Chapter Five

Summary and future prospects

The nucleus remains a complex physical entity whose internal workings remain obscure, mainly due to the difficulty in linking in vivo relationships to observable biochemical relationships. One of the most challenging problems is relating nuclear structure and organization to biochemical functions (Gonzalez-Melendi et al., 2000). A prime example of this problem is our inability to explain the positive effects on gene expression seen when MARs are used to flank transgenes. Numerous models have described how MARs might influence gene expression in vivo, and how these in vivo roles are reflected in transgenic experiments (reviewed in Allen et al., 2000). Presumably, these effects relate to the ability of the MAR to associate with the matrix, though this link has not been positively established. MAR sequences themselves are variable in length and composition, without strict sequence requirements (Michalowsk et al., 1999). This variability in sequence complicates comparisons of effects seen from different MARs in different organisms using different transformation techniques. Each MAR sequence may contain an unknown number of regulatory elements (Boulikas, 1995; Allen et al., 2000). In addition, since MARs are excised from their natural chomosomal location and re-introduced into new locations, in transgenic

experiments they are most likely separated from any native cis-interacting factors that may influence their native contribution to nuclear regulation.

To better understand the interactions that underlay MAR effects, this work attempted to identify plant matrix proteins that interact with MAR DNA. The experiments presented here are based on a working hypothesis that MARs associate with the nuclear matrix to form chromatin loops. We further hypothesize that these attachments delineate separate chromatin domains, which may represent separately regulated units. I wished to identify protein components of the matrix that might mediate the association with DNA. By identification of the protein constituents of the hypothesized interaction, I hoped to provide a basis for understanding regulatory properties ascribed to MAR DNA.

I have analyzed the protein composition of plant matrices, comparing the matrix structures revealed by a sodium chloride based protocol with those produced using a lithium diiodosalicylate protocol. These protocols represent the most common methods of matrix preparation, and use theoretically unrelated means to extract loosely bound nuclear protein from the matrix. The sodium chloride method probably frees proteins by disrupting ionic interactions with high salt concentration. Lithium diiodosalicylate acts as a detergent to disrupt molecular interactions. Both of these reagents disrupt electrostatic interactions, but their dissimilar nature would be expected to effect individual interactions to differing degrees. If the same matrix proteins are preserved by these two protocols then this evidence supports the idea that these proteins represent an authentic nuclear structure. Experimental results presented

show that matrices prepared using either of these methods generate similar sets of proteins as seen on denaturing SDS polyacrylamide gel separations. Some differences in protein composition can be seen, but the overall gel banding patterns show a large and complex set of proteins. Dozens of proteins are present, and only a few appear to have significant differences in abundance between the two matrix preparation protocols.

DNA protein blotting experiments were conducted to identify protein bands from electrophoretically separated nuclear matrices with specific affinity for labeled MAR DNA probes. Given the insoluble nature of the nuclear matrix, and also given the large size of MAR DNA fragments, the DNA protein blotting experiments seemed to offer the best opportunity to identify MAR binding matrix proteins. Three or more strong bands were routinely seen on protein blots prepared from tobacco matrices. These bands interacted with both the RB7-6 MAR (Hall et al., 1991) and the S116-1.1B MAR (Michalowski et al., 1999). Non-MAR control DNA did not interact with these same bands.

Using limited amino acid sequence from the lowest apparent molecular weight band from the DNA protein blotting experiments, sequence database searches identified this protein as similar to proteins characterized as nucleolar proteins in yeast and mammals. I have cloned and expressed three related members of this nucleic acid binding protein family from the model plant Arabidopsis thaliana. These proteins appear to be evolutionarily conserved, with homologues in animal, insect, yeast and other organisms. The homologues are involved in vivo in processing of ribosomal

RNA, and may have multiple functions. In vitro binding studies have demonstrated an unexpected additive binding effect between these proteins, MAR DNA and snoRNA. The binding experiments presented here were limited in their scope, and it remains to be determined whether the effects seen here in vitro represent an interaction that occurs in vivo.

One of the Arabidopsis proteins identified (T6H22) is highly similar to NOP56, a protein first identified (and named SIK1) in yeast due to its ability to interfere with an artificial DNA binding construct (Morin et al., 1995). The mechanism by which NOP56 interfered with the DNA binding of their construct remains undetermined. Subsequent work implicating this protein in ribosome biogenesis may have shifted attention from NOP56 as a potential DNA interacting factor. The additive binding effect between the Arabidopsis NOP56 homolog, MAR DNA and snoRNA may be related to the effect seen by Morin et al.

Later reports have shown NOP56 homologs to be involved in ribosome biogenesis. Yeast NOP56 (a.k.a. SIK1) was identified by silent lethal interaction with yeast NOP1 (yeast homolog of fibrillarin) (Gautier et al., 1997). Temperature sensitive mutants of NOP56 are defective in processing pre-ribosomal RNA. Pulse-chase labeling, primer extension, and northern blotting experiments demonstrated that the 35S pre-rRNA is not efficiently cleaved (Gautier et al., 1997; Lafontaine and Tollervey, 2000). However, steady state levels of mature rRNAs were observed at the non-permissive temperature and in genetic depletion studies, so the processing defects probably are not the direct cause of lethality (Gautier et al., 1997; Lafontaine and Tollervey, 2000).

Arabidopsis proteins T12H1 and F108 are NOP58 homologs. NOP58 (a.k.a. NOP5) homologs from yeast have also been shown to be involved in ribosome biogenesis. Genetic depletion of the yeast NOP58 homolog gave results similar to the NOP56 experiments above. Northern blotting experiments showed that pre-ribosomal RNA is not efficiently cleaved. More specifically, primer extension analysis showed that cleavage of the 35S pre-rRNA is inhibited at the 5' external transcribed spacer, one of the earliest cleavages in the processing of pre-rRNA (Wu et al., 1998). Additional NOP58 genetic depletion experiments have described the specific cleavage defects, which specifically effect the production of 18S rRNA. Genetic depletion of the yeast NOP58 also has been shown to reduce the amounts of all tested box C/D snoRNAs, though the loss occurs to differing degrees in different box C/D snoRNA species. H/ACA snoRNAs are not co-depleted (Lafontaine and Tollervey, 1999).

In yeast, each of NOP56 and NOP58 are essential for viability (Gautier et al., 1997; Wu et al., 1997). Each has been shown by immunoprecipitation to associate with NOP1, the yeast fibrillarin homolog. NOP1 protein is also essential for viability (Venema and Tollervey, 1999), and has been shown to associate with all tested box C/D snoRNAs (Lafontaine and Tollervey, 1999). NOP1/NOP56/NOP58 complexes have been observed by immunoprecipitation (Gautier et al., 1997). However, the link between lethality of NOP56 and NOP58 mutants and the reductions in snoRNA and rRNA is not fully understood (Gautier et al., 1997; Lafontaine and Tollervey, 1999).

Based on these yeast experiments and similar results with mammalian homologs (Watkins et al., 1998; Newman et al., 2000), it seems likely that the Arabidopsis homologs of these proteins would also be involved in pre-rRNA processing. In support of that hypothesis, this work includes evidence that all three of these proteins do interact with the mouse box C/D U14 snoRNA construct. These experiments also demonstrate an interaction with MAR DNA and most interestingly, an increased interaction when the proteins are incubated with both MAR DNA and box C/D RNA. Additional experiments will be required to demonstrate whether the interaction is specific to this RNA, or also seen with different RNAs. The RNA/DNA/protein interaction has not been reported previously. This interaction could provide a framework for explaining the "SIK1" effect of over expression of the NOP56 homolog (Morin et al., 1995) or the lethality seen in knockouts of NOP56 and NOP 58 (Gautier et al., 1997). Formation of a DNA/RNA protein complex might provide an organizational basis for the in vivo functions of regulatory factors.

An important aspect of the NOP56/NOP58 story is their involvement in the processing of box C/D snoRNA. Many box C/D snoRNAs are transcribed and excised from introns of mRNA. The splicing occurs in the nucleoplasm. The NOP56 and NOP58 proteins have been hypothesized to bind to the spliced immature snoRNA in the nucleoplasm, and to be subsequently transported to the nucleolus. Recently it was reported that the yeast spliceosomal protein Snu13p is bound to box C/D snoRNAs as part of the NOP1/NOP56/NOP58 core complex. A snoRNP complex of the human homologs of these proteins was also found in HeLa cell extracts. Snu13 is an essential gene, and genetic depletion causes a reduction in box C/D snoRNAs, followed by a

loss of bulk cellular RNA. The Snu13 protein is required to maintain the steady state levels of the majority of cellular RNAs, resulting in inhibition of cell growth. The reduction could be due to a block in RNA transcription or processing, and/or a stimulation of degradation (Watkins et al., 2000).

Mouse snoRNPs have been affinity purified with box C/D snoRNA. Homologs to NOP56 and NOP58 were identified as part of the snoRNP complex. In addition, two more proteins, p50 and p55 were observed in these complexes. The mouse p50 and p55 were identified as closely related proteins that are localized in the nucleoplasm (Newman et al., 2000), and they may act as a heterodimer (Kanemaki et al., 1999). Homologs to p55 from rat have previously been identified as Tip49, a TBP interacting protein with DNA helicase activity (Kanemaki et al., 1999). The human homolog has also been identified as a target of the ß-catenin signaling protein (Bauer et al., 1998). The human homolog was separately identified as a ubiquitous nuclear matrix protein, NMP238 (Holzmann et al., 1998). Thus, these snoRNA binding proteins appear to be nuclear matrix proteins that are also involved in regulating gene expression.

The fact that snoRNA interacting proteins have been identified as nuclear matrix proteins involved in regulation of gene expression adds credibility to the notion that the NOP56/58 proteins have functions beyond their interactions with snoRNAs. It may turn out that the Arabidopsis homologs of NOP56 and NOP58 are solely involved in ribosome biogenesis. MAR NOP56 interaction were seen in this work and the NOP58 interaction was seen here and in similar experiments using MAR DNA and matrix protein from pea (Hatton and Gray, 1999). However, the additive binding seen

when snoRNA and MAR DNA are present with these proteins argues for a more complex role for these proteins.

Looking to the future, it would be useful to determine the specific binding sites on each of the proteins, on the MAR DNA, and on the snoRNA. Identification of separate protein binding sites for DNA and for RNA would further support the idea of a multifunctional role for the proteins. Mutation of these binding domains, as well as mutational study of each of the entire proteins, could allow us to explain the lethal effects of knockouts. Lethality of the yeast knockouts cannot be accounted for by the negative effects of gene depletion on pre-rRNA processing.

It would also be useful to identify additional proteins that interact with the Arabidopsis NOP56/58 proteins. These proteins might be identified by immunoprecipitation, by yeast two hybrid screens, or by other methods. Identification of binding partners might require the presence of MAR DNA, box C/D snoRNA, or both.

The ability of the pea and Arabidopsis NOP56/58 homologs to be expressed in vitro may be an important difference between the plant proteins and the yeast homologs. To date, the NOP56/58 homologs from yeast have not been stably expressed in vitro (J. Kuhn, personal communication). To test for inherent differences between yeast and Arabidopsis proteins, one could attempt to complement yeast mutants with the Arabidopsis homologs. The full set of Arabidopsis NOP56/58 proteins produced in this work could be a valuable tool for future investigations into the functional roles of these proteins.

The choice of Arabidopsis as the source for cloning these genes also opens the opportunity to use the genomic tools generated for use with this model plant, including collections of Arabidopsis knockouts. Analysis of the Cereon Genomics Single Nucleotide Polymorphism and small insertion/deletion collections (www.cereon.com) may identify important coding regions for these genes. Analysis of the regulation of these genes can also be facilitated by the existence of genomic data for Arabidopsis as well as for human, yeast and other organisms with homologous NOP56/58 genes.

Lastly, an interesting use for these proteins has also been suggested: their affinity for MAR DNA might be used to isolate large numbers of MAR DNA fragments (B. Ülker, personal communication). For example, the proteins could be bound to a solid support and incubated with genomic DNA fragments. Fragments preferentially bound to the proteins could be released, cloned and confirmed as MARs by the exogenous MAR binding assay.