Characterization of the Molecular Basis of Group II Intron RNA Recognition by CRS1-CRM Domains*†

Received for publication, December 24, 2007, and in revised form, May 23, 2008 Published, JBC Papers in Press, June 17, 2008, DOI 10.1074/jbc.M710488200

Ido Keren1, Liron Klipcan1,2, Ayenachew Bezawork-Geleta, Max Kolton, Felix Shaya, and Oren Ostersetzer-Biran3

From the Institute of Plant Sciences, Agricultural Research Organization, Bet Dagan 50250, Israel

CRM (chloroplast RNA splicing and ribosome maturation) is a recently recognized RNA-binding domain of ancient origin that has been retained in eukaryotic genomes only within the plant lineage. Whereas in bacteria CRM domains exist as single domain proteins involved in ribosome maturation, in plants they are found in a family of proteins that contain between one and four repeats. Several members of this family with multiple CRM domains have been shown to be required for the splicing of specific plastidic group II introns. Detailed biochemical analysis of one of these factors in maize, CRS1, demonstrated its high affinity and specific binding to the single group II intron whose splicing it facilitates, the plastid-encoded atpF intron RNA. Through its association with two intronic regions, CRS1 guides the folding of atpF intron RNA into its predicted “catalytically active” form. To understand how multiple CRM domains cooperate to achieve high affinity sequence-specific binding to RNA, we analyzed the RNA binding affinity and specificity associated with each individual CRM domain in CRS1; whereas CRM3 bound tightly to the RNA, CRM1 associated specifically with a unique region found within atpF intron domain I. CRM2, which demonstrated only low binding affinity, also seems to form specific interactions with regions localized to domains I, III, and IV. We further show that CRM domains share structural similarities and RNA binding characteristics with the well known RNA recognition motif domain.

The CRM (chloroplast RNA splicing and ribosome maturation) domain is a recently recognized RNA-binding domain (~10 kDa) of ancient origin that has been retained in eukaryotes only within plant and algal genomes (1–4). Structural analysis of bacterial members in this group has revealed an α-β-α-β-α-β fold that is closely related to the C-terminal domain of the translation-initiation factor-3 protein (5–7). These, together with several other nucleic acid-binding proteins, including the archaeal Alba protein, DNase I, the ribosomal protein S8, RNA 3′-terminal phosphate cyclase, C-terminal domain of prolyl-tRNA synthetase, and the THUMP domain, have been included into a protein family, known as the Alba superfamily (8).

In prokaryotes, CRM domains are encoded by a single stand alone open reading frame; YhbY, the Escherichia coli member of this family, binds precursors to 50 S ribosomal subunits and is likely to be required for their maturation (4, 9). In plants, CRM domains are found in a family of proteins (16 in Arabidopsis and 14 in rice) (4) containing between one and four repeats of the domain. Forward and reverse genetics have implicated several proteins with multiple CRM domains (CAF1, CAF2, CRS1, and CFM2)4 in the splicing of plastid-encoded group I and II introns (1, 2, 10–13). CAF1 and CAF2 contain two CRM domains each and function in complexes with the peptidyl-tRNA hydrolase homolog (CRS2) to promote the splicing of nine group II introns (2, 11, 14). CRS1 has three CRM domains (see Fig. 1) and is required specifically for the splicing of a single subgroup IIA intron, the atpF intron RNA (1, 11, 12). Through its tight and specific association with two intronic regions within domains I and IV, CRS1 either stabilizes or induces the folding of atpF intron RNA, into its postulated “catalytically active” form (12). CFM2, which is closely related to CRS1, harbors four CRM domains and is involved in the splicing of both group I and group II introns in maize and Arabidopsis chloroplasts (13). Recently, we have shown that an individual CRM domain from CRS1 and E. coli YhbY protein are able to associate with RNA (4). Thus, the RNA binding specificities and functions in intron folding and/or splicing of CRS1, CAF1, CAF2, and CFM2 are likely to be mediated by their CRM domains.

To address this possibility, we characterized the RNA binding activities associated with each individual CRM domain in CRS1. Although CRM3 demonstrated high affinity binding to RNA but lacked sequence specificity, CRM1 formed specific interactions with a unique region within domains I, the 111-nt sequence, of the atpF intron ligand of the intact CRS1 protein. CRM2, which demonstrated low binding affinity (in the mM range), also seems to form specific interactions with regions localized to domains I, III, and IV.

In silico and biochemical analyses of CRS1 CRM domain RNA binding activities also revealed structural similarity to the

© 2008 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.
RNA Binding Characteristics of CRM Domains

RNA recognition motif (RRM) and suggest a common mechanism of RNA recognition by these two evolutionarily distinct RNA-binding domains.

EXPERIMENTAL PROCEDURES

Expression and Purification of Glutathione S-Transferase (GST) Fusion CRS1-CRM Domains—The maize crs1 open reading frame, cloned into a PGEX-2TK plasmid (12), was used as a template for the production of GST fusion proteins that included each single CRM domain from CRS1, including CRM1, CRM2, CRM3, and CRM3 “βαββ core” or the following mutations: CRM1 KKAG(23–28)GRRG; CRM2 GRNT(25–28)GRRG; CRM3 GRRG(23–26)ARRG; CRM3 GRRG(23–26)GARG; CRM3 GRRG(23–26)GARR; CRM3 GRRG(23–26)AAAA; CRM3 F30A, WKHK(38–41)AAAA; and CRM3 YRG(87–89)AAP. The different oligonucleotides used to generate these constructs are summarized in supplemental Table S1. Mutations in CRS1 CRM domains were introduced by a QuikChange® site-directed mutagenesis kit (Stratagene) with Pyrobest DNA-polymerase enzyme (Takara), according to the manufacturer’s instructions. Alternatively, mutations were also generated by extension PCRs; the upstream mutated strand was generated with forward wild-type primer and a reverse primer encoding the mutated residues; the overlapping downstream fragment was generated with a forward mutated primer and a reverse wild-type oligonucleotide. The “complete” mutated DNA fragment was generated by a third PCR, using these two partially overlapping PCR products as templates to each other, together with the wild-type forward and reverse primers of the domain. The wild-type and mutant domains PCR products were digested with BamHI and EcoRI and subcloned into the same sites in pGEX-2TK (Amersham Biosciences), in such a way that the each sequence encoding the domains was fused in-frame to GST. Similarly, CRM3 βαββ core was subcloned into pGEX-4T1 with Sall and NotI sites, in-frame to GST.

The resulting recombinant GST-CRM fusion proteins were expressed in E. coli XL-1-Blue strain (Stratagene), grown in 1 liter of LB medium at 37 °C to an A600 of ~0.8, and induced with 1.0 mM isopropyl β-D-thiogalactopyranoside for 16 h at 22 °C (or with 1.0 mM isopropyl β-D-thiogalactopyranoside for 3 h at 37 °C, in the case of CRM3 βαββ core). The cells were then pelleted, resuspended in 40 ml of ice-cold phosphate-buffered saline, and lysed twice in a French press manifold (Thermo). Nonsoluble materials were removed by 15 min of centrifugation at 12,000 × g (4 °C). The clear lysate was applied to 1 ml of glutathione beads (GE Healthcare); the bound protein was washed twice with cold phosphate-buffered saline and once with high salt buffer (50 mM Tris-HCl, 750 mM NaCl, 0.1% (v/v) Triton X-100, pH 8.0) to minimize nucleic acid contamination and protein aggregation. The salt was removed by phosphate-buffered saline wash, and the tagged protein was eluted (as attempts to release the CRM proteins from the GST tag resulted in the proteins aggregating on the beads) with 20 mM reduced glutathione (Sigma) in a minimal volume (0.5–2 ml) of buffer containing 100 mM Tris-HCl, 100 mM NaCl, pH 8.0, dialyzed against 50% (v/v) glycerol, 50 mM HEPES-KOH, 500 mM KCl, 0.1% Triton X-100, 5 mM β-mercaptoethanol, pH 7.0, all quoted into 100-μl fractions, and stored at −20 °C until use. Under these conditions, the proteins retained their full activity for a few days, during which the binding experiments were performed. SDS-PAGE analyses of the purified GST-CRM proteins were in good agreement with expected sizes and confirmed that the purity of the eluted protein was higher than 95% (data not shown). Mutated CRM1, CRM2, and CRM3 domains were purified as described above, generally resulting in similar protein yields.

In Vitro Transcription of RNA Templates and RNA Binding Assays—RNAs used for binding studies are outlined in supplemental Table S1. atpF intron was generated by PCR with genomic maize DNA, whereas the psal mRNA was obtained by reverse transcription-PCR with maize chloroplast RNA. The mutated DNA fragment of atpF domain I 111-nt region was generated by PCRs with two complementary DNA oligonucleotides overlapping the entire 111-nt sequence, containing the consensus T7 promoter site upstream to the 111-nt RNA sequence. In vitro transcription of “body-labeled” RNAs was carried out with T7-RNA polymerase (Promega), with 0.5 mM each of ATP, GTP, and CTP and 0.05 mM UTP in the presence of 20 μCi of [α-32P]UTP (3000 Ci/mmol). The in vitro transcribed RNAs were treated with RQ1 DNase (RNase-free) (Promega) to remove contaminating DNAs, purified by a RNase® kit (Qiagen), and stored in 50 μl of double distilled H2O at −20 °C for several days till their use in the RNA binding experiments.

Filter binding assays were performed essentially as described by Ostersetzer et al. (12). Before their use in the binding assay, the in vitro transcribed RNAs were renatured by heating to 95 °C for 2 min, in a small volume of 10 mM Tris, pH 7.0, 1 mM EDTA buffer, after which the RNAs were transferred to ice. The RNAs were allowed to refold in the presence of KCl (0.15 m) and MgCl2 (to 10 mM) at 37 °C for 5 min and were then transferred to ice until their use in the binding assays. Binding reactions (20 μl) contained 25 pmol labeled RNA, 20 mM Tris-HCl, pH 7.0, 150 mM KCl, 5 mM MgCl2, 5 mM DTT, 10 μg/ml bovine serum albumin, 1 unit/μl RNase inhibitor (Fermentas or Promega). Protein was added to the “renatured RNA” and incubated for 15 min at 25 °C. Following the binding, the reactions were chilled on ice and immediately passed through a sandwich of nitrocellulose (Protran, 0.2 μm; Schleicher & Schuell) and charged nylon membranes (Nytran SuperCharge nylon; Schleicher & Schuell) by vacuum filtration, using a slot-blot manifold (Hybri-slot; Invitrogen). The membranes were washed twice with 200 μl of 10 mM Tris-HCl, pH 7.0, 150 mM KOAc, 5 mM MgCl2 buffer, dried for 5 min at room temperature, and exposed to PhosphorImager screen (Fuji). The data were quantified by ImageQuant software (Version 5.1; Molecular Dynamics), and the fraction of bound RNA was calculated as the ratio between RNA captured by the nitrocellulose and total RNA signal by both membranes.

RNA Mobility Shift Assay—For gel retardation assays, binding reactions were performed by incubating increasing amount of purified GST-CRM1 (1–1,000 nM) with 25 pm 32P-labeled 111-nt RNA, in 10 μl of reaction volume containing 50 mM HEPES, pH 7.0, 4% glycerol, 66 mM KCl, 1 mM MgCl2, 0.1 mM EDTA, 5 mM DTT for 15 min at room temperature. Supershifts
were induced by anti-GST monoclonal antibodies (Santa-Cruz Biotech), which were added to the binding reaction mix and incubated for 15 min at room temperature. The binding reactions were transferred to ice and loaded onto native pre-electrophoresed 5% polyacrylamide gel (37.5:1 acrylamide:bis) and electrophoresed at 4°C, in 0.25× TBE, containing 2.5% glycerol. The gels were dried, and the RNA bands were visualized by PhosphorImager (Fuji).

Hydroxyl Radical RNA Footprinting—RNA footprinting analysis was performed as described by us previously (12). Prior to each assay, a nonradioactive T7-transcribed 111-nt region RNA was purified and extracted using the RNeasy® kit and folded as above. Binding reactions (50 μl) were performed by incubating 10 nM "refolded" 111-nt RNA template with 500 nM purified GST-CRM1 protein, in buffer containing 50 mM HEPES-KOH pH 7.0, 150 mM KOAc (in the absence of DTT or β-mercaptoethanol) at specified concentrations of MgCl₂, at 25°C for 10 min. Hydroxyl radical cleavage was induced as described previously (12). Cleavage sites were visualized by primer extension with Omniscript-reverse transcriptase (Qiagen) with 5'-32P-labeled reverse 111-nt primer (supplemental Table S2). Sequencing ladder was generated with untreated RNA by primer extension reactions with ddNTPs (ddATP and ddTTP at 0.15 mM, and ddCTP and ddGTP at 0.1 mM concentrations). Molar weight standard was generated by 5'-32P-labeling of a DNA mass ladder (Fermentas). Primer extension reactions were stopped by the addition of two volumes of 90% (v/v) formamide in 0.5× TBE, boiled for 3 min, and resolved (5–10 μl) by electrophoresis in 6% polyacrylamide, 7 M urea gels (under 15 Watts for ~2 h). After electrophoresis the gels were dried and imaged with a PhosphorImager (Fuji).

RNA Structure Probing—The secondary structure of the maize atpF intron RNA 111-nt region was analyzed by an enzymatic method, generally as described by Warf and Berglund (15). The in vitro transcribed 111-nt RNA was purified by a RNeasy® kit (Qiagen) and folded by denaturation and renaturation procedure, as described above. Limited RNA hydrolysis was initiated by mixing 0.7 μg of the in vitro transcribed 111-nt RNA fragment with ribonuclease V1 (0.01, 0.1, and 0.2 unit), ribonuclease T1 (1, 10, and 100 units), or mung bean nuclease (1, 8.5, and 85 units), in the appropriate digestion enzyme buffer, for 2 min at room temperature (22–25°C). The reactions were stopped by adding an equal volume of stop solution buffer, containing 7.5 M urea, 20 mM EDTA, followed by phenol extraction and EtOH precipitation. For primer extensions, a reverse 111-nt primer (5'-GGGGATTGT- GTTTGCTC-3'), labeled with 32P[γATP], was combined with 111-nt fragment RNA. The products of the RNA cleavage and primer extension were loaded onto 6% polyacrylamide gels containing 7 M urea in 1× TBE, along with a sequencing ladder obtained by primer extensions with ddNTPs (see above) and a 5'-labeled DNA ladder. Following electrophoresis (under 15 Watts for ~2 h), the gels were dried and imaged with a PhosphorImager.

In Silico Analyses—Multiple alignments of sequences were performed with ClustalX (16) and T-coffee programs (17). MEGA program (version 3) and ClustalX were used for the construction of phylogenetic trees (18). Predicted three-dimensional structures of individual CRM domains were constructed by homology modeling, using the SwissModel server (19) and established bacterial CRM structures (5–7). ProtSkin program (20) was used to extrapolate multiple sequence alignment on the three-dimensional structure of the proteins, thereby allowing us to predict, map, and highlight selected residues on the predicted three-dimensional structure. The visualization and manipulation of obtained models were arranged by the PyMol package. RNA-binding sites on each protein were predicted using the RNABindR model and folded as above. Binding stoichiometry was determined by RNA secondary structures were analyzed by the RNA Alifold (22) prediction server using multiple sequence alignments with ClustalX of orthologous atpF 111-nt sequences,
obtained from various plants; the secondary structure was plotted with Mfold software (23).

The manuscript is supported by supplementary data.

RESULTS

Individual CRM Domains from CRS1 Bind with Different Affinities and Specificities to Different Fragments of the atpF Intron—Individual CRS1-CRM domains were expressed and purified as recombinant proteins fused to GST, and each protein was assayed for its binding to atpF intron RNA in vitro. By fitting the binding data with the Hill model (Origin 7.5 software; Microcal Software Inc., Northampton, MA), the calculated dissociation constant value ($K_D$) for CRS1-CRM1 domain binding to atpF intron RNA was $65 \pm 15$ nM, with an apparent Hill coefficient ($n_H$) of $0.82 \pm 0.11$ (Fig. 1), suggesting non-cooperative binding activity. Under these conditions, GST alone demonstrated no activity (Fig. 1). GST-CRM3 demonstrated the highest affinity, with a calculated dissociation constant of $27 \pm 7$ nM and an apparent Hill coefficient of $0.93 \pm 0.14$ (Fig. 1), whereas GST-CRM2 demonstrated only low binding activity, with a calculated dissociation constant of $0.32 \pm 0.10$ mM, ~10,000-fold higher than the $K_D$ values of CRM1 or CRM3, and a calculated Hill coefficient value of 0.4 (Fig. 1), suggesting negative binding cooperativity.

We also assayed the binding specificities associated with each of the CRM domains of CRS1 to different fragments of atpF intron RNA (Fig. 2A); these included a unique 111-nt sequence found within atpF domain I (the 111-nt region) and domain IV, two previously identified CRS1-binding sites (12), and several other fragments that are not recognized by the intact CRS1 protein, including atpF domains II, III, and V-VI and the plastid psaI mRNA (12).

GST-CRM3 bound with similarly high affinity to all of these RNAs (affinities ranging between 50 and 80 nM; Fig. 2B). However, GST-CRM1 demonstrated specificity in its binding only to the 111-nt region in domain I (Fig. 2B). Although GST-CRM1 $K_D$ value to the 111-nt fragment was ~80 nM (Fig. 2B), which is similar to the binding affinity of CRM1 to atpF intron RNA (~65 nM), its
calculated dissociation constants to other RNA fragments corresponding to *atpF* domains II-III, IV, and V-VI or to *psal* mRNA were significantly lower (ranging between 200 and 300 nM) (Fig. 2B). Its specific association with the 111-nt region was also demonstrated in gel mobility shift assays (Fig. 2C). Although the resolution of this assay was limited, a single band corresponding to GST-CRM1/111-nt RNA ribonucleoprotein complex was formed (Fig. 2C). Moreover, in the supershift assay, the addition of anti-GST antibodies decreased the mobility of the ribonucleoprotein GST-CRM1/111-nt RNA complex (Fig. 2C, lanes 9–13), further supporting the specific association of CRM1 to the 111-nt sequence.

In contrast to CRM1 and CRM3, GST-CRM2 demonstrated only low affinity to *atpF* intron, with a calculated dissociation constant value in the low millimolar range (Fig. 1). GST-CRM2 was found to be unstable, because under high concentrations the protein tended to aggregate upon its elution from the glutathione beads. Although we were unable to make any firm conclusions about its RNA-binding sequence specificity, results from RNA binding assays suggested that CRM2 form specific interactions with several regions localized to *atpF* intron fragments corresponding to domains I, III, and IV (data not shown).

**CRM1 Associates Specifically with Short Single-stranded Sequences within the 111-nt RNA**—To identify regions in the 111-nt region that are directly bound by CRM1, we performed hydroxyl radical footprinting with GST-CRM1 and the 111-nt RNA (Fig. 3). First, the “native” secondary structure of the re-arranged *atpF* intron domain I 111-nt RNA fragment was assayed by limited RNA hydrolysis, mediated by either single-stranded (T1 and mung bean) or double-stranded (V1) ribonucleases. The positions of the nuclease-induced cleavage sites were visualized by primer extension assays, using a 5′-32P-labeled oligonucleotide complementary to the 111-nt RNA sequence, because direct 5′ or 3′ labeling of the RNA fragment failure to be informative. As predicted from its primary nucleotide sequence, many regions were likely to form stable base pair interactions (including nucleotides 456–466, 484–496, 498–504, 505–508, 514–517, and 529–535), because these were all found to be susceptible to V1 nuclease treatment (Fig. 3A), whereas nucleotides 474–478 and 518–524, were found more sensitive to mung bean degradation (Fig. 3A, marked with dashed lines), suggesting that these are found as single strands. These also included the Gly520 residue that is highly sensitive to T1-induced degradation (Fig. 3A, marked with an asterisk). Fig. 3B summarizes the predicted secondary structure of the 111-nt region, generated by Mfold software (24), according to the enzymatic method (Fig. 3A) and Alifold predictions (23), based

---

**FIGURE 2. CRM binding specificity to different *atpF* intron fragments.** A, schematic representation of the group IIA *atpF* RNA and fragments used in the binding assays. B, filter binding assays were performed with 25 pm of the indicated RNA (atpF RNA fragments or *psal* mRNA), incubated with increasing concentrations (0–1,000 nM) of GST-CRM1 or GST-CRM3 in standard binding buffer in the presence of 50 µg/ml (E. coli) tRNA, as competitor RNA. The data points are the average of five experiments; the error bars represent the corresponding standard deviations. Kd and n values are the means ± S.D. of these experiments, as determined by Origin software. C, RNA-based gel mobility shift assays were performed by incubating 25 pm of labeled 111-nt transcript with increasing concentrations (0–1,000 nM) of purified GST-CRM1 or GST-CRM3, in 10-µl reactions containing 50 mM HEPES, pH 7.0, 4% glycerol, 66 mM KCl, 1 mM MgCl2, 0.1 mM EDTA, 5 mM DTT. Supershifts (lanes 9-13) were induced by monoclonal anti-GST antibodies (at various concentrations, as indicated), in reaction mixtures containing 1,000 nM GST-CRM1 protein and 25 pm of labeled RNA, prior to their loading on the gel. The samples were run on 7% polyacrylamide gel (37.5:1 acrylamide:bis) and electrophoresed in 0.25× TBE, containing 2.5% glycerol, at 4°C. The gels were then dried and visualized with PhosphorImager (Fuji). The arrows indicate unbound RNA (Free RNA), GST-CRM1/111-nt ribonucleoprotein complex (RNP complex), and supershifted complex induced by anti-GST antibodies (Super-shift).
RNA Binding Characteristics of CRM Domains

on multiple sequence alignments of orthologous plant atpF 111-nt sequences.

Specific RNA residues in the 111-nt region that are bound by CRM1, were determined by hydroxyl radical footprinting, with GST-CRM1 and the in vitro transcribed renatured 111-nt RNA sequence (Fig. 3C). The unlabeled 111-nt intron RNA fragment (10 nx) was pre-incubated with 10 mM Mg2+ in the presence or absence of 500 nM GST-CRM1. The ribonucleoprotein complex was subjected to hydroxyl radical cleavage as previously described (12), and cleavage positions were visualized by primer extension assays, using a 5’-32P-labeled DNA oligonucleotide complementary to the “linker region” between domains I and II. To estimate the degree of protection, intensities of individual bands were quantified and calculated from the ratio between the lanes, normalized with several “strong stop” bands (Fig. 3C, marked with asterisks) and compared with the unfolded RNA form (in the absence of Mg2+; data not shown).

Overall, the highest protection sites (over 5-fold; Fig. 3C, cross-hatched bars) were localized to single-stranded regions in the 111-nt RNA (as predicted from the nuclease method and in silico analyses; Fig. 3B). These are likely to arise by stable and specific interactions between CRM1 and the atpF 111-nt region RNA. The positions of GST-CRM1-induced protection sites are highlighted on the secondary structure of the 111-nt RNA (Fig. 3B); lines (Fig. 3B) and arrows (Fig. 3, A and C) point to Gly residues within the 111-nt region.

The specific association of CRM1 with the UUUUU (nt 476 – 480), UUGAAA (residues 518 – 522), and AGCAAA (residues 538 – 543) sequences within the 111-nt region (highlighted in Fig. 3B) was assayed in a series of binding experiments. These included a mutation in CRM1 putative binding sites within the 111-nt sequence (Fig. 4A), the addition of DNA oligonucleotides, complementary to the binding site into the reaction mix containing CRM1 and the renatured 111-nt RNA (Fig. 4B), and competition assays with various RNA oligonucleotides (Fig. 4C).

Because deletion of the sequences within the 111-nt region was expected to alter the folding of the 111-nt RNA (Mfold software), CRM1 putative binding sites were mutated, in which each of these sequences was replaced by short adenosine stretches (i.e. UUUUU 476 – 480 into AAAAA; UUGAA 518 – 522 to AAAAAA; and AGCAAA 538 – 543 to AAAAAA). Although this mutation had no obvious effect on the overall secondary structure of the mutated 111-nt RNA, as determined by limited RNA hydrolysis (data not shown), the binding activity of GST-CRM1 to the mutated 111-nt RNA was significantly reduced (Fig. 4A). Moreover, DNA oligonucleotides complementary to nucleotides 476 – 480, 518 – 522, and 538 – 543 (AAAAA, TTTCAC, and TTGTGT, respectively), which were added to the reaction mix containing the renatured 111-nt RNA (to minimize the risk of RNA misfolding by the DNA oligonucleotides), led to a significant reduction in CRM1 binding to the intact 111-nt RNA fragment (Fig. 4B). The specific association of CRM1 to the UUUUU, UUGAAA, and AGCAAA sequences was further supported by competition experiments (Fig. 4C). Specific RNA oligonucleotides repre-senting the putative sites of CRM1 within the 111-nt region (i.e. nucleotides 476 – 480, 518 – 522, and 538 – 543, respectively) competed very effectively with the intact intron fragment for the binding to GST-CRM1 (Fig. 4C), whereas hexameric poly(A), poly(U), poly(G), or poly(C) RNA oligomers had less effect on the binding of CRM1 to the 111-nt RNA. In these assays, the binding of GST-CRM1 to the 111-nt RNA was...
affected to a higher extent by the addition of purines rather than pyrimidines (Fig. 4C).

Taken together with the footprinting data, these results (i.e. mutations within the 111-nt region, the addition of DNA oligonucleotides and competition assays) strongly support the association of CRM1 to three short single-stranded sequences within *atpF* domain I 111-nt RNA: UUUUUU 476–480, UUGAAA 518–522, and AGCAA 538–543.

**CRM3 Binding Activity Is Mediated through a GXXG Sequence and Several Conserved Aromatic Residues**—Multiple sequence alignments, RNA-binding prediction servers (21) and the predicted structures of CRS1 CRM domains, based on the known structures of bacterial CRM proteins (5–7), have indicated several sites that may bind RNA. These include a GRRG sequence within CRS1-CRM3 domain (4, 12) and several aromatic residues, which are highly conserved among the CRM lineage. The GXXG sequence is a highly conserved motif found in an unlooped region between the first β-strand and the second α-helix that is found in all CRM proteins (4). Interestingly, a GXXG motif has been shown to play an important role in the RNA binding activity of the nonrelated nucleic acid-binding K homology family (24–29). Although at least one CRM domain in each CRM protein contains the GXXG motif, it is typically conserved in just one of the domains in proteins with multiple CRM domains (4, 12). In the CRS1 subfamily of the plant CRM domain family, the GXXG sequence is typically conserved as G(R/I)RG within the third domain (see supplemental Fig. S1). In most cases, the first CRM domain, in these triple CRM proteins, lacks the first glycine in the GXXG sequence but retains a basic residue (typically lysine) in the second position and glycine in the fourth position, whereas the second CRM domain in each protein contains the first two residues of the GXXG motif (glycine followed by an arginine) but lacks the two sequential RG residues.

The variable nature of the GXXG motif within each of the CRS1 CRM domains may therefore account for the differences in their binding affinities and specificities (Fig. 1). To test this idea, we mutated this region within the CRS1–CRM3 domain. Mutations in the GXXG sequence, in which CRM3 GRRG was replaced with AAAA, significantly reduced the binding activities of both intact CRS1 (data not shown) and GST-CRM3 to *atpF* intron RNA (Fig. 5A) (4). When each of the amino acids in the GRRG of GST-CRM3 were replaced individually by alanine, only mutations of the arginine residues reduced the activity significantly (approximately four to five times lower affinities; Fig. 5A).

CRM1 and CRM2 of CRS1 lack a GXXG motif. To determine whether introduction of this sequence would alter their RNA binding properties, we introduced the motif into the CRS1 CRM1 and CRM2 domains; these included the insertion of glycine instead of lysine in the first position and arginine instead of alanine in the third position (KKAG into GKRG) in CRM1 domain, and the insertion of glycine and arginine instead of asparagine and threonine, respectively, in the first and fourth positions (NTRG to GRRG) in CRM2 domain. However, although it had a large effect on CRM3 activity, insertion of the GXXG motif into CRM1 or CRM2 had no obvious effect on their binding affinities or specificities to different RNAs tested (data not shown). Thus, the association of RNA to the CRS1 CRM1 and CRM2 domains is likely to be dependent upon various other interactions.

In addition to the GXXG sequence, three other mutations were also found to have a significant effect on CRM3 RNA binding activity. Interestingly, in each case these mutations included conserved aromatic residues. The first site included a phenylalanine residue (Phe30), present in the second loop (L2) in adjacent to the GXXG motif, which is highly conserved among CRS1-CRM3 homologous sequences (supplemental Fig. S1). A mutation in this site, in which phenylalanine 30 was replaced with alanine (F30A), had a profound effect on CRM3 binding, resulting in ~15-fold lower binding affinity to *atpF* intron RNA (Fig. 5B). The second region included a cluster of an aromatic and positively charged residues at the end of helix 1 (i.e. WKHK 39–42) that is highly conserved in all CRM domains (supplemental Fig. S1). As observed in Fig. 5B, when this WKHK region was replaced with AAAA, a mutation that is unlikely to break the helix formation, CRM3 RNA binding affinity was reduced 4–5-fold. The third site also included an aromatic residue that is found within a highly conserved YRG sequence (YRP in bacteria; amino acids 87–89) at the end of the fourth strand of each CRM domain, which has also been predicted by RNABINDR program (21) to be a putative RNA-binding site. To avoid a mutation that might affect CRM3 folding (proline along with glycine are commonly found in sharp turns connecting β-strands, known as β-bends), the YRG sequence was mutated to AAP. As with the Phe30 residue and the WKHK sequence, a mutation within this region resulted in a severalfold reduction in RNA binding activity (by ~5-fold; Fig. 5B), further supporting the role of this region in RNA binding. The conserved GXXG motif and the aromatic residues that were found to affect GST-CRM3 binding to *atpF* intron are highlighted on the on the putative three-dimensional structure of the maize CRS1-CRM3 ortholog (Fig. 5C).

**DISCUSSION**

CRM is a recently discovered RNA-binding domain of ancient origin that was recruited in higher plants to function in the splicing of their highly degenerated organellar-encoded group II introns (4). Of eight nuclear-encoded splicing factors identified to-date, four have been identified as CRM proteins, including CRS1, CAF1, CAF2, and CFM2 (2, 4, 11, 12, 14, 30–32). A detailed biochemical analysis of one of these factors in maize (CRS1) has suggested that CRS1 protein acts as a dimer; via its tight and specific association with regions in the 111-nt region and domain IV, CRS1 either stabilizes or induces the formation of tertiary interactions in the *atpF* intron RNA (12). These activities are likely to be mediated by the three CRM domains of CRS1.

To address this possibility, we analyzed the RNA binding activities associated with individual CRS1-CRM domains in *atpF* intron RNA binding specificity and intron folding. Although they were all found to associate with *atpF* intron RNA *in vitro* (Fig. 1), the CRS1-CRM domains varied markedly in their binding characteristics. Although CRM3 demonstrated high affinity (Fig. 1) but low specificity (Fig. 2) in its
binding to various fragments of \textit{atpF} intron RNA, CRM1 was found to associate specifically with three short single-stranded sequences within the 111-nt sequence: nucleotides UUUUU 476–480, UUGAAA 518–522, and AGCAAA 538–543 (Fig. 3B), as was evident by filter binding and RNA mobility shift assays (Fig. 2) and footprinting analyses (Fig. 3). Its specific association with these nucleotides was further demonstrated by competition assays (Fig. 4C) and RNA binding experiments with a mutated 111-nt RNA, in which each of the nucleotides in the putative binding sites of CRM1 was replaced with adenosine (Fig. 4A). Furthermore, DNA oligonucleotides complementary to CRM1-binding sites, which were added to the reaction mix containing the renatured 111-nt, significantly reduced the binding of CRM1 to the intact 111-nt RNA (Fig. 4B).

The association of CRM1 with a unique sequence inserted within \textit{atpF} intron domain I (i.e. the 111-nt region), which is not found in other plastidic group II introns (12), may therefore explain CRS1 specificity for the \textit{atpF} intron RNA both \textit{in vivo} and \textit{in vitro} (1, 12). However, we cannot rule out the possibility that CRM1 might also form weak interactions with various other regions within the \textit{atpF} intron (Fig. 2B).

Although specific activities have been associated with CRS1 CRM1 and CRM3 domains (i.e. binding specificity and strong affinity sites, respectively), the function of CRS1-CRM2 domain is least understood from our results. RNA binding assays suggested that CRM2 form specific interactions with \textit{atpF}, mainly with regions localized to domains I, III, and IV (data not shown). Yet CRM2 demonstrated only low binding activity ($K_D = \sim 0.3 \text{ mM}$) (Fig. 1) and was found to be unstable, because the protein tended to aggregate upon its elution from the glutathione beads. Unlike CRM1 and CRM3, the potential CRM2 electrostatic surface (PyMol package) revealed a hydrophobic $\alpha$-helical surface (see supplemental

---

**FIGURE 5.** CRM RNA binding activity is mediated through several conserved aromatic residues and an unlooped GXXG motif. CRM3 binding activity was assayed by mutational analysis within conserved residues found in the GRRG motif (A) or within highly conserved aromatic residues (B), as indicated. Binding reactions were performed as described under “Experimental Procedures.” The data points and calculated $K_D$ values are the means $\pm$ S.D. of at least three independent experiments. C, schematic representation of a putative three-dimensional structure of the CRM3 domain, modeled on \textit{E. coli} CRM protein (annotated as YhbY; Protein Data Bank code 1LN4) and generated by PyMol software. The conserved GXXG motif and the aromatic residues are highlighted on the three-dimensional structure.
RNA Binding Characteristics of CRM Domains

Fig. S2). Taken together, these observations may indicate that, in addition to its role in RNA binding, CRM2 functions in protein-protein interactions as well, as was also shown in the case of several RRM domains in “multiple RRM” proteins (33–36). However, because no experimental data existed until now to support CRM2 roles in protein-protein interactions (i.e. gel filtration analyses of CRM2 or CR51 mutant that is lacking CRM2 domain proteins), these speculative views must be considered with care.

In addition to the characterization of CRS1-CRM binding to atpF intron, we also analyzed the molecular basis of RNA recognition by the CRM domain. Interestingly, beside their overall structural organization, CRM also seem to share a common RNA-binding mechanism with the RRM domain. Both molecules possess a basic β-strand RNA-binding platform, in which the binding to RNA is further supported by a small number of conserved aromatic residues (Fig. 5; and reviewed in Refs. 37–39). Their similar architecture, in which several α and β domains are attached to a basic βαββ core, namely βα-(βαββ) in the CRM/Alba family (5–8) and (βαββ)-αβ in RRM domains (37–39), may suggest either convergent evolution to carry out similar functions (e.g. the binding or folding of large catalytic RNAs), or co-evolution of the CRM and RRM domains from an ancient RNA-binding βαββ-type protein ancestor, as has been previously suggested for αβ-type RNA-binding proteins (40, 41). Indeed, when assayed for its binding activity, the CRM3 βαββ core was found to associate with atpF intron RNA in the low micromolar range (supplemental Fig. S3), further supporting its role in RNA binding.

In addition to RRM, CRM also seem to share RNA binding characteristics with the K homology domain, through the role of a highly conserved unlooped GXXG sequence (28, 42, 43). However, the conservative appearance of this motif is not dictated by any structural necessity and thus is more likely to reflect “functional selective pressure.” Mutations within this region resulted in reduced binding activities of both CRS1 protein (data not shown) and the CRM3 domain (Fig. 5A) (4). In contrast to the K homology domain proteins, where the Gly residues in the GXXG are postulated to have an important role in RNA binding (28, 42, 43), the association of RNA to the GXXG motif in CRM3 seems to be largely conferred by its two arginine residues (Fig. 4A). Nevertheless, despite the importance of this region for CRM3 binding activity and its degenerated appearance in domains CRM1 and/or CRM2, the GXXG motif by itself cannot explain the observed differences in RNA binding between CRS1 CRM domains (Figs. 1 and 2), because the insertion of GXXG into CRM1 (KKAG into GKRG) or CRM2 (NTRG into GRRG) did not have any noticeable effect on the binding affinities or specificities of these domains (data not shown). The molecular basis of CRM1 or CRM2 binding specificity has yet to be established. RNA recognition by these domains is likely to be mediated by various other interactions, such as salt bridges, formed by basic residues with the phosphodiester backbone and by aromatic residues that stack with the ribonucleobase.

In summary, CRM is a novel RNA-binding domain that shares structural similarities and RNA binding characteristics with the well known RRM domain. These two αβ-sandwich-like proteins may have arisen from an ancient βαββ RNA-binding protein ancestor. Although in bacteria, CRM and RRM domains exist as single domain proteins, in plants these are found in a family of proteins containing between one and four repeats (4, 33). In multiple-domain proteins, CRM and RRM domains are separated by a variable length of linker sequences, which may provide a repertoire of binding affinities and specificities (44). In CRM, these seem to include a conserved GXXG motif, found in an unlooped region between the first β-strand and the second α-helix in CRM3, and several conserved aromatic residues.

Acknowledgments—We greatly appreciate Prof. Alice Barkan’s help with manuscript preparation. We also thank Bryan M. Warf and Dr. Andy Berglund for help with RNA structure probing and Dr. Hagai Abeliovich, Dr. Yoram Eyal, Dr. Arthur A. Schaffer, and Prof. Uri Lavi for useful discussions regarding the manuscript.

REFERENCES

1. Till, B., Schnitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. (2001) RNA 7, 1227–1238
2. Ostheimer, G. J., Williams-Carrier, R., Belcher, S., Osborne, E., Gierke, J., and Barkan, A. (2003) EMBO J. 22, 3919–3929
3. Barkan, A. (2004) in Molecular Biology and Biotechnology of Plant Organelles (Danielli, J., and Chase, C., eds) pp. 281–308, Kluwer Academic Publishers, Norwell, MA
4. Barkan, A., Klipcan, L., Ostertzer, O., Kawamura, T., Asakura, Y., and Watkins, K. P. (2007) RNA 13, 55–64
5. Liu, D., and Wyss, D. F. (2004) J. Biol. Chem. NMR 29, 391–394
6. Ostheimer, G. J., Barkan, A., and Matthews, B. W. (2002) Structure 10, 1593–1601
7. Willis, M., Krajewski, W., Chalamasetty, V., Reddy, P., Howard, A., and Herzberg, O. (2002) Proteins 49, 423–426
8. Aravind, L., Iyer, L., and Anantharaman, V. (2003) Gen. Biol. 4, R64
9. Jiang, M., Sullivan, S. M., Walker, A. K., Strahler, J. R., Andrews, P. C., and Maddock, J. R. (2007) J. Bacteriol. 189, 3434–3444
10. Asakura, Y., and Barkan, A. (2006) Plant Physiol. 142, 1656–1663
11. Jenkins, B. D., Kulhanek, D. J., and Barkan, A. (1997) Plant Cell 9, 283–296
12. Ostertzer, O., Cooke, A. M., Watkins, K. P., and Barkan, A. (2005) Plant Cell 17, 241–255
13. Asakura, Y., and Barkan, A. (2007) Plant Cell 19, 3864–3875
14. Jenkins, B. D., and Barkan, A. (2001) EMBO J. 20, 872–879
15. Warf, M. B., and Berglund, J. A. (2007) RNA 13, 2238–2251
16. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) Trends Biochem. Sci. 23, 403–405
17. Notredame, C., Higgins, D. G., and Heringa, J. (2000) J. Mol. Biol. 302, 205–217
18. Kumar, S., Tamura, K., and Nei, M. (2004) Brief Bioinform. 5, 150–163
19. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) Nucleic Acids Res. 31, 3381–3385
20. Deprez, C., Lloubes, R., Gavioli, M., Marion, D., Guerlesquin, F., and Blanchard, L. (2005) J. Mol. Biol. 346, 1047
21. Terrilobini, M., Lee, J.-H., Yan, C., Jernigan, R. L., Honavar, V., and Dobbs, D. (2006) RNA 12, 1450–1462
22. Hofacker, I. L., Fekete, M., and Stadler, P. F. (2002) J. Mol. Biol. 319, 1059–1066
23. Zuker, M. (2003) Nucleic Acids Res. 31, 3406–3415
24. Peled-Zehavi, H., Berglund, J. A., Rosbash, M., and Frankel, A. D. (2001) Mol. Cell. Biol. 21, 5232–5241
25. Lukong, K. E., and Richard, S. (2003) Biochim. Biophys. Acta 1653, 73–76
26. Zhou, Y., Mah, T.-F., Greenblatt, J., and Friedman, D. I. (2002) J. Mol. Biol. 318, 1175–1188
27. Huang, Y. J., Swapna, G. V. T., Rajan, P. K., Ke, H., Xia, B., Shukla, K., Inouye, M., and Montelione, G. T. (2003) J. Mol. Biol. 327, 521–536
RNA Binding Characteristics of CRM Domains

28. Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D., and Clore, G. M. (2002) *Nature* **415**, 1051–1056
29. Brykalo, M. A., Corbett, A. H., and Fridovich-Keil, J. L. (2007) *Nucleic Acids Res.* **35**, 1108–1118
30. Schmitz-Linneweber, C., Williams-Carrier, R. E., Williams-Voelker, P. M., Kroeger, T. S., Vichas, A., and Barkan, A. (2006) *Plant Cell* **18**, 2650–2663
31. Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. (2005) *Plant Cell* **17**, 2791–2804
32. Watkins, K. P., Kroeger, T. S., Cooke, A. M., Williams-Carrier, R. E., Friso, G., Belcher, S. E., van Wijk, K. J., and Barkan, A. (2007) *Plant Cell* **19**, 2606–2623
33. Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) *Cell* **106**, 595–605
34. Lau, C.-K., Diem, M. D., Dreyfuss, G., and Van Duyne, G. D. (2003) *Curr. Biol.* **13**, 933–941
35. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003) *Mol. Cell* **11**, 965–976
36. El-Antak, L., Tzakos, A. G., Locker, N., and Lukavsky, P. J. (2007) *J. Biol. Chem.* **282**, 8165–8174
37. Maris, C., Dominguez, C., and Allain, F. H. T. (2005) *FEBS J.* **272**, 2118–2131
38. Lunde, B. M., Moore, C., and Varani, G. (2007) *Nat. Rev. Mol. Cell. Biol.* **8**, 479–490
39. Klein, D. J., Moore, P. B., and Steitz, T. A. (2004) *J. Mol. Biol.* **340**, 141–177
40. Efimov, A. V. (1997) *Proteins Struct. Funct. Genet.* **28**, 241–260
41. Anantharaman, V., Koonin, E. V., and Aravind, L. (2002) *Nucleic Acids Res.* **30**, 1427–1464
42. Lewis, H., Musumuru, K., Jensen, K., Edo, C., Chen, H., Darnell, R., and Burley, S. (2000) *Cell* **100**, 323–332
43. Braddock, D. T., Baber, J. L., Levens, D., and Clore, G. M. (2002) *EMBO J.* **21**, 3476–3485
44. Shamoo, Y., Abdul-Manan, N., and Williams, K. R. (1995) *Nucleic Acids Res.* **23**, 725–728