nature structural & molecular biology

RNA recognition by the Vts1p SAM domain

Philip E Johnson[1] & Logan W Donaldson[2]

The putative yeast post-transcriptional regulator Vts1p and its related protein Smaug, from *Drosophila melanogaster*, each use a sterile alpha motif (SAM) domain to bind an RNA hairpin termed the Smaug recognition element (SRE). Here, we present the NMR structures of the Vts1p–SRE complex and the free SRE. Structural highlights include the direct recognition of a guanine base and the formation or stabilization of a base pair in the SRE loop.

In yeast, the protein Vts1p has been implicated in vesicular transport[1] and sporulation[1]; however, its precise role remains unknown. Vts1p is a homolog of the *Drosophila* protein Smaug, a translational repressor that mediates body patterning during embryogenesis[2]. The RNA sequence bound by Smaug in *nanos* transcripts, its native substrate, is a hairpin termed the Smaug recognition element (SRE). Deletion analyses of Smaug and Vts1p have demonstrated that SRE binding is exclusively provided by their SAM domains[3,4], a motif typically associated with protein-protein recognition[5]. A previous mutagenesis study has identified the 5′-CUGGC-3′ pentaloop of the SRE as the primary binding site for both the Vts1p and Smaug SAM domains[3]. The SAM domains from both proteins bind with high affinity ($K_d \approx$ 30 nM), and they share the same RNA-recognition sequence. No sequence preference was observed for the base-paired SRE stem. From these results, it has been hypothesized that recognition is limited to a few bases presented in a specific conformation in the pentaloop. In sequence conservation and structure-based mutagenesis studies, the Vts1p and Smaug SRE-binding sites have been defined as a correspondingly small and shallow pocket[3,4].

We have determined the NMR-derived structure of Vts1p SAM domain in complex with a 19-nucleotide SRE RNA containing the *nanos* 5′-CUGGC-3′ pentaloop (**Fig. 1**, **Supplementary Fig. 1** and **Supplementary Table 1** online). The characteristic five helices of the SAM domain in the complex are apparent (H1, residues 456–463; H2, 466–472; H3, 477–480; H4, 485–491; H5, 496–515). Three short helices (residues 444–447, 450–454 and 520–522) build on the SAM domain and complete the hydrophobic core. In Smaug, in lieu of these short helices, a pseudo HEAT repeat analogous topology (PHAT) domain is directly fused to the SAM domain[4] (**Supplementary Fig. 2** online).

From $^{13}C,^{15}N$-edited NOESY spectra and ^{12}C-filtered, ^{13}C-edited NOESY spectra, 20 intermolecular NOEs were identified between the Vts1p SAM domain and the SRE RNA (**Supplementary Table 2** and **Supplementary Methods** online). Several NOEs originating from H1 on base G10 serve to anchor it among a number of hydrophobic (Leu465, Tyr468, Ala495, Leu496, Ala498) and basic (Lys467) amino acid residues that cluster at the junction of helices H1 and H5 (**Supplementary Fig. 3** online). The remaining intermolecular NOEs were assigned to the ribose sugars of C8, U9 and G10 (**Fig. 1b**). NOEs from Leu496 and Gly497 complete the binding site. As the complex was determined at pH 7.8, we speculate that a high degree of solvent exchange precluded our ability to observe NOEs from guanidino and amino groups of arginine and lysine. Of the five arginines in Vts1p, the Hε resonances of Arg464 and Arg500 show the largest shift changes, suggesting a possible interaction with the SRE[6].

Before RNA structure determination, the Vts1p SAM domain was titrated into samples of unlabeled and $^{15}N,^{13}C$-isotopically labeled SRE RNA. Upon addition of Vts1p, the imino resonance of U7 showed minor chemical shift changes, emphasizing the limited role of the SRE stem in Vts1p binding (**Fig. 2a**,**b**). The observation that most surprised us was the emergence of two new imino resonances from G10 and G11 that saturated in intensity upon formation of a 1:1 Vts1p–RNA complex.

As no intermolecular NOEs to G11 were detected, the most plausible explanation for its diminished exchange with water is the

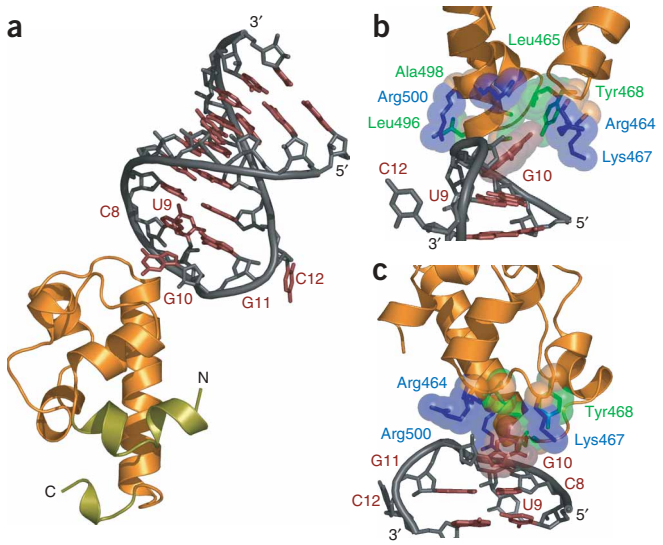

Figure 1 The Vts1p–SRE complex. (**a**) Three secondary structural elements (olive) augment the hydrophobic core of the Vts1p SAM domain (orange). The high-affinity RNA interaction draws upon base-specific contacts to G10 and sugar contacts to C8, U9 and G10. (**b**,**c**) The SRE-binding site of Vts1p.

[1]Department of Chemistry and [2]Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3. Correspondence should be addressed to P.E.J. (pjohnson@yorku.ca) or L.W.D. (logand@yorku.ca).

Received 5 October 2005; accepted 15 November 2005; published online 22 January 2006; doi:10.1038/nsmb1039

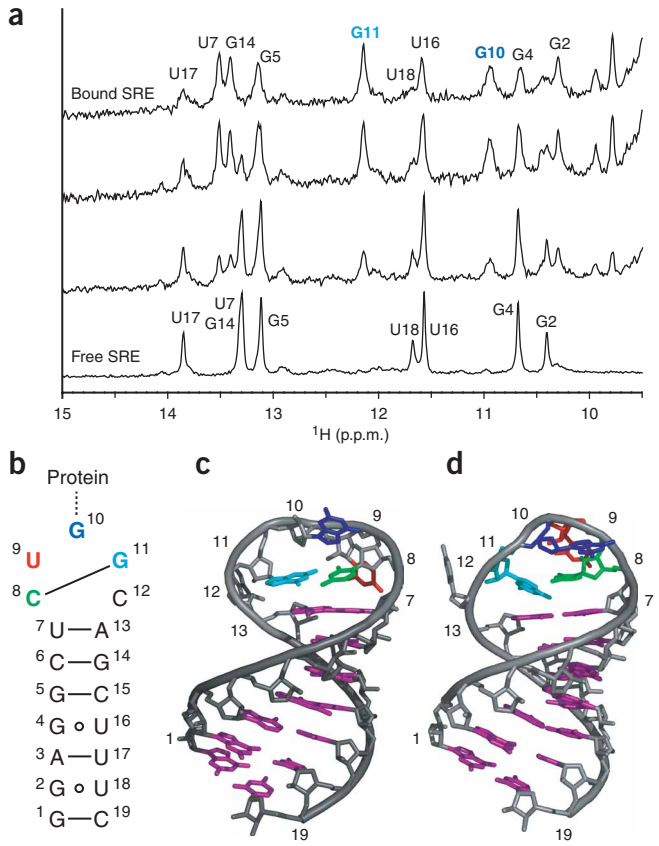

Figure 2 Structural changes in the SRE RNA upon protein binding. (**a**) Successive titration of Vts1p SAM domain reveals two new imino resonances in the bound molecule's spectrum. (**b**) Secondary structure of the SRE. (**c**,**d**) The SRE RNA in the Vts1p–SRE complex (**c**) and the free SRE (**d**).

formation of a new base pair with C8. In the free SRE RNA structure, a C8-G11 base pair may be present but not sufficiently protected from water exchange for the G11 imino resonance to be visible in NMR spectra. Notably, substitution of any nucleotide for C8 or G11 that disrupts a Watson-Crick base pair also disrupts protein binding[3]. Together, the mutational and NMR data demonstrate a functional requirement for the formation or stabilization of a base pair at positions 1 and 4 in the pentaloop.

In the Vts1p–SRE complex, G10 is oriented somewhat parallel to the C8-G11 base pair with its imino proton sequestered in the Vts1p binding site. At 3.4 Å from the imino proton of G10, the hydroxyl group of Tyr468 may serve as a direct hydrogen bonding partner or may hydrogen bond indirectly through a bridging water molecule. In support of a functional role for the Tyr468 hydroxyl group, an analogous Y613F mutation in Smaug decreased binding by ten-fold[3].

In both the free and Vts1p-bound SRE, the RNA stem adopts a standard A-form helix for 7 base pairs leading into the pentaloop (**Fig. 2c,d**, **Supplementary Fig. 4** and **Supplementary Table 3** online). Despite many loop resonances shifting upon protein binding (**Supplementary Table 4** online), both SRE forms share a similar structure supported by a similar pattern of NOEs. The general topology of the SRE is described as an ordered tetraloop with the bases of C8, G10 and G11 situated in the major groove. Together, these three nucleotides form the site for protein binding. In the free and bound forms of the SRE, C12 is excluded from the pentaloop and disordered, as demonstrated by characteristically sharp NMR resonances. The Vts1p-SRE interaction is not affected by substitution or elimination of C12 (ref. 3). U9 is excluded from base stacking and packs against the loop on the face opposite from the protein-binding side. Consequently, substitution of U9 does affect Vts1p binding[3].

The NMR structure of the Vts1p–SRE complex defines a previously unobserved mode of RNA recognition in which a pentaloop sequence presents one nucleotide base in a specific structural context. The unrelated winged-helix protein SelB from *Moorella thermoacetica* also binds an exposed guanine base in the pentaloop of its SECIS RNA partner[7]. Topologically, the SRE and SECIS RNA exclude the second and fifth nucleotides to preserve the A-form helix throughout the hairpin and stack the third nucleotide in a nearly parallel orientation with a base pair between the first and fourth positions. The greatest difference is the positioning of the excluded base at second position. In the SRE, U9 is packed against the loop and is not involved in binding, whereas in the SECIS RNA, G23 is exposed for direct recognition by a cleft in SelB that is much deeper than the binding site of Vts1p.

As only a small portion of the SAM domain is occupied by the SRE RNA, it is possible that the SAM domain may participate in additional intra- or intermolecular protein partnerships. Through these partnerships, target selection may be refined to larger and therefore less diverse sequences in the yeast mRNA pool.

Accession codes. Protein Data Bank: Coordinates have been deposited with accession codes 2B6G (Vts1p–SRE complex) and 2B7G (free SRE).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We are grateful to A. Amborski, J. Kwan, M. Motamed and X. Zhao for their assistance. NMR datasets at 800 MHz were acquired at the National High Field NMR Centre and the Québec/Eastern Canada High Field NMR Facility. This work was supported by funding from the Canadian Institutes of Health Research to P.E.J. and L.W.D.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Dilcher, M., Kohler, B. & von Mollard, G.F. *J. Biol. Chem.* **276**, 34537–34544 (2001).
2. Smibert, C.A., Lie, Y.S., Shillinglaw, W., Henzel, W.J. & Macdonald, P.M. *RNA* **5**, 1535–1547 (1999).
3. Aviv, T. *et al. Nat. Struct. Biol.* **10**, 614–621 (2003).
4. Green, J.B., Gardner, C.D., Wharton, R.P. & Aggarwal, A.K. *Mol. Cell* **11**, 1537–1548 (2003).
5. Qiao, F. & Bowie, J.U. *Sci. STKE [online]* **286**, re7 (2005) (doi:10.1126/stke.2862005re7).
6. Aviv, T. *et al. J. Mol. Biol.* published online 7 December 2005 (doi:10.1016/j.jmb.2005.11.066).
7. Yoshizawa, S. *et al. Nat. Struct. Mol. Biol.* **12**, 198–203 (2005).