Chapter Four

Cloning and Characterization of Arabidopsis MAR binding NOP56/58 homologs

Introduction

To increase our understanding of the organization and regulation of genes in the plant nucleus, we have undertaken detailed studies of protein composition of the nuclear matrix. Previous chapters in this dissertation have demonstrated that the plant nuclear matrix is a stable and reproducible structure. Plant matrices have been prepared by extraction of loosely bound proteins and putative "loop" DNA using combinations of either lithium diiodosalicylate and restriction enzymes, or DNase I and sodium chloride. Matrices prepared by these two fairly dissimilar methods generate very similar sets of proteins. This work has further demonstrated that certain matrix proteins bind specifically to MAR DNA, via an in vitro DNA protein blotting technique. To investigate whether the interaction seen in the DNA protein blotting technique represents an interaction in vivo, we needed more information about these MAR binding proteins.

The first step in further investigating these proteins was to establish the identity of the proteins and the genes that encode the proteins. Although our previous studies of the

plant matrix had used wheat and tobacco nuclei, advances in the sequencing of plant genomes suggested using a third species for isolation of individual genes encoding MAR binding proteins. In 1990, the Multinational Coordinated Arabidopsis thaliana Genome Research Project was launched by an international group of scientists. In 1996 the genomic sequencing effort was boosted by the infusion of funding from the Department of Energy (DOE), the National Science Foundation (NSF) and the Department of Agriculture (USDA)for the systematic, large-scale genome sequencing of Arabidopsis thaliana. While the initial goal of this project was to complete sequencing of this genome by 2004, it became clear during the course of this work that enough genomic data was available that the use of sequencing data available from the Arabidopsis plant would accelerate the identification of plant MAR binding proteins initially seen in wheat and tobacco.

Peptides derived from the lowest molecular weight protein of the group of proteins with apparent molecular weight near 70 kDa were sequenced. Amino acid sequence data allowed identification and cloning of three related proteins from Arabidopsis. Analysis of Arabidopsis EST sequence data and genomic sequence data suggests that we have cloned all of the expressed genes of this family of proteins from this plant. Homologs to the two most closely related proteins of these three have previously been reported as MAR binding proteins (Hatton and Gray, 1999).

Yeast and mammalian homologs of all three proteins have been described as nucleolar proteins involved in processing of ribosomal RNA (reviewed in Venema andTollervey, 1999). Specifically, the yeast homologs of these proteins are thought to

be involved in the processing of C/D box class of small nucleolar RNA (snoRNA). These snoRNAs in turn target specific bases on preribosomal RNA for 2'-O-ribose methylation. Recent evidence indicates that the role of snoRNAs in ribosome biogenesis may be just one of many possible interactions. Whereas initial evidence distinguished two main classes of snoRNAs as being involved in separate modifications of ribosomal RNA, new evidence indicates that there are additional processes in which these snoRNAs participate (Jady and Kiss, 2001; Watkins et al, 2000; Pederson, 1999). Protein components of the snoRNP complexes are also found as components of other complexes, including spliceosomes and telomerase. Particularly in light of the multifunctional roles of the snoRNAs and associated proteins, we cannot discard the MAR binding interaction of the closely related plant proteins.

We have expressed and purified all three of the Arabidopsis proteins as *E coli* expression proteins. These fusion proteins were used to conduct binding assays with both MAR DNA and with C/D box snoRNA. Both types of nucleic acid bound to the fusion proteins. Competition between the MAR DNA and C/D box snoRNA revealed a surprising result. Rather than compete with each other, the combination of nucleic acids showed a significant incremental binding effect, with more of each labeled nucleic acid being bound when present with the other nucleic acid.

Materials and methods

Peptide sequencing

Tobacco nuclear matrix protein was separated by SDS polyacrylamide (8%) gel electrophoresis. The gel was stained briefly in Coomassie blue R-250. Protein bands, which corresponded to the MAR-binding bands, were excised from the gel as approximately 1mm X 1mm X 50mm slices. The slices were washed twice with 50% acetonitrile, the liquid removed and the gel slice frozen at –70° C. The gel slices were shipped on dry ice to the Harvard Microchemistry Facility for sequence determination. Tryptic fragments were analyzed by HPLC to select peptides for further analysis. Two peptides were chosen for sequence analysis. Sequence of the peptides was determined by both matrix-assisted laser desorption time-of-flight (MALDI-TOF) mass spectroscopy performed on a Finnigan Lasermat 2000, and by Edman degradation. The two peptide sequences were compared to conceptual translations of DNA sequences from Genbank, using the BLAST similarity search programs, to tentatively identify homologous proteins.

Cloning and sequencing of Arabidopsis cDNAs and generation of expression constructs

Several degenerate PCR primers were synthesized based on the amino acid sequences determined for the tobacco protein. PCR was conducted using an Arabidopsis thaliana (ecotype Columbia) lambda cDNA library (CD4-7 Newman PRL2 library, Arabidopsis Biological Resource Center) as template. This allowed use of defined primer sites at the border of each cDNA insert. Positive PCR products were identified by hybridization with a 525 bp PCR fragment generated by PCR from Arabidopsis

genomic BAC clone TM21B4 using primers specific for an internal fragment of a predicted gene with close homology to the two sequenced peptides. The positive PCR products were gel isolated and cloned. DNA sequencing of several clones revealed three separate partial cDNAs. Full length cDNA was generated by PCR using primers designed from sequence of the partial cDNAs supplemented with data from Arabidopsis sequence data available through Genbank. The full length PCR products were cloned into the vector pCR2.1TOPO (Invitrogen) and transformed into *E coli* (TOP10) cells, following the manufacturer's (Invitrogen) instructions. The identity of the inserts was confirmed by cycle sequencing (Iowa State University DNA Sequencing & Synthesis Facility). Plasmids were digested with EcoRI and the inserts gel isolated. Inserts were ligated in frame at the EcoRI site of pMAL c2x (New England Biolabs) for expression of the protein in *E coli*. Plasmids were transformed into *E coli* (strain TB1) cells made competent by calcium chloride method (Sambrook et al., 1989). The pMAL c2x plasmid contains the coding sequence for maltose binding protein 5' of the insert site. Correct orientation of the fusion constructs was determined by PCR.

Identification of homologous proteins and sequence analyses

Protein homologs for the proteins enconded by each Arabidopsis clone were identified using the BLAST sequence similarity search tool (Altschul et al., 1990). Arabidopsis NOP56 and NOP58 proteins were aligned with each other using BLAST. Multiple alignments of NOP56 and NOP58 protein homologs were conducted using the CLUSTALW Multiple Sequence Alignment Program (version 1.81). EST data was obtained from the TIGR *Arabidopsis* Gene Index (AtGI). Genomic mapping information for the three cloned genes, as well as additional predicted Arabidopsis

homologs was obtained from The Arabidopsis Information Resource (www.arabidopsis.org).

Expression and purification of fusion protein

E coli containing the fusion construct was grown overnight in liquid culture at 37°C, shaking at 200 rpm. The culture was diluted 100 fold and grown for an additional four hours. IPTG (0.3 mM) was added to induce production of the fusion protein, and the culture grown an additional two hours. Cells were harvested by centrifugation and were washed in column buffer (0.1M NaCl, 20 mM Tris-HCl, pH 7.5, 0.5mM PMSF). Cells were lysed by sonication, and the lysate cleared by centrifugation. The cleared lysate was applied to a 2.5 ml amylose column, and washed with 12 column volumes of column buffer. The purified fusion protein was eluted with column buffer supplemented with 10 mM maltose. Fractions were collected and protein concentration determined by Bradford assay (BioRad). Fractions were also analyzed by PAGE to assess purity.

Probe labeling

DNA probes were radioactively end labeled with ^{32}P αdCTP by a Klenow fill-in reaction (Sambrook et al., 1989). RNA probes were transcribed from gel isolated, linearized plasmid template containing the mouse U14 ΔAV construct (Watkins et al., 1996;1998) in the presence of ^{32}P αCTP, using the Riboprobe labeling system (Promega). Unincorporated nucleotides were removed from both DNA and RNA probes using Sephadex G-50 spin column purification.

Slot blot DNA/RNA protein binding assays

Equal amounts of each column purified protein were loaded onto Immobilon PVDF membranes using a Schleicher and Schuell slot blot manifold. Protein was diluted into

binding buffer and 0.5 ml loaded per well. Protein was fixed to the membrane by drying, according to manufacturer's instructions. Membranes were rewet and then blocked for at least one hour in binding buffer (10 mM Tris-HCl, 50 mM NaCl, 2 mM EDTA, pH 7.5) supplemented with 5% (w/v) nonfat dry milk. Membranes were rinsed in binding buffer plus 0.5% (w/v) nonfat dry milk. For the standard protein DNA (RNA) binding assay, membranes were incubated for 2 hours with gentle agitation at room temperature in 3 ml binding buffer, with radioactively labeled probe and cold competitor as indicated in figure legends. Equal counts of probe were used, at approximately 10^6 cpm (Cerenkov counts) per binding reaction for both the DNA and RNA probes. Approximately 1ng DNA or 30 pg RNA was used in each binding reaction. Membranes were washed four times in binding buffer supplemented with 0.5% nonfat dry milk. Membranes were exposed in phosphoimager cassettes, and data collected and analyzed using a Molecular Dynamics Storm imaging system and ImageQuant data analysis software. Data were graphed using SigmaPlot (version 4.0).

Results

Identification and cloning of MAR binding proteins

Amino acid sequencing of the (apparent) lowest molecular weight MAR binding proteins from tobacco was done by the Harvard University Microchemistry Facility. Sequence for two tryptic peptides was generated:

Peptide 1 IVNDNYLYAK

Peptide 2: YGLIFHSSFIGR

Comparison of these peptide sequences to DNA and protein sequences deposited at Genbank revealed similar proteins from yeast and from human. Two putative homologs were predicted in Arabidopsis, based on preliminary sequencing of the genomic clone T21B4. The positions of these two peptide sequences in the three homologous proteins cloned from Arabidopsis are shown in in Figures 4-3A and 4-3B. The T21B4 BAC clone contained two closely related DNA sequences, spaced about 2 kb apart, that could be translated as peptides containing regions homologous to the two tobacco peptides. A single PCR primer pair was designed to amplify these DNA sequences. PCR using these primers and the BAC template generated products migrating on TAE agarose gels at approximately 525 base pairs. This band represented both products, since the identity of one of the products could be determined by EcoRV restriction site. I was unable to identify clones from either tobacco or Arabidopsis cDNA libraries using this genomic PCR product as a probe. The same primers were subsequently used to generate a product of approximately 475 bp from Arabidopsis cDNA. Restriction digestion of the PCR product showed that only one cDNA was amplified, rather than two, as would be expected if both predicted

genes were present as cDNA (data not shown). I was also not able to identify clones in the tobacco and Arabidopsis cDNA libraries using the labeled cDNA PCR products as probe.

Several degenerate primers were designed based on the two tobacco peptide sequences above. Attempts to generate useful products with pairs of degenerate primers were unsuccessful, largely due to the lack of specificity inherent in degenerate primers. Several potential products were observed using pairs of degenerate primers, but none contained the expected products.

A more sensitive PCR strategy using single degenerate primers paired with either the M13 Universal or M13 Reverse sequencing primers was employed as an attempt to generate partial cDNAs from the two libraries. This strategy offered several advantages, in addition to the gain in specificity due to the sequencing primers. Longer PCR products could be obtained, and some should contain full 5' or 3' cDNA sequence. The "M13" end of the PCR products would have a stretch of known sequence, which facilitates both cloning and sequencing. This known sequence could also be used to design nested PCR primers, if necessary.

Both the Arabidopsis and tobacco cDNA libraries employ vectors with M13 primer sites flanking the cDNA insertion site. The M13 based PCR strategy generated numerous products in most reactions using the Arabidopsis cDNA library (Figure 4-1), but only a few bands from the tobacco cDNA library. To identify PCR products containing the correct cDNA, PCR products were separated on TAE agarose (1%)

transferred to nylon, and probed with the 525 bp TM21B4 (genomic DNA) PCR product described above. Several positive products were identified, gel isolated, cloned and sequenced. The sequence data indicated three closely related cDNAs. The 5' and 3' ends of the three separate cDNAs were generated by RACE, and the full length cDNAs cloned using primers homologous to the ends of the coding sequence of each of the three cDNAs. Figure 4-2 presents the agarose gel analysis of the full length PCR products and the cloned products. The inserts from these clones were sequenced, and the accuracy confirmed by comparison to genomic DNA sequence published in Genbank. The sequence of the three cDNAs has been deposited in Genbank (accession numbers AF302490, AF302491 and AF302492).

Amino acid sequence comparisons

The three Arabidopsis cDNAs, were cloned based on their homology to tobacco MAR binding protein peptide sequences from a single protein. As expected, these three clones are closely related. However, two of the three clones are much more homologous to each other than to the third clone, at both the amino acid and DNA levels. Arabidopsis clones F108 and T12H1 are 84% identical, and 90% similar, when the proteins are aligned with each other (Figure 4-3A). However, when either of these proteins is aligned with the third clone, the homologies drop off considerably, as shown in Figure 4-3B. The alignment of F108 with T6H22 shows 42% identity and 58% similarity. Similarly, alignment of T12H1 with T6H22 shows 46% identity and 66% similarity at the amino acid level (data not shown).

Conceptual translations of the three cloned Arabidopsis proteins were compared to nucleotide sequences in Genbank using a Blast 2.0 protein query (tblastn). This search compares a protein query sequence against a nucleotide sequence database translated in all reading frames (Altschul et al., 1990). Searches conducted using the cloned Arabidopsis protein T6H22 or with the pair of closely related Arabidopsis proteins F108 and T12H1 each produced homologs from a diverse group of eukaryotic species, including fungus, insect and mammal. Detailed phylogenetic analysis and sequence comparisons of the mouse homologs of both NOP56 and NOP58 has been recently published (Newman et al., 2000). Therefore, the Blast results and ClustalW scores for the Arabidopsis proteins have been summarized in Table 4-1 and Table 4-2. The Blast results and detailed ClustalW alignments have been included for reference in Appendix B.

The best homologies for the Arabidopsis T6H22 protein were members of the NOP56/SIK1 family of proteins. This protein was first reported as a yeast protein that could overcome growth inhibition in yeast containing a GAL4-I kappa B fusion protein or a LEXA- I kappa B fusion protein. Yeast containing the GAL4-I kappa B fusion protein did not grow. When the NOP56 homolog was overexpressed, normal growth was restored. The researchers called the protein Suppressor of I kappa B 1, or SIK1 (Morin et al., 1995). The mechanism of this suppression of growth inhibition has not yet been resolved, but did not involve a direct interaction of SIK1 with the fusion protein.

The best homologies for the two closely related Arabidopsis proteins F108 and T12H1 were members of the NOP58 family of proteins. The NOP 56 and NOP 58 proteins are distinct but closely related proteins in all organisms in which homologs for each protein have been reported (Newman et al., 2000). The first member of the NOP58 protein family was also reported in yeast. In a screen for proteins interacting with NOP1, the yeast homolog of the nucleolar protein fibrillarin, both NOP 56 and NOP 58 were identified. The authors showed by gene knockouts that each protein is essential for viability in yeast. These proteins were shown to be required for proper assembly of the large ribosomal subunit (Gautier et al., 1997).

As a result of the sequencing of the Arabidopsis genome, it is possible to map each of the three cDNA clones. Figure 4-4 shows the chromosomal map positions for each of the three genes detected in this study. Each of the three genes are located on separate chromosomes. T6H22 is located on chromosome I, T12H1 is located on chromosome III, and F108 is located on chromosome V. There are two additional sequences with homologies to the three genes detected in this work. On chromosome III, a sequence on the genomic clone MBK21 is predicted to encode a protein which would be most similar to the NOP56 family of proteins. On chromosome V. a sequence predicted to encode a NOP58 homolog is located just 2 kb from the F108 gene detected in this work. However, based on the results from this work, as well as from comparisons of the predicted genes to EST sequences compiled by The Institute for Genomic Research (TIGR) database, only the three cDNAs cloned have been detected as expressed gene products.

Based on Arabidopsis EST sequences in Genbank, the F108 gene is expressed and the cDNAs is present in pools of cDNAs from multiple tissue types (Genbank accession numbers T43058, C99910, AA39515). The T12H1 gene product has been detected in leaf (BE037630, AI998474) and flower buds (F15123). The T6H22 gene product has been detected in seedlings (BE529903, AV53652), roots (AV551261, AV547230), green siliques (AV558805), leaf (AI998805), and is seen in several cDNAs from pooled tissues (T88657, AV441682, AA650935). Despite the large number of Arabidopsis EST clones (over 100,000) that have been reported in Genbank, none represent products of the two remaining predicted genes from Arabidopsis. Based on this information and the results from this work, the predicted genes on genomic clones MBK21 and T21B4 are not active genes.

Bacterial expression and purification of Arabidopsis proteins

To further study the proteins encoded by the three cDNAs, the full length cDNA coding sequences were subcloned into the pMAL c2x expression vector (New England Biolabs). This vector contains coding sequence for the maltose binding protein (MBP) upstream of the multiple cloning site. The MBP tag facilitates purification of the fusion proteins. *E coli* cultures containing the expression plasmids were grown in 100 ml liquid cultures and induced to express the fusion proteins. The protein was harvested, and purified by liquid chromatography on amylose columns. Figure 4-5 presents a comparison of the crude lysates and column purified proteins on an SDS polyacrylamide gel. The purified fusion proteins migrate at approximately 105 kDa. Note that there are some contaminants still present in the column purified protein extracts. These contaminants did not interfere with subsequent experiments, but

complicated the quantitation of the purified protein. Note also that purified fusion proteins F108 and T12H1 contain a doublet at 105 kDa. It is likely that these doublets are two bands that each represent the fusion protein. However, the data from these experiments did not resolve the identity or activity of each of the two bands. Amylose column fractions containing the purified protein were analyzed on SDS polyacrylamide gels (data not shown). Total protein concentration of the fractions with the fusion protein were done by Bradford assay (data not shown).

DNA protein blotting experiments were conducted to see that the *E coli* expressed proteins exhibited MAR binding activity, and that any contaminants present did not exhibit such activity. *E coli* expressed fusion protein was separated by SDS polyacrylamide gel electrophoresis, transferred to PVDF membranes and probed with the RB7-6 MAR. Figure 4-6 presents results from these experiments. All three fusion proteins exhibit MAR binding activity. In the results shown, only the fusion proteins show strong MAR binding activity, whether from the total lysate or the column purified fractions. A small amount of MAR binding activity can be seen in lower molecular weight bands in the first column of this figure, for the NOP58-like protein F108. These smaller peptides probably represent degradation products, and are not seen in at lower protein concentrations (columns 4, 5 and 9). The smaller peptides are present, but only in small amounts relative to the full length fusion protein.

Slot blot DNA/RNA protein binding assays

The binding of nucleic acids to membrane bound proteins in the standard DNA protein blotting experiment used to identify plant MAR binding proteins introduces several

variables to each experiment. In the standard DNA protein blotting experiment, proteins are denatured for separation by electrophoresis, then transferred electrophoretically to PVDF membranes. Renaturation presumably occurs when the membrane is rewet prior to hybridization. During the course of separation, transfer and renaturation, it is likely both that unequal amounts of protein are lost and that unequal amounts of activity are lost. These variations contribute to large experiment to experiment differences between DNA protein blotting experiments. Similarly, anomalies in the electorphoretic separations sometimes result in differences in the size of the protein bands transferred to the membrane. These variations make comparisons of binding activities difficult.

The production of the Arabidopsis proteins as *E coli* fusion proteins allowed binding studies to be conducted with proteins that do not undergo denaturation or electrophoretic transfer. Use of a slot blot vacuum manifold allows loading of equal amounts of protein in reproducible positions. Denaturation of the proteins was also avoided, since no separation beyond the column purification of the fusion proteins was required.

Initial experiments were conducted to determine the amount of each protein to be loaded on the membranes to obtain detectable DNA binding activity (data not shown). In addition, some slot blots were stained with Coomassie Blue or loaded with bromophenol blue dye in the binding buffer to assess whether equal amounts of protein were attached to the membranes, and that the protein was distributed evenly and within each slot. These initial experiments were conducted with both

nitrocellulose membranes and PVDF membranes. The nitrocellulose membranes occasionally suffered from diffusion outside the slot. For this reason, PVDF membranes were chosen for the binding experiments.

Homologs of the Arabidopsis proteins cloned in this work have been reported as nucleolar proteins that specifically interact with the C/D box class of snoRNAs. The C and D boxes are highly conserved between the multitude of snoRNAs within a species and between species (Leader et al, 1998). C/D box snoRNAs are characterized by conserved box C (UGAUGA) and D (CUGA) elements (Smith and Steitz, 1997). Gel shift experiments have demonstrated that a modified mouse U14 snoRNA interacts strongly with mouse putative homologs of the Arabidopsis proteins (Watkins et al., 1996; 1998; Newman et al., 2000). This RNA is modified by the deletion of about 40 nucleotides between box C and D. This construct, U14ΔAV, was used as both a competitor and as a probe in the slot blot binding experiments. The RB7-6 MAR from tobacco was used as the MAR, and restriction digested *E coli* genomic DNA was used as non-specific competitor in these experiments.

In the first set of binding experiments, the three *E coli* expressed Arabidopsis proteins were bound to PVDF membranes by slot blotting. RB7-6 MAR DNA was used as probe, and the binding was done in the presence of competition by either cold *E coli* DNA, by cold RB7 MAR DNA, or by cold ΔAV U14 RNA. Three outcomes to these experiments were envisioned:

1. MAR DNA competes more efficiently than C/D box RNA for protein binding,
2. C/D box RNA competes more efficiently than MAR DNA for protein binding,
3. RNA and DNA binding is independent.

Unexpectedly, a fourth result was observed. Instead of a reduction in binding by MAR DNA in the presence of snoRNA, an increase in binding was seen. The results of these binding experiments are presented graphically in Figure 4-7. RB7-6 MAR probe was competed by itself, as expected, but the competitor RNA gave a consistent increase in binding by the labeled MAR DNA probe. Given the small number of trials for each set of binding conditions (n=3), it remains possible that this effect is a random variation. The means for two of the three binding reactions for each experimental condition showed an overlap in the standard deviation between the "no competitor" condition and the snoRNA condition. It should be noted that in all of the "no competitor" binding reaction, as well as in all other binding reactions, 1 μg restriction digested *E coli* DNA was present as a non-specific competitor.

To further address this surprising phenomenon, the experiment was repeated with a new set of slot blotted protein. In this second set of experiments, we also incorporated a reversal of the binding situation used in the experiments shown in Figure 4-7. In this second set of experiments, we also used the ΔAV U14 snoRNA as probe. In addition, we increased the amount of ΔAV U14 snoRNA competitor from 250ng to 500 ng. Figure 4-8 presents the results of these experiments. Once again, the presence of both

RB7-6 MAR DNA and ΔAV U14 snoRNA in a binding reaction produced an increase in binding relative to binding in the presence of either no competitor (other than 1 μg *E coli* DNA) or in the presence of self competition, regardless of whether the RB7-6 MAR DNA or ΔAV U14 snoRNA was used as probe. In these experiments, with one exception, the increase in the means for each set of competitors fell outside the standard deviations. The exception is for the T6H22 fusion protein, where there is some overlap of standard deviations. However, even in this set of data, the combined presence of both ΔAV U14 snoRNA and RB7-6 MAR DNA does give an increase in binding.

Discussion

In this work we have established the identity of three genes from Arabidopsis encoding proteins that bind to MAR DNA in vitro. Analysis of genomic data and of gene expression data from Arabidopsis sequencing projects suggests that these three genes represent the complete set of the expressed genes of this gene family. Comparisons of the sequences from the Arabidopsis genes show a high level of homology to the NOP56 and NOP58 genes from yeast and other eukaryotes, including mammals. These genes and gene products have been identified initially by their involvement in the processing of ribosomal RNAs. cDNA sequences encoding similar plant proteins have been reported in pea and in chickpea. These plant sequences are much more closely related to each other than to the homologs in yeast or mammals. It remains to be determined whether the plant homologs of the yeast NOP56/58 proteins perform identical functions. A recent report of work done concurrently to the work in this dissertation describes the similar proteins from pea as MAR binding proteins (Hatton and Gray, 1999).

All three proteins determined to be expressed in Arabidopsis have been cloned into expression constructs. Protein from each of these constructs has been successfully produced in *E coli* as fusions with a maltose binding protein. The fusion proteins have been purified and shown to retain the MAR binding function we had predicted based on their homology to a tobacco protein identified by DNA protein blotting experiments. These fusion proteins were used to conduct further experiments to

compare the relative binding of MAR DNA ligands versus a modified C/D box snoRNA ligand.

Surprisingly, the data from these experiments suggests that these ligands do not compete for a single binding site. Instead, when the MAR DNA and snoRNA are both present in the binding reaction, an increase in binding occurs. The increase in binding is not seen when additional amounts of self competitor on non-specific competitor is present. These experiments do not provide insight into the mechanism by which this increased binding might occur, but one could speculate that the binding of one ligand causes a conformational change in the protein, allowing stronger binding of the second ligand. While it is possible that the binding occurs at a single site, or a multiple equivalent sites, the data from these binding experiments do not support such an idea.

Published data from experiments using the yeast *Saccharomyces cerevisiae* have shown that the NOP56 and NOP58 (also known as NOP5) proteins are required for normal ribosome biogenesis. Deletions and disruptions of the genes show that each is essential for growth in yeast. In addition, immunolocalization has shown that these proteins are concentrated in the nucleolus in *S. cerevisiae*. Immunoprecipitation experiments demonstrate that NOP1, the yeast fibrillarin homolog, forms a complex with NOP 56 and NOP 58 (Gautier et al., 1997; Wu et al., 1998). Therefore, these snoRNP complexes are thought to be the basic components of riboprotein complexes required for the 2'-O-ribose methylation of specific rRNA sequences. These proteins form complexes with a specific class of snoRNAs, the C/D box RNAs. C/D box snoRNAs are characterized by conserved box C (UGAUGA) and D (CUGA) elements

(Smith and Steitz, 1997). Additional proteins are hypothesized to interact with the basic complex to facilitate the modifications made at specific locations throughout the rRNA before is integrated into mature ribosomal particles.

The primary role for snoRNAs appears to be as guides for the modification of their target RNA. The vast majority of snoRNAs can be matched to sites of modifications on ribosomal RNA transcripts via regions of putative base pairing between the snoRNAs and specific segments near the target modification sites on rRNA transcripts. The matching of individual snoRNAs to their target sites has been done for almost all of the 2'O-methlyation sites in yeast. However, this work did not attempt to identify snoRNAs that might target other RNAs (Lowe and Eddy, 1999).

In order for the snoRNA guided modifications of the rRNA to take place, the snoRNA must associate with a number of proteins, including the enzymes responsible for the modifications. We currently have only a basic understanding of the framework for how these snoRNPs might form. The enzymes catalyzing the 2'-O-ribose methylation have not been identified, though NOP1/fibrillarin is hypothesized to fulfill that role. The remaining snoRNA associated proteins are expected to be quite varied in order to provide specificity, but only a few of those proteins have been identified (Venema and Tollervey, 1999).

In eukaryotes, it has recently been revealed that snoRNAs do not act solely on rRNA. A number of cellular and viral RNAs transit through the nucleolus during maturation and at least one of these, the spliceosomal snRNA U6, is a substrate for snoRNA

guide-directed methylation (Omer et al., 2000). Vertebrate spliceosomal snRNAs (small nuclear RNAs), are similar to snoRNAs in that they are also modified by pseudouridylation and 2'-O-methylation (Tycowski, et al., 1998). Proteins associated with snoRNAs, have also been shown to be associcated with RNA components of the spliceosome including U5 and U4 snRNAs (Jady and Kiss, 2001; Watkins et al, 2000). The interaction of snoRNAs with their protein partners appears to be due to similarity of the secondary structure of the RNAs (Watkins et al.,2000).

The activities of the nucleolus and of the nucleus remain mysterious. Similar to the cell as a whole, the nucleus is compartmentalized, and subnuclear structures including the nucleolus can be readily seen under a microscope. However, these subnuclear structures are not membrane bound, and may represent both sites of concentrated activities, such as ribosome synthesis, or storage locations, as may be the case for Cajal (coiled) bodies. Our knowledge of interactions between different components of the nucleus continues to grow, but much remains to be deciphered.

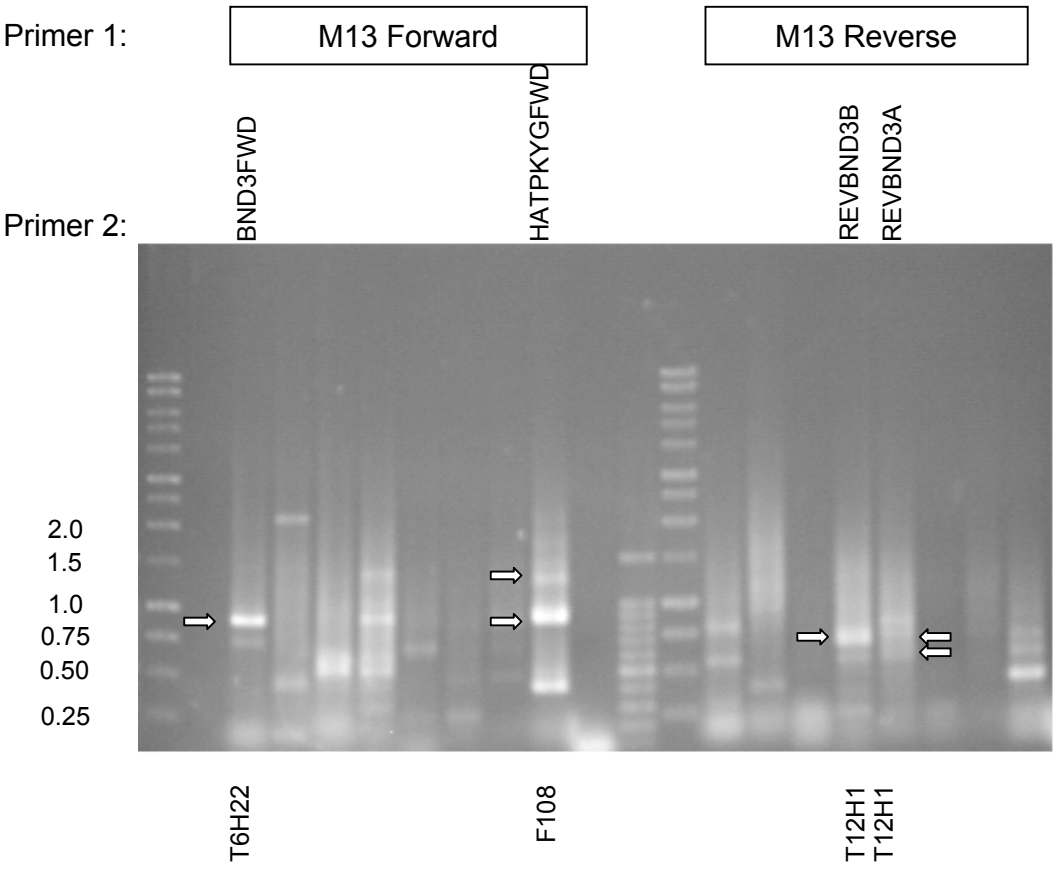

Figure 4-1. PCR of partial Arabidopsis cDNAs using degenerate primers. A lambda-cDNA library (CD4-7) was used as template for PCR with degenerate primers from one direction and M13 sequencing primers from the opposite direction. Products were separated by electrophoresis through 1% agarose, stained with ethidium bromide, and photographed. Positive products (arrows) were determined by Southern hybridization, then cloned and sequenced. At top, primers for products successfully cloned and confirmed by sequencing are shown. At bottom, identity of full length clone is noted.

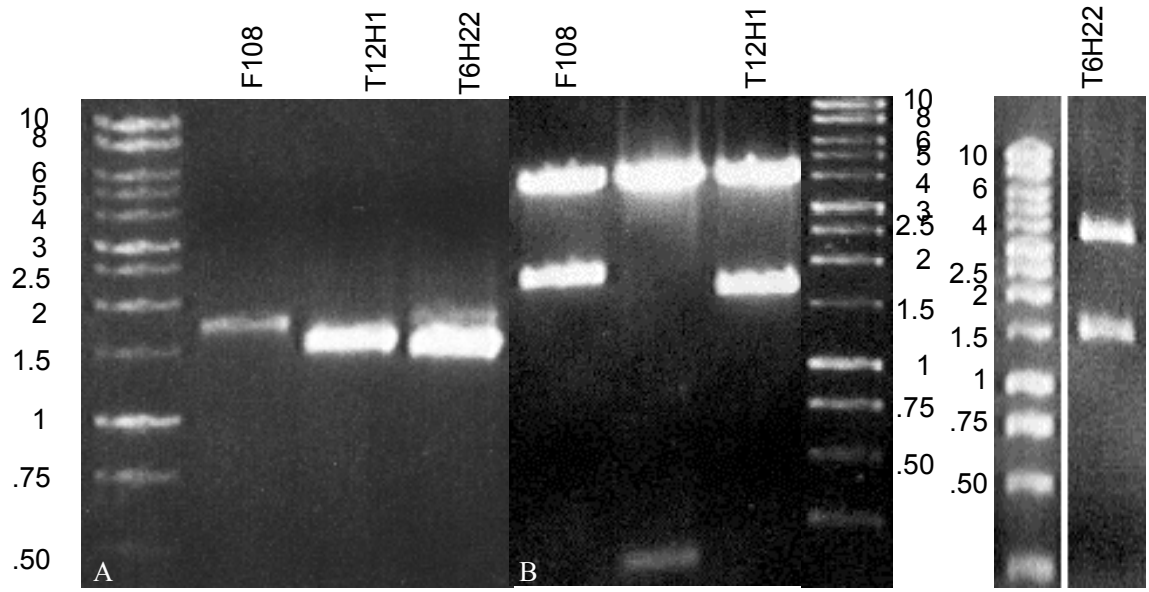

Figure 4-2. PCR and cloning of cDNA for putative Arabidopsis MAR binding proteins. (A) Specific primers for the 5' and 3' ends of the coding sequence were used to amplify first strand cDNA for putative Arabidopsis MAR binding proteins F108, T12H1 and T62H22 from Arabidopsis seedlings. (B) cDNA encoding the putative Arabidopsis MAR binding proteins was cloned into the vector pCR2.1TOPO (Invitrogen). Individual clones were isolated and DNA extracted. Clones with the correctly sized inserts were identified by digestion with EcoRI. Sizes of the markers used on these gels is indicated in kb.

TABLE 4-1. Alignment Scores for NOP58 homologs
(CLUSTAL W Multiple Sequence Alignments)

	Arabidopsis-F108	Arabidopsis-T12H1	PeaSARBP1	PeaSARBP2	Rice SARBP	Human NOP5/58	Rat	D. melanogaster	S. pombe
Arabidopsis-F108 AAG40838									
Arabidopsis-T12H1 AAG40837	82								
PeaSARBP1 AAC16330	74	73							
PeaSARBP2 T06379	73	73	86						
Rice SARBP BAA31260	74	74	77	75					
Human NOP5/58 AAD27610	52	55	56	55	54				
Rat AAF05769	54	55	55	55	56	95			
D. melanogaster AAF52455	52	52	53	52	53	60	61		
S. pombe CAB72231	54	52	54	52	52	49	48	50	
S. cerevisiae NP_014955	41	40	41	40	39	38	37	37	39

Table 4-1. Alignment and sequence comparison of NOP58 proteins from a number of diverse species was performed with ClustalW (1.81) multiple sequence alignment program. Accession numbers for each protein included in this comparison are shown below the sequence name in the left-hand column. Alignment scores between each amino acid sequence are listed as a percentage. Detailed Clustal alignments can be found in Appendix C.

TABLE 4-2. Alignment Scores for NOP56 homologs
(CLUSTAL W Multiple Sequence Alignments)

		Arabidopsis-T6H22	Cicer arietinum	S. pombe	Human	D.obscura
Arabidopsis-T6H22	AAG40836					
Cicer arietinum	CAA10127	71				
S. pombe	CAA22814	55	51			
Human	CAA72789	51	57	51		
D.obscura	CAB92783	52	52	51	49	
S. cerevisiae	AAB67431	49	48	58	49	49

Table 4-2. Alignment and sequence comparison of NOP56 proteins from a number of diverse species was performed with the ClustalW (1.81) multiple sequence alignment program. Accession numbers for each protein included in this comparison are shown below the sequence name in the left-hand column. Alignment scores between each amino acid sequence are listed as a percentage. Detailed Clustal alignments can be found in Appendix C.

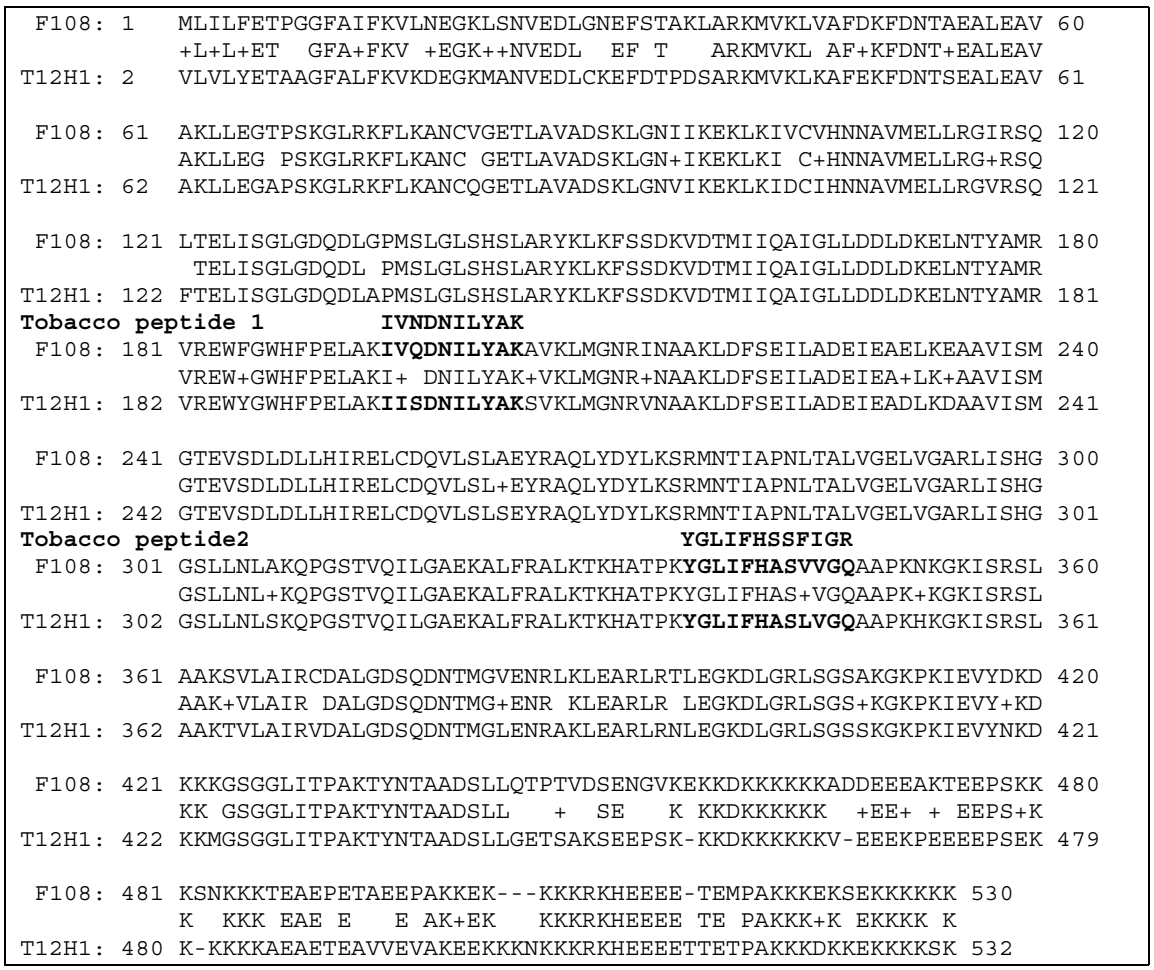

Figure 4-3A. Alignment of Arabidopsis NOP58 homologues F108 and T12H1. Sequences of the proteins deduced from cDNA sequences F108 and T12H1 were aligned using the BLAST local alignment tool (Tatiana and Madden, 1999). Both proteins have a predicted length of 533 amino acids. Regions homologous to the tobacco peptides sequenced are shown in bold. Identical amino acids are shown in the line between the two homologues, and similar (positives) amino acids are denoted by a "+". Identities = 453/534 (84%), positives = 486/534 (90%), Gaps = 7/534 (1%).

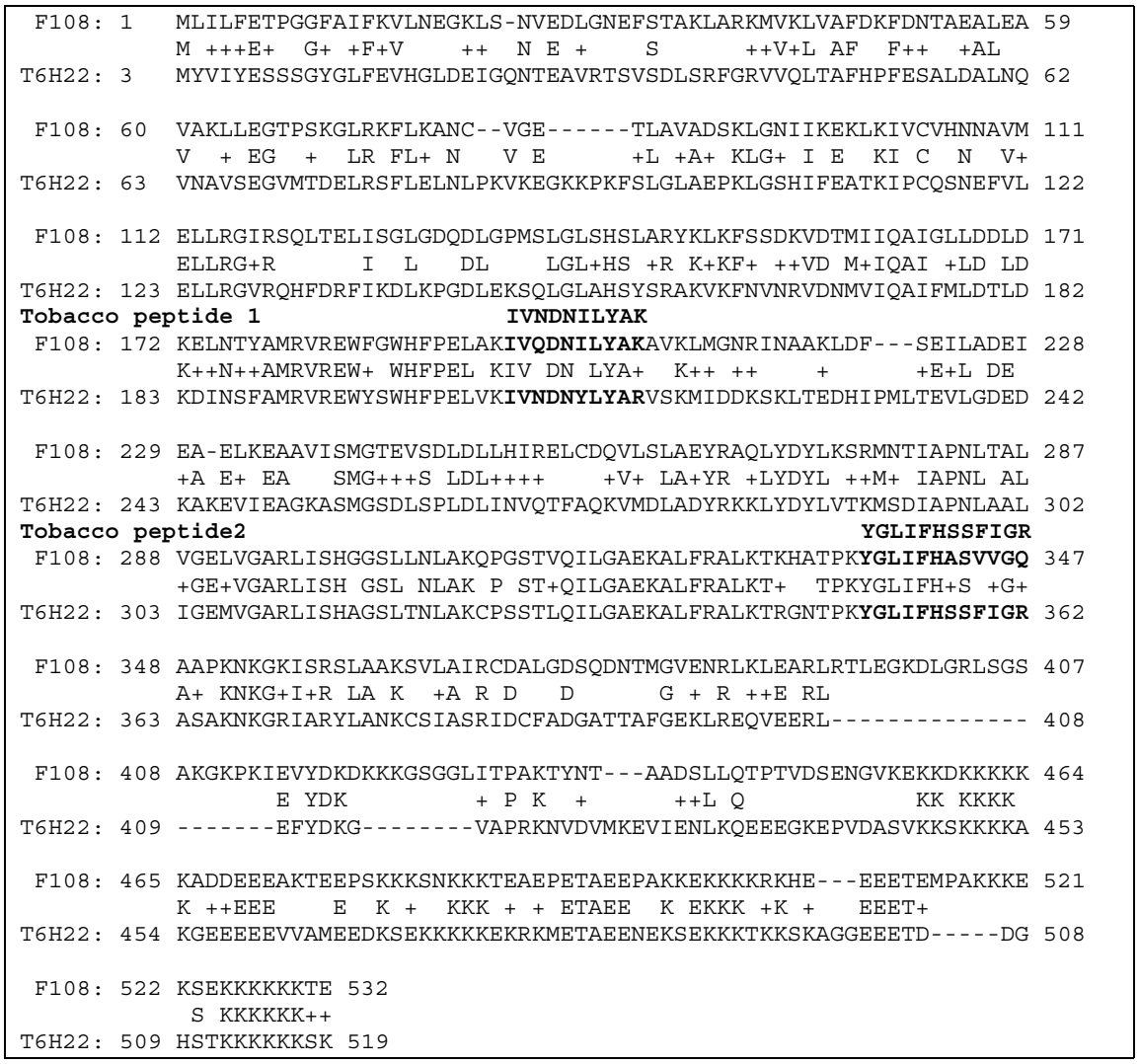

Figure 4-3B. Alignment of Arabidopsis NOP58 (F108) and NOP56 (T6H22). Sequences of the proteins deduced from cDNA sequences F108 and T6H22 were aligned using the BLAST local alignment tool (Tatiana and Madden, 1999). Proteins have a predicted length of 533 and 522 amino acids, respectively. . Regions homologous to the tobacco peptides sequenced are shown in bold. Identical amino acids are shown in the line between the two homologues, and similar (positives) amino acids are denoted by a "+". Identities = 234/551 (42%), Positives = 324/551 (58%), Gaps = 53/551 (9%). A similar level of homology is seen between T12H1 and T6H22.

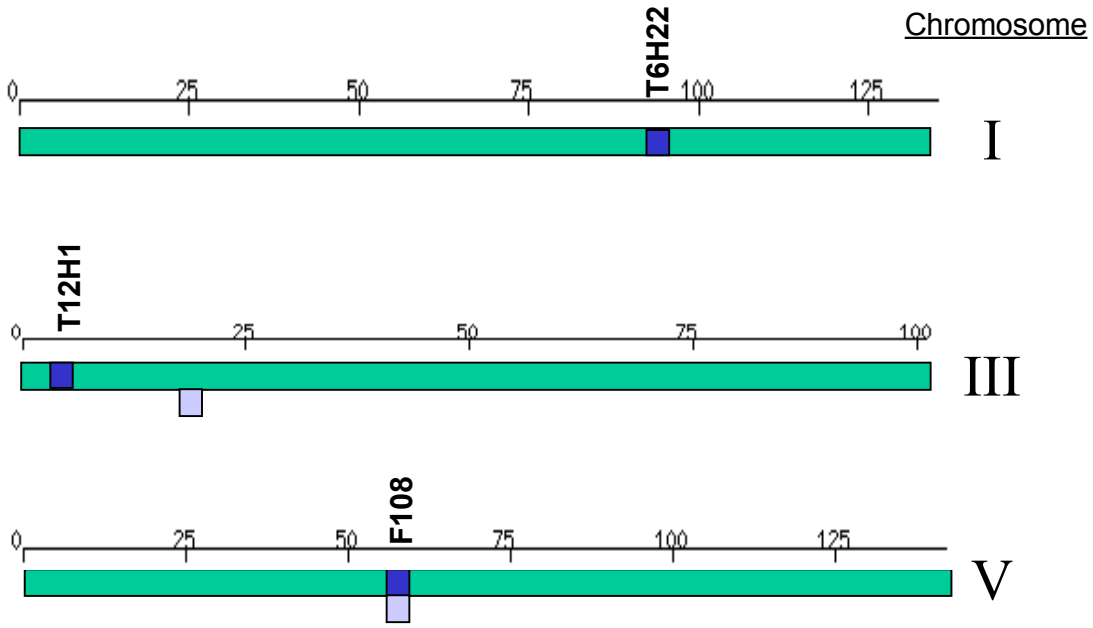

Figure 4-4. Locations of genes for the Arabidopsis NOP56/58 homologs in the Arabidopsis genome. Dark boxes indicate the location of BAC clones containing the genomic DNA coding for these genes. Genetic distances are denoted in centiMorgans above each chromosome map. Additional predicted genes on chromosome III (P1 clone MBK21) and chromosome V (BAC T21B4) were not detected in this work, and may not code for functional gene products.

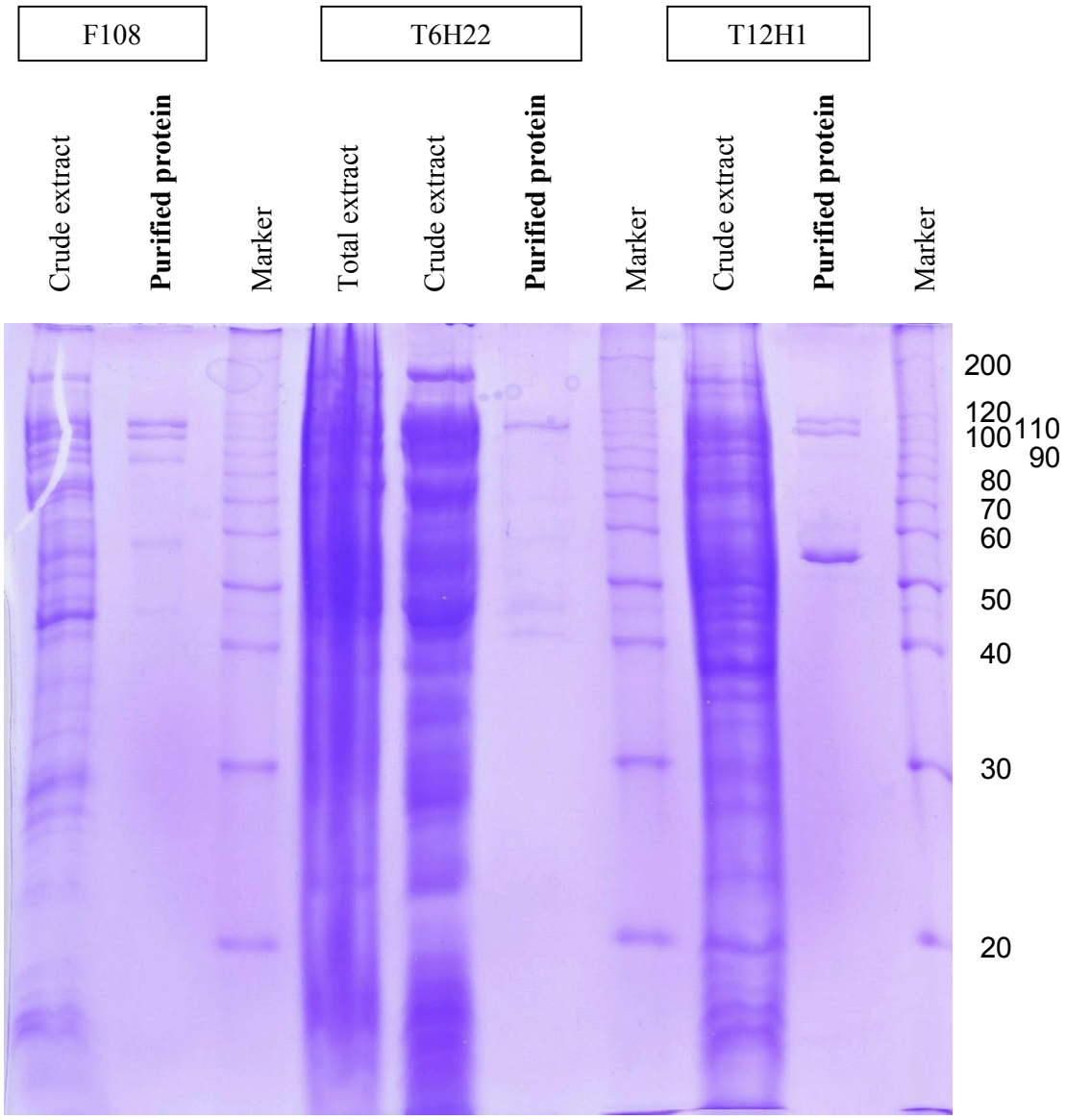

Figure 4-5. SDS polyacrylamide gel analysis of purified *E Coli* expressed fusion proteins. cDNAs for Arabidopsis proteins F108, T6H22, and T12H1 were cloned into the pMAL c2x expression vector. Proteins were purified on amylose columns using the maltose binding protein tag incorporated at the N terminus of the fusion proteins. Crude extract was loaded on amylose columns and eluted with column buffer plus 10mM maltose. Proteins were separated by electrophoresis on an SDS polyacrylamide (8%) gel. The recombinant proteins migrate at approximately 105 kDa.

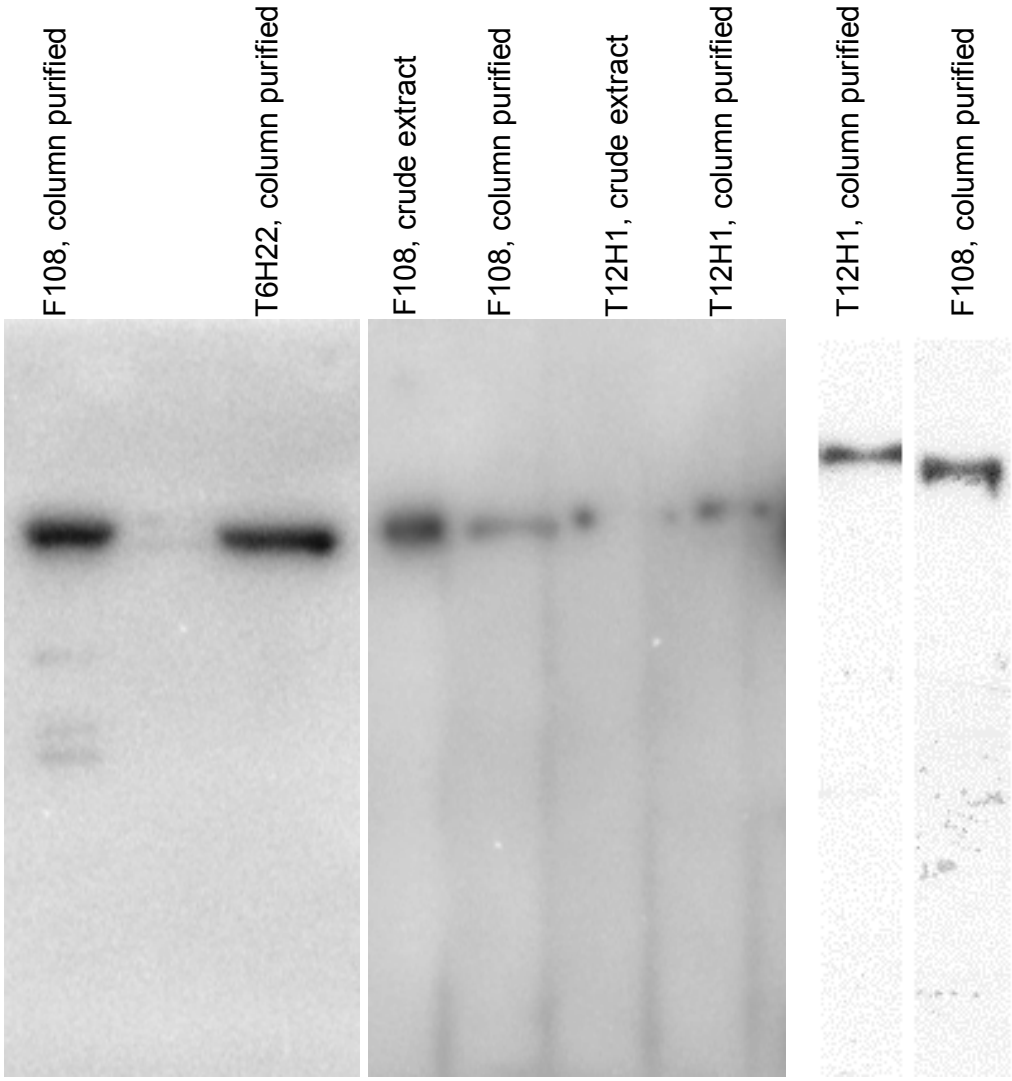

Figure 4-6. MAR binding activity of expression clones tested by Southwestern blotting. Expression clones judged positive for correct orientation of insert were grown in liquid culture and expression was induced with IPTG. Sonicated extracts were separated by electrophoresis through SDS polyacrylamide gels and transferred to PVDF membranes. The membranes were probed with 25 ng of end labeled RB7- MAR DNA in the presence of 200 fold *E. coli* competitor DNA. Results from three separate experiments are shown.

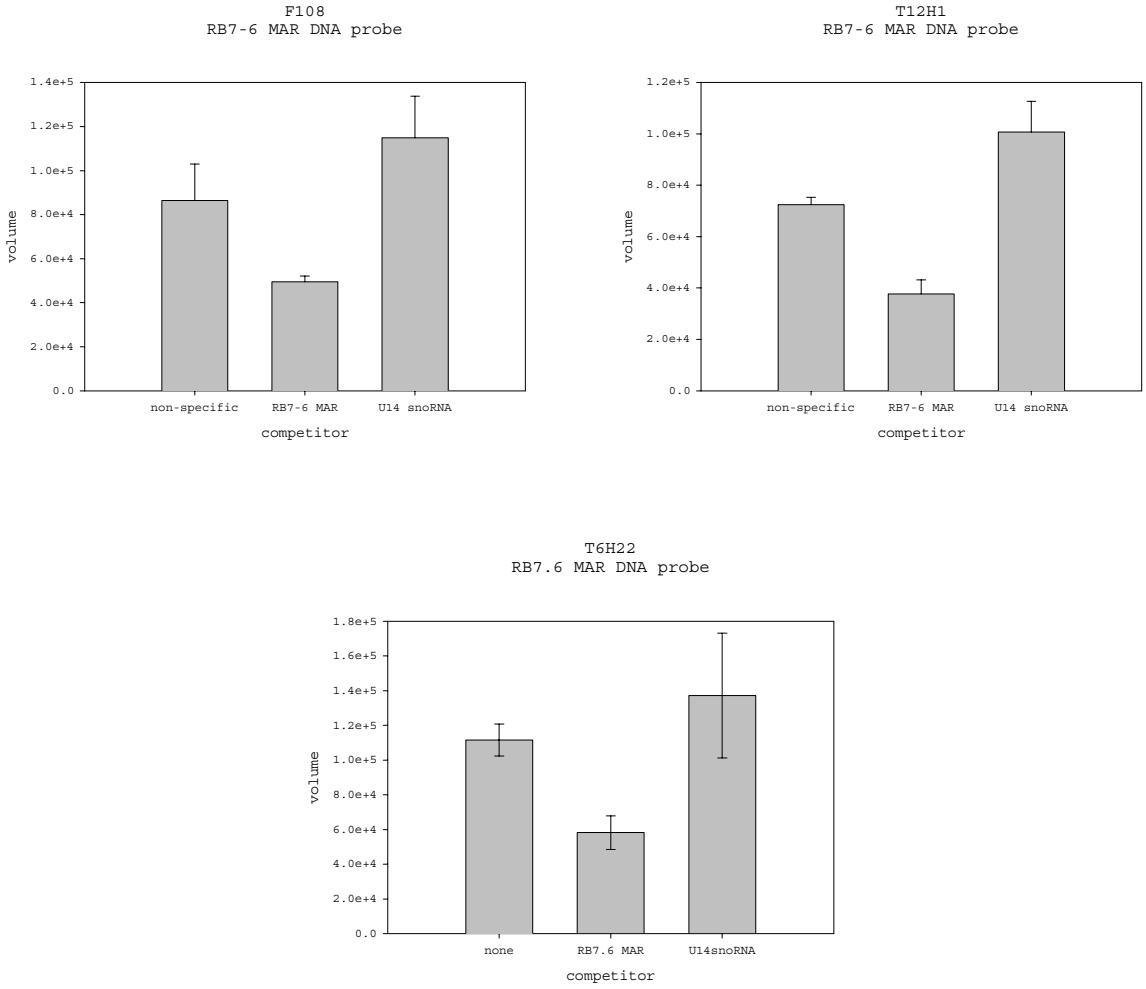

Figure 4-7. Arabidopsis NOP56/NOP58 homologs show increased binding to MAR DNA when C/D box snoRNA is present. Fusion proteins representing Arabidopsis homologs to mammalian NOP 58 (F108 and T12H1) and NOP 56 (T6H22) were bound to PVDF membranes by slot blotting. All binding reactions used 1 ng MAR DNA as probe and 1µg *E. coli* DNA as non-specific competitor. An increase in binding is seen whenever both of the nucleic acids were present together in the binding reaction. Specific competitors: RB7-6: 1 µg; snoRNA: 250 ng. Error bars represent the standard deviation (n=3).

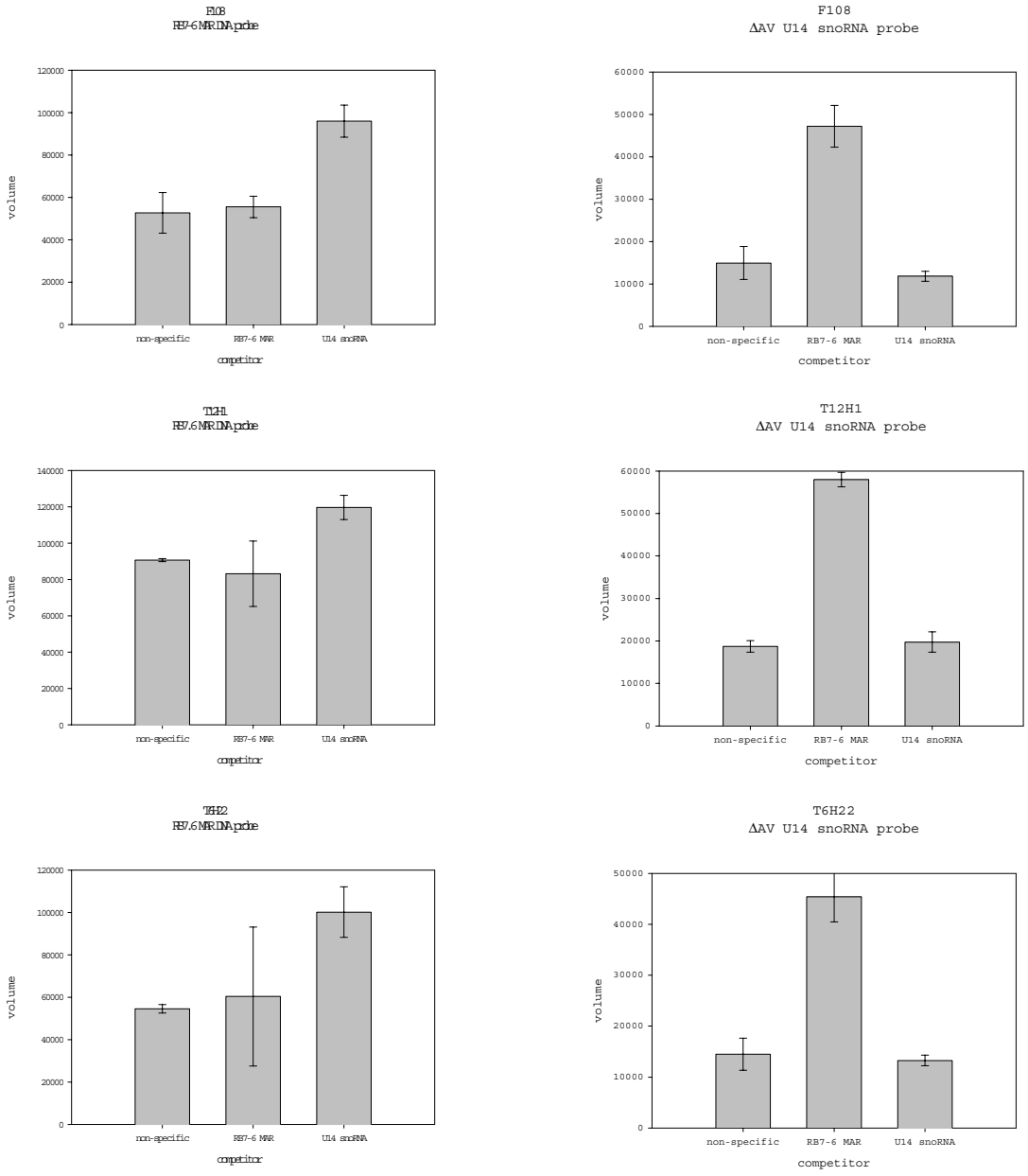

Figure 4-8. Arabidopsis NOP56/NOP58 homologs show increased binding when both C/D box snoRNA and MAR DNA are present. Fusion proteins representing Arabidopsis homologs to mammalian NOP 58 (F108 and T12H1) and NOP 56 (T6H22) were bound to PVDF membranes by slot blotting. Binding reactions used 1 ng RB7-6 MAR DNA (left column) or a mouse U14 C/D box snoRNA transcript as probe. An increase in binding was seen whenever both the nucleic acids were present together in the binding reaction. 1 µg *E. coli* DNA was included in each reaction. Error bars represent the standard deviation (n=2).