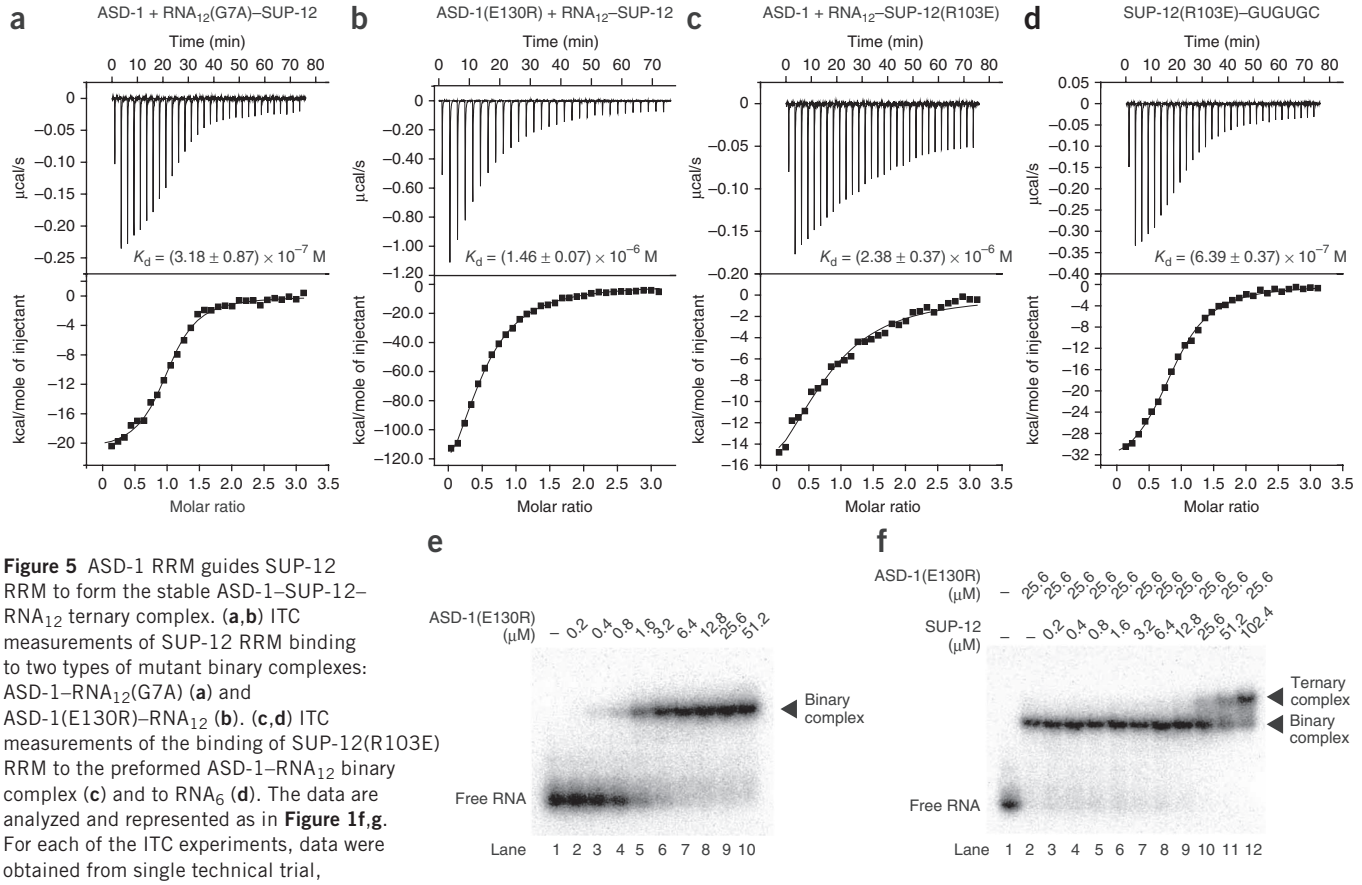

Figure 5 ASD-1 RRM guides SUP-12 RRM to form the stable ASD-1–SUP-12–RNA_{12} ternary complex. (**a**,**b**) ITC measurements of SUP-12 RRM binding to two types of mutant binary complexes: ASD-1–RNA_{12}(G7A) (**a**) and ASD-1(E130R)–RNA_{12} (**b**). (**c**,**d**) ITC measurements of the binding of SUP-12(R103E) RRM to the preformed ASD-1–RNA_{12} binary complex (**c**) and to RNA_6 (**d**). The data are analyzed and represented as in **Figure 1f**,**g**. For each of the ITC experiments, data were obtained from single technical trial, owing to the difficulty of the preparation of the purified binary complex. (**e**,**f**) EMSA experiments using a radiolabeled RNA_{12} probe with a dilution series of ASD-1(E130R) RRM alone (**e**) and with a dilution series of SUP-12 RRM in the presence of 25.6 µM ASD-1(E130R) RRM (**f**). For each of the experiments, a representative image from two independent experiments is shown.

compared to binding to the wild-type ASD-1–RNA_{12} binary complex (**Fig. 1f**). All these data supported the idea that the G7 base was crucial for ASD-1 RRM and SUP-12 RRM to cooperatively recognize the *egl-15* pre-mRNA and form the stable and rigid ternary complex.

Although we did not observe a direct NOE between ASD-1 RRM and SUP-12 RRM, NOEs from the G7 and U8 nucleotides to Ile132 and Arg177 of ASD-1 and to Tyr44 and Arg103 of SUP-12 fixed the spatial relationship between ASD-1 RRM and SUP-12 RRM (**Supplementary Table 3**). We confirmed the orientation of the two RRM domains by residual dipolar coupling (RDC) experiments (**Supplementary Fig. 4b**). Correspondingly, the side chains of ASD-1 Ile132 and SUP-12 Tyr44 and the G7 base were well converged (**Supplementary Fig. 4e**).

From the structural information, we expected direct electrostatic interactions between Asp128 and Glu130 of ASD-1 and Arg103 of SUP-12 in the ASD-1–SUP-12–RNA_{12} complex (**Figs. 2b** and **3e** and **Supplementary Fig. 4e**). However, we did not confirm the interactions with the NMR data because carboxyl groups in the side chains of Asp128 and Glu130 had no protons for NOE measurements.

In order to elucidate the requirement for the interaction between ASD-1 Glu130 and SUP-12 Arg103, we prepared a mutant protein ASD-1(E130R) RRM and performed EMSA and ITC experiments. The EMSA experiments demonstrated that ASD-1(E130R) RRM by itself exhibited the same RNA binding activity to RNA_{12} (**Fig. 5e**) as wild-type ASD-1 RRM (**Fig. 1b**). However, the formation of the ASD-1(E130R)–SUP-12–RNA_{12} ternary complex was diminished (**Fig. 5f**) compared to formation of the wild-type ternary complex (**Fig. 1d**). The ITC experiment revealed that the K_d value of SUP-12 RRM for binding to the preformed ASD-1–RNA_{12} complex was weakened from 42.4 nM for the wild-type complex (**Fig. 1f**) to 1.46 µM for the E130R-mutant complex (**Fig. 5b**). These results indicated that ASD-1 Glu130 was essential for the intermolecular interaction between the ASD-1–RNA_{12} complex and SUP-12 RRM and not for the RNA binding activity of ASD-1 RRM.

SUP-12 Arg103 was also involved in RNA recognition in distinct ways in the binary and the ternary complexes (**Figs. 2** and **3e**,**f**). ITC experiments with a mutant protein SUP-12(R103E) RRM revealed that the mutation in SUP-12 Arg103 more profoundly affected the binding of SUP-12 RRM to the ASD-1–RNA_{12} binary complex than to RNA_6 (**Fig. 5c**,**d**), confirming that SUP-12 Arg103 is one of the key residues for the cooperative formation of the ASD-1–SUP-12–RNA_{12} complex.

Conclusively, the present structural studies delineated three characteristic cooperative effects that stabilize the ASD-1–SUP-12–RNA_{12} ternary complex: the intrusion of Arg177 of ASD-1 RRM to interact with the G6–U8 stretch, the sandwich-like recognition of the G7 base and the electrostatic interaction between ASD-1 and SUP-12. In cooperation with ASD-1, which specifically recognizes the UGCAUG stretch, SUP-12 accurately recognized the GUGUGC stretch with higher affinity.

The G7 base was crucial for splicing regulation *in vivo*

In order to confirm the structural model in the actual regulation of alternative splicing *in vivo*, we performed mutation analyses of fluorescent *egl-15* splicing reporter (**Fig. 6**). RFP was predominantly expressed from the wild-type minigene (**Fig. 6a**), thus reflecting exon 5A selection in body-wall muscles[26,27], whereas GFP was

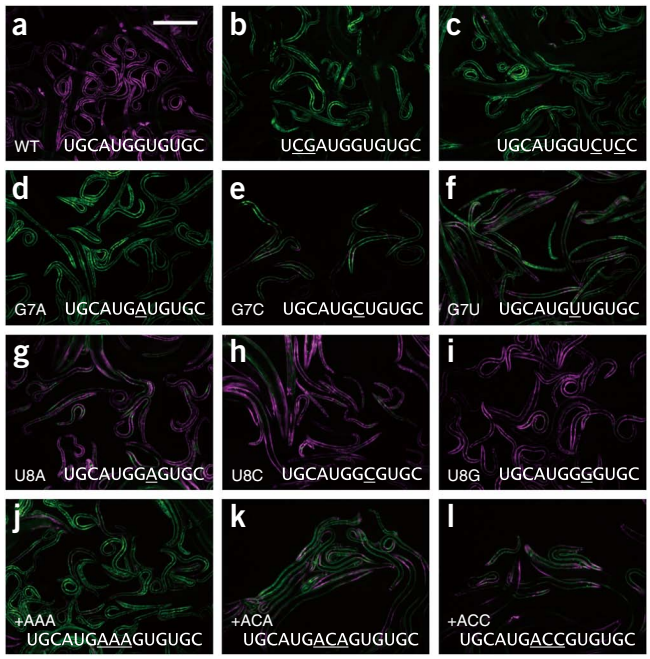

Figure 6 The sandwiched base G7 and not U8 is crucial for regulation of muscle-specific alternative splicing *in vivo*. (**a**–**l**) Microphotographs of *egl-15* reporter worms carrying wild-type (**a**) and mutant (**b**–**l**) *egl-15BGAR* minigenes, acquired with a dual-band-pass filter. (**b**,**c**) Substitution mutations of the RBFOX site (**b**) and the SUP-12 site (**c**). (**d**–**f**) Substitution mutations of the G7 base: G7A (**d**), G7C (**e**) and G7U (**f**). (**g**–**i**) Substitution mutations of the U8 base: U8A (**g**), U8C (**h**) and U8G (**i**). (**j**–**l**) Insertion mutations with three extra bases between the elements: AAA (**j**), ACA (**k**) and ACC (**l**). The sequences of the *cis* elements are indicated with the altered residues underlined. Scale bar, 200 μm. The images were processed so that the images in the red channel appear in magenta. These microphotographs are representative of three or more transgenic lines (**a**,**c**,**d**,**f**,**g**,**j**–**l**) and of one or two transgenic lines (**b**,**e**,**h**,**i**).

predominantly expressed from mutant minigenes lacking the consensus for either RBFOX or SUP-12 (**Fig. 6b**,**c**), thus indicating exon 5B selection. When the G7 base was replaced with the other bases (i.e., A, C or T) in the minigenes, the color of the worms also changed to green (**Fig. 6d**–**f**), a result confirming that the G7 base was crucial for the muscle-specific splicing regulation of the *egl-15* gene *in vivo*.

However, when the T8 base was replaced with the other bases (i.e., A, C or G), the body-wall muscles remained red or orange (**Fig. 6g**–**i**), as we had speculated would happen, on the basis of the solution structure. Indeed, the T8 base of the tandem elements is not fully conserved in the *egl-15* gene of related nematode species but is replaced with C in *Caenorhabditis briggsae* (**Supplementary Fig. 7a**). These data confirmed that the U8 base was not crucially involved in splicing regulation.

To investigate the importance of the distance between the respective *cis* elements, we inserted three nucleotides (AAA, ACA or ACC) between G6 and G7 in the minigene. These insertions drastically changed the color of the splicing-reporter worms (**Fig. 6j**–**l**), indicating that proper positioning of the *cis* elements for the two RRM proteins is crucial for the splicing regulation of *egl-15* exon 5 *in vivo*.

Conservation of the cooperative splicing regulation

We speculated that a UGCAUG stretch immediately followed by a GYGUG stretch in introns flanking an alternative exon may predict cooperative regulation of the exon by the RBFOX family and SUP-12 in *C. elegans*. To test this idea, we searched the *C. elegans* genome for the UGCAUGGYGUG sequence. We found that 14 out of 47 UGCAUGGYGUG sites in the genome reside in the sense strand of protein-coding genes including *egl-15*. Among them, the *cle-1* gene, encoding the sole homolog of vertebrate type XV and type XVIII collagen, has a highly conserved UGCAUGGUGUG stretch in the downstream intron of its cassette exon (**Supplementary Fig. 7b**). Consistently with the observations that the RBFOX family and the UGCAUG stretch(es) in the downstream introns tend to promote inclusion of cassette exons in mammals[19,20] and nematodes (H.K., unpublished data), we confirmed that the inclusion level of *cle-1* exon 16 was decreased in the *asd-1; fox-1* double mutant and was highly dependent on SUP-12 (**Fig. 7**), thus confirming that *cle-1* exon 16 was a common target for the RBFOX family and SUP-12 *in vivo*. To our knowledge, this is the first successful prediction and validation of a new target pre-mRNA for SUP-12 based on the combination of the splicing-regulatory elements, and our results emphasize the importance of cooperative splicing regulation by the RBFOX family and SUP-12.

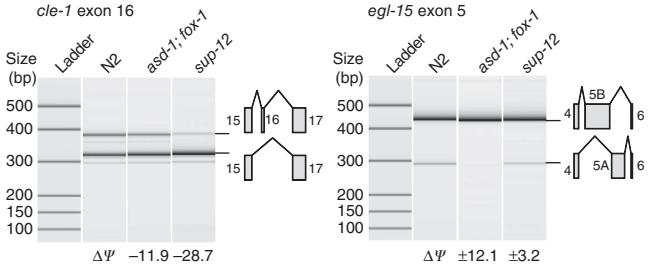

Figure 7 A highly conserved UGCAUGGUGUG stretch in *cle-1* intron 16 successfully predicts the activation of exon 16 by the RBFOX family and SUP-12. Reverse-transcription PCR analyses of endogenous *cle-1* exon 16 (left) and *egl-15* exons 5A and 5B (right) in the wild-type strain (N2), the *asd-1; fox-1* double mutant and the *sup-12* mutant. Changes in percentage spliced in (ΔΨ) in molar concentration as compared to N2 are indicated. The results shown are from one of two technical replicates.

DISCUSSION

In this study, we elucidated the cooperative RNA-recognition mode of ASD-1 and SUP-12. When ASD-1 RRM binds to the UGCAUG stretch, SUP-12 binds to the juxtaposed GUGUG stretch with higher affinity and specificity, to form a tight ternary complex with the G7 base sandwiched between them.

So far, two types of cooperative and specific RNA recognition by multiple RRM domains have been elucidated. One is the intermolecular protein-protein interaction between Hrp1 and Rna15 on the scaffold protein Rna14 for tight RNA binding in yeast. In this case, the RRM domains of Hrp1 and Rna15 separately recognize their own target sequences without any cooperation[29]. Another is the intramolecular cooperation between two RRM domains in a single molecule, such as *Drosophila* Sxl[30]. Intermolecular cooperation between RRM domains for specific base recognition, as we observed between ASD-1 and SUP-12, had not been reported previously, to our knowledge. We identified the combinations of the amino acid residues in ASD-1 and SUP-12 involved in the cooperative recognition (**Fig. 3e**). We could not find the combinations of residues among all of the 705 human RRM domains in the Pfam database (release 26.0). Thus, we consider the sandwich-like RNA recognition by the RBFOX family and SUP-12 to be unique among the RRM proteins.

The muscle-specific RNA-binding protein SUP-12 confers muscle specificity in the RBFOX-mediated splicing regulation of *egl-15* exon 5 (ref. 27). Similarly, a neuron-specific CELF family RNA-binding protein, UNC-75, confers neuron specificity in RBFOX-mediated splicing regulation of *unc-32* exon 7 (refs. 31,32). In mammals, some cassette exons with UGCAUG stretches in their flanking introns are oppositely regulated in brain and muscle tissues[19]. These facts suggest a general model in which specific binding of the RBFOX family to the UGCAUG stretch(es) is necessary yet insufficient for proper splicing regulation and that other tissue-specific factors are required for combinatorial splicing regulation. One possible reason why the high-affinity binding of the RBFOX family is insufficient might be that occupancy of most of the UGCAUG stretches by the RBFOX family proteins is low, owing to their low concentration *in vivo*, and this is overcome by cooperative recognition, as shown in the present study. This view is supported by splicing autoregulation of the RBFOX family to strictly maintain RBFOX expression level in mammals[33] and *C. elegans* (H.K., unpublished data).

We also demonstrated that the cooperation relies on the appropriate distance of the *cis* elements for the RBFOX family and SUP-12. In the case of muscle-specific alternative processing of *unc-60* pre-mRNA, the target sequences for SUP-12 and ASD-2, a signal transduction and activation of RNA (STAR) family protein, are separated by three or more nucleotides[34–36], thus implying that the spatial relationship between the *cis* elements directing the cooperative recognition can vary according to the combination of splicing-regulatory factors. Future structural studies on the combinatorial recognition of target sequences by multiple RNA-binding proteins together with the confirmation experiments with multichromatic fluorescence splicing reporters will contribute to deciphering the detailed splicing codes *in vivo*.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Chemical shifts of the SUP-12–RNA_6 complex and those of the ASD-1–SUP-12–RNA_{12} complex have been deposited in the Biological Magnetic Resonance Data Bank under accession numbers 11517 and 11518, respectively. Atomic coordinates for the ensembles of 20 energy-refined NMR conformers, representing the solution structures of the SUP-12–RNA_6 and ASD-1–SUP-12–RNA_{12} complexes, have been deposited in the Protein Data Bank under accession codes 2RU3 and 2MGZ, respectively.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank H. Kurokawa for technical assistance and T. Imada, K. Ake and T. Nakayama for help with manuscript preparation. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses (to Sh.Y.) of the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), Grants-in-Aid for Scientific Research on Innovative Areas "RNA Regulation" (no. 20112004 to H.K.; nos. 21112522 and 23112723 to Y.M.) and "Transcription Cycle" (no. 25118506 to H.K.) from MEXT, Grants-in-Aid for Scientific Research (B) (no. 23370080 to Y.M.; no. 26291003 to H.K.) from the Japan Society for the Promotion of Science (JSPS) and Precursory Research for Embryonic Science and Technology (PRESTO) from Japan Science and Technology Agency (JST) to H.K.

AUTHOR CONTRIBUTIONS

K.K. and M.T. performed the structure determinations of the SUP-12–RNA_6 complex and the ASD-1–SUP-12–RNA_{12} complex by NMR. S.U. performed the analytical ultracentrifugation experiments. Se.Y., M.S., T.I. and A.T. assisted with sample preparation. K.T., F.H., N.K., P.G. and Sh.Y. assisted with the structural determination. H.K. performed the *in vivo* splicing assays. The project was directed by M.H. and Y.M. All authors contributed to the preparation of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
2. Cáceres, J.F. & Kornblihtt, A.R. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* **18**, 186–193 (2002).
3. Faustino, N.A. & Cooper, T.A. Pre-mRNA splicing and human disease. *Genes Dev.* **17**, 419–437 (2003).
4. Busch, A. & Hertel, K.J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* **3**, 1–12 (2012).
5. Chen, Y.I. *et al.* Proteomic analysis of *in vivo*-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.* **35**, 3928–3944 (2007).
6. Ruskin, B., Zamore, P.D. & Green, M.R. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 207–219 (1988).
7. Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
8. Wu, S., Romfo, C.M., Nilsen, T.W. & Green, M.R. Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. *Nature* **402**, 832–835 (1999).
9. Zhuang, Y. & Weiner, A.M. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. *Cell* **46**, 827–835 (1986).
10. Kuroyanagi, H. Fox-1 family of RNA-binding proteins. *Cell. Mol. Life Sci.* **66**, 3895–3907 (2009).
11. Auweter, S.D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163–173 (2006).
12. Castle, J.C. *et al.* Expression of 24,426 human alternative splicing events and predicted *cis* regulation in 48 tissues and cell lines. *Nat. Genet.* **40**, 1416–1425 (2008).
13. Das, D. *et al.* A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing. *Nucleic Acids Res.* **35**, 4845–4857 (2007).
14. Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I. & Conboy, J.G. The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res.* **33**, 714–724 (2005).
15. Sugnet, C.W. *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2**, e4 (2006).
16. Takeuchi, A., Hosokawa, M., Nojima, T. & Hagiwara, M. Splicing reporter mice revealed the evolutionally conserved switching mechanism of tissue-specific alternative exon selection. *PLoS ONE* **5**, e10946 (2010).
17. Underwood, J.G., Boutz, P.L., Dougherty, J.D., Stoilov, P. & Black, D.L. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* **25**, 10005–10016 (2005).
18. Kim, K.K., Adelstein, R.S. & Kawamoto, S. Identification of neuronal nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors. *J. Biol. Chem.* **284**, 31052–31061 (2009).
19. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550–2563 (2008).
20. Yeo, G.W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* **16**, 130–137 (2009).
21. Goodman, S.J., Branda, C.S., Robinson, M.K., Burdine, R.D. & Stern, M.J. Alternative splicing affecting a novel domain in the *C. elegans* EGL-15 FGF receptor confers functional specificity. *Development* **130**, 3757–3766 (2003).
22. Birnbaum, D., Popovici, C. & Roubin, R. A pair as a minimum: the two fibroblast growth factors of the nematode *Caenorhabditis elegans*. *Dev. Dyn.* **232**, 247–255 (2005).
23. Burdine, R.D., Branda, C.S. & Stern, M.J. EGL-17(FGF) expression coordinates the attraction of the migrating sex myoblasts with vulval induction in *C. elegans*. *Development* **125**, 1083–1093 (1998).
24. Burdine, R.D., Chen, E.B., Kwok, S.F. & Stern, M.J. egl-17 encodes an invertebrate fibroblast growth factor family member required specifically for sex myoblast migration in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **94**, 2433–2437 (1997).
25. DeVore, D.L., Horvitz, H.R. & Stern, M.J. An FGF receptor signaling pathway is required for the normal cell migrations of the sex myoblasts in *C. elegans* hermaphrodites. *Cell* **83**, 611–620 (1995).
26. Kuroyanagi, H., Kobayashi, T., Mitani, S. & Hagiwara, M. Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms *in vivo*. *Nat. Methods* **3**, 909–915 (2006).

27. Kuroyanagi, H., Ohno, G., Mitani, S. & Hagiwara, M. The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing *in vivo*. *Mol. Cell. Biol.* **27**, 8612–8621 (2007).
28. Anyanful, A. *et al.* The RNA-binding protein SUP-12 controls muscle-specific splicing of the ADF/cofilin pre-mRNA in *C. elegans*. *J. Cell Biol.* **167**, 639–647 (2004).
29. Leeper, T.C., Qu, X., Lu, C., Moore, C. & Varani, G. Novel protein-protein contacts facilitate mRNA 3′-processing signal recognition by Rna15 and Hrp1. *J. Mol. Biol.* **401**, 334–349 (2010).
30. Handa, N. *et al.* Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585 (1999).
31. Kuroyanagi, H., Watanabe, Y. & Hagiwara, M. CELF family RNA-binding protein UNC-75 regulates two sets of mutually exclusive exons of the *unc-32* gene in neuron-specific manners in *Caenorhabditis elegans*. *PLoS Genet.* **9**, e1003337 (2013).
32. Kuroyanagi, H., Watanabe, Y., Suzuki, Y. & Hagiwara, M. Position-dependent and neuron-specific splicing regulation by the CELF family RNA-binding protein UNC-75 in *Caenorhabditis elegans*. *Nucleic Acids Res.* **41**, 4015–4025 (2013).
33. Damianov, A. & Black, D.L. Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA* **16**, 405–416 (2010).
34. Kuroyanagi, H. Switch-like regulation of tissue-specific alternative pre-mRNA processing patterns revealed by customized fluorescence reporters. *Worm* **2**, e23834 (2013).
35. Ohno, G. *et al.* Muscle-specific splicing factors ASD-2 and SUP-12 cooperatively switch alternative pre-mRNA processing patterns of the ADF/Cofilin gene in *Caenorhabditis elegans*. *PLoS Genet.* **8**, e1002991 (2012).
36. Ohno, G., Hagiwara, M. & Kuroyanagi, H. STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing *in vivo*. *Genes Dev.* **22**, 360–374 (2008).

ONLINE METHODS

Preparation of the recombinant proteins for SUP-12 RRM and ASD-1 RRM. The cDNAs encoding the RRM domains of SUP-12 (SUP-12 RRM; Ser20–Gln121, NCBI accession number NP_508674.1) and ASD-1 (ASD-1 RRM; Asp97–Gly190, NCBI accession number NP_497841.1) were cloned into pET-15b (Novagen) and pGEX6P-1 (GE Healthcare), respectively. In both constructs, a TEV protease–cleavage site was placed between the tag and the protein sequences. Point mutants were introduced into ASD-1 RRM and SUP-12 RRM by PrimeSTAR Mutagenesis Basal Kit (TaKaRa) according to the manufacturer's instructions.

Escherichia coli strain BL21 (DE3) was transformed with the recombinant plasmids and grown at 37 °C, in LB medium supplemented with 50 mg/l ampicillin for the nonlabeled samples, and in modified minimal medium[37] supplemented with 50 mg/l ampicillin for the ^{15}N-^{13}C –labeled samples. IPTG was added to the culture to a final concentration of 1 mM, to induce protein expression. After 3–4 h of cultivation, the cells were harvested.

In order to prepare SUP-12 RRM, the harvested cells were lysed by sonication in 20 mM Tris-HCl buffer, pH 7.0, containing 300 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol and protease-inhibitor cocktail (Nacalai Tesque). The lysate was applied to an Ni^{2+}-NTA SuperFlow column (Qiagen) eluted with an imidazole gradient from 20 mM to 250 mM, and the tag was removed by incubation with TEV protease overnight at room temperature. The tag-free SUP-12 RRM was further purified by RESOURCE S column chromatography according to the manufacturer's instructions (GE Healthcare).

In order to prepare ASD-1 RRM, the harvested cells were lysed by sonication in phosphate-buffered saline containing 1 mM DTT and protease-inhibitor cocktail (Nacalai Tesque). The lysate was applied to a glutathione Sepharose 4 Fast Flow column (GE Healthcare) eluted by the addition of glutathione, and the tag was removed by incubation with TEV protease overnight at room temperature. The tag-free ASD-1 RRM was further purified by RESOURCE S column chromatography according to the manufacturer's instructions (GE Healthcare).

NMR spectroscopy. For NMR measurements, the samples were concentrated to 0.1–1 mM in 20 mM *d*-Tris-HCl buffer, pH 7.0 or pH 5.0, containing 100 mM NaCl, 1 mM 1,4-DL-dithiothreitol-d_{10} (*d*-DTT), 1 μL of RNase inhibitor SIN-101 (TOYOBO) and 0.02% NaN_3 (in 90% H_2O/10% D_2O and 99.99% D_2O), with an Amicon Ultra-15 filter (3,000 MWCO, Millipore). NMR experiments were performed at 298 K for SUP-12 RRM and at 288 K, 298 K and 303 K for the SUP-12–RNA_6 and ASD-1–SUP-12–RNA_{12} complexes on Bruker 600, 700, 800 and 900 MHz spectrometers (Bruker AV600 and AV700 equipped with cryoprobes and Bruker AV800 and AV900 equipped with normal probes). The RNA oligomers were purchased from Dharmacon. The ^{1}H, ^{15}N, and ^{13}C chemical shifts were referenced relative to the frequency of the ^{2}H lock resonance of water. Backbone and side chain assignments of the proteins were obtained with a combination of standard triple-resonance experiments. 2D ^{1}H-^{15}N HSQC and 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH spectra were used for the ^{1}H, ^{15}N, and ^{13}C assignments of the protein backbone. The ^{1}H and ^{13}C assignments of the nonaromatic side chains including all prolines were obtained with 2D ^{1}H-^{13}C HSQC, and 3D HBHA(CO)NH, H(CCCO)NH, (H)CC(CO)NH, HCCH-COSY, HCCH-TOCSY and (H)CCH-TOCSY spectra. Assignments were checked for consistency with 3D ^{15}N-edited ^{1}H-^{1}H NOESY and ^{13}C -edited ^{1}H-^{1}H NOESY spectra. The ^{1}H and ^{13}C spin systems of the aromatic rings of phenylalanine, tryptophan, histidine and tyrosine were identified with 3D HCCH-COSY and HCCH-TOCSY experiments, and 3D ^{13}C-edited ^{1}H-^{1}H NOESY was used for the sequence-specific resonance assignment of the aromatic side chains. NOESY spectra were recorded with mixing times of 80 ms and 150 ms. For the assignments of the RNA molecules, 2D filtered NOESY (mixing times of 150 and 400 ms) and 2D filtered TOCSY (mixing time of 45 ms) spectra were used. The sugar-ring conformation was identified by the intensity of the cross-peaks between H1′ and H2′ in the 2D TOCSY spectra. The NMR data were processed with NMRPipe[38]. Analyses of the processed data were performed with NMRView[39] and KUJIRA[40].

Structure calculations. The three-dimensional structures of the SUP-12–RNA_6 and the ASD-1–SUP-12–RNA_{12} complexes were determined by a combination of automated NOESY cross-peak assignment and structure calculations with torsion-angle dynamics, implemented in CYANA 2.1 (ref. 41). Dihedral-angle restraints for ϕ and ψ were obtained from the main chain and $^{13}C^{\beta}$ chemical-shift values with TALOS[42], and by analysis of the NOESY and HNHA spectra. The χ^1 angles of the protein side chains were estimated by inspection of the patterns of inter- and intra-NOE intensities in conjunction with the 3D HNHB and HN(CO)HB spectra. For the determination of the three-dimensional structures of the RNA molecules, the intermolecular protein-RNA NOEs were manually assigned with 2D NOESY spectra, with mixing times of 150 and 400 ms. The distance restraints for the protein-RNA NOEs were set as follows: the NOEs derived from the RNA molecule in the 2D NOESY spectra with a mixing time of 80 ms were divided into four groups with upper distance restraints of 3.5, 4.0, 4.5 and 5.0 Å, according to their intensity. Upper distance restraints of 6.0 Å were applied for the intermolecular NOEs that could be identified from only 2D NOESY spectra with a mixing time of 400 ms. The information about the 2′- or 3′-end configuration of RNA sugars and *syn-anti* configuration of bases included in the structural calculation, were derived from the 1′-2′ cross-peak intensities in TOCSY experiments and from the intra-residue NOE pattern between the bases and the sugars.

The structure calculations started from 200 randomized conformers, and the standard CYANA simulated annealing schedule was used, with 40,000 torsion-angle dynamics steps per conformer. We select the calculated structure with the reasonable relative intensities of intra-residue NOEs for main chain resonances to reduce the number of amino acid residues appearing in the disallowed region in the final structures. The 40 conformers with the lowest final CYANA target-function values were further refined with AMBER9 (ref. 43), with a generalized Born solvation model and an AMBER 2003 force field, as described previously[44,45]. RDC information was incorporated into the Amber calculation as restraints. RDC restraints for the protein backbone were applied with the force constant of 0.1 kcal/mol/Hz^{-2} for the calculation of the ASD-1–SUP-12–RNA_{12} complex. An error restraint of 1.0 Hz was used for all RDCs. From the 40 refined structures, the 20 conformers with the lowest AMBER energy and violation were selected, to form the final ensemble of structures. From the structures determined by the AMBER calculation, we calculated the ^{1}H-^{15}N RDC values for this ordered region with PALES[46]. **Supplementary Figure 7** shows a comparison between the calculated and experimental RDC values. The 20 conformers that were most consistent with the experimental restraints were then used for further analyses. PROCHECK-NMR[47] and MOLMOL[48] were used to validate and to visualize the final structures, respectively.

NMR titration experiments. For the amide chemical-shift titration experiments, the RNA oligonucleotide 5′-GUGUGC-3′ (Dharmacon) was dissolved in 20 mM *d*-Tris-HCl buffer, pH 7.0, containing 100 mM NaCl and 1 mM *d*-DTT, to make a 6 mM solution. 2D ^{1}H-^{15}N HSQC spectra were recorded while the concentration of the RNA was increased relative to that of the SUP-12 RRM solution (200 μM), to a final 1:2 ratio of SUP-12 RRM/RNA. The perturbation values were obtained from the ^{1}H-^{15}N HSQC spectrum. The absolute values of the chemical-shift change $\Delta\delta(^{15}N + {}^1H_N)$ were calculated as follows: $\Delta\delta(^{15}N + {}^1H_N) = ((\delta_{15N} \times 0.15)^2 + (\delta_{1H})2)^{1/2}$. The baseline of the amide perturbation was defined as the average of the smallest 75% (0.10 p.p.m.).

RDC measurements. The ^{15}N-^{1}H RDC measurement of the ASD-1–SUP-12–RNA_{12} complex was performed by comparison of coupled two-dimensional ^{1}H-^{15}N IPAP HSQC spectra, obtained in the absence of orienting medium, against spectra obtained in the presence of acrylamide gel medium (9%), with a Bruker 700 MHz spectrometer equipped with cryoprobes. The RDC values were measured at the probe temperature of 298 K. All data were normalized to the ^{15}N-^{1}H data. On the basis of the final ASD-1–SUP-12–RNA_{12} structure, the RDC values were calculated with the PALES program[46].

Measurements of dynamic parameters. Measurements of the nitrogen relaxation times T_1 and T_2 and the proton-nitrogen heteronuclear NOEs were performed on a Bruker 600 MHz spectrometer with cryoprobe (Bruker AV 600) at 25 °C, with the ^{15}N, ^{13}C -labeled ASD-1 RRM and SUP-12 RRM in a 1:1:1 complex of ASD-1–SUP-12–RNA_{12} at a 500 μM concentration. Eight different values for the relaxation time were recorded for the ^{15}N T_1 (T_1 delays = 5, 65, 145, 246, 366, 527, 757 and 1,148 ms) and ^{15}N T_2 (T_2 delays = 32, 48, 64, 80, 96, 112, 128 and 144 ms) relaxation experiments. The ^{15}N T_1 and ^{15}N T_2 values were extracted with a curve-fitting subroutine included in Sparky (SPARKY 3, http://www.cgl.ucsf.edu/home/sparky/). The proton-nitrogen heteronuclear NOE values were

calculated as the ratio between the cross-peak intensities with (I) and without (I_0) ^{1}H saturation (I/I_0). The errors were estimated from the root mean square of the baseline noise in the two spectra. The overall correlation time (τ_c) was obtained by fitting the experimental data, with a model-free approach. Residues that exhibited overlapped resonance peaks and slightly insufficient resonance qualities for the magnetic decay analyses are not shown except for the following residues in ASD-1 RRM (*T*1: Tyr111, Gly137, Gln147, Ala155, Glu158, Gly161, Thr163, Asn172, Thr175 and His179; *T*2: Ser105, Phe120, Val123, Glu130, Ile132, Asn134, Arg136, Ser138, Gly142, Val144, Gln147, Ala155, Glu158, Thr163, Asn172 and Thr175; NOEs: Asn106, Tyr111, Ala118, Glu121, Gly124, Val127, Arg136, Gly137, Ser138, Lys139, Phe143, Thr145, Gln147, Ala155, Glu158, Asn160, Gly161, Thr163 and Thr175), SUP-12 RRM (*T*1: Thr32, Ser47, Leu51, Glu53, Glu56, Val81 and Lys84; *T*2: Phe34, Ile37, Gly41, Leu42, Ser47, Glu53, Glu56, Thr68, Asp69, Gly77, Lys84, Ile99, Ala105, Asn106, Val107 and Leu112; NOEs: Thr32, Ser47, Lys49, Leu51, Glu53, Glu56, Phe58, Val81, Ile99, Val107 and Leu109), ASD-1 RRM (*T*1: Met119, Phe120, Asp128, Ile131, Glu135, Glu170 and Arg177; *T*2: Leu102, His103, Lys110, Val123, Asp128, Glu135, Arg136, Lys139, Asp151, Arg156, Glu158 and Thr175; NOEs: Arg101, His103, Lys110, Phe120, Val123, Val126, Asp128, Glu135, Arg136, Gln147, Asp151, Arg156 and Glu170), and SUP-12 RRM (*T*1: Va26, Ser29, Arg30, Lys36, Ile37 and His45; *T*2: Val26, Ser29, Arg30, Phe34, His45, His52, Arg91, Asn97, Ile99 and Leu112; NOEs: Val26, Ser29, Arg30, Met33, Phe34, Tyr44, His45, Glu56, Arg91, Leu109, Leu112 and Lys115).

ITC measurements. ITC measurements were performed at 25 °C, with a MicroCal VP-ITC and Auto-iTC$_{200}$ calorimeter. Samples were buffered with 20 mM Tris, pH 7.0, containing 100 mM NaCl, and were thoroughly degassed before use. At first, 2.0-ml portions of the SUP-12 RRM solution (5 μM) and the ASD-1 RRM solution (20 or 25 μM) were prepared. Then 20-fold higher–concentrated RNA solutions (RNA$_6$, 5′-CUUUGUUUCAG-3′, 5′-CUUUGUU-3′ and RNA$_{12}$ for SUP-12 RRM, and RNA$_6$ and 5′-CUUUGUUUCAG-3′ for ASD-1 RRM) were injected into the protein solutions. In the case of 5′-UGCAUGG-3′, 2 mL RNA solutions (0.5 μM for ASD-1 RRM and 40 μM for SUP-12 RRM) were prepared in the cell chamber. The 20-fold higher–concentrated ASD-1 RRM and SUP-12 RRM solutions were then injected into the RNA solutions. For the ITC measurements of the ASD-1–RNA$_{12}$ binary complexes titrated with SUP-12 RRM or SUP-12(R103E) RRM, the ASD-1–RNA$_{12}$ binary complexes were purified by gel filtration after mixing of ASD-1 RRM or ASD-1(E130R) RRM with the with RNA$_{12}$ or RNA$_{12}$(G7A). The data were analyzed with MicroCal ORIGIN, with a binding model assuming a single site of interaction.

Electrophoretic mobility shift assays. For the EMSA experiments, RNA$_{12}$ and RNA$_{12}$(G7A) were subjected to the kinase reaction with [^{32}P]ATP, and the ^{32}P-labeled RNA oligonucleotides were purified with spin columns. The ^{32}P-labeled RNA oligonucleotides (10 nM) were mixed with proteins in 20 mM Tris-HCl buffer, pH 7.0, containing 60.7 mM NaCl, 1 mM DTT and 1 mg/mL yeast tRNA at 25 °C for 30 min and were applied to 8% PAGE gels under non-denaturing conditions with 0.5× Tris-borate-EDTA buffer (8.9 mM Tris base, 8.9 mM boric acid, and 0.25 mM EDTA).

Reporter minigene construction. The mutant *egl-15BGAR* splicing-reporter minigenes were constructed as described previously[49]. The sequences of the primers used in plasmid construction are available from H.K. upon request.

Worm culture and microscopy. Worms were cultured with standard methods. The *C. elegans* strains used were N2 (wild type), KH1234: *asd-1 (yb978) III; fox-1 (e2643) X* and KH1667: *sup-12 (yb1253) X*. Transgenic worms were generated as described previously[27,49]. Images of desynchronized fluorescence-reporter worms were captured with a fluorescence stereoscope (M205FA, Leica) equipped with a color, cooled CCD camera (DFC310FX, Leica). No randomization or blinding was used; images of many worms were captured instead. The color images are processed with Photoshop (Adobe).

RT-PCR. Total RNA was extracted from synchronized L1 larvae of N2, the *asd-1; fox-1* double mutant and the *sup-12* mutant, with an RNeasy Mini kit and DNase I (Qiagen). RT-PCR of the *cle-1* and *egl-15* mRNAs was performed essentially as described previously[27,49]. RT-PCR products were analyzed with BioAnalyzer (Agilent), and the sequences of the RT-PCR products were confirmed by direct sequencing. The primers used for *cle-1* were 5′-GGTGCTGAAGGTTCGGG TAC-3′ and 5′-CATGAAGTCCTGGAGCACCA-3′.

37. Kuwasako, K. *et al.* Solution structures of the SURP domains and the subunit-assembly mechanism within the splicing factor SF3a complex in 17S U2 snRNP. *Structure* **14**, 1677–1689 (2006).
38. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
39. Johnson, B.A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352 (2004).
40. Kobayashi, N. *et al.* KUJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. *J. Biomol. NMR* **39**, 31–52 (2007).
41. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).
42. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
43. Case, D.A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
44. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).
45. Tsuda, K. *et al.* Structural basis for the dual RNA-recognition modes of human Tra2-β RRM. *Nucleic Acids Res.* **39**, 1538–1553 (2011).
46. Zweckstetter, M. NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **3**, 679–690 (2008).
47. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
48. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55, 29–32 (1996).
49. Kuroyanagi, H., Ohno, G., Sakane, H., Maruoka, H. & Hagiwara, M. Visualization and genetic analysis of alternative splicing regulation *in vivo* using fluorescence reporters in transgenic *Caenorhabditis elegans*. *Nat. Protoc.* **5**, 1495–1517 (2010).

doi:10.1038/nsmb.2870