1. Grant Specifications

Title: Regulation of Embryonic Development in Higher Plants
Number: DE-FG02-03ER15392
Dates: July 1, 2003 - December 31, 2010

2. Objectives

The overall goal of the project was to define the cellular processes that underlie embryo development in plants at a mechanistic level. Embryo development can be divided into two distinct but temporally overlapping phases. During the morphogenesis phase that occurs early in seed development, the basic body plan of the embryo is established, and the rudimentary tissue and organ systems are generated. The maturation phase occurs late in seed development. During this time, storage macromolecules accumulate in vast amounts, and the embryo acquires the ability to withstand desiccation. Our studies focused on a critical regulator of embryo development, Arabidopsis LEAFY COTYLEDON (LEC1). Loss-of-function mutations show that *LEC1* is required for several critical aspects of embryo development, including the maturation phase. Ectopic expression of LEC1 confers maturation characteristics to seedlings and induces somatic embryogenesis in vegetative cells. Therefore, LEC1 is a central regulator of embryogenesis sufficient to cause cells to alter their fate and enter into an embryonic program of development.

The three major objectives of the grant were as follows. First, we proposed to determine how LEC1 regulates the transcription of genes that promote embryo development at a mechanistic level. LEC1 is a novel HAP3 subunit of CCAAT binding transcription factor (CBF, aka NF-Y) that is found primarily in seed plants. Our working hypothesis was that the CBF containing LEC1 regulates the transcription of genes that promote distinct aspects of embryo development. Second, we proposed to identify target genes directly regulated by the LEC1 CBF and determine their function. Given the sufficiency of LEC1 to induce embryo development, some of these target genes are likely to be regulators of embryo development. Analyzing target genes that encode transcription factors may provide insights into regulatory circuits controlling embryo development. Third, given that LEC1 is a plant-specific isoform of HAP3 subunits, we asked when a LEC1-type subunit first appeared during evolution of the land plants.

We accomplished these major objectives during the tenure of the grant. Our findings are summarized below.

3. How Does LEC1 Regulate Genes Transcriptionally at a Mechanistic Level

Our phylogenetic analyses showed previously that seed plants have two types of HAP3 subunits: the LEC1 type found primarily in seeds of higher plants and non-LEC1 type (NLH3) most closely related to HAP3 subunits found in virtually all other eukaryotes. In addition to structural differences between the two types of subunits, they differ functionally in that ectopic expression of LEC1-type subunits but not NLH3-type subunits induces embryonic characteristics in vegetative tissues. We hypothesized that CBF complexes containing LEC1 regulates the transcription of genes required for embryo development whereas those possessing NLH3

subunits do not.

We analyzed CBF complexes containing LEC1 versus NLH3 subunits to understand the basis for their different functions. We first showed that both types of HAP3 subunits form CBF complexes in vitro and in yeast. This result suggested that LEC1 and NLH3 subunits are able to interact with the other CBF subunits, HAP2 and HAP5, to form complexes.

We compared the DNA binding specificity of CBF complexes containing LEC1 and NLH3 subunits and showed that both the LEC1 CBF and the NLH3 CBF bind DNA sequences *in vitro* that correspond to the consensus CBF binding sites in mammals. We compared DNA binding interactions of the two complexes by measuring the equilibrium dissociation constants for DNA binding to the complexes. Similar K_d values for LEC1 CBF and the NLH3 CBFs were measured using two different CCAAT-containing DNA fragment. These experiments suggest that the LEC1 CBF and NLH3 CBF do not exhibit significant differences in the specificity or relative affinity of binding with the CCAAT DNA motif in vitro.

We next explored interactions between the LEC1 CBF and NLH3 CBFs with DNA by measuring their rate constants for dissociation (k_{off}) in off-time electrophoretic mobility shift assays experiments. We found that the CCAAT DNA probe was competed much more rapidly from LEC1-CBF than NLH3-CBF. The calculated k_{off} indicated that, although the DNA binding affinities of the LEC1 and NLH3 CBFs are similar, the half-life of the LEC1 CBF-DNA complex is significantly shorter than that of the NLH3-CBF complex.

We analyzed DNA conformational change induced by LEC1 CBF and NLH3 CBF complexes. We showed previously that the LEC1 B domain contains an Asp residue at position 28 that is essential for LEC1 function and that differs from a conserved Lys residue that is present in the B domains of NLH3 subunits. This Lys residue lies in the L1 loop of the histone-fold motif that has been implicated to make contact with DNA and potentially mediate bending of the DNA. Given differences in the off rates for LEC1 CBF and NLH3 CBFs, we performed circular permutation assays to determine if the two complexes bend DNA differently. The experimental data showed that both the LEC1 CBF and NLH3 CBF caused DNA bending, although no appreciable difference in the bending angle was detected. These results suggest that LEC1 CBF and NLH3 CBF do not bend DNA to significantly different extents.

One key to resolving questions about functional differences between the LEC1-CBF and NLH3-CBF complexes is to compare DNA sequences bound by the two complexes in planta. We conducted chromatin immunoprecipitation coupled with DNA tilling array hybridization (ChIP-chip) experiments to define all of the LEC1 binding sites in the Arabidopsis genome. We used transgenic plants containing LEC1 tagged with the FLAG peptide and fused with the steroid binding domain of glucocorticoid receptor (*35S:FLAG-LEC1-GR*) or *35S:FLAG-NLH3*. ChIP experiments were done on *35S:FLAG-NLH3* seedlings grown for eight days and on eight-day *35S:FLAG-LEC1-GR* seedlings treated with dexamethsone, an inducer of LEC1 activity, for 4 h. DNA immunoprecipitated with anti-FLAG antibodies or with non-specific anti-GST

antibodies as a control was amplified and hybridized with the Affymetrix Arabidopsis Tiling 1.0R Array GeneChip. We then analyzed the data using stringent statistical tests, and validation experiments using quantitative PCR suggested that this analytical method reported few false positive results.

Our analyses suggested that the LEC1 CBF and NLH3 CBF have shared and unique binding sites in planta. We currently estimate that there are approximately 2300 LEC1 CBF-specific sites, 400 NLH3 CBF-specific sites, and 1200 sites bound by both complexes. Our experiments using quantitative PCR to validate these results confirm the existence of these three types of binding sites in the genome. Although analyses of these data are continuing, the finding that LEC1 CBF-specific and NLH3 CBF-specific binding sites are detected in planta but not in vitro with reconstituted complexes may indicate that these CBFs interact with other transcription factors that influence DNA binding specificity. Experiments relevant to this point are described below.

4. What Genes are Directly Regulated by LEC1

The LEC1 CBF complex is putatively a transcription factor that serves as a central regulator of embryogenesis sufficient to induce embryo development. Therefore, identifying target genes transcriptionally regulated by the LEC1 CBF provides the opportunity to gain insight into the regulatory networks controlling embryo development. We defined LEC1 target genes as those whose activities are induced by LEC1 activity and are bound by LEC1 in the genome.

We used an inducible form of LEC1, LEC1 fused with the steroid binding domain of glucocorticoid receptor, LEC1-GR, to identify mRNAs upregulated and downregulated following a short term induction of LEC1 activity for one hour. We identified genes that displayed at least a two-fold, statistically significant (FDR < 0.05) change in mRNA levels following the induction of LEC1 activity. By these criteria, we identified 408 and 197 genes, respectively, that were upregulated and downregulated after LEC1 induction. Changes in RNA levels for representative LEC1-regulated genes were validated using quantitative RT-PCR experiments. We also reanalyzed a dataset of mRNAs induced four days after LEC1 induction (Mu et al. 2008. Plant Physiol 148:1042-1054) to compare target genes induced early and late following the induction of LEC1 activity. We showed that 527 and 417 mRNAs, respectively, were up and down-regulated by this long-term induction of LEC1 activity.

To identify genes bound by LEC1, we performed ChIP-chip experiments with *35S:FLAG-LEC1-GR* plants to identify genes bound by LEC1 following a short-term (4 hour) induction of LEC1 activity as discusssed above. ChIP-chip experiments were also done following a long-term induction (8 days) of LEC1 activitiy. Immunoprecipitated DNAs were amplified and hybridized with the Affymetrix Arabidopsis Tiling 1.0R Array GeneChip. Our statistical analyses of enriched DNA sequences showed that LEC1 was bound within 2 kb upstream of 3,838 and 6,489 genes, respectively, 4 hours (short-term induction) and 8 days (long-term induction) following LEC1 induction.

Integration of the databases of RNAs that are induced by LEC1 activity and genes that are bound by LEC1 permitted the identification of genes putatively directly regulated by LEC1. We identified 103 and 52 putative target genes, respectively, that were upregulated and downregulated following a short term induction of LEC1 activity, and 236 and 73 target genes, respectively that were upregulated and downregulated following a long-term induction. Surprisingly, there was almost no overlap in the target genes that were identified following short-term and long-term LEC1 induction. Only two target genes were in common among those upregulated following short-term or long-term induction of LEC1 activity. Similarly, only two target genes were in common among those downregulated following short-term or long-term induction. We showed that genes directly regulated by LEC1 following long-term induction were enriched for Gene Ontology terms related to the processes that occur during seed maturation but the short-term induction target genes were not. Together, these results suggest that LEC1 regulates different sets of genes following short-term and long-term induction of activity and that target genes following long-term induction are involved in maturation processes.

To obtain insight into the mechanisms that mediate this shift in target gene selection by LEC1, we compared DNA sequences bound by LEC1 following short-term and long-term induction of LEC1 activity. We analyzed regions bound by LEC1 in the ChIP experiments and identified enriched DNA sequence motifs. Surprisingly, the CCAAT DNA sequence motif, which we showed previously to be the binding site for the LEC1 CBF in vitro, was not overrepresented among DNA fragments bound by LEC1 compared with randomly selected DNA sequences. Rather, several DNA sequence motifs containing a G box core DNA sequence (CACGTG), including ABF, ABRE, and DPBF1&2, were significantly enriched among the DNA sequences bound by LEC1 ($p < 0.0001$). Furthermore, de novo motif discovery analyses identified G box DNA sequence motifs in the fragments bound by LEC1 both in short-term and long-term induction studies.

We developed a working model to explain why LEC1 has different target genes following short-term and long-term induction and how LEC1 regulates the maturation phase. We propose that LEC1 is able to interact with different bZIP transcription factors and that each bZIP targets LEC1 to different sets of target genes. Consistent with these hypotheses, Yamamoto et al. (2009 Plant J. 58: 843-856) reported that LEC1 interacts with bZIP67 that binds DNA motifs containing the G box core element. Furthermore, ectopic expression of LEC1 and bZIP67 caused activation of a 12S storage protein gene, a gene that is normally expressed during the maturation phase. Thus, we hypothesize that LEC1 interacts with a yet undetermined bZIP transcription factor early in seed development to regulate a set of target genes. Immediately prior to the initiation of the maturation phase, bZIP67 accumulates and interacts with LEC1, and together they regulate genes involved in the maturation phase. Experiments to test this working model are ongoing.

5. When did LEC1 First Appear Among Land Plants?

We showed previously that LEC1 is the founding member of a novel class of HAP3 subunit that

accumulates primarily in seeds and that specific amino acid residues in the conserved B domain could differentiate LEC1-type and NLH3-type HAP3 subunits. We exploited plant genome sequences to determine when in land plant evolution a LEC1-type HAP3 subunit was first observed. We searched DNA sequencing reads and identified a gene encoding a LEC1-type subunit in the lycophyte, *Selaginella moellendorffii*, but we were unable to identify a homolog in the bryophyte, *Physcomitrella patens*. The *SmLEC1* suppressed the Arabidopsis *lec1* mutation when expressed under the control of the endogenous Arabidopsis promoter. This result suggests that *SmLEC1* is orthologous with *LEC1*. Moreover, we showed that *SmLEC1* RNA accumulated at highest levels in bulbils, shoot apices that had ceased elongation, undergone thickening and accumulated lipid reserves. This finding opens the possibility of a common function of SmLEC1 and Arabidopsis LEC1 in lipid biosynthesis.