The KH domains of *Xenopus* Vg1RBP mediate RNA binding and self-association

ANNA GIT and NANCY STANDART

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, United Kingdom

ABSTRACT

Xenopus Vg1 mRNA is localized to the vegetal cortex during oogenesis in a process involving microtubules and microfilaments and proteins that specifically recognize the vegetal localization element (VLE) within the 3' untranslated region. One of the best characterized VLE-binding proteins is Vg1RBP or Vera. Primary sequence analysis of Vg1RBP and its homologs suggests that most of its open reading frame is occupied by RNA-binding modules, including two RRM and four KH domains, arranged as three pairs of didomains. In the first detailed domain analysis of Vg1RBP, we show that the interaction of Vg1RBP with the VLE requires both KH didomains, but not the RRM didomain, and moreover that the KH didomains contribute cooperatively to RNA binding. In the full-length protein, individual KH domains display significant redundancy, and their relative importance appears to vary with the RNA target. We also demonstrate that the KH34 didomain mediates Vg1RBP self-association, which is stabilized by RNA, and occurs in vivo as well as in vitro. Altogether, our findings highlight the importance of multiple KH domains in mediating RNA–protein and protein–protein interactions in the formation of a stable complex of Vg1RBP and Vg1 mRNA.

Keywords: 3' untranslated region; maternal mRNA; RNA localization; RRM; vegetal cortex

INTRODUCTION

Vg1RBP (also known as Vera) is a protein implicated in localizing Vg1 mRNA to the vegetal cortex during *Xenopus* oogenesis (Deshler et al., 1998; Havin et al., 1998; Zhang et al., 1999; Kwon et al., 2002). One of the best characterized localized maternal RNAs in *Xenopus*, Vg1 mRNA encodes a TGF- β -like factor involved in mesoderm induction and left–right axis formation. Vg1 mRNA translocation to and anchoring at the vegetal cortex involves microtubules and microfilaments, respectively, and is mediated by the 340-nt 3' UTR vegetal localization element (VLE) defined in microinjection assays (Yisraeli et al., 1990; Mowry & Melton, 1992; reviewed in King et al., 1999; Palacios & St. Johnston, 2001; Yaniv & Yisraeli, 2001; Kloc et al., 2002). Vg1 mRNA is not translated until the mRNA is localized, reflecting the tight control of RNA translation and localization shown to be critical for pattern formation in *Drosophila* (Palacios & St. Johnston, 2001; Kloc et al., 2002). The 250-nt UA-rich element that represses Vg1 mRNA translation is also located in the 3' UTR, distinct from and down-

stream of the VLE (Wilhelm et al., 2000; Otero et al., 2001).

Several proteins interact specifically with the VLE (Mowry, 1996), including Vg1RBP/Vera (Deshler et al., 1998; Havin et al., 1998; Zhang et al., 1999), hnRNP I/PTB (Cote et al., 1999), and PrrP (Zhao et al., 2001), all of which colocalize with Vg1 mRNA and are thought to play a role in the localization process. Vg1RBP/Vera and hnRNP I bind the pyrimidine-rich VM1 and E2 motifs which are repeated three to five times throughout the VLE; mutation or deletion of these elements abolishes protein-binding and RNA localization (Deshler et al., 1998; Havin et al., 1998; Cote et al., 1999; Bubunencko et al., 2002; Kwon et al., 2002). Vg1RBP mediates the association of Vg1 RNA with microtubules in *Xenopus* oocytes (Elisha et al., 1995), whereas PrrP has been reported to bind to ligands involved in actin polymerization (Zhao et al., 2001). Although these data strongly implicate these proteins in mediating Vg1 mRNA localization, the actual binding sites of Vg1RBP/hnRNP I/PrrP in the VLE remain incompletely resolved, nor is it known how these factors transport the mRNA along the cytoskeleton.

Interestingly, Vg1RBP is a homolog of the chicken zipcode-binding protein (ZBP-1) implicated in localizing β -actin mRNA to the leading edge of fibroblasts

Reprint requests to: Nancy Standart, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, United Kingdom; e-mail: nms@mole.bio.cam.ac.uk.

(Ross et al., 1997) and to growth cones in neurites (Zhang et al., 2001). The conserved role of this protein in targeting different mRNAs in oocytes, fibroblasts and neurons strongly implies that Vg1RBP is a common component of diverse localization pathways. Other members of the Vg1RBP family include the mouse c-myc coding region determinant-binding protein, CRD-BP, implicated in regulating RNA stability (Doyle et al., 1998), and three human insulin-like growth factor II mRNA-binding proteins (IMP1-3) that repress translation in late development (Nielsen et al., 1999). Vg1RBP homologs are overexpressed in a large number of cancers, including KOC (KH domain containing protein overexpressed in cancer) in pancreatic carcinoma lines (Müller-Pillasch et al., 1997), though the functional significance of this overexpression in carcinogenesis is not yet known (Yaniv & Yisraeli, 2002). Yisraeli and colleagues have recently proposed that this family of RNA-binding proteins be termed the VICKZ family, in acknowledgment of the founding members (Vg1RBP/Vera, IMP, CRD-BP, KOC, and ZBP-1; Yaniv & Yisraeli, 2002).

VICKZ family members are composed almost entirely of potential RNA-binding domains. The highly conserved vertebrate homologs (77–98% identity, see Fig. 1) contain the unique combination of two RRM (RNA recognition motif) and four KH (hnRNP K-homology) domains, whereas the putative *Drosophila* homolog lacks the two N-terminal RRM. The RRM is the most widespread and well-characterized RNA-binding domain, consisting of two short consensus sequences, RNP1 (octamer) and RNP2 (hexamer) that are part of a structurally conserved domain about 80 amino acids long, with a $\beta\alpha\beta\beta\alpha\beta$ topology. RNA recognition occurs on the four-stranded antiparallel β -sheet surface in which the RNP1 and RNP2 motifs are juxtaposed in the two middle strands (Varani & Nagai, 1998). The KH (hnRNP K homology) motif with $\beta\alpha\alpha\beta\beta\alpha$ topology, approximately 70 amino acids long, has a characteristic pattern of aliphatic residues with the most conserved IGxxGxxI core located roughly in the middle of the domain (Adinolfi et al., 1999; Grishin, 2001). The KH consensus sequence is in direct contact with nucleic acids, as the recent crystal structure of Nova-2 KH domain bound to a 20mer RNA hairpin con-

firmed (Lewis et al., 2000). Both RRM and KH domains may be present in single or (more usually) multiple copies and are present in a wide variety of proteins—from bacteria to humans—that function in close association with RNA. In some proteins, RRM and KH domains, in addition to their role in RNA binding, also support protein–protein interactions, mediating homodimerization or association with other proteins. Examples include RRM2 of PABP-binding eIF4G (Kessler & Sachs, 1998; Otero et al., 1999), RRM of PTB and Sxl-mediated homodimerization (Perez et al., 1997; Samuels et al., 1998), and the KH domains of hnRNP E2 and hnRNP K-binding hnRNPs E2, K, I, and L (Kim et al., 2000). In addition to the RNA-binding domains, the presence of potential NES and NLS motifs in Vg1RBP homologs has also been noted, but their functional significance is not known (Ross et al., 1997; Havin et al., 1998).

Here we present a detailed characterization of the domains of Vg1RBP responsible for its high affinity binding to VLE RNA. The RRM is inert in a variety of RNA-binding assays, whereas the KH domains, as two didomains, contribute cooperatively to RNA binding. We also show that recombinant and native forms of Vg1RBP self-associate, principally through the KH34 didomain, and that this self-association is stabilized by RNA binding.

RESULTS

The RNA-binding domains of Vg1RBP are arranged as didomains

We aligned Vg1RBP with all its homologs, superimposing the consensus domain definitions for the two RRM and four KH domains, which were assigned by a range of secondary structure prediction programs (Fig. 1A). We noted that all the variable regions in this protein family lie outside the predicted domain boundaries, consistent with the notion that it is the functional domains that are conserved in evolution. Notably, the *Drosophila* homolog lacks the RRM regions, as reported previously (Nielsen et al., 2000), and ends with an unmatched long tail, probably unstructured. To test whether the *Drosophila* protein is a true homolog, or simply a pro-

FIGURE 1. Vg1RBP and its homologs contain three RNA binding didomains. **A:** The following sequences were aligned using the “pileup” command of the Wisconsin-GCG package: Allelic variants A (AF064633) and D (AF064634) of the *Xenopus* (x) Vg1RBP; human (h) HCCA (AF057352), IMP 1-3 (AF117106, AF117107, AF117108; IMP3 and KOC (U76705) are identical), CRD-BP (AF198254) and mouse (m) CRD-BP (AF061569); chicken (c) ZBP-1 (AF026527); zebrafish (z) (AF161270) and *Drosophila* (d) Vg1RBP (AF161270). Residues identical in six or more proteins are highlighted in gray. Secondary structure prediction was performed using the Jpred website (<http://jura.ebi.ac.uk:8888>) and a consensus prediction of the various algorithms was calculated manually. *Xenopus* Vg1RBP domain definitions are underlined: wave: RRM, thick: KH, thin: additional residues included in constructs. The RNP1 and RNP2 elements in the RRM are noted. Inter-domain spacers are in boldface. The conserved core of the KH domain iGxxGxxi (i: aliphatic residue; k: preference for basic residue; x: any residue) is indicated with * denoting position of I304N-like mutations introduced. **B:** A “pileup” multiple alignments of individual KH domains from *Xenopus* Vg1RBP (allelic variant A) and *Drosophila* Vg1RBP was adapted by ClustalW for graphical representation as an unrooted tree by “TreeView,” where the similarity between two proteins is inversely related to the length of the shortest line connecting them.

A	hHCCA	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET	RRM1	RRM2	120
	hIMP2	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	hCRD-BP	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	hIMP1	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	mCRD-BP	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	cZBP1	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	hIMP3	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	xVRBP-A	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	xVRBP-D	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	zVRBP	MMNKLYIGNLSPAVTADDRLRGLFGDKRLPLAGQVLLKSGYAFVDYDQNWAIKAIETLISGKVELBGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENCEQVNTDDET			
	hHCCA	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ	RRM1	RRM2	240
	hIMP2	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	hCRD-BP	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	hIMP1	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	mCRD-BP	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	cZBP1	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	hIMP3	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	xVRBP-A	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	xVRBP-D	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	zVRBP	AVVNVVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPQRAQRGDHSSREQGHAPG...GTSQARQ.....IDFPLRLVPTQPVGAIIGKEGLTIKNTKQTQ			
	hHCCA	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK	KH1	KH2	360
	hIMP2	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	hCRD-BP	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	hIMP1	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	mCRD-BP	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	cZBP1	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	hIMP3	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	xVRBP-A	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	xVRBP-D	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	zVRBP	SKVDIHRKENSAGAAEFVTHATPEGTSEACRMILEIMQKEADETKLAETIPLKILAHNGLVGRILIGKEGRNLKKIEHETGKRTITSSLDQLSYNPERTITTVKGTEACASABIEIMKK			
	hHCCA	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA	KH3	KH4	480
	hIMP2	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	hCRD-BP	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	hIMP1	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	mCRD-BP	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	cZBP1	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	hIMP3	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	xVRBP-A	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	xVRBP-D	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	zVRBP	IREAFENDMLAVN.....THSGYFSSLYPHHQGFPPHHHSYPEQIVNLFPTQAVGATICKKGAHIKOLARFAGASIKIA			
	hHCCA	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE	KH1	KH2	600
	hIMP2	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	hCRD-BP	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	hIMP1	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	mCRD-BP	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	cZBP1	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	hIMP3	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	xVRBP-A	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	xVRBP-D	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	zVRBP	P...AEGPDVSR...RMVITGPPPAQFAQGRIFGKLEENFFPKKEEVKLEAHIRVPSSTAGRVIGKGGKTVNLQNLTSARVIVPRDQTPDENDQV...VKIIGHFPAQTAQRKIRE			
	hHCCA	IVQQVKEQEQKYPQGVASQRSK	KH3	KH4	720
	hIMP2	IVQQVKEQEQKYPQGVASQRSK			
	hCRD-BP	IVQQVKEQEQKYPQGVASQRSK			
	hIMP1	IVQQVKEQEQKYPQGVASQRSK			
	mCRD-BP	IVQQVKEQEQKYPQGVASQRSK			
	cZBP1	IVQQVKEQEQKYPQGVASQRSK			
	hIMP3	IVQQVKEQEQKYPQGVASQRSK			
	xVRBP-A	IVQQVKEQEQKYPQGVASQRSK			
	xVRBP-D	IVQQVKEQEQKYPQGVASQRSK			
	zVRBP	IVQQVKEQEQKYPQGVASQRSK			

B

FIGURE 1. See caption on facing page.

tein with similar domain composition and spacing, individual KH domains from the two open reading frames yielding the lowest similarity score (*Xenopus* Vg1RBP, allelic variant A, and *Drosophila* Vg1RBP—54% identity) were subjected to multiple sequence alignment and presented as an unrooted tree (Fig. 1B). The results of this analysis unequivocally cluster each KH domain of the *Xenopus* Vg1RBP sequence with its equivalent from the *Drosophila* protein. This indicates a consistently higher degree of similarity between equivalent KH domains from different proteins than between KH domains within the same protein, and strongly suggests that individual KH domains have distinct functions.

A closer inspection of the distribution of the highly conserved residues revealed that they appear as three clusters, each spanning two RNA-binding domains and the spacer between them. Thus, the spacers separating RRM1 from RRM2, KH1 from KH2, and KH3 from KH4 are short (11, 1, and 8 residues, respectively) but are also as conserved as the domains surrounding them, suggesting they are a functional sequence, rather than just a hinge. In contrast, the long spacers separating RRM2 from KH1 and KH2 from KH3 are highly variable. This observation led us to propose that Vg1RBP comprises three domain pairs—didomains—rather than six independent RNA-binding domains. Indeed, experimental confirmation of the didomain arrangement was obtained from RNA-binding experiments (see below), in which each pair of KH domains appears to function as one unit. Furthermore, preliminary NMR steady-state $\{^1\text{H}\}$ - ^{15}N NOE experiments on a uniformly ^{15}N -labeled construct spanning the KH1 and KH2 domains (G. Guler-Gane, A. Git, N. Standart, & B. Broadhurst, unpubl.) show that the peptide backbone is generally well structured in each domain and that there is significant flexibility at the N- and C-termini, in the spacer between KH1 and KH2, and in the disordered loop between strands $\beta 2$ and $\beta 3$ in both domains. However, the degree of backbone flexibility in the interdomain spacer region is no greater than that observed in the loops within each KH domain.

Based on these domain definitions, a series of truncated variants, all C-terminally His-tagged, was overexpressed in bacteria and purified (Fig. 2A,B). Variants are named after the didomains they span: RRM1 and RRM2 (R), KH1 and KH2 (K12), KH3 and KH4 (K34), separated by hyphens representing the variable spacers. Some ambiguity arose in the definition of the C-terminal boundary of didomain K12 because conserved sequences extended beyond the predicted end of the KH fold. Hence, two end points were assigned, resulting in two sets of proteins. The first was based on the consensus KH domain definition (thick-underlined sequences in Fig. 1A, proteins R-K12 and K12), and the other based on evolutionary conservation included an additional 14 residues (thin-underlined sequences in Fig. 1A, proteins R-K12+ and K12+). We did not find

any difference between these two protein sets in any assay. Full-length Vg1RBP (R-K12-K34) expresses to high levels and is almost entirely soluble under standard conditions. Only variants spanning intact didomains expressed to similarly high levels and were soluble. In agreement with the theoretical prediction of didomains, variants whose end points lay within a didomain (including single domains) expressed poorly and were found mostly as inclusion bodies (data not shown), suggesting that they are misfolded compared to the full-length counterpart.

In addition to the truncated variants, seven full-length variants were constructed that contained point-mutated KH domains (individually, as didomains, or all four; see Fig. 2), named so that an (^m) follows the number of the mutated KH domain (see position of mutations in Fig. 1A). The mutation introduces an asparagine residue instead of a conserved aliphatic residue from the KH domain core, adjacent to the RNA binding platform (Lewis et al., 2000). In FMR1, the gene product mutated in the fragile X mental retardation syndrome, which contains 2 KH domains, this naturally occurring mutation in the second KH domain (I304N) severely impairs RNA binding (Siomi et al., 1994). Equivalent mutations in the KH domains of hnRNP K and Sam68 also abrogate RNA binding (Siomi et al., 1994; Chen et al., 1997). We noted that some of the I304N-like mutations introduced into KH domains of Vg1RBP have an adverse effect on the solubility of the protein: R-K12-K3^m4^m is partially soluble whereas the quadruple mutant R-K1^m2^m-K3^m4^m is almost completely insoluble and is included in this study only as an in vitro translated product (Fig. 2).

An important issue to consider is whether the addition of a tag, even one as short as six His residues, may affect the folding and hence the activity of the fused protein. Two independent results indicate that the added tag does not affect RNA binding. First, in vitro-translated native and His-tagged Vg1RBP variants bind poly(C)-agarose to the same extent (Fig. 3D). Similarly, in UV crosslinking analysis, recombinant Vg1RBP added into oocyte extract shows binding to ^{32}P -labeled VLE RNA that is indistinguishable from its native counterpart (data not shown). These experiments suggest that neither the overexpressed nor the in vitro-translated Vg1RBP are adversely affected by the presence of a 6×His tag.

RNA binding by Vg1RBP requires both KH didomains but not the RRMs

Vg1RBP was originally detected in UV-crosslinking reactions with ^{32}P -labeled VLE (e.g., Deshler et al., 1997; Havin et al., 1998). We compared the ability of full-length and deletion/mutation variants to crosslink to ^{32}P -labeled VLE and noted that only proteins containing two active KH didomains (R-K12-K34 and K12-

and found that the variants fall into three categories (Fig. 3B). First, proteins spanning both KH didomains (R-K12-K34 and K12-K34) show high affinity binding to the VLE with an apparent K_D of approximately 15 nM (7–20 nM variations between protein batches; data not shown). Second, proteins containing only one KH didomain (R-K12, K12, -K34, K34) show much lower affinity for the VLE with an apparent K_D of approximately 300 nM (100–300 nM variation). Finally, the variant containing only an RRM didomain and no KH didomains (R–) demonstrates very poor binding to the VLE with an apparent K_D of 750 nM (700–1,000 nM variation). Moreover, as the removal of the RRMs (compare R-K12-K34 with K12-K34; R-K12 with K12) does not

FIGURE 3. Both KH didomains but not the RRM didomain are required for RNA binding. **A:** Three hundred nanograms of recombinant Vg1RBP variants were UV crosslinked to ³²P-labeled VLE. RNase-digested samples were analyzed using SDS-PAGE followed by autoradiography and the labeled protein were quantitated on a PhosphorImager screen. Binding of full-length Vg1RBP is set as 100%. Shown are typical results obtained in at least two independent experiments. **B:** Twofold serial dilutions (0.5–1,000 nM) of recombinant Vg1RBP variants were incubated with ³²P-labeled VLE and filtered through a nitrocellulose membrane. The protein-bound fraction of the RNA retained by the membrane was quantitated using a PhosphorImager screen. **C:** R-K12-K34 at 10 nM was incubated with uniformly labeled VLE in the absence or presence of homopolymeric competitors as indicated at 10-, 100- or 1,000-fold w/w excess over the labeled RNA. The protein-bound fraction of the VLE was analyzed as in **B**. Binding in the absence of competitor is set as 100%. **D:** In vitro-translated native Vg1RBP or His-tagged R-K12-K34 were mixed with control Sepharose CL-6B or with poly(C)-agarose (duplicate samples) and the resin-bound proteins were analyzed alongside an untreated sample equivalent to 15% binding (L) using SDS-PAGE followed by autoradiography. **E:** In vitro translated Vg1RBP variants were mixed with control Sepharose CL-6B or with poly(C)-agarose (duplicate samples) and the resin-bound proteins were analyzed alongside an untreated sample equivalent to 20% binding (L) using SDS-PAGE followed by autoradiography and quantitation on a PhosphorImager screen. Binding of full-length Vg1RBP is set at 100%.

seem to affect the interaction of the KH domains with the VLE, we conclude that the RRM portion of Vg1RBP does not interact with VLE to any significant extent. Taken together, these assays suggest that the interaction of Vg1RBP with the VLE requires both KH didomains, but not the RRM didomain. The contribution of the KH didomains appears to be cooperative, in that the joint affinity of both KH didomains exceeds the combined affinity of the individual KH didomains approximately 10-fold.

The exact combination of binding sites of Vg1RBP in Vg1 mRNA remains unclear (see Introduction), and it seems more than likely that Vg1RBP has additional mRNA targets. To examine how Vg1RBP may bind RNA other than the VLE, we first examined the homopolymer preference of Vg1RBP. Using filter binding, we observed that pyrimidine homopolymers (either poly(C) or poly(U)) are efficient competitors for VLE binding, and, conversely, that purine homopolymers (poly(A) and poly(G)) competed poorly (Fig. 3C). We then analyzed

Roles of Vg1RBP KH domains

the contribution of Vg1RBP KH didomains to poly(C) binding. In these experiments, Vg1RBP variants were labeled by translation in vitro in the reticulocyte lysate and bound to poly(C)-agarose beads. First, we used this assay to assure that the His tag does not affect the RNA binding activity of the protein (Fig. 3D); we then examined the RNA-binding activity of point-mutated full-length variants (Fig. 3E). Again, the variants spanning both KH didomains (R-K12-K34 and K12-K34) show the best binding to poly(C). This binding is halved by point mutations affecting either of the didomains, and completely abolished by mutating all four KH domains, reinforcing the importance of the KH domains in RNA binding. Indeed, the quadruple mutant is probably severely misfolded, as indicated both by its insolubility in a recombinant form, and its reproducible binding to Sepharose CL-6B, typical of proteins with exposed hydrophobic sequences. Interestingly, the contribution of the KH didomains to poly(C)-binding seems additive, rather than cooperative, as seen previously with the VLE. Whether this difference results from a higher RNA-to-protein ratio in this assay compared to filter binding or UV crosslinking analyses, or whether it reflects variation in the mode of binding to homopolymeric RNA compared to the VLE sequence is at present unknown. Irrespective of this difference, all three assays point to the critical roles played by both KH didomains of Vg1RBP in RNA binding.

Individual KH domains show a redundant contribution to RNA binding

To analyze the contribution of individual KH domains to the RNA-binding activity of Vg1RBP, we analyzed full-length variants bearing single mutations in KH domains in filter-binding, UV crosslinking, and poly(C)-binding assays (Fig. 4). Although none of the individual mutations significantly alters the overall affinity of Vg1RBP for the VLE (batch- and experiment-dependent variation of up to twofold in all K_D values was observed, data not shown) in a filter-binding assay (Fig. 4A), the proteins show different crosslinking efficiencies to the same RNA (Fig. 4B). Here, KH domains 1 and 3 seemingly play the greater role, with some contribution from KH4. Conversely, it is mutations in KH domains 1 and 2 that individually halve the binding of Vg1RBP to poly(C). It appears that, at least in binding to poly(C)-agarose, the relationship of the KH domains within a didomain differs: KH1 and 2 are both required for the activity of didomain K12, and their combined mutation is no worse than their individual ones (compare Figs. 3E and 4C), whereas KH 3 and 4 are redundant for the activity of didomain K34, so that the mutation of either individually does not affect the protein, but when combined, they abolish the activity of didomain K34. Overall, we found that the KH domains of Vg1RBP show significant redundancy in all assays, so that no single mutation

FIGURE 4. Individual KH didomains show redundant contribution to RNA binding by Vg1RBP. **A:** Twofold serial dilutions (0.5–600 nM) of recombinant Vg1RBP variants were analyzed by filter binding as in Figure 3B. **B:** Three hundred nanograms of recombinant Vg1RBP variants were subjected to UV-crosslinking as in Figure 3A. Crosslinking of wild-type Vg1RBP is set at 100%. Shown are typical results obtained in at least two independent experiments. **C:** In vitro translated Vg1RBP variants were mixed with Sepharose CL-6B or poly(C)-agarose and the resin-bound proteins were analyzed as in Figure 3E.

has a drastic effect on RNA binding, and that their relative importance may vary according to the RNA target.

Vg1RBP can self-interact in the absence of RNA

As a rapidly increasing number of RNA-binding proteins are reported to oligomerize (see Introduction), we

examined the ability of Vg1RBP and its variants to self-interact in chemical crosslinking and pull-down assays.

Vg1RBP variants were crosslinked with dimethyl suberimidate (DMS), a bifunctional crosslinking reagent, capable of attacking free amino groups, such as those found on the side chain of lysines, which are usually accessible on the surface of a protein (Thomas, 1989). Following phosphate-SDS-PAGE and silver staining, the gels were scanned and the increase in the proportion of multimer species larger than the monomer resulting from DMS addition was estimated for each protein (Fig. 5A). All KH-containing variants show an ability to form multimers, albeit with different efficiency: Proteins spanning didomain K34 (R-K12-K34, K12-K34, -K34, K34, K3^{m4m}) self-associate much more efficiently than those spanning only K12 (R-K12, R-K12+, K12, K12+), whereas the RRM alone was inactive. Importantly, the variance observed in the crosslinking of the truncated variants is not simply a reflection of their lysine content, as R, K12, and K34 didomains span 13, 17, and 16 lysine residues respectively. Moreover, the ability of didomain K34 to crosslink is not dependent on its ability to bind RNA, as shown by the efficient multimerization of K3^{m4m}. The self-interaction of Vg1RBP is dynamic, as formation of heterodimers between full-length HA-tagged Vg1RBP and K34 but

not with K12 can be demonstrated when mixed proteins are crosslinked and analyzed by western blot analysis using anti-HA antibodies (Fig. 5B). We conclude that Vg1RBP self-associates principally through the K34 didomain, and that this interaction can occur in the absence of RNA, as all the recombinant proteins were free of detectable bacterial RNA contamination following purification in a high-salt buffer (see Materials and Methods).

Stable self-association of Vg1RBP requires RNA

We next examined the behavior of Vg1RBP variants in pull-down assays, where recombinant His-tagged variants bound to Ni-NTA beads were used as bait for untagged in vitro-translated variants (³⁵S-prey). In this assay, the in vitro translated variants were present at low concentrations, and the reticulocyte lysate provided a more competitive environment for analysis of self-association, in contrast to chemical crosslinking, which relied entirely on purified recombinant proteins. First, several truncated and mutated variants of Vg1RBP were tested for their ability to bind an in vitro-translated untagged full-length prey (Fig. 6A). Here, recombinant variants containing wild-type didomain K34 (R-K12-

FIGURE 5. Vg1RBP can self-interact in vitro. **A:** Recombinant Vg1RBP variants were incubated in the presence (+) or absence (-) of dimethyl suberimidate, resolved using SDS-PAGE and visualized using silver staining. The increase in the proportion of multimeric forms upon addition of DMS is indicated below the gel. (Because of a polyacrylamide precipitate, crosslinking of R-K12+ could not be reliably quantitated.) Sizes of molecular weight markers (M) in kilodaltons are indicated on the left. **B:** R-K12-K34-HA aliquots were crosslinked by DMS in the absence (-) or presence of one-, two-, or fivefold molar excess of either K12 or K34. Samples were then resolved using SDS-PAGE and visualized by western blot analysis using anti-HA antibodies followed by chemiluminescence.

FIGURE 6. Stable self-association of Vg1RBP requires RNA. **A, B:** Two micrograms of each recombinant bait (indicated above the figure) or buffer alone (–) were bound to Ni-NTA matrix and mixed with in vitro-translated untagged prey (indicated on the left). Matrix-bound proteins were analyzed using SDS-PAGE alongside an untreated sample equivalent to 20% binding (L), followed by autoradiography and quantitation on a PhosphorImager screen. Efficiency of interaction is indicated by ++, +, or –. **C:** The interaction between full-length His-tagged R-K12-K34 and in vitro-translated untagged R-K12-K34 was carried out either as above (–) or treated with RNase A either before (pre) or after (post) the binding step. Matrix-bound proteins were analyzed as in **A**.

K34, K12-K34, -K34, K34, R-K12^m-K34) bind the ³⁵S-labeled prey efficiently, and those bearing only didomain K12 show weak or no binding (R-K12, R-K12+ and K12, K12+, respectively). However, the two proteins bearing a mutated didomain K34 (R-K12-K34^m, K34^m) differ in their behavior: the full-length variant binding strongly whereas the short variant does not interact at all. This extreme discrepancy, given the efficient multimerization exhibited by K34^m with DMS (Fig. 5A), could be attributed to the different assays employed. Although crosslinking captures all interactions, transient and stable alike, the pull-down assay is more stringent, and captures only interactions withstanding competition with additional proteins and several washes. Thus, we hypothesized that the RNA-binding activity, absent in K34^m, is required to transform transient interactions into stable ones. The RNA-binding activity does not have to reside in didomain K34, as exemplified by the efficient interaction of recombinant R-K12-K34^m with its full-length counterpart prey (Fig. 6A).

A prediction of this hypothesis is that Vg1RBP variants with different affinities to RNA would exhibit concentration-dependent variations in their abilities to self-associate. To examine this, pull-down assays were performed with two bait proteins (R-K12-K34 and K34) and several in vitro-translated prey proteins. And indeed, as seen in Figure 6B, variants spanning two intact KH didomains that possess high affinity to RNA (R-K12-K34 and K12-K34; Fig. 3) can interact with full-

length recombinant Vg1RBP even when found at low concentrations as in vitro-translated preys. This interaction is due to the KH domains and not the extreme C-terminal glutamine-rich stretch (Fig. 1A), as shown by the truncation variant R-K12-K34ΔBbsI (Fig. 6B). Conversely, variants spanning only one KH didomain (R-K12, R-K12+, K12, K34) interact very poorly as in vitro-translated proteins, although some are capable of significantly stronger association in a recombinant form (see R-K12, R-K12+, K34 in Fig. 6). Moreover, RNase A treatment, either before or after the binding step, abolishes the pull-down of full-length ³⁵S-labeled prey by recombinant full-length Vg1RBP bait, demonstrating the involvement of RNA in forming stable interactions (Fig. 6C).

Vg1RBP self-associates in vivo

Finally, to examine the ability of Vg1RBP to self-associate in vivo, we subjected a mixed stage oocyte extract to DMS crosslinking and analyzed the polymeric state of Vg1RBP using western blot analysis (Fig. 7). With increasing concentrations of DMS, we observe a dose-dependent decrease of the Vg1RBP species comigrating with the recombinant monomer. In the presence of DMS, a species comigrating with the recombinant Vg1RBP dimer appears at lower concentrations of DMS (0.3–0.9 mg/mL) and is gradually replaced by a high-molecular weight species, presumably a large complex, at the highest concentrations of DMS

FIGURE 7. Vg1RBP can self-associate in vivo. Mixed-stage *Xenopus* oocyte extract was crosslinked with buffer alone (–) or 0.3–1.5 mg/mL DMS. Samples were resolved using SDS-PAGE alongside mock treated (–) or crosslinked (+) samples of recombinant R-K12-K34 and subjected to western blot analysis using α -Vg1RBP antiserum. The migration of Vg1RBP monomers, dimers, and the large complex is indicated on the right.

(0.9–1.5 mg/mL). Interestingly, we note the presence of Vg1-RBP complex intermediates between the dimer species and the large complex, and the absence of intermediates between the monomer and dimer forms, suggesting that Vg1RBP dimerization is the first step in a formation of a large complex.

DISCUSSION

Members of the conserved VICKZ family of RNA-binding proteins play important roles in RNA localization, translation, and stability in a variety of cell types and organisms. They all share (with the exception of the *Drosophila* homolog; see below) the potential to bind RNA via six motifs, including two RRM and four KH domains, but the contributions of the individual domains remain to be elucidated. In this study, we carried out a systematic analysis of the structural elements important for RNA binding and for self-association of *Xenopus* Vg1RBP.

We first show that the KH domains of Vg1RBP mediate high affinity RNA binding. Both the full-length

Vg1RBP protein and the K12-K34 variant bind VLE with a similar K_D of about 15 nM. Moreover, mutagenesis of the critical I304 residue (FMR-1 numbering) in all four KH domains, in the full-length protein, completely abolishes RNA binding. Second, we show that both the KH didomains are required to bind RNA, and that their contribution is cooperative in the case of VLE (Fig. 3). Loss of any one KH domain, by mutagenesis, only somewhat reduces RNA binding, implying that no one domain is absolutely essential, and that the individual KH domains are partially redundant (Fig. 4). The observation that high affinity RNA binding requires all four KH domains very likely reflects the nature of the Vg1RBP targets in *Xenopus* oocytes. Recently, the interaction of Vg1RBP/Vera with two mRNAs, encoding Vg1 and VegT, which both localize to the vegetal cortex during late stages of oogenesis, has been characterized in some detail. In both cases, the 3' UTRs of approximately 300–400 nt that are necessary and sufficient for localization of reporter RNAs are characterized by the critical clustering of redundant protein-binding motifs known as VM1 (typically UUUCUA) and E2 (typically UUCAC). Loss of protein binding, by deletion or mutagenesis of these motifs, correlates with localization deficiencies (Bubunencko et al., 2002; Kwon et al., 2002). We propose, therefore, that different KH domains interact with the pyrimidine-rich elements scattered throughout these localization elements, in a cooperative manner. Indeed, most KH-containing proteins contain multiple copies of the domain and/or in combination with other RNA-binding domains, implying that more than one domain is required to form specific and stable interactions. In the case of hnRNP K and FMR1, the multiple RNA binding domains are all required to optimally associate with RNA (Siomi et al., 1994), whereas an array of S1 and KH domains in transcription factor NusA cooperate to create an extended specific RNA-binding surface (Worbs et al., 2001).

The crucial role of the four KH domains of Vg1RBP in RNA binding is echoed by a recent study of IMP-1 in NIH 3T3 cells, in which the authors demonstrate that all four KH domains, but not the RRM, are required for cytoplasmic trafficking and RNA binding. Interestingly, in the case of this homolog, the individual KH didomains do not bind RNA, implying that the contributions of the separate didomains may vary with the nature of the target RNA (Nielsen et al., 2002).

The two tandem RRM in Vg1RBP do not appear to contribute to RNA binding. The K_D value of the RRM didomain for Vg1 VLE (~ 750 nM) is very close to the estimated concentration of Vg1RBP in the oocyte (~ 1 μ M, data not shown), strongly arguing that the RRM are inactive in binding mRNA in vivo. As far as we can tell, the Vg1RBP RRM display the expected sequence features, including canonical RNP2 and RNP1 motifs normally associated with RNA binding (Fig. 1), and note that RNA binding can, in fact, be supported by

noncanonical RRM, such as those in PTB (Conte et al., 2000). At very high protein concentrations, the RRM displayed a striking preference for poly(U) relative to poly(C), (A) and (G) (data not shown); though the physiological implications of this observation are not clear. That the RRM is dispensable for the RNA-binding activity of Vg1RBP agrees with their complete absence from the *Drosophila* homolog (Fig. 1; Nielsen et al., 2000), suggesting that they are required by the vertebrate homologs to support either binding to as yet unknown RNAs or additional functions, such as protein–protein interactions. Indeed, the RRM appears to enhance K12 self-association (Fig. 6).

Finally, we also show that Vg1RBP self-associates, via the K34 didomain, in a manner that is greatly stabilized by RNA, both as recombinant protein, and in oocyte lysates (Figs. 6 and 7). The notion that KH domains mediate oligomerization was first indirectly implied by the finding that KH3 from human Nova-1 antigen forms an intermolecular tetramer when crystallized (Lewis et al., 1999). A recent NMR study of the Nova KH3 domain independently supports our findings that KH domains, although able to interact in the absence of RNA, are greatly stabilized in their protein–protein interactions by RNA. Here, homodimerization involves a specific protein–protein interface and results in a general stiffening of the protein such that the protein binds to RNA by a rigidification of the protein/RNA-binding surface mediated by protein dimerization (Ramos et al., 2002). However, the primary sequence of the Nova KH3 interacting surface is not conserved in Vg1RBP.

Members of the GSG/STAR subfamily of RNA-binding proteins, which contain a single KH domain embedded in a larger conserved domain of ~200 amino acids called the GSG domain, also oligomerize and bind RNA. Oligomerization of Qk1 is mediated by a coiled-coil in the N-terminal extension to the KH domain, whereas RNA binding requires the entire GSG domain. Importantly, a lethal mutation in mouse, *quaking*, was shown to be due to absence of self-association mediated by the GSG domain (Chen & Richard, 1998). Similarly, FMR-1, which targets G quartet mRNAs important for neuronal function (Darnell et al., 2001), associates with its close homologs FXR-1 and 2 family members through a coiled-coil motif located outside of the KH domains (Siomi et al., 1996). In the case of Vg1RBP, however, self-association is mediated principally by the K34 didomain, and to a lesser extent by K12, and the glutamine-rich stretch present C-terminal to KH4 is not required.

Although the ability to oligomerize presumably has important implications for the function of Vg1RBP, we do not know whether self-association is required for RNA binding because no dimerization mutant is available in the context of the full-length protein. An attractive model relating dimerization and high affinity RNA binding has recently been proposed (Harvey et al.,

2002), in which the forced engagement of a RNA/protein complex by a chemical inducer of dimerization is shown to modulate gene expression. On presentation of the chemical inducer of dimerization to an mRNA template by the protein, cooperative binding ensued, resulting in higher-affinity complexes, and inhibition of mRNA translation (Harvey et al., 2002).

On the basis of our observations, we propose that Vg1RBP self-associates in two steps. Initially, protein monomers interact via didomain K34 (and to a lesser extent via K12). This interaction is then stabilized by RNA binding. Hence, strong stable interactions require the presence of didomain K34 and efficient RNA binding in both partners. Proteins that possess only one KH didomain (in other words, have lower affinity for RNA) show intermediate RNA binding in the pull-down assay when present at high concentrations as recombinant proteins (150 nM), but poor RNA binding as in vitro-translated proteins (estimated maximum of 2.5 nM). Table 1 summarizes combinations of Vg1RBP variants, recombinant and in vitro translated forms, that were assessed for self-association. It is not an exhaustive list of all combinations analyzed (see Fig. 6), but the examples illustrate our principal observations: efficient self-association by proteins that bind RNA with a high affinity, lower levels of self-association by proteins that bind RNA with an intermediate affinity, and no self-association by proteins that do not bind RNA.

In summary, we propose that Vg1RBP self-associates initially through interactions mediated by the K34 didomain, which is subsequently stabilized by RNA binding, and furthermore, that self-association appears to be the first step in the formation of very much larger complexes. An important future avenue of investigation will be to address the functional importance of Vg1RBP self-association in the formation of a Vg1 mRNA localization complex.

MATERIALS AND METHODS

Construction of plasmids

pVLE

A fragment spanning positions 1383–1833 of the Vg1 cDNA (Accession No. M18055) was amplified using cloned *Pfu* polymerase (Stratagene) and the oligonucleotide primers: 5'-CGAATTCACCTACCATCAGTGTTTC-3' and 5'-CCCAAGCTTAGATCTATCACTGGCCAG-3', containing *EcoRI* and *HindIII* sites, respectively (underlined). The amplified fragment was cloned in between the *EcoRI* and *HindIII* sites of pGEM1 and verified by sequencing.

Recombinant Vg1RBP truncation variants

Fragments of the Vg1RBP open reading frame were amplified from cDNA clone D (Accession No. AF064634; Havin

TABLE 1. Summary of self-association efficiency of Vg1 RBP variants.^a

KH didomains spanned	Examples	RNA binding exhibited	Association with R-K12-K34
K34 and K12	R-K12-K34 K12-K34 R-K12-K34ΔBbsI	High	++
K34 only	-K34 K34	Intermediate (recombinant, 150 nM) Low (in vitro-translated, 2.5 nM)	++ +
K12 only	R-K12(+) K12(+)	Intermediate (recombinant, 150 nM) Low (in vitro translated, 2.5 nM)	+ or –
None	R-	Low	–
Any	K3 ^m 4 ^m	None	–

^aVg1 RBP variants were classified according to the KH didomains they span and the RNA binding they exhibit. Efficiency of interaction with a full-length counterpart is presented on an arbitrary scale: ++, +, –.

et al., 1998) using native *Pfu* polymerase (Stratagene) and combinations of the upstream and downstream oligonucleotide primers listed below. All upstream and downstream primers carried *Nco*I and *Xho*I sites, respectively (underlined). The amplified fragments were cloned in between the *Nco*I and *Xho*I sites of pET21d (Novagen) resulting in a C-terminal fusion of the Vg1RBP open reading frame segments to a 6×His tag. Because of the use of the *Nco*I site, all open reading frames start with ATGG, sometimes changing the residue following the initiation Met. Upstream primers (5′ → 3′): CATGCCATGGACAAGCTGTATATT (R-K12-K34; R-; R-K12; R-K12+); TGCCATGGAGACAGAGGTTCCG (K12-K34, K12, K12+); TGCCATGGCACATTTGATTCCTGG (-K34); TGCCATGGAGTCAGAGACCGTTCA (K34). Downstream primers (5′ → 3′): CGCTCGAGCGGAACCTCTGTTG (R-); CGCTCGAGCTGGAGATTCATAGC (R-K12, K12); CGCTCGAGGCCAATGCATTCAG (R-K12+, K12+); ACCGCTCGAGTTTCTTCTCGGTTG (R-K12-K34, K12-K34, -K34, K34).

R-K12-K34-HA

A triple copy of the hemagglutinin (HA) tag was amplified from GTEPI (a gift from Scott Ballantyne) using primers carrying *Xho*I sites (CGCTCGAGATCTTTACCCATAC and CGCTCGAGGCACTGAGCAGCGTA) and introduced in-frame into the *Xho*I site of R-K12-K34.

KH point mutant variants

Mutagenesis of ²²⁸Ile (KH1), ³⁰⁹Leu (KH2), ⁴⁴¹Ile (KH3), and ⁵²³Val (KH4) to Asn was achieved using the QuickChange kit (Stratagene) and the following pairs of oligonucleotide primers, respectively (5′ → 3′; mutated positions in lower case):

KH1: GAAAGAGGGTGCTACCAaCAGGAACATCACTAAAC and GTTTAGTGATGTTCTGtGGTAGCACCCCTCTTC,

KH2: GTAAAGAAGGAAGGAACaaTAAGAAAATTGAGCAAG and CTTGCTCAATTTTCTTAgtGTTCTTCCTTCTTTAC, KH3: GAAAGCAAGGACAACACAaCAAACAACATATCAC and GTGATAGTTGTTTGtTGTGTTGTCCTTGCTTCT, KH4: CAAAGGAGGAAAAACAaacAATGAACCTCAGAATTT AAC and GTTAAATTCTGAAGTTCATTgtTGTTTTCTCTTTG.

Expression and purification of Vg1RBP variants

Two hundred milliliters of overnight culture of BLR(DE3)pLysS cells (Novagen) in LB medium were diluted 10-fold in LB medium and left to grow at 37 °C until they had reached an A₆₀₀ of 0.5. Recombinant protein overexpression was induced by addition of 1 mM IPTG for 4 h at 37 °C. Following centrifugation, the cell pellet was resuspended in 30 mL HNTA (1 M NaCl, 50 mM sodium phosphate buffer, pH 7.8, 1% Triton X-100) and complete lysis of the bacteria was achieved in a French® pressure-cell press. Following centrifugation, the soluble fraction was mixed with 1 mL of HNTA-equilibrated Ni-NTA resin (QIAGEN) and left rotating at 4 °C for 1 h to allow protein binding. Following washes in HNTA and PBS, and one wash in PBS containing 75 mM imidazole, the recombinant protein was eluted off the Ni-NTA beads into 1 mL of PBS containing 500 mM imidazole and dialyzed against buffer M (150 mM NaCl, 50 mM sodium phosphate buffer, pH 7.8, 3.33 mM DTT, 15% glycerol, 0.1% Triton X-100). Protein concentration was determined by SDS-PAGE analysis of protein dilutions against BSA and RNase A standards. It is important to note that all recombinant proteins were purified on Ni-NTA beads in a high-salt buffer (1 M NaCl), resulting in protein preparations essentially free of bacterial RNA (data not shown). Indeed, in the absence of a high-salt or RNase treatment of the bacterial lysate, the recombinant proteins bound the Ni-NTA column very inefficiently, implying that the His tag is inaccessible when Vg1RBP binds bacterial RNA.

In vitro transcription

To generate a uniformly ^{32}P -labeled VLE probe, 0.5 μg of *MscI*-linearized pVLE (position 1816 in the Vg1 cDNA) was transcribed using T7 RNA polymerase in the presence of [α - ^{32}P]-UTP as described (Walker et al., 1996). To generate capped RNAs encoding Vg1RBP variants, plasmids were linearized with *XhoI* (resulting in removal of the His tag) or *BbsI* (resulting in an open reading frame truncation), and transcribed using T7 RNA polymerase in the presence of m⁷G(5')ppp(5')G and trace amounts of [α - ^{35}S]-UTP for quantitation purposes, essentially as described in Dasso and Jackson (1989). All in vitro-transcribed RNAs were purified by centrifugation through ProbeQuant G-50 Micro columns (Amersham-BioSciences) and quantitated by precipitation with CTAB (Dasso & Jackson, 1989).

In vitro translation

As described by Jackson and Hunt (1983), 150–350 ng of capped RNA encoding Vg1RBP variants were translated in 10 μL micrococcal nuclease-treated rabbit reticulocyte lysate (Promega) in the presence of 5 μCi [^{35}S]-Met. After 90 min at 30 °C, translation products were either subjected to binding assays (poly(C) binding assay or Ni-NTA pull-down assay) or analyzed by SDS-PAGE and autoradiography. In vitro synthesis of native and His-tagged variants (Fig. 3D) was achieved by coupled transcription–translation of Vg1RBP cDNA clone D (Havin et al., 1998) and R-K12-K34 using T3 and T7 RNA polymerases, respectively.

RNA-binding assays

Filter-binding assay

The procedure was adapted from Black et al. (1998). Fifty-microliter reactions in FBA buffer (100 mM NaCl, 50 mM Tris, pH 8.0, 1 mM MgCl_2) containing 0.1 mg/mL yeast tRNA (Roche), 10^4 cpm ^{32}P -labeled VLE probe, and varying concentrations of recombinant protein were incubated at room temperature for 20 min and filtered through a Hybond-C membrane (Amersham Biosciences) using a dot-blot vacuum apparatus. Following three washes with 300 μL of FBA buffer, the membrane-bound fraction of the probe was quantitated on a PhosphorImager screen (Molecular Dynamics). For poly(N) competition experiments, the binding reaction was performed using 10 nM full-length recombinant protein, supplemented with increasing concentrations of the four homopolynucleotides (Amersham Biosciences), and analyzed as above.

UV crosslinking

Ten-microliter reactions containing 300 ng recombinant protein and 10^5 cpm of uniformly ^{32}P -labeled RNA probe in UV buffer (100 mM KCl, 10 mM HEPES, pH 7.2, 3 mM MgCl_2 , 5% glycerol, 1 mM DTT, 5 mg/mL heparin, 0.1 mg/mL *E. coli* rRNA) were incubated at room temperature for 15 min. The RNA-bound proteins were then covalently crosslinked to the probe by exposing the samples on ice to 1.72 J/cm² of UV radiation using a XL1500 Stratallinker (Spectronics Corp.).

Unbound RNA, as well as the portions of bound RNA not protected by protein, were digested by addition of RNase A to 0.3 mg/mL and RNase T1 to 4 U/ μL for 1 h at 65 °C. Radio-labeled proteins were then analyzed by SDS-PAGE followed by autoradiography.

Poly(C)-agarose binding assay

In vitro translated Vg1RBP variants (10 μL) were diluted in 200 μL BB (200 mM KCl, 50 mM HEPES, pH 7.4, 0.2% Tween 20, 5 mM imidazole) and mixed with 10 μL poly(C)-agarose (Sigma) for 2 h at 4 °C. After three washes in 500 μL BB, the bead-bound fractions were analyzed by SDS-PAGE followed by autoradiography and quantitated on a Phosphor-Imager screen.

Self-association assays

Preparation of mixed *Xenopus* oocyte extract

A collagenase-treated buffer-free ovary was crushed by a 20 min spin at 10k rpm in a Sorvall HB-4 swing-out rotor. The entire aqueous phase was transferred and supplemented with Complete Protease Inhibitors (Roche) to a final concentration of 1 $\times$. Remaining debris was removed from the extract by four to five sequential spins of the supernatant at 14k rpm. Aliquoted extract was stored at –80 °C and thawed at most twice.

Chemical (DMS) crosslinking

Recombinant Vg1RBP variants (10 μL , 18 μM) were supplemented with heparin to 1 mg/mL and crosslinked by addition of 5 μL fresh 20 mg/mL dimethyl suberimidate (DMS) in 1 M boric acid, pH 11, as adapted from Thomas (1989). Control reactions were supplemented with 1 M boric acid, pH 11. Following a 1 h incubation at room temperature, excess cross-linking reagent was quenched by 2.5 μL of 600 mM L-lysine and samples containing 1 μg protein were analyzed by phosphate-SDS-PAGE followed by silver staining. Cross-linking of mixed oocyte extract was performed similarly using 0.3–1.5 mg/mL DMS and analyzed using western blot analysis.

Western blot analysis

Proteins were resolved using phosphate-SDS-PAGE and blotted onto Immobilon P membrane (Amersham BioSciences). Membranes were blocked overnight in TBST (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 0.1% Tween-20) containing 2% Marvel milk powder (TBST-M). Vg1RBP was detected with rabbit polyclonal antisera (1:10,000; Havin et al., 1998) in TBST-M, followed by horseradish peroxidase coupled donkey-anti-rabbit antibodies (Jackson ImmunoResearch, 1:10,000) in TBST-M. R-K12-K34-HA was detected with rat monoclonal HA antibody (Roche, 1:1,000), followed by horseradish peroxidase coupled donkey-anti-rat antibodies (Jackson ImmunoResearch). Visualization was by ECL.

Pull-down assay

Ten microliters Ni-NTA beads (Qiagen) were equilibrated with DB (100 mM KCl, 50 mM HEPES, pH 7.4, 0.2% Tween 20, 5 mM imidazole) and mixed with 30 pmol of recombinant His-tagged Vg1RBP variants for 1 h at 4°C. The matrix-bound proteins were then mixed with 10 μ L in vitro-translated untagged Vg1RBP variants in 200 μ L DB for 2 h at 4°C. After three washes in 500 μ L DB, the bead-bound fraction was analyzed by SDS-PAGE followed by autoradiography and quantitated on a PhosphorImager screen.

ACKNOWLEDGMENTS

We thank Richard Benton for KH domain mutagenesis, Scott Ballantyne for the GTEPI plasmid, Jean Thomas for technical advice on DMS crosslinking, Andy Raine and Bill Broadhurst for unpublished NMR data, and Joel Yisraeli for discussions. We are grateful to the British Council and the Wellcome Trust, who provided funding for this work.

Received June 26, 2002; returned for revision
July 15, 2002; revised manuscript received
August 1, 2002

REFERENCES

- Adinolfi S, Bagni C, Castiglione Morelli MA, Fraternali F, Musco G, Pastore A. 1999. Novel RNA-binding motif: The KH module. *Bio-polymers* 51:153–164.
- Black DL, Chan R, Min H, Wang J, Bell L. 1998. The electrophoretic mobility shift assay for RNA binding proteins. In Smith CWJ, ed. *RNA-protein interactions, a practical approach*. Oxford: Oxford University Press.
- Bubunenko M, Kress TL, Vempati UD, Mowry KL, King ML. 2002. A consensus RNA signal that directs germ layer determinants to the vegetal cortex of *Xenopus* oocytes. *Dev Biol* 248:82–92.
- Chen T, Damaj BB, Herrera C, Lasko P, Richard S. 1997. Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: Role of the KH domain. *Mol Cell Biol* 17:5707–5718.
- Chen T, Richard S. 1998. Structure function analysis of Qk1: A lethal point mutation in mouse *quaking* prevents homodimerization. *Mol Cell Biol* 18:4863–4871.
- Conte MR, Grune T, Gluman J, Kelly G, Ladas A, Matthews S, Curry S. 2000. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J* 19:3132–3141.
- Cote CA, Gautreau D, Denegre JM, Kress TL, Terry NA, Mowry KL. 1999. A *Xenopus* protein related to hnRNP I has a role in cytoplasmic RNA localization. *Mol Cell* 4:431–437.
- Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. 2001. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* 107:489–499.
- Dasso MC, Jackson RJ. 1989. On the fidelity of mRNA translation in the nuclease-treated rabbit reticulocyte lysate system. *Nucleic Acids Res* 17:3129–3144.
- Deshler JO, Highett MI, Abramson T, Schnapp BJ. 1998. A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates. *Curr Biol* 8:489–496.
- Deshler JO, Highett MI, Schnapp BJ. 1997. Localization of *Xenopus* Vg1 mRNA by Vera protein and the endoplasmic reticulum. *Science* 276:1128–1131.
- Doyle GA, Betz NA, Leeds PF, Fleisig AJ, Prokipcak RD, Ross J. 1998. The c-myc coding region determinant-binding protein: A member of a family of KH domain RNA-binding proteins. *Nucleic Acids Res* 26:5036–5044.
- Elisha Z, Havin L, Ringel I, Yisraeli J. 1995. Vg1 RNA binding protein mediates the association of Vg1 RNA with microtubules in *Xenopus* oocytes. *EMBO J* 14:5109–5114.
- Grishin NV. 2001. KH domain: One motif, two folds. *Nucleic Acids Res* 29:638–643.
- Harvey I, Garneau P, Pelletier J. 2002. Forced engagement of a RNA-protein complex by a chemical inducer of dimerization to modulate gene expression. *Proc Natl Acad Sci USA* 99:1882–1887.
- Havin L, Git A, Elisha Z, Oberman F, Schwartz SP, Standart N, Yisraeli JK. 1998. RNA binding protein conserved in both microtubule and microfilament-based RNA localization. *Genes & Dev* 12:1593–1598.
- Jackson RJ, Hunt T. 1983. Preparation and use of nuclease treated rabbit reticulocyte lysate for the translation of eukaryotic mRNAs. *Methods in Enzymol* 96:50–72.
- Kessler SH, Sachs AB. 1998. RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. *Mol Cell Biol* 18:51–57.
- Kim JH, Hahm B, Kim YK, Choi M, Jang SK. 2000. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J Mol Biol* 298:395–405.
- King ML, Zhou Y, Bubunenko M. 1999. Polarizing genetic information in the egg: RNA localization in the frog oocyte. *BioEssays* 21:546–557.
- Kloc M, Zearfoss NR, Etkin LD. 2002. Mechanisms of subcellular RNA localization. *Cell* 108:533–544.
- Kwon S, Abramson T, Munro TP, John CM, Kohrmann M, Schnapp BJ. 2002. UUCAC and Vera dependent localization of VegT RNA in *Xenopus* oocytes. *Curr Biol* 12:558–564.
- Lewis HA, Chen H, Edo C, Buckanovich RJ, Yang YY, Musunuru K, Zhong R, Darnell RB, Burley SK. 1999. Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure Fold Des* 7:191–203.
- Lewis HA, Musunuru K, Jansen KB, Edo C, Chen H, Darnell RB, Burley SK. 2000. Sequence-specific RNA binding by a Nova KH domain: Implications for paraneoplastic disease and the fragile X syndrome. *Cell* 100:323–332.
- Mowry K. 1996. Complex formation between stage-specific oocyte factors and a *Xenopus* mRNA localization element. *Proc Natl Acad Sci USA* 93:14608–14613.
- Mowry KL, Melton DA. 1992. Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in *Xenopus* oocytes. *Science* 255:991–994.
- Müller-Pillasch F, Lacher U, Wallrapp C, Micha A, Zimmerhackl F, Hameister H, Varga G, Friess H, Büchler M, Beger HG, Vila MR, Adler G, Gress TM. 1997. Cloning of a gene highly overexpressed in cancer coding for a novel KH-domain containing protein. *Oncogene* 14:2729–2733.
- Nielsen FC, Nielsen J, Kristensen MA, Koch G, Christiansen J. 2002. Cytoplasmic trafficking of IGF-II mRNA-binding proteins by conserved KH domains. *J Cell Sci* 115:2087–2097.
- Nielsen J, Christiansen J, Lykke-Andersen J, Johnsen AH, Wewer UM, Nielsen FC. 1999. A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol Cell Biol* 19:1262–1270.
- Nielsen J, Nielsen FC, Jakobsen RK, Christiansen J. 2000. The biphasic expression of IMP/Vg1-RBP is conserved between vertebrates and *Drosophila*. *Mech Dev* 96:129–132.
- Otero LJ, Ashe MP, Sachs AB. 1999. The yeast poly(A)-binding protein Pab1p stimulates in vitro poly (A)-dependent and cap-dependent translation by distinct mechanisms. *EMBO J* 18:3153–3163.
- Otero LJ, Devaux A, Standart N. 2001. A 250-nucleotide UA-rich element in the 3' untranslated region of *Xenopus laevis* Vg1 mRNA represses translation both in vivo and in vitro. *RNA* 7:1–15.
- Palacios IM, St Johnston D. 2001. Getting the message across: The intracellular localization of mRNAs in higher eukaryotes. *Annu Rev Cell Dev Biol* 17:569–614.
- Perez I, McAfee JG, Patton JG. 1997. Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* 36:11881–11890.
- Ramos A, Hollingworth D, Major SA, Adinolfi S, Kelly G, Muskett FW, Pastore A. 2002. Role of dimerization in KH/RNA complexes: The example of Nova KH3. *Biochemistry* 41:4193–201.

- Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH. 1997. Characterization of a β -actin mRNA zipcode-binding protein. *Mol Cell Biol* 17:2158–2165.
- Samuels M, Deshpande G, Schedl P. 1998. Activities of the sex lethal protein in RNA binding and protein:protein interactions. *Nucleic Acids Res* 26:2625–2637.
- Siomi H, Choi M, Siomi MC, Nussbaum RL, Dreyfuss G. 1994. Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* 77:33–39.
- Siomi M, Zhang Y, Siomi H, Dreyfuss G. 1996. Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. *Mol Cell Biol* 16:3825–3832.
- Thomas JO. 1989. Chemical cross-linking of histones. *Methods Enzymol* 170:549–571.
- Varani G, Nagai K. 1998. RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* 27:407–445.
- Walker J, Dale M, Standart N. 1996. Unmasking mRNA in clam oocytes: Role of phosphorylation of a 3' UTR masking element-binding protein at fertilization. *Dev Biol* 173:292–305.
- Wilhelm J, Vale R, Hegde R. 2000. Coordinate control of translation and localization of Vg1 mRNA in *Xenopus* oocytes. *Proc Natl Acad Sci USA* 97:13132–13137.
- Worbs M, Bourakev GP, Bartunik HD, Huber R, Wahl MC. 2001. An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol Cell* 7:1177–1189.
- Yaniv K, Yisraeli JK. 2001. Defining cis-acting elements and trans-acting factors in RNA localization. *Int Rev Cytol* 203:521–539.
- Yaniv K, Yisraeli JK. 2002. The involvement of a conserved family of RNA binding proteins in embryonic development and carcinogenesis. *Gene* 287:49–54.
- Yisraeli JK, Sokol S, Melton DA. 1990. A two-step model for the localization of maternal mRNA in *Xenopus* oocytes: Involvement of microtubules and microfilaments in the translocation and anchoring of Vg1 mRNA. *Development* 108:289–298.
- Zhang HL, Eom T, Oleynikov Y, Shenoy SM, Liebelt DA, Dichtenberg JB, Singer RH, Bassell GJ. 2001. Neurotrophin-induced transport of a β -actin mRNP complex increases β -actin levels and stimulates growth cone motility. *Neuron* 31:261–275.
- Zhang Q, Yaniv K, Oberman F, Wolke U, Git A, Fromer M, Taylor WL, Meyer D, Standart N, Raz E, Yisraeli JK. 1999. Vg1 RBP intracellular distribution and evolutionarily conserved expression at multiple stages during development. *Mech Dev* 88:101–106.
- Zhao WM, Jiang C, Kroll TT, Huber PW. 2001. A proline-rich protein binds to the localization element of *Xenopus* Vg1 mRNA and to ligands involved in actin polymerization. *EMBO J* 20:2315–2325.

RNA

A PUBLICATION OF THE RNA SOCIETY

The KH domains of *Xenopus* Vg1RBP mediate RNA binding and self-association.

Anna Git and Nancy Standart

RNA 2002 8: 1319-1333

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>