Molecular Characterization of a Maize Regulatory Gene

Progress Report

Dr. Susan Wessler

University of Georgia Research Foundation
Athens, Georgia 30602-4222

July 1989 through 1990

PREPARED FOR THE
U.S. DEPARTMENT OF ENERGY
UNDER CONTRACT DE-FG09-86ER13621

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Progress Report for Grant #DE-FG09-86ER13621
Report covers July 1989 through March 1990

Progress Report

The cloning and preliminary characterization of the R-nj allele is described in the reprint of Dellaporta et al., 1988. Most of this progress report represents a summary of two additional manuscripts, Ludwig, et al. 1989, PNAS, in press, describes our characterization of the Lc gene and Ludwig et al., 1989, (in preparation for Science) describes a transient assay for R function using the particle gun (in collaboration with Ben Bowen and Larry Beach of Pioneer Hi-bred).

A. Cloning of R locus sequences

R locus sequences were cloned by transposon tagging from a strain harboring the Ac R-nj:m1 allele (13). An Ac element, unmethylated at its internal PvuII sites in the genome of strains harboring the Ac R-nj:m1 allele, was isolated on a 15 kb Bam fragment (Figure 1). Since there is a Bam site near one terminus of the Ac element (25), this clone was flanked on one side only by non-Ac sequences. A 700 bp BglII-HincII fragment was subcloned from these flanking sequences (Figure 1) and used as a probe in Southern analysis to confirm that this was in fact from the R locus. These analyses (reprint, Figures 4 and 5) demonstrated that (1) the probe is homologous with a genomic fragment that is 4.56 kb larger in the Ac R-nj:m1 strain than in strains harboring either the R-nj progenitor or a phenotypic revertant of the Ac R-nj:m-1 allele, (2) The P, S, and Lc genes are all homologous with the 700 bp probe and (3) hybridization with strains harboring the P+S, P+S+Lc, P or S genes is consistent with the genetic description of these strains. All of these results support the contention that the 700 bp fragment contains R locus sequences (analysis of R transcription presented below reveals that this probe is from the extreme 5' end of the R-nj gene. The HincII site is within intron 1 while the BglII site is in 5' flanking sequences).

Restriction maps of λ Ac-28, λ R-nj:1 and plasmid subclone. λ Ac-28 contains a 15 kb BamHI fragment with DNA homologous to *Ac* shown as the rectangle. The pAcH1.6 probe contains the internal 1.6 kb Hind III fragment of *Ac* that corresponds to the stippled region of the rectangle. A flanking 0.7 kb DNA segment from the Bgl II to Hinc II sites (black box) was subcloned into pUC119 and designated pR-nj:1.

B. The R gene encodes a 2.5 kb transcript

As a first step in understanding what an *R* gene encodes and to determine if transcripts from diverse *R* genes were similar, northern blot analysis was performed. Additionally, this analysis was used to identify the source of the most abundant *R* transcript for subsequent cDNA isolation.

To maximize our chances of identifying an *R* gene product, RNA was isolated from organs displaying high levels of anthocyanin pigmentation. To identify mRNA encoded by the *R* genes designated P, *Lc*, and S, poly (A)+ RNA was isolated from male spikelets, female spikelets or kernels, respectively and analyzed on northern blots. The blot was probed with the 0.7 kb BglII - HincII fragment isolated from the *R-nj* genomic clone (Fig. 1 above). The predominant transcript encoded by each gene is approximately 2.5 kb in length (Fig.2). The results of several other northern blots using various combinations of mutant *R* genes indicated that the transcript observed was from the *R* gene that is indicated. For example, strains harboring P but not S contained a 2.5 kb transcript in anthers but not aleurone, whereas strains with S but not P contained the transcript in aleurone and not anther samples.

Since the *R-nj* gene was already cloned, we wanted to isolate its cDNA, however, we were not able to detect an *R-nj* transcript in aleurone, roots or anther tissue. This is not surprising considering the amount of anthocyanin conditioned by this allele is almost 50 fold less than that conditioned by standard *R-r* (26).

kb
4.4 -

2.4 -

1.4 -

1 2 3

Fig. 2: Northern blot of 5ug poly A+ RNA from: (1) P, male spikelets; (2) *Lc*, female spikelets, 10 dap; (3) S, kernels 35 dap. Probed with pR-nj:1.

C. Cloning of the Lc gene and a full length cDNA

We decided to focus our analysis on the Lc gene which encodes a relatively abundant mRNA (Fig.2). However, first an Lc gene had to be identified and isolated. Southern blot analysis of genomic DNA, isolated from isogenic lines that differed by the presence or absence of the Lc gene, was used to identify a 3.7 kb HindIII fragment in the strain containing Lc (Fig. 3, lanes 1 and 2). This HindIII fragment was cloned and used to probe a cDNA library prepared from female spikelet poly(A)+ RNA (displayed in Fig 2, lane 2). Of 300,000 plaques screened, only one clone hybridized to the 3.7 kb Lc genomic fragment. This was purified, subcloned and sequenced. In addition, the 3.7 kb HindIII genomic fragment was sequenced and compared with the cDNA sequence. It was determined that the 3.7 kb genomic fragment contained only two small exons (228 bp and 159 bp in length) homologous with the 5' end of the cDNA (Figure 5). To identify the genomic fragment which encoded the rest of the transcription unit, the Southern blot displayed in Fig.3 was stripped and probed with the 3' end of the cDNA (position 880 to 1772, Fig. 4). This Southern blot revealed the presence of a 6 kb HindIII fragment specific to genomic DNA containing Lc (Fig. 3, lanes 3 and 4). This fragment was isolated from a genomic library containing size fractionated HindIII digested genomic DNA cloned into lambda ZAP.

Fig. 3: Southern blot analyses of isogenic lines with and without the Lc gene, digested with HindIII. With Lc, lanes 1,3; without Lc, lanes 2,4. Lanes 1,2 were probed with pR-nj:1; lanes 3,4 were probed with a 3' Lc cDNA probe (nucleotides 880 to 1772).

D. Characterization of the Lc transcription unit

Comparison of the sequence of the cDNA clone with the sequence of the two genomic fragments, revealed that the cDNA was almost certainly derived from Lc because the exonic sequences

were identical. The transcription unit was found to span approximately 7.5 kb with introns at positions 228-229, 387-388, 648-649, 745-746, 760-761, 817-818, 1566-1567 and 1996-1997 (Fig. 4, shown schematically in Fig. 5). All introns contained the consensus splice junction sequence (5' GT-- --AG 3').

Primer extension was used to determine the start of transcription (27). Two major bands were observed which correspond to positions one and three (fig. 4). An RNase protection experiment confirmed this as the start of transcription. The 5' end of the Lc cDNA was determined to be 20 bp from the transcription start site.

A putative TATA box (TATATATA) is correctly positioned upstream of the start of transcription at position -31 to -24 (Fig. 4). When translated the cDNA sequence contains a 610 amino acid open reading frame beginning with the fourth AUG at nucleotide 236 and ending with a stop codon at nucleotide 2066. The first three AUG's do not lead to a significant ORF (in experiments described below, we were able to delete the first three AUG's and still produce a functional protein). To determine if a protein could be translated from this cDNA, synthetic mRNA produced from the full length cDNA was translated in a rabbit reticulocyte system in the presence of ³⁵S-methionine. The 88 kd polypeptide synthesized in vitro is larger than the approximately 67 kd polypeptide predicted from the 610 amino acid open reading frame (Fig. 6). Similar discrepancies have been observed for several proteins synthesized in vitro, including the Ac transposase (28), GCN4 (1) and the regulatory proteins encoded by o2 (Robert Schmidt, personal communication) and c1 (Karen Cone, personal communication).

E. Features of the Lc encoded protein

The putative protein encoded by the 610 amino acid open reading frame has the features of transcriptional activators (29). It has a large acidic domain and a smaller basic domain (amino acids 188-318 and 415-508, respectively, Fig. 7A). The activating regions of transcriptional activating proteins frequently contain stretches of amino acids with a significant net negative charge (10,29). The Lc acidic domain contains 33 acidic and 8 basic amino acids for a net charge of -25.

To further characterize the putative Lc protein, the predicted protein sequence was analysed for homology to Genbank DNA translated in all three reading frames. The basic domain of the protein (amino acids 420 to 462) was found to have homology to the "myc homology region" (5) shared by the nuclear myc oncogenes [L-myc shown, (30)], a myoblast determination factor [MyoD1, (31)], daughterless [da, (32)], genes of the achaete-scute complex [gene T3 of AC-S shown, (33)], twist (34), and E12 (5) (Fig. 7B).

Fig. 7: Characteristics of the putative Lc protein. (a) average net charges over 30 amino acids, measured at 10 amino acid intervals. (b) Amino acid comparison of the helix-loop-helix myc homology region of the Lc protein (a.a. 420-462), L-myc (a.a. 297-334), MyoD1 (a.a. 118-161), da (a.a. 563-608), T3 of AS-C (a.a. 62-146), twist (a.a. 365-409) and E12 (a.a. 344-390). Identities that include Lc are shaded and a consensus sequence containing amino acids that occur in more than 50% of the protein is presented.

B.

Lc	SRKRREKLMEHIVLKSLLPSIR	VKKASIDAEIAYLKEIR
L-myc	EDKRRLDLSRFLARDHPTLASCSE	APKVVLSKALEYLGALV
MyoD1	RCRRRLSKVWCAFEYLRCTSSHPNR	LPEVEILRNHAIYIEGLQ
da	REAIRRDINEALKEELGRMOKHLKSECP	QTKLGLHQAIVEVINTLE
T3 AC-S	RLRNEVKOVNNGEVNLRQH(PQTVVNSLSNGGREGSKLSKYDTLRIVAYEYIRGLQ	
twist	REKORTOSLMDATKSLQGIITLPSDK	LSEIGTLKATRYIDFLC
E12	RELRVRDINEALKEELGRMOKHLKSECP	QTELLHQAIVSVILMLE
CONSENSUS	RER R M F L P	K IL A TI L

HELIX ----- LOOP ----- HELIX

Fig. 8: Northern blot of strains with and without a functional R gene. Lanes 1,3,5,7, RNA from kernels 35 dap with the S gene; lanes 2,4,6,8, without S. Probes: lanes 1,2, Lc cDNA; lanes 3,4, c2 (pc46E); lanes 5,6, al (pALC2); lanes 7,8, maize actin (pMAc1).

F. RNA Analysis of R, C2, and A1 mRNA

Previous biochemical studies indicated that an R gene was required for A1, C2, and Bz enzymatic activity (22,23,35). Because the sequence data indicated that Lc may encode a DNA binding-transcriptional activator, we investigated the possibility that R may influence C2 and A1 steady state transcript levels. To do this analysis correctly, the B-b allele present in all stocks was replaced with a null allele. This precaution was taken because of the homology between B and R (V. Chandler, S. Dellaporta, personal communication).

Poly(A)+ RNAs isolated from kernels (35 dap) containing or lacking the S gene were analysed by RNA blot hybridizations (Fig. 8). It was observed that S, C2, and A1 transcripts were present only in the RNA containing the S gene. The cloned maize actin gene (pMAc1) was used as a control and was present in equal amounts in both lanes (Fig. 8, lanes 7 and 8). Similar experiments using RNA from male spikelets containing or lacking the R gene P indicated that P must be present in order to detect C2 and A1 transcripts in anthers (data not shown).

G. Different R genes are highly conserved

We have sequenced over 9 kb of the Lc gene (Ludwig, Hebara and Wessler, unpublished), a full length Lc cDNA (9) and several kb of the R genes R-nj and P (Ludwig, Hebara and Wessler, unpublished). All comparisons indicate that the genes are remarkably similar in their exons, introns and even 5' flanking regions. For example, comparison of Lc and R-nj genomic sequences are as follows: 360 bp 5' flanking - 95%; exon 1, 2, intron 1 - 100%; and 1852 bp of the 4 kb intron 2 - 96%. Interestingly, many of the changes arise from short sequence duplications. We have hypothesized that most R genes encode functionally equivalent proteins and phenotypic differences reflect promoter diversity. If this is indeed the case then the subtle sequence differences between these two genes in the 5' flanking regions may be responsible for the altered tissue specificity. Alternatively, unsequenced regions more distant from the transcription start may influence expression.

Karen Cone has just sequenced a full length cDNA of S. A comparison with our sequence reveals over 96% homology of the DNA sequence. Virtually all of the changes are in the untranslated exon 1 - the protein coding region is almost identical. Finally, Vicki Chandler has just told us that her partial sequence of a B cDNA also indicates about 90% conservation of nucleotide sequence.

H. Introduction of the Lc cDNA into maize by microprojectiles:

From the above data, we propose that all or most R genes encode functionally equivalent proteins. Further, we propose that the diverse patterns of anthocyanin pigmentation conditioned

by different R genes probably reflects differences in the R promoters rather than their gene products (see the discussion section of the PNAS manuscript). To prove this, we have fused the full length Lc cDNA to the CMV 35S promoter and introduced this construct into a variety of maize tissues via the microprojectile technology (36). In most cases, a similar construct containing GUS rather than the Lc cDNA was co-precipitated on the microprojectiles with the Lc construct and served as a control. This work represents a collaboration with Larry Beach and Ben Bowen of Pioneer Hibrid. Our results to date are as follows:

1. The Lc construct can induce pigmentation in (a) aleurone [either immature or dried (imbibed for 24 hrs.)] (Fig. 9); (b) both inner and outer layers of immature pericarp (12-25 dap); (c) epidermal cells of coleorhiza; (d) scutellar node (Fig. 9); (e) mesocotyl (Fig 9); (f) coleoptile (Fig. 9); (g) leaf blade; (h) leaf sheath; (i) tassel glume; (j) husk; (k) roots (Fig. 9); (l) hairs or trichomes on leaf blade margins (Fig. 9).
2. Pigmentation is first visible within 12 hrs after shooting, it plateaus at about 36 to 48 hrs and is stable for a least 3 weeks.
3. No pigmentation has been seen in the inner endosperm (GUS controls were positive). There are no R genes that pigment this tissue either.
4. Some red cells have been seen in vascular tissue below the leaf epidermis, indicating particle penetration several cell layers deep.
5. Red cells are most frequently single and non-adjacent. However, most of the tissues shot are no longer dividing. Clusters of up to six adjacent cells have been seen in tunica tissues in which cells would be expected to be still dividing in the interval after shooting (Fig. 9)
6. Most expression is probably transient. This can be deduced from the following experiment. When W22::Stadler seedling tissue (C r-g pl B-b) was bombarded, expression was shown to be light dependent. Red cells were only seen in tissue that had been grown in the light, bombarded and maintained in the light. No red cells were seen when tissue was grown in the dark, bombarded then put in the light or when tissue was grown in the dark, bombarded then put in the light. Consistent with the genetics, strains harboring Pl showed light independence.
7. Unlike GUS expression, we always see discrete pigmented cells (GUS usually produces diffused patches). This may be because Lc does not act directly, presumably the Lc protein is transported into the nucleus where it induces the other genes in the pathway. Maybe only healthy cells are capable of this.

8. Pigmentation as a marker has several advantages over GUS. No stain is required to observe Lc+ phenotypes. It may be possible to observe lineages into the progeny.

9. By decreasing the concentration of the Lc construct we see a proportional increase in the number of spots. Since taking the photographs, we now have aleurones with over 400 pigment spots and only a single GUS spot. Experiments are now underway to see when we "bottem out". Apparently too much R protein inhibits induction.

Although there are many directions this work can (and will) go, the immediate significance of these findings for this proposal are two-fold. Firstly, it provides support for the notion that R proteins are functionally equivalent. The Lc gene does not condition aleurone pigmentation. Yet, by fusing Lc cDNA to the 35S promoter, it is capable of inducing aleurone pigmentation. Second, the success of these experiments provides a very simple transient assay for either natural or constructed R mutations. The relative ease of this assay coupled with literally overnight results are important prerequisites to the proposed experiments involving a structure-function analysis of the R protein.

Fig. 9: W22::Stadler r-g tissue was shot in a-f. (a) imbedded kernel, germinated a few days, expression in scutellar node and mesocotyl. (b) aleurone expression, damaged cells are golden. (c) expression in coleoptile, possible dividing cells. (d) root cell. (e) leaf margin hair.

a

b

c

d

e