Protein-RNA and Protein-Protein Recognition by Dual KH1/2 Domains of the Neuronal Splicing Factor Nova-1

Marianna Teplova,1,7 Lucy Malinina,1,7,9 Jennifer C. Darnell,3,7 Jikui Song,1 Min Lu,6 Ruben Abagyan,5 Kiran Musunuru,3,4,8 Alexei Teplov,1 Stephen K. Burley,2,4,10 Robert B. Darnell,2,3 and Dinshaw J. Patel1,*

1Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA
2Howard Hughes Medical Institute
3Laboratory of Molecular Neuro-Oncology
4Laboratory of Molecular Biophysics
Rockefeller University, New York, NY 10021, USA
5Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
6Biochemistry Department, Weill Medical College of Cornell University, New York, NY 10021, USA
7CIC bioGUNE, Technology Park of Bizkaia, 48160 Derio-Bilbao, Spain
8Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
9These authors contributed equally to this work
10Present address: Research & Development, Eli Lilly & Co, Indianapolis, IN 46285, USA
*Correspondence: pateld@mskcc.org
DOI 10.1016/j.str.2011.05.002

SUMMARY

Nova onconeural antigens are neuron-specific RNA-binding proteins implicated in paraneoplastic opso-clonus-myoclonus-ataxia (POMA) syndrome. Nova harbors three K-homology (KH) motifs implicated in alternate splicing regulation of genes involved in inhibitory synaptic transmission. We report the crystal structure of the first two KH domains (KH1/2) of Nova-1 bound to an in vitro selected RNA hairpin, containing a UCAG-UCAC high-affinity binding site. Sequence-specific intermolecular contacts in the complex involve KH1 and the second UCAC repeat, with the RNA scaffold buttressed by interactions between repeats. Whereas the canonical RNA-binding surface of KH2 in the above complex engages in protein-protein interactions in the crystalline state, the individual KH2 domain can sequence-specifically target the UCAC RNA element in solution. The observed antiparallel alignment of KH1 and KH2 domains in the crystal structure of the complex generates a scaffold that could facilitate target pre-mRNA looping on Nova binding, thereby potentially explaining Nova’s functional role in splicing regulation.

INTRODUCTION

The generation of a versatile repertoire of functionally diverse proteins is critically dependent on alternate splicing of pre-mRNAs. Alternate splicing of the subunits of all the main neurotransmitter receptors can influence their localization, ligand-binding, signal-transducing, and electrophysiological properties. Aberrant splicing of mRNAs in highly specialized cells, such as neurons, can impact proteins critical for neuronal function, leading to the onset of neurological disease (Dredge and Darnell, 2003; Dredge et al., 2001; Licatalosi and Darnell, 2006). The proteins of the Nova (neuro-oncological ventral antigen) family, which are exclusively expressed in central nervous system (CNS) neurons, are important regulators of neuronal RNA metabolism. The Nova proteins are target antigens in the autoimmune disorder paraneoplastic opso-clonus-myoclonus-ataxia (POMA) (Musunuru and Darnell, 2001). POMA is a neurodegenerative syndrome that originates when systemic malignant tumors express proteins normally sequestered in the central nervous system (Albert and Darnell, 2004). The immune system recognizes these antigens to be nonself, and the ensuing response results in neuronal degeneration. Homozygous Nova-1 knockout mice appear to exhibit normal gestation and development, but the resulting litters show apoptosis of motor neurons in the brain stem and spinal cord and become incapacitated by progressive motor system failure (Jensen et al., 2000a).

K-homology (KH) domains are RNA-binding elements (Grishin, 2001; Siomi et al., 1993), identified in hnRNP proteins involved in RNA stabilization and translational control (Burd and Dreyfuss, 1994; Ostareck-Lederer et al., 1998; Valverde et al., 2008). Nova-1 and Nova-2 proteins each contain three proteolytically stable KH domains (Figure 1A), connected by flexible linkers. The crystallographic structures of Nova KH3 domains exhibit a three-stranded antiparallel β sheet backed by three α helices (Lewis et al., 1999). RNA targets for the Nova KH3 domain (and adjacent C-terminal residues) have been identified by in vitro selection experiments (Jensen et al., 2000b). The optimal RNA target exhibits a YCAY tetranucleotide element (Y is a pyrimidine) as part of an accessible 12-residue loop segment within the context of a RNA hairpin. The crystal structure of this complex has been determined at 2.4 Å resolution, with the majority of the intermolecular contacts restricted to the extended UCAC segment of the RNA loop, which is pinned between the invariant Gly-X-X-Gly motif and a variable loop of the KH domain.
This CA segment, which is most intolerant to mutations, is recognized by Watson-Crick-like hydrogen bonding patterns from protein residues, whereas the flanking nucleotides are restricted to pyrimidine residues.

Full-length Nova proteins containing three KH domains have been shown to target and regulate alternative splicing events within neuronal pre-mRNAs that contain repeats of the (YCAY) binding element (Buckanovich and Darnell, 1997; Dredge and Darnell, 2003; Dredge et al., 2005; Licatalosi et al., 2008; Ule et al., 2006; Zhang et al., 2010). RNA aptamers generated against full length Nova protein adopt a RNA hairpin loop scaffold containing a conserved [YCAY-(N)0-2]3 sequence in the loop segment (Buckanovich and Darnell, 1997; Yang et al., 1998). The published structure of the Nova-1 KH3-RNA hairpin complex (Lewis et al., 2000) does not address how multiple KH domains of Nova proteins might cooperate in recognizing tandem YCAY repeat modules. To address this issue in a systematic manner, we set out to determine the structure of a Nova complex containing adjacent KH1 and KH2 domains (designated KH1/2; Figure 1B) targeted to a pair of tandem YCAY RNA sites. Such an opportunity was available because in vitro selection experiments had identified a RNA hairpin architecture (Figure 1C), containing a potential 15-mer loop within which is nestled an UCAG-UCAC segment, which binds the Nova-1 KH1/2 construct (Figure 1B) with high affinity (Musunuru and Darnell, 2004). Mutagenic analysis established that both CA steps within the UCAG-UCAC segment are most intolerant to mutations and therefore represent key residues involved either in protein recognition, and/or in buttressing the recognition interface.

We now report the crystal structure of the Nova-1 KH1/2 construct bound to the UCAG-UCAC segment-containing RNA hairpin. Strikingly, the RNA hairpin is targeted solely by the KH1 domain, with both loop UCAG and UCAC tetranucleotide sites contributing to recognition. Both the crystal structure and binding studies establish that the KH2 domain does not interact with the UCAG-UCAC-containing RNA hairpin. In the crystal, the RNA-binding surface of the KH2 domain instead engages in protein-protein interactions with symmetry-related KH2 domain. By contrast, Nuclear magnetic resonance (NMR) chemical shift perturbations, gel-shift, and ITC binding studies demonstrate that the KH2 domain does target UCAY tetranucleotide-containing RNAs in a sequence specific manner in solution. Nova’s functional role in splicing regulation could be explained by the observed antiparallel alignment of KH1 and KH2 domains in the crystal structure of the complex, thereby providing an ideal scaffold to induce target pre-mRNA looping on Nova binding.

RESULTS

Crystallization and Structure Determination of Complex

A Nova-1 KH1/2-expression construct (Figure 1B) was engineered lacking a 24-amino acid segment between the KH1 and
Table 1. Crystal Data, Data Collection, and Refinement Statistics for Nova-1 KH1/2-RNA Hairpin Aptamer Complex

<table>
<thead>
<tr>
<th>Crystal data</th>
<th>Type I</th>
<th>Type II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C 2</td>
<td>C 2</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>155.3</td>
<td>159.5</td>
</tr>
<tr>
<td>b (Å)</td>
<td>37.2</td>
<td>37.9</td>
</tr>
<tr>
<td>c (Å)</td>
<td>34.8</td>
<td>38.5</td>
</tr>
<tr>
<td>β</td>
<td>94.4</td>
<td>96.6</td>
</tr>
<tr>
<td>Asymmetric unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA bases</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>KH1+KH2 residues</td>
<td>148</td>
<td>155</td>
</tr>
<tr>
<td>Solvent: H$_2$O / ion</td>
<td>77/3 (K$^+$, 2Mg$^{2+}$)</td>
<td>155/3 (K$^+$, 2Mg$^{2+}$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source detector</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Last shell</td>
</tr>
<tr>
<td>Unique reflections</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Mosaicity</td>
</tr>
<tr>
<td>R-merge$^a$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement (F&gt;0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work/free reflections</td>
</tr>
<tr>
<td>R-factor/R-free</td>
</tr>
<tr>
<td>Rmsd</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
</tr>
<tr>
<td>Average B-factors (Å$^2$)</td>
</tr>
</tbody>
</table>

Values in parentheses correspond to the last resolution shell. $^a$R-merge = $\Sigma_{i}$$\Sigma_{[hkl]}$(I$[hkl]$ - <I$[hkl]$>)/$\Sigma_{i}$$\Sigma_{[hkl]}$$<I$[hkl]$> over i observations.

The crystal structure of two Nova-1 KH1/2-RNA hairpin complex molecules related by 2-fold crystallographic symmetry is shown in Figure 1E. The KH1 (gold) and KH2 (cyan) domains are aligned in an antiparallel orientation. The UCAG-UCAC-containing RNA hairpin (green) is targeted solely by the KH1 domain, whereas the canonical RNA-binding surface of KH2 engages in protein–protein interactions with symmetry-related KH1. The packing arrangement of KH1/2-RNA hairpin complexes in the crystal and the details of KH2-KH2 interface are outlined in the Supplemental Experimental Procedures and Figure S2.

In contrast, analytical ultracentrifugation measurements indicate that KH1/2 exists as a monomer in solution at 10–200 µM concentration range, and the KH1/2-RNA complex sediments as a 1:1 complex under these conditions (for details see Supplemental Experimental Procedures and Figure S3).

RNA Hairpin Architecture in the Bound State

The entire UCAG-UCAC-containing RNA hairpin (Figure 2A) can be traced in the crystal structure of the type II complex, whereas the 5’-C1-G2 segment is disordered in the crystal structure of the type I complex. The simulated annealing 2Fo-Fc omit map of the UCAC RNA site at 1σ level for the type I complex is shown in Figure S1C. The conformation of the UCAG-UCAC is of great interest, due to the potential of each tetranucleotide element serving as a KH domain binding site. All four-hairpin residues within the gold-colored U9 to G12 segment (Figure 2B) form a stacked array. By contrast, residues U13 and C14 loop out, whereas residues A15 and C16 form a stacked pair within the cyan-colored U13 to C16 segment (Figure 2B). The conformation of the bound RNA hairpin is shown in stereo in Figure 2C and establishes extensive stacking interactions that propagate from the paired stem segment into both arms of the hairpin, with the stacking discontinuities restricted to the G12 to A15 segment.

The Watson-Crick paired stem of the RNA hairpin zippers up through formation of G6-A20, G7-A19, and A8-C18 noncanonical pairs (Figure 2D). The two noncanonical G-A pairs are of the sheared type, whereas the A8-C18 noncanonical pair is buckled and involves bifurcated and water-mediated hydrogen bonds. The small twist angle at the G6-G7 step and the large twist angle at the G7-A8 step, results in a pronounced cross-strand stack between residues A8 and A19 (Figures 2E and 2F). This results in a continuous helical stack that spans G25 to A19 and A8 to G12. The other continuous helical stack spans...
C1 to G7 and C18 to A15. The zippered-up helical segment involving the three noncanonical pairs, is further stabilized by a pair of hydrated divalent Mg²⁺ ions and a monovalent K⁺ ion (with partial occupancies), as shown in stereo in Figure 2G.

Structure of Interface between Nova KH1 and KH2 Domains in Complex

The KH1 and KH2 domains align in an antiparallel orientation across the KH1/2 interface, with interdomain contacts involving segments of α3 and β1 within each domain (Figure 3A, pink). The three-stranded antiparallel β sheets are on the same side of the protein, thereby forming a continuous six-stranded β sheet (Figure 3B). The β1 strands are highly twisted relative to each other and the resultant β sheet between them is stabilized by NH•••O main chain hydrogen bonds involving Ile11(KH2)•••Lys9(KH1) and Leu11(KH1)•••Lys9(KH2) (Figure 3C). The KH1/2 interface buries 1720 Å² of protein surface, with the interface dominated by van der Waals contacts between conserved hydrophobic and aromatic residues, plus a salt bridge between the side chains of Glu53(KH1) and Arg4(KH2).

The linker segment (residues 81–99) connecting KH1 and KH2 is disordered in the crystal structure of the KH1/2-RNA complex, consistent with its sensitivity to proteases (Lewis et al., 1999). Pronase cleaves the linker at the Ser90-Ile91 step permitting comparative gel shift studies of complex formation for intact and cleaved KH1/2 with the UCAG-UCAC-containing RNA hairpin (Figure 3D). These studies establish that complex formation is independent of the cleavage status of the linker connecting KH1 and KH2, suggestive of linker-independent association across the KH1/2 interface.

Surface Interaction Propensities of Nova KH Domains

To independently evaluate the interaction propensity of the surface regions of Nova KH domains (Figures 3E–3G), we applied the “Optimal Docking Area” (ODA) algorithm (Fernandez-Recio et al., 2005). The calculated energy changes reflect the effect of replacing the water environment for a lower dielectric (but still polar) medium, such as the surface of another protein. The highest protein-protein interaction propensities span adjoining β1/α3 elements for Nova KH1 (Figure 3E) and KH3 (Figure 3G) domains. These results are consistent with the β1/α3 elements of Nova KH1 forming contacts with KH2
Figure 3. Ribbon Representations of KH1/2 Interface in the KH1/2-RNA Hairpin Complex and Distribution of the Protein-Protein Interaction Propensities over the Surface of Nova KH Domains

(A) A global overview of the KH1(gold)/KH2(cyan) interface, with the segments constituting the interface, colored in pink.

(B) The three-stranded β sheets of the KH1 and KH2 domains form a continuous twisted six-stranded β sheet.

(C) Details of the KH1/2 interface, highlighting interactions between side chains of residues involved in interfacial contacts.

(D) SDS-PAGE electrophoresis (left) showing pronase cleavage resulting in two protein fragments (cleavage at Ser90-Ile91 step) with molecular mass of ~9 and 10 kDa. Native PAGE (middle and right) establishing complex formation between KH1/KH2 and RNA, regardless of cleavage of the linker.
Structure

Structure of Nova-1 KH1/2-RNA Complex

(Figures 3A) in the Nova KH1/2-RNA hairpin complex, and with the β1/α3 elements of Nova KH3 forming contacts with symmetry-related KH3 in the crystal (Lewis et al., 2000). By contrast, the highest surface interaction propensities span adjoining β2/α2 elements, the loop between α1- and α2-helices, and the C terminus of α1-helix for Nova KH2 (Figure 3F). These elements are located on the canonical RNA-binding surface of KH domains and it is this surface of KH2 that participates in protein-protein interactions with symmetry-related KH2 in the crystal (Figure S2B).

Recognition between KH1 Domain of Nova KH1/2 and RNA Hairpin

Protein-RNA recognition in the Nova-1 KH1/2-RNA hairpin complex involves only one of the two KH domains and both UCAG-UCAC tetranucleotide elements within the RNA hairpin. Specifically, the KH1 domain primarily targets the U13 to C16 segment, and the C10 base of the preceding U9 to G12 segment (Figure 4A), with C14 and A11 buttressing each other through hydrogen bonding interactions (Figure 4B). The U13 base loops out of the hairpin and interacts with the KH1 domain through both its base and sugar residues (Figure 4A). The U13 base stacks over the Gly18-Ser19 peptide bond, with O5' of U13 forming van der Waals contact with C16 of Ser19 and O2' of U13 forming a weak hydrogen bond with main chain NH of Lys23 (Figure 4C). In addition, the 2'-OH of the U13 sugar forms a bifurcated hydrogen bond with the main chain O atoms of Gly18 and Gly22. The C14 base forms two hydrogen bonds through its Watson-Crick edge with the guanidinium group of Arg54, with C14 buttressed in position through a network of base-sugar/sugar-base-phosphate hydrogen bonds with A11 of the adjacent U9 to G12 segment (Figure 4B). The next residue, A15 forms a pair of hydrogen bonds to the main chain NH and O atoms of Leu41, and is further buttressed in place by a bifurcated water-mediated hydrogen bond from its base (N3) and sugar (2'-OH) atoms to the side chain of Gln32. The network of hydrogen bonds and water-mediated bridges (w1 and w2) that direct the observed intermolecular recognition and buttressing events, are specific for cytосine at position 14 and adenine residues at positions 15 and 11 in the complex (Figure 4B). The Lys40 and Lys43 side chains are directed toward the C16 and C10 bases, with the specific intermolecular contacts involving these lysine side chains differing between the type I and type II complexes (Figure 4A). The Lys43 ε-amino group forms a hydrogen bond with either N3 atom of C16 or O2 atom of C10. When Lys43 is occupied in the interaction with C10, the N3 atom of C16 is hydrogen bonded to the Lys40 ε-amino group, with the rest of the Lys40 side chain forming van der Waals contacts with O2 atom of C16 (Figure 4D).

Binding of Nova KH1/2 Domains to UCAG-UCAC-Containing RNA Hairpin

We have used filter-binding assays to measure the binding affinities of the individual KH2 and KH3 domains, KH1/2 dual (E–G) Distribution of the protein-protein interaction propensities over the protein surface as calculated by the optimal docking area (ODA) approach for three different KH domains. Ribbon representation of KH1 (E) (this study), KH2 domains of the Nova-1 KH1/2 protein (F) (this study), and Nova-2 KH3 domain (G) (Lewis et al., 2000). Structures are colored by the absolute magnitude of the ODA signal from the strongest in red, through medium in orange and weak in yellow, to the weakest in white. See also Figure S5.
domains, and full-length KH1/2/3 Nova protein to a panel of three RNA hairpin targets. These include the KH3-binding RNA aptamer containing a single UCAC site (Figure 5A), the KH1/2-binding RNA aptamer containing a UCAG-UCAC site (Figure 5B), and the KH1/2/3-binding RNA aptamer containing a (YCAY-N)_3 site (Figure 5C). Thus, although the intact KH1/2/3 Nova protein (Figure 5D, open circles) and KH1/2 dual domains (Figure 5D, black circles) bind the UCAG-UCAC-containing RNA aptamer target (Figure 5B) with μM affinity, the KH3 domain (Figure 5D, green circles) exhibits reduced binding affinity, whereas the KH2 domain (Figure 5D, red circles) exhibits no measurable binding affinity.

We have used filter-binding assays to compare the binding affinities of wild-type (KD = 0.42 ± 0.11 μM; Figure 5E, black circles) and mutant KH1/2 dual domains for the KH1/2-binding UCAG-UCAC-containing RNA hairpin (Figure 5B). Replacement of Ser14 in KH1 by its Glu counterpart observed at this position in KH3, within the context of the KH1/2 construct, resulted in an ~4-fold decrease in RNA hairpin binding affinity to 1.6 ± 0.2 μM (Figure 5E, green circles). Replacement of Lys40 and Lys43 in KH1 by their Gln counterparts observed at these positions in KH2, within the context of the KH1/2 construct, resulted in no measurable RNA hairpin binding affinity (Figure 5E, red circles).

### Binding of Nova KH Domains to (YCAY)_n-Containing RNA Hairpins

Filter-binding assays have also been used to measure the binding affinities of the KH2 and KH3 individual domains (KH1 is unstable as an individual domain) and KH1/2/3 dual domains to RNA aptamer hairpin targets containing (YCAY)_n sites, where n = 1, 2, 3 (Figures 5A–5C). The KH1/2 dual domains bind most tightly to the RNA aptamers containing UCAG-UCAC (KD = 0.42 ± 0.11 μM; Figure 5F, black circles) and (YCAY-N)_3 (KD = 1.9 ± 0.3 μM; Figure 5F, green circles), with weaker binding affinity to the RNA aptamer containing a single UCAC site (Figure 5F, red circles).

The KH2 domain binds very weakly to all three RNA aptamers (Figure 5G), with the residual binding observed for the RNA aptamer containing a (YCAY-N)_3 site (Figure 5G, green circles) being nonspecific, because a similar binding curve is observed for an unrelated FMRP KH2-domain-binding RNA aptamer, kc1 (Figure 5G, open circles).

Finally, the KH3 domain binds most tightly to its own RNA aptamer containing a single UCAC site (KD = 0.58 ± 0.09 μM; Figure 5H, red circles), less tightly to the RNA aptamer containing a (YCAY-N)_3 site (KD = 3.4 ± 0.6 μM; Figure 5H, green circles), and weakest to the RNA aptamer containing a UCAG-UCAC site (Figure 5H, black circles).
NMR Chemical Shift Mapping of RNA-Binding Surface of Nova KH2 Domain

Structural similarities between all three Nova-1 KH domains, as well as the conservation of a key recognition residue Arg54, suggests that the KH2 domain could recognize the YCAY sequence in a manner similar to that observed for KH1 (this study) and KH3 (Lewis et al., 2000) domains in their respective crystal structures. Nevertheless, we detect no binding of the KH2 domain to (YCAY)n-containing RNA hairpins in either crystal (Figure 1E) or solution (Figure 5G) states.

To elucidate whether Nova-1 KH2 is capable of YCAY-containing single-stranded RNA binding, we performed chemical shift perturbation studies of the KH2 domain in the free state and bound to the UCAC RNA tetranucleotide by NMR spectroscopy. We could assign 96% of backbone residues (excluding prolines) in a 1H-15N-HSQC spectrum of the complex, with slow exchange observed between free and bound states (Figure 6A). The largest weighted 1H/15N chemical shift changes on complex formation map to the \( \beta_2 \) and \( \beta_3 \) strands and their connecting loop, as well as to the residues located on the \( \alpha_1 \) and \( \alpha_2 \) helices near the GXXG loop (Figure 6B). These changes when mapped onto the backbone structure of the KH2 domain (Figure 6C), show that KH2 residues Arg54, Leu41, Lys40, Gly22, Ser19, and Gly18, whose KH1 equivalents mediate intermolecular contacts with 13-UCAC-16 motif in the crystal structures of KH1/2-RNA complexes (Figure 4), are located either within or adjacent to these regions (Figure 6C).

Nova Tandem and Single KH Domains Bind to [UCAU(N)2]3 Motif

In view of the observation that the KH2 domain is able to specifically recognize an isolated UCAC tetranucleotide RNA sequence, we next studied binding of tandem KH1/2 and single KH2 to in vivo Nova-1 target—an intronic UCAU repeat element of glycine receptor \( \alpha_2 \) (GlyR \( \alpha_2 \)) pre-mRNA (Figure 7A). Previous studies have identified that Nova-1 binds to the GlyR \( \alpha_2 \) 50-nt sequence in vitro with \( \sim \)20 nM affinity via the [UCAU(N)2]3 motif (Buckanovich and Darnell, 1997), and enhances utilization of an
adjacent 3′-splice site of an alternatively spliced exon in vivo (Jensen et al., 2000a). KH1/2 binds GlyR2 20-nt RNA element (designated GlyR20; Figure 7A) with 1:1 stoichiometry, whereas KH2 binds this RNA with 2:1 stoichiometry, as determined by electrophoretic mobility shift (gel-shift) assay (Figure 7B). According to analytical ultracentrifugation measurements over a 10-fold range of protein concentrations, the apparent molecular mass of KH2 in the free state (10–100 μM concentration range) is 17.1 kDa (Figure 7C), indicating that KH2 forms a dimer in solution. Consistently, the gel-filtration column elution volume of KH2 also indicates that it is dimeric in solution (data not shown). Single RNA band shifts observed on KH1/2 binding at 1:1 protein/RNA molar ratio and KH2 binding at 2:1 protein/RNA molar ratio (Figure 7B) suggest similar modes of KH1/KH2 intramolecular pseudodimer and of KH2/KH2 homodimer binding to GlyR20, with each KH domain of the dimer bound to one of three UCAU motifs of the RNA. We have monitored the binding affinities of Nova-1 KH1/2, KH2, or KH3 to GluR20 by isothermal titration calorimetry (ITC). The dissociation constant ($K_D$) values are 0.13 μM for the KH1/2 (Figures 7D and 7G),
0.78 μM for the KH2 (Figures 7E and 7G) and 0.30 μM for the KH3 (Figures 7F and 7G) complexes with GlyR20. The n value (a number of binding sites per protein monomer) estimated from ITC curves using a single site binding model is 0.89 for the KH1/2-RNA complex, whereas n = 0.43 for KH2-RNA complex, consistent with 1:1 and 2:1 protein/RNA stoichiometries of these respective complexes (Figures 7D, 7E, and 7G). The n value of 0.38 estimated from ITC curve (Figures 7F and 7G) for the KH3-RNA complex indicates that stoichiometry of KH3-GlyR20 complex is either 2:1 or 3:1. These experiments demonstrate that individual Nova KH2 and KH3 domains bind [UCAU(N)]2 motif of single stranded RNA with comparable (2.6-fold difference) binding affinity, while dual KH1/2 domain binds this ssRNA with 6-fold higher affinity compared to a single KH2 domain.

We have also monitored binding of KH1/2 to GlyR20 RNA by NMR spectroscopy. We observed significant chemical shift perturbations between free RNA and RNA-bound states (Figure S4A). The number of KH1/2 resonances that experienced chemical shift changes is at least twice the number observed for individual KH2 (Figure 6B), suggesting that both KH2 and KH1 domains are bound to GlyR20 in the complex. For example, in addition to Gly22 and Gly25 of KH2 domain, there are two other cross-peaks in the same region of the KH1/2 1H-15N-HSQC spectrum that undergo significant chemical shift changes on RNA binding (compare the selected regions of the spectra of KH2 and KH1/2 in Figures S4B and S4C, respectively).

DISCUSSION

KH domains, which constitute one of the most abundant RNA-binding motifs (Messias and Sattler, 2004), are of particular interest because of their role in regulating gene expression at both transcriptional and translational levels, in alternate splicing, and in the maturation of mRNA (Valverde et al., 2008). Eukaryotic KH domains bind their mRNA targets with sequence specificity and affinities spanning the μM to nM range, and target tetranucleotide sequences where the first and last bases are pyrimidines, with structures of complexes available where the third base is an adenine or cytosine. Our current understanding of the principles underlying KH-RNA recognition emerged following determination of the crystal structure of Nova-2 KH3 domain bound to a UCAC element within the hairpin segment of a RNA stem-loop (Lewis et al., 2000), the NMR structure of the STAR (signal transduction and activation of RNA) member KH domain augmented by a QUA1 (quaker homology 1) domain of SF1 (splicing factor 1) bound to a UAAC segment in a single-stranded RNA context (Liu et al., 2001), the complex between NusA, a key regulator of bacterial antitermination, and RNAs derived from the antitermination region (Beuth et al., 2005), and the crystal structures of KH1 domain of pol(Y)-binding protein 2 (PCBP2) bound to C-rich telomeric DNA and RNA (Du et al., 2007). In each case, the RNA tetranucleotide is gripped in a molecular vise composed of a hydrophobic α-helix/β strand platform, an invariant GXGXG motif and a variable loop, thereby aligning the Watson-Crick edge of the adenine at the third position for hydrogen-bonding with the peptide backbone.

The current contribution reports the crystal structure of dual eukaryotic KH domains of Nova-1 targeted to tandem RNA tetranucleotide repeats, where KH1 and KH2 domains form an intramolecular pseudodimer, with KH1 solely involved in RNA recognition and KH2 involved in protein-protein contacts. These results highlight the ability of KH domains to participate in protein-RNA and protein-protein recognition and outline how RNA-RNA buttressing interactions expand on our current understanding of KH-RNA tetranucleotide recognition events.

Comparison of Nova KH1/2 Interface with Other KH-KH Interfaces

Our structure of the complex establishes that KH1 and KH2 have extensive interdomain interactions, forming an intramolecular pseudodimer. Two recently reported structures revealed a similar arrangement of tandem KH domains in the RNA-free state: the NMR structure of KH1/2 domains of human PCBP2 (Figure SSA) (Du et al., 2008), and the crystal structure of IMP1 mRNA-Binding Protein 1 (also called ZBP1) KH/4 domains (Figure SSB), whereas three other structures of tandem KH domains in the free state revealed different modes of arrangement of their respective KH domains (Figures SSC–SSE) (Diaz-Moreno et al., 2010; Gopal et al., 2001; Valverde et al., 2007; Worbs et al., 2001). A detailed comparison of the Nova KH1/2 interface with these KH-KH interfaces is given in the Supplemental Information.

Novak KH1/2-RNA Recognition Involving Tandem YCAY Repeats

The three KH domains of Nova proteins have been shown to target [YCA(N)]3 sites in natural α-2 glycine receptor subunits, GABAA and Nova-1 pre-mRNAs (Buckanovich and Darnell, 1997; Dredge and Darnell, 2003; Dredge et al., 2005) and [YCA(N)]2 sites in RNA hairpin aptamers identified through in vitro selection (Buckanovich and Darnell, 1997). Nevertheless, it was not clear whether individual KH domains target one or more tandem YCAY sites or whether one or more KH domains target a single YCAY site. The structure of the Nova-1 KH1/2-RNA hairpin complex reported in the present study addresses this issue by revealing a novel recognition mode, where the KH1 domain differentially recognizes adjacent tandem UCAG and UCAC sites, mediated by specific buttressing interactions between the two sites. The intermolecular interactions are primarily between the KH1 domain and all four nucleotides of the second UCAC tetranucleotide (U13–C16) site, with a few additional interactions with C10 and its phosphate of the first UCAG tetranucleotide (U9–G12) site in the complex (Figure 4A). Both UCAG and UCAC sites are involved in sculpting the RNA scaffold due to the extensive base-sugar and base-phosphate interactions associated with the mutual buttressing of A11 of the first tetranucleotide repeat and C14 of the second tetranucleotide repeat in the complex (Figure 4B).

The recognition elements between the KH1 domain and the second UCAC tetranucleotide (U13–C16) site in the Nova-1 KH1/2-RNA hairpin complex reported in this study are very similar to those reported previously between the KH3 domain and the single UCAC site in the Nova-2 KH3-RNA hairpin complex (Lewis et al., 2000). In both complexes, the sequential uracil, cytosine and adenine residues of the UCAC segment are positioned atop an α/β platform and pinioned in place by interactions with the invariant Gly-X-Gly motif and the variable loop. The
cytosine pairs with R54 using Watson-Crick-like hydrogen bonding alignments and the adenine pairs with the backbone using Watson-Crick-like hydrogen bonding alignments (Figure 4B).

We highlight below a common intermolecular recognition feature that has been observed in the eukaryotic Nova-2 KH3-RNA hairpin (Lewis et al., 2000) and Nova-1 KH1/2-RNA hairpin (this study) complexes, as well as in the splicing factor 1-intron branch site RNA complex (Liu et al., 2001) and the bacterial NusA KH1/2 complex with a GAACUCAAUAG RNA sequence (Beuth et al., 2005). All these complexes have one key element in common, namely that the main chain of the β2-strand forms two hydrogen bonds with the Watson-Crick edge of the adenine base (Figure 4B). A more detailed comparison of Nova KH1/2-RNA hairpin and NusA KH1/2-ssRNA complexes is provided in the Supplemental Experimental Procedures.

Comparison of RNA Hairpin Topologies in Nova KH1/2 and KH3 Complexes

The Nova KH1/2 and KH3 domains bind preferentially their respective RNA hairpin aptamers, whereas they bind each other’s aptamers significantly weaker (Figures 5F and 5H), implying that not only the presence of primary UCAY recognition site but also the RNA hairpin loop structure determine binding affinity and specificity of individual Nova KH domains. The charge distributions are somewhat different on the canonical RNA-binding surfaces of the three Nova KH domains (Figure 8). The electrostatics of the RNA-binding KH1 surface in the KH1/2-RNA hairpin complex (Figures 8A and 8B) appears to be designed to favor binding by a compact RNA loop (contains nine bases), as reflected in a sharp turn facilitated by the Lys45 side chain. The conformation of the primary U13-C14-A15-C16 target site (Figure 4A, cyan) together with the sharp turn results in the mutual buttressing of C14 by A11 of the neighboring site (Figure 4B). At the same time, two appropriately located positively charged residues Lys40 and Lys43, which adopt finger-like conformations, recognize cytosine bases C16 and C10, respectively (Figure 4A). This network of intermolecular and buttressing interactions results in both the primary UCAC target site and the CA step of the preceding nine bases), as reflected in a sharp turn facilitated by the Lys45 side chain. The conformation of the primary U13-C14-A15-C16 target site (Figure 4A, cyan) together with the sharp turn results in the mutual buttressing of C14 by A11 of the neighboring site (Figure 4B). At the same time, two appropriately located positively charged residues Lys40 and Lys43, which adopt finger-like conformations, recognize cytosine bases C16 and C10, respectively (Figure 4A). This network of intermolecular and buttressing interactions results in both the primary UCAC target site and the CA step of the preceding U9-C10-A11-G12 site participating in KH1 domain recognition and stabilization.

The electrostatics of the RNA-binding KH3 surface in the KH3-RNA hairpin complex (Lewis et al., 2000) appears to be designed to favor a less-restrictive RNA loop (contains 10 bases) topology, as reflected in no sharp turns in the loop trajectory (Figure 8E). Two important differences between the electrostatics of the RNA-binding surfaces of KH1 (Figure 8A) and KH3 (Figure 8E) are the presence of Glu14 and Lys44 in the latter complex. This distribution of charged amino acids appears to contribute to the separation of the 5’- and 3’-halves of the loop in the KH3 complex, such that the loop topology adopts a more gradual turn trajectory (Figure 8E). The positively charged residues Arg83 and Arg52 together with the negatively charged Glu14 side chain direct the RNA backbone toward the C terminus of the KH3 domain (Figure 8E). No buttressing interactions are
observed across the loop in the KH3-RNA hairpin complex, due to the increased separation between the opposite sides of the loop.

The importance of the electrostatics of the RNA interacting surface of the Nova KH domains to recognition is reinforced from the observed reduction in binding affinities associated with the KH1/2-RNA hairpin complex for the KH1 Ser14Glu single mutant and the KH1 Lys40,43Gln dual mutant (Figure 5E). In the single mutant, neutral serine in KH1 is replaced by negatively-charged glutamate observed at this position in KH3, whereas in the dual mutant, positively-charged lysines observed at these positions in KH1 are replaced by neutral glutamines observed at these positions in KH2.

The KH2 domain does not participate in RNA recognition in the Nova-1 KH1/2-RNA hairpin complex (Figure 1E), and it does not appear to bind either UCAG-UCAC- or UCAC-containing RNA hairpin aptamers with high affinity and binds very weakly to a RNA hairpin containing (YCAY-N3) sites (Figure 5G). The charge distributions are somewhat different on the canonical RNA-binding face of the KH2 domain (Figures 8C and 8D), due to Gin40 and Gin43 replacing their lysine counterparts in the KH1 domain (Figures 8A and 8B). We attempt to explain why the canonical RNA-binding surface of the KH1 domain as part of a KH1/2 construct (Figure 5D, black circles), but not the KH2 domain (Figure 5D, red circles), targets the Nova-1 KH1/2-RNA hairpin. We put forward two potential contributors to such discriminative recognition following a detailed structural analysis of protein-RNA contacts in the complex, as outlined in the Supplemental Experimental Procedures.

All Nova KH Domains Bind UCAY Repeat Single-Stranded RNA

We demonstrated that all three Nova-1 KH domains bind [UCAU(N)]3 repeats of a single-stranded RNA, with comparable binding affinities between individual KH2 and KH3 domains (Figure 7). NMR chemical shift perturbation studies (Figures 6A and 6B) establish that Nova KH2 binds the UCAC RNA tetranucleotide in slow exchange, a feature characteristic of a tightly bound complex in solution. These results on a single-stranded RNA that spans the KH domain binding pocket contrasts with no measurable binding when the same tetranucleotide element, or its tandem repeats, is part of the loop segment of RNA hairpins (Figure 5G). The chemical shift changes in the KH2 domain on UCAC binding are mainly distributed within the nucleic acid-binding channel (Figure 6C), and their magnitude is in the range reported for Nova KH3 binding to ssRNA pentanucleotide sequences containing UCAN (N is either pyrimidine or purine) elements (Beuth et al., 2007).

The ability of KH2 to recognize UCAC motif is fully expected based on structural similarity of its RNA-binding surface to that of the other two Nova KH domains, as well as the conservation of a key recognition residue Arg54. A 2.6-fold lower binding affinity of KH2 than KH3 for [UCAU(N)]3 repeat ssRNA is consistent with less positively charged RNA binding surface of KH2 domain compared to that of KH3 domain (Figures 8C–8F). We have established that KH2 domain in an isolated state forms a dimer in solution. Because hydrophobic interactions dominate the KH1/KH2 interface of the dual KH1/2 domain, the KH2 domain in the absence of KH1 binding partner likely forms the KH2/KH2 homodimer through the same hydrophobic β1/α3 surface. Such very similar dimeric interfaces have been previously observed between KH1 domains of PGBP2 in the crystal structures of its complexes with poly(C) nucleic acids (Du et al., 2007) and between the tandem KH1 and KH2 domains of the protein (Du et al., 2008) (Figure S5A). Therefore, it is possible that KH2/KH2 homodimer and KH1/KH2 intramolecular pseudodimer bind RNA in a similar way, such that each KH domain of a dimer bind the two terminal UCAU motifs of GlyR20, consistent with 1:1 (RNA/dimer) stoichiometry determined for each complex. The NMR chemical shift perturbations observed for the intramolecular KH1/2 pseudodimer on binding GlyR20 (Figure S4) are also consistent with the involvement of both KH domains in the interaction with the RNA. This model is stereochemically possible as the 8 nucleotide intervening sequence is of sufficient length to connect the two terminal UCAU motifs simultaneously bound by the two KH domains of a dimer.

Nova KH Domain Surface Mimicry Associated with Protein and RNA Recognition

The same canonical RNA-binding surface of the KH domain participates in protein-protein interactions in the crystal (KH2 domain) (Figure 9A, top left) or in RNA recognition (KH1 domain) (Figure 9B, top left) in the Nova-1 KH1/2-RNA hairpin complex. A competing molecular recognition duality has also been reported for the Y14 protein, whose canonical RNA-binding surface, is also involved in protein-protein interactions with Mago, with both Y14 and Mago being components of the exon junction complex (Fribourg et al., 2003; Lau et al., 2003; Shi and Xu, 2003).

Equally striking is the extension of the structural mimicry to an element of the interacting partner. Thus, the indole ring of a Trp residue mediates the protein-protein contacts involving KH2 domains in the crystal (Figure 9A), whereas the purine ring of the key adenine base occupies the same interfacial position in the KH1-RNA interaction (Figure 9B) in the Nova-1 KH1/2-RNA hairpin complex.

Alternate Models Involving RNA Looping on the Surface of Nova-1 KH1/2

Our ultracentrifugation studies imply that Nova-1 KH1/2 forms a monomer in the free state (Figure S3A) and when bound to its RNA aptamer target (Figure S3B) in dilute solution. The KH1 and KH2 domains interact with each other through an extensive interface (Figures 3A and 3C), thereby positioning their canonical RNA-binding surfaces at opposite ends of the molecule and accessible for recognition. Thus, an RNA containing a pair of sequence elements capable of targeting KH1 (with high affinity) and KH2 (most likely with lower affinity) and separated by a linker segment of sufficient length, should be capable of complex formation through a RNA looping mechanism (Figure 9C), as initially proposed for splicing regulation from the structure of tandem RNA-binding domains RBBD3/4 of polypyrimidine tract-binding protein bound to RNA (Oberstrass et al., 2005). An RNA looping mechanism has also been proposed for RNA recognition based on the structure of tandem zinc-finger domains of Muscleblind-like protein bound to RNA (Teplova and Patel, 2008), KH1/2 domains of polyC-binding protein in...
the free state (Du et al., 2008) and KH3/4 domains of IMP1 protein in the free state (Chao et al., 2010).

An alternate possibility for an RNA looping mechanism is based on the X-ray structure of the complex (Figure 1E), where the RNA-binding surfaces of the KH2 domains pack against each other in the crystal lattice. Thus, an RNA containing a pair of sequence elements capable of targeting KH1 and separated by a linker segment of sufficient length, could also be capable of complex formation through an RNA looping mechanism (Figure 9D).

We have no definitive evidence differentiating between models of RNA looping between the 1:1 KH1/2-RNA complex shown in Figure 9C and the 2:1 KH1/2-RNA complex shown in Figure 9D, especially in the context of full-length Nova protein containing KH1/2/3 domains and under cellular conditions, where additional binding proteins could modulate the recognition process. These models are proposed in the spirit of stimulating further research in this area.

**Biological Implications**

Nova proteins can either enhance or inhibit exon inclusion depending on the position of its binding sites relative to the regulated exon (Ule et al., 2006). If the KH domains of Nova bind intronic YCAY clusters, it could enhance exon inclusion, whereas if it binds YCAY clusters located within the exon or within intronic region immediately adjacent to an alternative exon, it could inhibit exon inclusion. Our structural models of target pre-mRNA looping induced by Nova binding (Figures 9C and 9D) could provide one explanation of Nova’s function in splicing regulation. Looping of the intron induced by Nova binding might help to bring 5' and 3' splice cites in close proximity and thereby enhance spliceosome assembly and exon inclusion (Figure 10A). By contrast, looping out an entire exon or adjacent regions could prevent assembly of the spliceosome components on pre-mRNA, subsequently resulting in exon skipping (Figure 10B).

The current structural data provides a satisfying mechanistic explanation for the functional interaction maps generated that describe Nova actions on target pre-mRNAs (Licatalosi and Darnell, 2010; Licatalosi et al., 2008; Ule et al., 2006). The looping model discussed above not only suggests explanations for these position-dependent interactions, but may be extended in several additional directions. Hints of a Nova-dependent positional map in which regulation of alternative polyadenylation (Licatalosi et al., 2008) may have analogous explanations in the looping model described. Moreover, we have found that Nova has functional interactions with other RNABPs. These include inhibitory interactions with PTBP2 (Polydorides et al., 2000), suggesting that some protein-protein interactions may inhibit Nova-RNA-looping bridges from forming. In contrast, interactions with the Fox1/2 RNA binding protein (Zhang et al., 2010) are agonistic, suggesting that dimer/heterodimer formation may indeed play an important role providing sufficient RNA binding affinities to allow loop formation.
Figure 10. Models of RNA Looping Induced by Nova KH1-3 Domains Binding to Splicing Enhancers and Splicing Silencers Predicted by Nova RNA Map of Splicing Regulation

(A) Model based on the examples of Nova upregulation of exon inclusion by binding to intronic splicing enhancer elements (red circles) to enhance spliceosome assembly.

(B) Model based on the examples of Nova inhibition of exon inclusion by binding to exonic splicing silencing element (blue circle) and blocking U1 snRNP (U1) assembly on the pre-mRNA, and Nova inhibition of exon inclusion by binding to intronic splicing silencing element immediately upstream of alternative exon by blocking recognition of the 3' splice site by U2 snRNP (U2). KH1, KH2, and KH3 are colored gold, cyan, and purple, respectively. The linkers between KH1 and KH2, and between KH2 and KH3 are represented by the dotted gray lines. The green line represents RNA loop.

EXPERIMENTAL PROCEDURES

Detailed procedures for protein and RNA preparation, ITC measurements, crystallization and data collection, structure determination and refinement, NMR sample preparation, NMR spectroscopy and chemical shift assignments, NMR relaxation measurements, gel electrophoretic mobility shift binding assays, and filter binding assays are listed in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Atomic coordinates and structure factors for the type I and type II Nova-1 KH1/KH2-RNA hairpin complexes have been deposited in the Protein Data Bank (www.rutgers.edu/pub) under accession codes 2ANN and 2ANR, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at doi:10.1016/j.str.2011.05.002.

REFERENCES


ACKNOWLEDGMENTS

D.J.P. received support from NIH grant CA49982, J.C.D. from NIH grant HD40647, R.B.D. from NIH grant NS34388, and K.M., S.K.B., and R.B.D. from NIH grant NS40955 and the Howard Hughes Medical Institute. K.M. received support from the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD program and NIH MSTP grant GM07739.

Received: February 1, 2011
Revised: April 29, 2011
Accepted: May 2, 2011
Published: July 12, 2011


