RNA Recognition via the SAM Domain of Smaug

Justin B. Green,1 Cary D. Gardner,2 Robin P. Wharton,2,* and Aneel K. Aggarwal1,*
1Structural Biology Program
Department of Physiology and Biophysics
Mount Sinai School of Medicine
Box 1677
1425 Madison Avenue
New York, New York 10029
2Howard Hughes Medical Institute
Department of Molecular Genetics & Microbiology
Box 3657
Duke University Medical Center
Durham, North Carolina 27710

Summary

The Nanos protein gradient in Drosophila, required for proper abdominal segmentation, is generated in part via translational repression of its mRNA by Smaug. We report here the crystal structure of the Smaug RNA binding domain, which shows no sequence homology to any previously characterized RNA binding motif. The structure reveals an unusual makeup in which a SAM domain, a common protein-protein interaction module, is affixed to a pseudo-HEAT repeat analogous topology (PHAT) domain. Unexpectedly, we find through a combination of structural and genetic analysis that it is primarily the SAM domain that interacts specifically with the appropriate nanos mRNA regulatory sequence. Therefore, in addition to their previously characterized roles in protein-protein interactions, some SAM domains play crucial roles in RNA binding.

Introduction

Control of mRNA translation is a common posttranscriptional mechanism for regulating growth and development (Gray and Wickens, 1998; Willis, 1999). In particular, signals in the 3' untranslated region (3' UTR) of maternal mRNAs play key roles in the embryogenesis of many organisms (Curtis et al., 1995; Macdonald and Smibert, 1996). For instance, a cascade of translation control events, where the product of one mRNA activates or represses the translation of another, is instrumental in organizing the anterior-posterior axis in Drosophila. A prominent example of translational regulation in early fly embryogenesis is the repression of maternal nanos (nos) mRNA translation by Smaug (Smg) (Dahanukar et al., 1999; Smibert et al., 1999). At fertilization, nos mRNA is distributed in the bulk cytoplasm and the posterior pole plasm, but only the latter pool of mRNA is translated due to Smg-dependent repression. As a result, a gradient of Nos protein emanates from the pole plasm. Nos protein, in combination with Pumilio (Pum) and Brain Tumor (Brat), in turn represses the translation of hunchback (hb) mRNA in the posterior to allow abdominal segmentation (Chagnovich and Lehmann, 2001; Sonoda and Wharton, 2001).

The signals necessary to regulate nos mRNA reside in a 184 nucleotide translation control element (TCE) in its 3' UTR (Dahanukar and Wharton, 1996; Gavis et al., 1996; Smibert et al., 1996). Deletion of the TCE from nos mRNA leads to improper accumulation of Nos in the bulk cytoplasm and defects in head development. For repression, the TCE contains a pair of redundant hairpins, each bearing the loop sequence CUGGC. Smg was identified in embryonic extracts as a 120 kDa factor capable of binding wild-type TCE (Smibert et al., 1996). Subsequent studies of embryos lacking normal smg function and in vitro translation experiments demonstrated that Smg is indeed required for repression of nos mRNA (Dahanukar et al., 1999; Smibert et al., 1999).

Although Smg is a large protein (999 amino acids), a domain consisting of residues 583–763 is sufficient to recognize the TCE hairpins (Dahanukar et al., 1999). Surprisingly, this RNA binding domain (RBD) shows no sequence homology to previously characterized RNA binding motifs, such as the RRM or the KH domain (Nagai, 1996). Curiously, one portion of the RBD is homologous to the SAM (sterile alpha motif) domain, a common protein-protein interaction module (Schultz et al., 1997; Thanos et al., 1999). The SAM domain has been identified in wide variety of eukaryotic proteins, including the signal transducing EphB2 and EphA4 receptors, the transcriptional repressors TEL and polyhomeotic (Ph), and splice variants (p73 and p63) of the tumor suppressor p53, among others (Thanos et al., 1999; Stapleton et al., 1999; Kim et al., 2001, 2002; Levreto et al., 2000). The SAM domain is believed to mediate homo- and hetero-oligomerization, as well as heterotypic interactions with proteins lacking SAM domains (Serra-Pages et al., 1995).

We report here the crystal structure of Smg RBD, which shows the SAM domain packed against a pseudo-HEAT repeat analogous topology (PHAT) domain. Through a combination of structural and genetic analyses, we show that it is primarily the SAM domain that interacts specifically with the TCE. Thus, in addition to their previously characterized roles in protein-protein interactions, some SAM domains play crucial roles in RNA binding.

Results and Discussion

Defining the RNA Binding Domain

The Smg RBD previously was mapped in a relatively coarse deletion analysis to residues 584–763 (Dahanukar et al., 1999). To more precisely define the minimal domain, we took two approaches. First, we prepared the N- and C-terminal truncations shown schematically in Figure 1A, expressed each in yeast, and assayed their RNA binding activity (SenGupta et al., 1996) and stability. The general scheme of the yeast three-hybrid RNA bind-
Figure 1. Minimal RNA Binding Domain

(A) The Smg RNA binding domain (RBD) delineated in this study is shown in green, while the boundaries mapped by Dahanukar et al. (1999) are indicated by dashed lines. The six deletion constructs (Δ1–6) used to define the minimal binding domain are shown schematically (not to scale) along with their ability to bind the TCE. “+” indicates binding activity indistinguishable from Smg 583–764. Some derivatives indicated by an asterisk bear a C-terminal TRLV extension of vector-encoded residues.

(B) Schematic of the yeast three-hybrid RNA binding assay and representative results of HIS3 reporter activity in the presence of 5 mM 3-AT.

(C) Western blot of extracts from yeast transformed with plasmids that direct the synthesis of the indicated Smg derivatives (or empty pAct2 vector, “−” as a control). Note that Δ3 is stable but not functional, whereas Δ6 is barely detectable, presumably due to instability.

The results of these experiments suggest that the minimal domain for high-affinity binding lies between residues 600 and 762. Truncation to residue 602 substantially reduced binding activity (Figure 1B), and truncation at residue 760 destabilized the protein sufficiently that it was barely detectable (Figure 1C) and showed no binding activity at 5 mM 3-AT.

To determine whether residues 600–762 comprise an independently folded domain, we undertook limited proteolysis on a fragment of the protein consisting of residues 583–860. Prolonged digestion with chymotrypsin, papain, or trypsin generated stable fragments of 19–23 kDa (Green et al., 2002). Mass spectrometry and N-terminal sequencing of a trypsin-derived peptide showed that the protease-resistant domain consists of residues...
Table 1. Data Collection, Phasing, and Refinement Statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Se-Remote</th>
<th>Se-Edge</th>
<th>Se-Peak</th>
<th>Nat-X-25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.94207</td>
<td>0.97941</td>
<td>0.97927</td>
<td>0.9188</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20–2.2</td>
<td>20–2.2</td>
<td>20–2.2</td>
<td>50–1.8</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>125847</td>
<td>124990</td>
<td>118393</td>
<td>151643</td>
</tr>
<tr>
<td>Unique</td>
<td>10606</td>
<td>10578</td>
<td>10581</td>
<td>19192</td>
</tr>
<tr>
<td>Data coverage (%)</td>
<td>96.4 (86.5)</td>
<td>95.9 (84.9)</td>
<td>96.3 (85.1)</td>
<td>95.5 (74.6)</td>
</tr>
<tr>
<td>R_merge (%)a,b</td>
<td>7.0 (28.7)</td>
<td>8.3 (24.8)</td>
<td>10.4 (23.7)</td>
<td>3.6 (33.0)</td>
</tr>
<tr>
<td>I/</td>
<td>H9268</td>
<td>15.6 (4.2)</td>
<td>12.2 (5.1)</td>
<td>9.4 (5.4)</td>
</tr>
<tr>
<td>MAD phasing statistics:</td>
<td>Number of Sites</td>
<td>5</td>
<td>0.71</td>
<td>0.9972</td>
</tr>
<tr>
<td>FoM 2.7 Åc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FoM (DM) 1.8 Åc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phasing power</td>
<td>–</td>
<td>1.90</td>
<td>0.93</td>
<td></td>
</tr>
</tbody>
</table>

Reefinement statistics

<table>
<thead>
<tr>
<th>Reflection range (Å)</th>
<th>40–1.8</th>
<th>18277</th>
</tr>
</thead>
<tbody>
<tr>
<td>R cryst (%)e</td>
<td>22.9</td>
<td>24.4</td>
</tr>
<tr>
<td>R free (%)f</td>
<td>22.9</td>
<td>24.4</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>1360</td>
<td>163</td>
</tr>
<tr>
<td>Protein (non-hydrogen)</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>Water</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Avg. B factor (Å2)</td>
<td>42.9</td>
<td>42.9</td>
</tr>
</tbody>
</table>

Ramachandran plot quality

| Most favored (%)     | 93     |
| Generously allowed (%)| 7     |
| Disallowed (%)       | 0      |

Values for outermost shell are given in parentheses.

<sup>a</sup>Rmerge = \[ \frac{\sum_{i} |I_i - \langle I_i \rangle|}{\sum_{i} I_i} \], where I is the integrated intensity of a given reflection.

<sup>b</sup>FoM = Mean figure of merit computed to 2.7 Å.

<sup>c</sup>FoM = Overall mean figure of merit at 1.8 Å after density modification.

<sup>d</sup>R cryst = \[ \frac{\sum_{i} |Fo - Fc|}{\sum_{i} Fo} \] for crystalline data and a solvent-flattened electron density map calculated at that resolution. The current model includes Smg residues 596–763, and 163 water molecules.

596–764, in excellent agreement with the boundaries determined by deletion analysis in yeast. Accordingly, we prepared a plasmid that directs the expression of this minimal region (henceforth, the Smg RBD) for crystallographic analysis.

Structure Determination

The Smg RBD was expressed, purified, and crystallized as described previously (Green et al., 2002). In brief, triangular crystals were obtained from solutions containing polyethylene glycol and ammonium sulfate and were used to measure native and multiwavelength anomalous dispersion (MAD) data. The MAD data were collected from a selenomethionine (SeMet) substituted derivative of the protein and used to compute experimental phases to 2.7 Å resolution (Table 1). The phases were then extended to 1.8 Å using the higher resolution native data and a solvent-flattened electron density map calculated at that resolution. The current model includes Smg residues 596–763, and 163 water molecules.

Overall Architecture

The RBD has the shape of a bent muscular arm, where the upper and lower parts of the arm are defined by two distinct domains at an approximate right angle to each other. Each domain is comprised entirely of α helices and several short 310 helices (Figure 2). Helices α1–α4 and h2 fold into a SAM domain at the N terminus (the upper arm, residues 596–657), while helices α7–α13 comprise a PHAT (for pseudo-HEAT analogous topology) domain at the C terminus (the lower arm, residues 658–764) (Figure 2). The SAM and PHAT domains are connected through helices α5 and α6, respectively. Both of these helices could be considered as portions of a single “α5–α6” helix that breaks in continuity at the “elbow” of the arm (essentially, a small deviation from helicity at residue E655). Approximately 1148 Å² of solvent accessible surface area is buried at the SAM-PHAT interface, but there is no significant intercalation of side chains across the boundary.

SAM Domain

The Smg SAM domain is globular in shape, composed of one long α helix (α5), three short α helices (α1, α3, and α4) and one 310 helix (h2). The five amino acid segment preceding helix α1 lacks secondary structure and is referred hereafter as the N-terminal arm. The Smg SAM domain shows striking structural similarity to the corresponding domain of EphB2 tyrosine kinase receptor (EphB2) (Thanos et al., 1999) (Figure 3A). Although there is little sequence similarity between the two proteins outside the hydrophobic core, the helices (α1–α4, h2)
Figure 2. Smg RBD Structure


align with a root mean square deviation (rmsd) of ~1.5 Å (54 atom pairs). The close structural similarity to EphB2 is surprising because the EphB2 SAM domain is believed to be involved in receptor oligomerization (Thanos et al., 1999), while the Smg SAM domain, we show below, is intimately involved in RNA binding. The structures do however show less overlap at sites implicated in protein-protein interactions in the EphB2 SAM domain. Thus, whereas the N-terminal arm of EphB2 SAM domain extends to a neighboring 2-fold related monomer in the EphB2 crystal structure (as well as in the EphA4 crystal structure), the equivalent arm in Smg is folded back toward the core of the domain. Moreover, residues such as Tyr913 or Met950 that mediate protein-protein interactions in the EphB2 SAM domain are different in Smg. Asn980 in EphB2 SAM domain, for instance, are directed away from the putative protein-protein interface by the kink in the central $\alpha$5-$\alpha$6 helix. Similarly, Smg lacks key apolar residues in the SAM domains of transcriptional repressors TEL (V112, L116) and Ph (A1544) that provide intermolecular contacts for self-polymerization into helices (Kim et al., 2002, 2001). (The structurally equivalent positions in the Smg SAM domain are occupied by K645, C649, and K624, respectively.)

The Smg SAM domain is instead characterized by an abundance of basic residues that define a RNA binding surface (shown below). Thus, whereas the EphB2 SAM domain contains six basic residues that are distributed more or less evenly over its surface, the Smg SAM domain contains seven lysines and two arginines, several of which are concentrated over one face of the domain (Figure 3C). This patch of electropositive potential, spread over helices $\alpha$1, h2, and $\alpha$5, stretches from the interface with the PHAT domain to the N terminus of $\alpha$5. There is no other SAM domain in the current protein database (www.rcsb.org/pdb) with an electropositive surface similar to that of Smg. Interestingly, however, the Smg SAM domain shares structural similarity to a domain within another nucleic acid binding protein, RuvA (Rafferty et al., 1996). Domain II of RuvA contains
Figure 3. Smg SAM Domain

(A) The Smg SAM domain (green) structurally aligns with the EphB2 SAM domain (red) with a rmsd of ~1.5 Å (at 54 Cα pairs). Note the divergence in the orientation of the N-terminal arms in the two structures. (B) Sequence alignment between Smg SAM domain and SAM domains of EphB2 and EphA4 receptors, transcription corepressor Polyhomeotic (Ph), and p73 (a splice variant of the tumor suppressor p53). Conserved residues are highlighted in blue, functionally similar residues are highlighted in orange. The consensus at the bottom lists residues that are over 50% identical or functionally similar among the proteins listed. (C) The electrostatic potential surfaces of the SAM domains of Smg and EphB2, and domain II of RuvA are compared. Increasing blue indicates increasing electropositivity, and increasing red indicates increasing electronegativity. This figure and Figure 5A were drawn using GRASP (Nicholls et al., 1991). (D) Smg SAM domain (green) is structurally aligned with RuvA domain II (gray). The labels correspond to helices in Smg; note that there is no equivalent of helix α4 in RuvA. The view is related to the orientation in (C) by a rotation of ~90° around the vertical axis.

Four α helices, described as a concatamer of two helix-hairpin-helix motifs, which can be aligned to helices α1, α2, α4 and α5 in the Smg SAM domain (rmsd ~1.8 Å, for 55 atom pairs). (Helix α3 in Smg SAM domain is substituted by a loop in RuvA domain II.) Importantly, RuvA domain II has an electropositive surface that significantly overlaps with the basic surface of Smg SAM domain (Figures 3C and 3D). This positive surface mediates the majority of DNA contacts made between domain II and a synthetic Holliday junction (Hargreaves et al., 1998).

PHAT Domain

The PHAT domain is a layer of three parallel helices (α6, α9, and α13) packed against a layer of two antiparallel helices (α7 and α11) into a cylindrical shaped five-helix
Smaug Recognizes RNA Primarily via Its SAM Domain

As a test of the idea from our structure that Smg primarily uses its SAM domain to contact RNA, we undertook a genetic interrogation of its surface. Following random mutagenesis of a smg gene by error-prone PCR, variants that bind RNA essentially as well as the wild-type protein were isolated using the yeast three-hybrid assay shown in Figure 1B. Collectively, the resulting variants bear 51 silent substitutions at 45 of the 168 residues of the minimal domain (highlighted in green in Figure 2B and listed in the legend). When displayed on the surface of the protein, substitutions are found essentially everywhere except the electropositive face of the SAM domain (Figure 5), consistent with the idea that this face of the protein most closely contacts the RNA. To more directly test the role of the conserved network of posi-
RNA Recognition via the SAM Domain of Smaug

Figure 5. Smg Recognizes RNA via Its SAM Domain

On the right, the electrostatic potential is mapped on the Smg RBD surface with increasing blue signifying increasing electronegativity and increasing red indicating increasing electrostriction. On the left, the 51 silent substitutions that do not significantly affect RNA binding are mapped on the surface in green; residues where no silent substitutions were recovered are white. The top view is related to the orientation in Figure 2A by a rotation of 90o about the vertical axis of the page. Note that the electropositive surface of the SAM domain is devoid of silent substitutions, consistent with the idea that it makes the majority of the RNA contacts.

Relatively charged residues in the SAM domain, we changed two of these, H611 and K640, to serine; as a control, we changed a nearby residue on the same face of the protein but in the PHAT domain, R739, to serine as well. As shown in Figure 6A, the H611S and K640S mutants are stable in yeast but fail to support growth in the presence of 5 mM 3-AT and thus have significantly reduced RNA binding (quantitated in Figure 6B). In contrast, RNA binding activity of the R739S mutant is indistinguishable from wild-type. Together, these results strongly imply that the front electropositive face of the SAM domain plays a predominant role in RNA binding.

To calibrate the RNA binding activity of the mutants described above, we purified three representative GST-Smg mutant fusion proteins and assayed their binding in gel mobility shift experiments. Two of the mutant proteins derive from the genetic screen (F632L+Q634L, N658D+Q665R+V724I+H749R, with substitutions in the PHAT domain), while the third bears the K640S substitution on the “front” surface of the SAM domain described above. When compared with the binding of wild-type GST-Smg, binding of the mutants bearing substitutions that are “silent” in the yeast screen is reduced at most 2-fold, whereas binding of the engineered mutant K640S is reduced 20- to 40-fold (Figure 6B). Therefore, we conclude that the large collection of substitutions throughout the PHAT domain and on the “back” surface of the SAM domain (Figure 5) have a negligible effect on RNA binding activity.

To further test the idea that the SAM domain is primarily responsible for RNA recognition, we asked whether the isolated SAM domain might bind the TCE hairpin specifically. To achieve this we sought to separate the SAM and PHAT domains of Smg by an appropriate truncation. Although the interface between the two domains is substantial, examination of the structure suggested that truncation after R656, which lies at the elbow, might release a form of the SAM domain with a solvent-exposed surface that is sufficiently hydrophilic to remain soluble. We tested the binding of such an isolated SAM domain, Smg residues E589–R656, fused to a transcriptional activation domain in yeast three-hybrid experiments. Assays of HIS3 and LacZ reporter activity reveal that the isolated SAM domain binds weakly to the wild-type TCE and less to a mutant TCE capped with a UGCG tetraloop (Figure 7B and Table 2). The isolated SAM domain is unstable both in yeast and bacteria (data not
Figure 6. Essential Role of the Electropositive Surface of the SAM Domain

(A) Substitutions at two of the residues on the “front” surface of the SAM domain, H611 and K640, significantly reduce RNA binding activity, assaying yeast HIS3 reporter activity in the presence of 5 mM 3-AT (left), without affecting stability (Western blot on the right). In contrast, substitution at a nearby residue in the PHAT domain (R739S) yields a protein whose activity is indistinguishable from wild-type at all concentrations of 3-AT tested (5 mM shown). Both SAM domain mutants have residual activity: yeast expressing H611S grow poorly in the presence of 1 mM 3-AT, and yeast expressing K640S grow poorly in the presence of 3 mM 3-AT and well in the presence of 1 mM 3-AT. “-” is an empty vector control.

(B) Gel mobility shift assays of purified GST-Smg fusion proteins. Based on these and other experiments, binding of K640S mutant (substitution on the “front” surface of the SAM domain) is reduced at least 20- to 40-fold, binding of the F632L/Q634L double mutant (substitutions on the “back” surface of the SAM domain) is indistinguishable from wild-type, and binding of the quadruple N658D/Q665R/V724I/H749R mutant (substitutions in the PHAT domain; derived by recombination between two double mutants isolated in the genetic screen described in Figure 5) is reduced approximately 2-fold. Reactions loaded in lanes 1–5 of each panel contained 0, 0.033, 0.11, 0.33, and 1.1 μg of protein, respectively.

(C) Residues mutated in (A) and (B) are shown on the “front” (top) and “back” (bottom) surfaces of Smg RBD (views are identical to those in Figure 5).

shown); thus, the binding experiments likely underestimate its affinity for the RNA.

While the SAM domain can discriminate between closely related RNAs, its binding is sufficiently weak that we cannot determine whether the isolated SAM domain retains the full specificity of the SAM+PHAT module. However, the data are consistent with the idea that the SAM domain is primarily responsible for reading the RNA sequence, with the PHAT domain contributing essential but non-specific interactions. This model is further supported by the analysis of Smg homologs described below.

Evolutionarily Conserved Function of the Smg SAM Domain

A search of the current database reveals a family of proteins similar to the Smg RBD in organisms from yeast to man (Figure 7A), among which homology is primarily restricted to the SAM domain. No significant homology with the PHAT domain is evident beyond the first 13–34 residues. This primary structural conservation suggests conserved function of the SAM domain in these proteins. To test this idea, we chose the mouse and frog homologs, prepared plasmids that direct the synthesis of activation domain fusions to either the SAM domain alone
RNA Recognition via the SAM Domain of Smaug

Figure 7. Homologous SAM Domains and Their Binding to RNA

(A) Sequence-based alignment of Smg-like SAM domains. Conserved residues are highlighted in blue, and functionally similar residues are highlighted in orange. In the consensus, strictly conserved residues are in bold, and residues with over 50% identity are in normal type. The residues highlighted in cyan correspond to basic residues (Arg, Lys, and His) that map to the putative RNA binding surface. The stars highlight H611 and K640, conserved residues where substitution to serine substantially reduces RNA binding (Figure 6). Accession numbers are as follows: fly (AF132213), bee (Bis53143), frog (AW643499), human (AW405595), zebrafish (BE605702), mouse (BF140280), C. elegans (AF078788), P. carinii (AW333644), S. cerevisiae (Z75267), C. albicans (AL033497), C. immitis (BF252128).

(B) Yeast binding assays of various proteins to chimeric RNAs with tandem copies of the wild-type (wt) TCE hairpin (sequence shown in [C]) or the tetra mutant (described in the text), in the presence of 0.5 mM 3-AT. On the left, yeast are transformed with pACT2 derivatives that express an activation domain fusion to the indicated SAM domain only (fly, E583–R656; frog, WLKS...LKER; mouse, WLKS...LRER). On the right, yeast express longer fragments containing the SAM domain (fly, E583–M763; frog, 459 amino acids [GTRP...TSTI]; mouse, 98 amino acids [WLKS...LLQG]). The complete frog cDNA sequence is available on request.

(C) Yeast RNA binding assays to singly mutant TCE hairpins. The plates shown contain 5 mM 3-AT. The complete frog cDNA sequence is available on request.

(C) Yeast RNA binding assays to singly mutant TCE hairpins. The plates shown contain 5 mM 3-AT. Note that binding of both proteins to C6U is indistinguishable from binding to wild-type, suggesting that Smg recognizes hairpins with CUGGY loops. Also, binding to U3C is undetectable at 10 mM 3-AT, whereas binding to wild-type TCE is detectable at concentrations up to 20 mM 3-AT. Each chimeric RNA accumulates to the same level in yeast as determined by Northern blots (data not shown). At the bottom is a schematic comparison of sequence identity between the fly (amino acids 583–763) and frog (459 amino acids) proteins used in the experiments, showing very limited homology outside the SAM domain.
Experimental Procedures

Smg-related SAM domains is likely to bind hairpins of nos with a panel of six singly mutant RNAs, even though structural context for additional experiments and a basis As shown in Figure 7C, the fly and frog proteins have plasm components or the corepressor posited by Crucs domain plus flanking residues) is robust, allowing us to is consistent with a specialized role in 7B and Table 2). Binding of the larger frog protein (SAM PHAT domain in Smg homologs from other organisms binding to the Smg RNA target in yeast. Much like the suggests that the role of the PHAT domain in RNA bind- or the SAM domain plus flanking residues, and assayed its surface has no appreciable effect on RNA bindingination regulation, including possible interactions with pole modification. This yielded an experimental electron density map or the SAM domain plus flanking residues, and assayed binding to the Smg RNA target in yeast. Much like the isolated fly SAM domain, the frog and mouse SAM do- or the SAM domain plus flanking residues, and assayed binding to the Smg RNA target in yeast. Much like the isolated fly SAM domain, the frog and mouse SAM do-mains bind weakly but preferentially to a wild-type Smg hairpin rather than the tetraloop mutant hairpin (Figure 7B and Table 2). Binding of the larger frog protein (SAM domain plus flanking residues) is robust, allowing us to test its specificity with a panel of singly mutant hairpins. As shown in Figure 7C, the fly and frog proteins have essentially identical binding specificities when surveyed with a panel of six singly mutant RNAs, even though homology between the two proteins is restricted mainly to the SAM domain. Thus, the conserved subfamily of Smg-related SAM domains is likely to bind hairpins of similar or identical structure.

Conclusions

We show here that the Smg SAM domain encodes an RNA binding motif. Several lines of evidence lead us to this conclusion. First, the charge distribution on the Smg RBD surface strongly suggests that the RNA primarily contacts the SAM domain. Second, randomly generated Smg RBD variants selected for retention of RNA binding reveal surface substitutions everywhere except the electropositive face of the SAM domain, consistent with this face of the SAM domain being the primary site of RNA recognition. Third, sequence similarity among Smg RBD homologs from yeast to man is restricted mainly to the SAM domain (and not the PHAT domain). Finally, we show that an isolated SAM domain of Smg is capable of recognizing the TCE hairpin specifically, and that this RNA binding function is preserved in homologous SAM domains from mouse and frog. Taken together, this structural, genetic, and evolutionary evidence identifies a subfamily of SAM domains with an RNA binding function.

The study of RNA binding proteins is progressing rap-idly (Draper, 1999; Williamson, 2000), but the number of known RNA binding motifs is still relatively small when compared to DNA binding domains. The SAM domain has been largely viewed as a protein-protein interaction module, and thus it may be the most surprising addition to the growing family of RNA binding domains. The Smg SAM domain appears to lack specific sidechains that allow other SAM domains in signal transducing and tran-scription repressing proteins to oligomerize; instead, similarity to the basic surface of RuvA may reflect commonalities in TCE hairpin and DNA Holliday junction binding. Intriguingly, this electropositive surface in Smg is capped by a tyrosine (Tyr613), the equivalent of which or the SAM domain plus flanking residues, and assayed binding to the Smg RNA target in yeast. Much like the isolated fly SAM domain, the frog and mouse SAM do-mains bind weakly but preferentially to a wild-type Smg hairpin rather than the tetraloop mutant hairpin (Figure 7B and Table 2). Binding of the larger frog protein (SAM domain plus flanking residues) is robust, allowing us to test its specificity with a panel of singly mutant hairpins. As shown in Figure 7C, the fly and frog proteins have essentially identical binding specificities when surveyed with a panel of six singly mutant RNAs, even though homology between the two proteins is restricted mainly to the SAM domain. Thus, the conserved subfamily of Smg-related SAM domains is likely to bind hairpins of similar or identical structure.

Conclusions

We show here that the Smg SAM domain encodes an RNA binding motif. Several lines of evidence lead us to this conclusion. First, the charge distribution on the Smg RBD surface strongly suggests that the RNA primarily contacts the SAM domain. Second, randomly generated Smg RBD variants selected for retention of RNA binding reveal surface substitutions everywhere except the electropositive face of the SAM domain, consistent with this face of the SAM domain being the primary site of RNA recognition. Third, sequence similarity among Smg RBD homologs from yeast to man is restricted mainly to the SAM domain (and not the PHAT domain). Finally, we show that an isolated SAM domain of Smg is capable of recognizing the TCE hairpin specifically, and that this RNA binding function is preserved in homologous SAM domains from mouse and frog. Taken together, this structural, genetic, and evolutionary evidence identifies a subfamily of SAM domains with an RNA binding function.

The study of RNA binding proteins is progressing rap-idly (Draper, 1999; Williamson, 2000), but the number of known RNA binding motifs is still relatively small when compared to DNA binding domains. The SAM domain has been largely viewed as a protein-protein interaction module, and thus it may be the most surprising addition to the growing family of RNA binding domains. The Smg SAM domain appears to lack specific sidechains that allow other SAM domains in signal transducing and transcriptions repressing proteins to oligomerize; instead, similarity to the basic surface of RuvA may reflect commonalities in TCE hairpin and DNA Holliday junction binding. Intriguingly, this electropositive surface in Smg is capped by a tyrosine (Tyr613), the equivalent of which

<table>
<thead>
<tr>
<th>Protein</th>
<th>wt RNA</th>
<th>Tetra Mutant RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fly SAM +</td>
<td>2250</td>
<td>58</td>
</tr>
<tr>
<td>Frog SAM +</td>
<td>320</td>
<td>14</td>
</tr>
<tr>
<td>Mouse SAM +</td>
<td>2.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Fly SAM</td>
<td>2.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Frog SAM</td>
<td>1.9</td>
<td>0.5</td>
</tr>
<tr>
<td>Mouse SAM</td>
<td>2.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Proteins as in Figure 7B.

Table 2. RNA-Binding Activity Assayed by LacZ Reporter Activation in Yeast (Miller Units)

Experimental Procedures

Protein Preparation and Crystallization

The Smg RBD (residues 596–764) was expressed in E. coli and purified as described in Green et al. (2002). Crystals were obtained by the hanging drop vapor diffusion (4 °C) against 17% polyethylene glycol 4000, 220 mM Na acetate, and 100 mM Tris·Cl (pH 8.0). (Best crystals are obtained by adding 100 mM ammonium sulfate to the crystallization drop, prior to equilibration.) Flat three-faceted triangular crystals grow to full size (0.3 × 0.3 × 0.05 mm) after 10–14 days and belong to space group R3 (a = b = 129.3 Å, c = 33.1 Å, α = β = γ) with one molecule in the asymmetric unit. For MAD phasing, a selenomethionine (SeMet) variant of the Smgtropositive face of the SAM domain, consistent with this conclusion. First, the charge distribution on the Smg RBD surface strongly suggests that the RNA primarily contacts the SAM domain. Second, randomly generated Smg RBD variants selected for retention of RNA binding reveal surface substitutions everywhere except the electropositive face of the SAM domain, consistent with this face of the SAM domain being the primary site of RNA recognition. Third, sequence similarity among Smg RBD homologs from yeast to man is restricted mainly to the SAM domain (and not the PHAT domain). Finally, we show that an isolated SAM domain of Smg is capable of recognizing the TCE hairpin specifically, and that this RNA binding function is preserved in homologous SAM domains from mouse and frog. Taken together, this structural, genetic, and evolutionary evidence identifies a subfamily of SAM domains with an RNA binding function.

The study of RNA binding proteins is progressing rapidly (Draper, 1999; Williamson, 2000), but the number of known RNA binding motifs is still relatively small when compared to DNA binding domains. The SAM domain has been largely viewed as a protein-protein interaction module, and thus it may be the most surprising addition to the growing family of RNA binding domains. The Smg SAM domain appears to lack specific sidechains that allow other SAM domains in signal transducing and transcription repressing proteins to oligomerize; instead, similarity to the basic surface of RuvA may reflect commonalities in TCE hairpin and DNA Holliday junction binding. Intriguingly, this electropositive surface in Smg is capped by a tyrosine (Tyr613), the equivalent of which in the EphB1 SAM domain (Tyr929) has been shown to be both phosphorylated in vivo and to serve as a binding site for a tyrosine phosphatase as well as the SH2 do-main of the adaptor protein Grb10 (Stein et al., 1998a, 1998b). Whether Smg is similarly phosphorylated in vivo remains to be seen, but the proximity of Tyr613 to the electropositive surface raises the possibility that phos-phorylation may serve to modulate TCE binding, particu-larly in the pole plasm to allow translation of localized nos mRNA. The Smg PHAT domain is essential for high-affinity RNA binding and contains two and a half pseudo-HEAT repeats folded into an unusual five-helix bundle. However, the observation that substitution of much of its surface has no appreciable effect on RNA binding suggests that the role of the PHAT domain in RNA bind- or the SAM domain plus flanking residues, and assayed its surface has no appreciable effect on RNA binding suggests that the role of the PHAT domain in RNA bind- or the SAM domain plus flanking residues, and assayed its surface has no appreciable effect on RNA binding suggestions that the role of the PHAT domain in RNA bind- or the SAM domain plus flanking residues, and assayed its surface has no appreciable effect on RNA binding suggestions that the role of the PHAT domain in RNA bind- or the SAM domain plus flanking residues, and assayed its surface has no appreciable effect on RNA binding suggestions that the role of the PHAT domain in RNA bind-
had an R factor of 44.6% (R\textsubscript{free}=45.0%), which quickly converged to 22.9% (R\textsubscript{free}=24.4%) after iterative rounds of refinement with CNS, model building with O (Jones et al., 1991), and the addition of water molecules. The final model includes Smg residues 596–763 and 163 water molecules. The model has excellent stereochemistry (Table 1), with 93% of the residues in the most favored regions in a Ramachandran plot, 7% in generously allowed regions, and none in the disallowed regions.

**Mutagenesis and TCE Binding Assays**

A gene encoding Smg residues 583–763 was mutagenized by error prone PCR essentially as described by Vidal et al. (1996). Yeast strain l-40 CP (SenGupta et al., 1996) was transformed with (1) pAD2297, which encodes a chimeric RNA containing a single copy of the Smg binding site, (2) a gapped derivative of pAct2 bearing a Kan\textsuperscript{r} gene, and (3) the PCR product. TCE binding was assayed by streaking transformants on media lacking His with 5 mM 3-aminotriazole (3-AT), and the mutated smg genes were recovered into E. coli by selecting Kan\textsuperscript{r}. Binding was generally assayed at concentrations of 1–5 mM 3-AT, except for the experiments in Figure 7B. In these, weak binding (at 0.5–1.0 mM 3-AT) is detected using pW2900 and pW2901, which encode chimeric RNAs with two tandem copies of the wild-type and tetra mutant Smg binding sites, respectively. The level of Smg derivatives in yeast was assayed using the MAB 22C10, which recognizes the vector-encoded E. coli epitope tag. The following DNA sequence was inserted at the BamH site in pBluescript II KS–1 to generate pAD2297 (SenGupta et al., 1996): GATCCGCAAGCCTGCGAATTCGATGC. pW2900 bears a tandem insertion of the sequence, and pW2901 bears a tandem insertion with the bases in italics substituted with TGCG. For the experiments of Figure 6B, GST fusion proteins bearing Smg residues 583–778 were purified by standard methods and incubated at various concentrations in the presence of 10 mM HEPES (pH 7.9), 1 mM DTT, 100 mM KCl, 5% glycerol, and radiolabeled RNA bearing a single copy of the hairpin shown in Figure 7C.

**Acknowledgments**

We thank Frank Rotella at beamline 19ID (APS) and Michael Becker at X25 (NSLS) for facilitating X-ray data collection; and Alan Capilli and Alex Kentsis for helpful comments. J.B.G. is supported by NIH grant GM47730. T.D.B. and J.B.G. are supported by a Ramachandran plot, 7% in generously allowed regions, and none in the disallowed regions.

**References**


Accession Numbers

The coordinates for the SAM domain of Smaug have been deposited in the Protein Data Bank under accession code 10XJ.