Sequence elements critical for efficient RNA editing of a tobacco chloroplast transcript \textit{in vivo} and \textit{in vitro}

Michael L. Hayes, Martha L. Reed, Carla E. Hegeman and Maureen R. Hanson*

Department of Molecular Biology and Genetics, Cornell University, Biotechnology Building, Ithaca, NY, 14853, USA

Received March 22, 2006; Revised June 7, 2006; Accepted June 19, 2006

\textbf{ABSTRACT}

In tobacco chloroplast transcripts 34 nt are efficiently edited to U. No common consensus region is present around all editing sites; however, sites can be grouped in clusters that share short common sequences. Transgene transcripts carrying either the wild-type −31/+22 or −31/+60 sequence near NTrpoB C473, an editing site within tobacco \textit{rpoB} transcripts, or three different mutated sequences, were all highly edited \textit{in vivo}. Endogenous transcripts of \textit{rpoB}, \textit{psbL} and \textit{rps14}, all of which contain common sequences S1, S2 and S3 5′ to NTrpoB C473, NTpsbL C2 and NTrps14 C80, were less edited in transgenic plants that over-express transcripts from NTrpoB C473 transgenes. Extent of reduction of endogenous editing differed between transgenic lines expressing mutated −31/+22 regions, depending on the abundance of the transgene transcripts. The −20/−5 sequence contains critical 5′ sequence elements. Synthetic RNA templates with alterations within this 5′ region were less efficiently edited \textit{in vitro} than wild-type templates, by either tobacco or maize chloroplast extracts. The tobacco chloroplast extract supports both RNA editing and processing of 3′ transcript termini. We conclude that within the −20/−5 region, sequences common to editing sites in the transcripts of \textit{rpoB}, \textit{psbL} and \textit{rps14} are critical for efficient NTrpoB C473 editing.

\textbf{INTRODUCTION}

RNA editing, a form of RNA processing, occurs in both nuclear and organelle transcripts of diverse organisms. In vascular plants, ∼30 C targets of editing typically exist in chloroplast transcripts, while >400 such targets have been observed in plant mitochondria (1–4). Tobacco chloroplasts are a particularly good model system for editing because deliberate alteration in editing substrates can be assayed \textit{in vivo} in chloroplast transgenic plants or \textit{in vitro} with chloroplast extracts (5–7). Tobacco chloroplasts have 34 known editing sites within the 155 939 bp organelle genome, and all are modifications from cytidine to uridine. Of the 34 identified editing sites, 32 sites are known to be efficiently edited, with 70–100% of transcripts modified from C-to-U at each edited position (8). In vascular plant chloroplasts, identified editing sites are almost exclusively within coding regions and occur most frequently at the second position of a codon (2). Editing of all but one of the C targets identified in tobacco results in a change of encoded amino acid (2,8). Edited codons commonly encode amino acids that are conserved among orthologous proteins of other plants (9–11). Defects in editing of some transcripts result in plants with severe phenotypes, producing dysfunctional proteins (12–15). Editing in plant organelles is likely a mechanism for the correction of genomic T-to-C mutations rather than for creation of protein diversity (16,17). The low number of C-to-U modifications, accompanied by the high extent of editing of C targets, suggests the existence of a highly efficient and specific editing mechanism within chloroplasts. \textit{In vivo} and \textit{in vitro} editing studies have focused on the sequence elements responsible for directing editing in chloroplasts as well as in mitochondria. Both organelles edit C-to-U and may share similar mechanisms for editing (18). Regions critical for editing are primarily located in nearby regions 5′ of the editing sites, and \textit{in vivo} studies have identified a number of editable substrates that carry only 20–40 nt 5′ and 10–20 nt 3′ to the C editing target.

The editing site NTrpoB C473 is within the tobacco \textit{rpoB} transcript; the C at position 473 from the A of the initiation codon is edited. This editing site has been referred to as \textit{rpoB}-2 previously, but because of the number of additional species in which chloroplast editing has been characterized, a previous nomenclature system (2) for chloroplast C editing sites has become unwieldy. We propose here to name editing sites by initials of genus and species, gene name, then nucleotides from the A of the closest gene’s initiation codon. C473 in \textit{tobacco} \textit{rpoB} is at the second position of the codon, altering the encoded amino acid from serine to leucine. The \textit{cis}-requirements for editing for NTrpoB C473 were examined previously \textit{in vivo} by expressing transgenes...
carrying a small portion of the \(\text{rpoB}\) gene surrounding the C editing target (5,19–21). A 27 nt sequence flanking the editing site NTrpoB C473 is sufficient for editing; however, a more highly edited template contained 92 nt around the editing site. The lower amount of editing observed in the smaller 27 nt sequence flanking the editing site \(\text{NTpsbL C2}\) is sufficient for editing; however, a template containing a region surrounding the C to better define the important \(\text{cis}\)-acting region in \textit{vivo} and \textit{in vitro}.

Although no consensus sequence is common to all editing sites, groups of sites with common sequences can be gathered into clusters of sites that may also share sequence-dependent specificity factors. Over-expression of sequences flanking NTrpoB C473 or NTndhF C290 in tobacco chloroplasts results in a reduction in editing of a small group of endogenous editing sites which contain some short common sequence elements (20). All of the known editing sites within tobacco can be grouped into clusters based on short common sequences. Some of these clusters exhibit similar changes in efficiency of editing depending on tissue type (8). Upon over-expression of a template containing a region surrounding NTrpoB C473 in tobacco chloroplast transgenic plants, endogenous \(\text{rpoB}\), \(\text{psbL}\) and \(\text{rps14}\) transcripts exhibit less editing at NTrpoB C473, NTpsbL C2 and Ntrps14 C80 than in wild-type plants. These three editing sites carry alterations that were generated by mutagenic PCR. Five different constructs were created for bombardment. Each construct was maintained for further analysis. Plants were assayed for homoplasmicity after selection through Southern blotting. All plants with integrated constructs remained homoplasmic throughout this investigation, except for R54. Despite continued rounds of selection, line R54 never achieved homoplasmicity and was analyzed as a heteroplasmic plant.

Expressed transcripts from all constructs include an ATG codon because of the use of the NcoI restriction site. Translation beginning at this AUG would be out of frame of \(\text{rpoB}\) and would proceed just 16 amino acids before reaching a stop codon.

**DNA blot analysis**

Total DNA was isolated from transgenic and wild-type leaves from shoots grown on RMOP media. DNA (1 \(\mu\)g) was digested using BamHI, electrophoresed on 1% agarose and blotted onto positively charged nylon (Amersham) using a turbo-blotter (Schleicher and Schuell). Oligonucleotides (PC1.1 and PC1.2) were used to amplify a 350 nt genomic probe from wild-type genomic DNA overlapping the insertion site. The probe was random labeled using the DECAprime II kit and \(^{32}\)P-dCTP and hybridized to the DNA blot for 24 h at 65°C.

**S1 nuclease assay**

A DNA probe was constructed by PCR using oligonucleotides \(\text{T}_{7,5}^{	ext{'}}\text{500s}\) and \(\text{500sreverse}_\text{Lg}\). The PCR product was designed to hybridize with the transgenic transcript and overlap the 3' end terminator sequence from tobacco \(\text{rps16}\). The DNA probe was restricted with NcoI (Invitrogen) and the 3' end of the antisense strand was labeled using a Klenow fill reaction. Labeled probe and 1, 10 or 25 \(\mu\)g of RNA were hybridized overnight. S1 nuclease (500 U/\(\mu\)l) (Promega) was added to the nuclease reaction for 1 h at 37°C and the products were electrophoresed on a 5% polyacrylamide gel.

**Immunoblotting**

Immunoblotting was performed as in Hegeman et al. (21). Total leaf protein was obtained from shoots grown on RMOP medium using homogenization buffer containing 50 mM Tris–HCl, 1 mM EDTA, 1 \(\times\) Protease Inhibitor cocktail (Complete, Roche) and 0.1% (v/v) Triton X-100. Protein was quantified using Bio-Rad Protein Assay kit and a BSA standard curve. Total proteins (20 \(\mu\)g) were boiled in SDS–PAGE, electrophoretically separated onto 10% acrylamide gels and transferred to nitrocellulose membranes (Pierce). Membranes were blocked overnight in blocking buffer (5% dried milk powder, 1%TBS-T) after which the primary antibody was added to 1:500 dilution from crude serum. The washed blots were incubated in secondary antibody (horse-radish peroxidase conjugated goat anti-rabbit; Amersham) diluted to 1:50 000 and proteins were visualized using SuperSignal West Dura Extended Duration Substrate as per manufacturer’s specifications (Pierce Biotechnology).

**Editing analysis**

Total RNA was isolated using Trizol (Invitrogen) for transgenic plants and wild-type leaves. Contaminating DNA was removed using Turbo DNase (Ambion) and cDNA was synthesized by reverse transcription (Omniscript;
transcripts with both 30 transcripts with 3 were amplified using PPrrn2 and either Trps16lg for Qiagen) using degenerate hexamers. Transgenic transcripts were amplified using PPrrm2 and either Trps16sh for amplification of transcripts with both 3 ends. Amplified transcripts were then assayed for editing extent using the poisoned primer extension assay as described previously (11,21,25).

Table 1. Oligonucleotides (Integrated DNA Technologies, Coralville, IA) used in experiments reported here.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5'–3'</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>500f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>500r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>501f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>501r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>502f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>502r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>503f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>503r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>504f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>504r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>505f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>505r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>506f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>506r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
</tbody>
</table>

Substrates for analysis in vitro

For substrates equivalent to the transgenic transcripts, DNA substrates were produced from PCR amplifications using primers T7_5'-500s and either Trps16sh or Trps16lg. For substrates with sequence alterations in the −31 to +2 region around NTrrpB C473, the respective mutagenic PCR primers

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5'–3'</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>500f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>500r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>501f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>501r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>502f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>502r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>503f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>503r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>504f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>504r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>505f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>505r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>506f</td>
<td>GATCCCCATGGGGCCACCATATCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
<tr>
<td>506r</td>
<td>CGGCTCTAGATTTTCTATCAATTTCGAGTTGGGAGG</td>
<td>Transgene construction</td>
</tr>
</tbody>
</table>

Substrates for analysis in vitro

For substrates equivalent to the transgenic transcripts, DNA substrates were produced from PCR amplifications using primers T7_5'-500s and either Trps16sh or Trps16lg. For substrates with sequence alterations in the −31 to +2 region around NTrrpB C473, the respective mutagenic PCR primers
were used. The bacterial sequences SK and KS were added to flank the region of \textit{rpoB} to prevent amplification from endogenous nucleic acids. A T7 sequence was added to the 5\textsuperscript{\prime} end of the substrate also by PCR amplification. RNA substrates were then produced using the PCR products as template by \textit{in vitro} transcription using the T7 MEGAscript kit (Ambion). RNAs were then purified using the RNA clean-up kit-5 (Zymo Research).

\textbf{Editing reactions in vitro}

The editing reactions were performed as described previously (25). RNA (0.1 fmol) was added to 80 \mu g of tobacco, competent, chloroplast extract (25) in assay conditions. Maize extracts were prepared from 7 to 10 day old maize plants grown in the same conditions as tobacco (25). Leaves were homogenized and plastids isolated using a Percoll (Amersham Biosciences) gradient. Intact chloroplasts were lyzed using Triton X-100, and dialyzed in Dialysis Buffer (25). Conditions for the maize \textit{in vitro} assay were identical to the conditions used for tobacco except only 20 \mu g of chloroplast extract was used. Editing of RNA substrates was analyzed using the poisoned primer extension assay (11,21,25).

\textbf{In vitro processing assay}

0.1 fmol of randomly- or 5\textsuperscript{\prime}-labeled RNA was incubated in tobacco chloroplast extract for 2 h under the \textit{in vitro} editing assay conditions described previously (25). RNA was purified by phenol:chloroform extraction and precipitated. Resuspended RNAs were separated on 6\% polyacrylamide gels.

\textbf{Semi-quantitative RT--PCR}

cDNA was synthesized for both short and long transcripts by reverse transcription (Sensiscrypt; Qiagen) from 50 ng of total RNA isolated from transplastomic plants and the Trps16sh primer. PCR amplification of the cDNA templates utilized the primers Trps16sh and PRtm2. RT--PCR products from different rounds of PCR were then separated on 3\% agarose gels. After 22 cycles differing quantities of RT--PCR products could be distinguished corresponding to the varying amounts of initial transcript. After 40 cycles all bands were of similar intensity and reactions without reverse transcriptase showed no specific amplification.

\textbf{RESULTS}

\textbf{Production of transgenic plants for further analysis of an \textit{rpoB} editing site \textit{in vivo}}

Previously, transgenic plants were analyzed that contained 27 nt surrounding the NTRpoB C473 editing site. Approximately 25\% of the transcripts carrying the wild-type tobacco sequence or a sequence altered at either \textit{\textminus}7 or \textit{\textplus}2 from the C target were edited, on average (19). However, transcripts carrying a T rather than the wild-type A at \textit{\textminus}20 were poorly edited; only 3 of 221 individually analyzed transcripts exhibited editing. Furthermore, a homologous sequence from black pine (which contains T rather than A at \textit{\textminus}20) was also not edited. Because a homologous 92 nt region from maize \textit{rpoB} had been observed previously to be highly edited (>50\%), we produced transgenic plants carrying a sequence larger than 27 nt, but smaller than 92 nt to further define sequence requirements for editing. A 54 nt region (\textminus31 to +22) from tobacco was inserted into vector pLAA24A, as well as a 92 nt tobacco sequence (\textminus31/+60) for use as a control analogous to the maize region expressed previously in tobacco (19). Several mutated versions were also produced. R92m1 carries 2 nt found in the black pine sequence at \textit{\textminus}20 and \textit{\textplus}25. R54m1 contains the \textit{\textminus}20 T change found iminical to editing in the 27 nt transgene. Because the region \textit{\textminus}16 to \textit{\textminus}21 was observed to be complementary to the \textit{\textplus}12 to +17 sequence, the \textit{\textplus}12 T was changed to A as a potential compensatory mutation for the \textit{\textminus}20A to T change (Figure 1).

The transformation vector was introduced to young tobacco leaves by biolistic delivery, and plastid genome composition was determined in regenerating plants by Southern blotting (Figure 2). A single R54-containing shoot did not reach homoplasmy despite lengthy efforts and this line was therefore examined as a heteroplasmic plant. Single transformation events of the four other lines achieved homoplasmy after repeated selection on antibiotic medium (Figure 2).

Regenerating shoots were treated with auxin and grown on rooting medium. Only R54m2 could not be induced to root,
all other plants were taken to soil (Figure 3). R54 as a heteroplastic plant exhibited sensitivity to light and leaves became bleached in exposure to levels of light equal to the normal growing conditions for the other plants. R54 plants did not survive in soil even in low light conditions. R54m1 plants thrived on soil but had severe leaf abnormalities. Leaves appeared malformed, with defects in venation, long trichomes and short internodes. Plants were severely stunted and did not flower. R92 plants were stunted and flowers were male sterile. R92m1 plants grew normally with no other defects observed from the transgene. Progeny resulting from an outcross using male pollen from wild-type plants and chloroplast transformant flowers with R92 were male sterile, but did not display any other significant phenotypes compared to wild-type tobacco plants (data not shown), suggesting that mutant phenotypes have resulted from the bombardment and tissue culture for plant regeneration.

Total protein was isolated from plants and RpoB protein levels determined through immunoblotting. There was no correlation between RpoB protein levels and severity of phenotype, and only R92m1 appeared to have reduced RpoB protein levels compared to total protein (Figure 4A). The transplastomic progeny of R92 and R92m1 plants did not display reduced RpoB protein levels (Figure 4B). We have not investigated the developmental phenotypes further, as our primary interest was to analyze the effect of the transgenes on editing of related sequences. We cannot conclude whether or not the rpoB transgene stimulated tissue culture mutations or is directly responsible for any of these phenotypes without obtaining a number of additional independently generated transgenic lines.

Assay of editing extent of transgene transcripts

Editing of transgene transcripts in leaves of the five transgenic lines were assessed by poisoned primer extension. R92m1 transcripts, which carry altered −23 and −25 nt were 65% edited compared to 67% for the R92 construct, the template with the same region of wild-type sequence. The −23 and −25 nt appear to be of little importance for the editing process. Surprisingly, even the lines carrying the −20A to T change that prevented editing in transgene transcripts of shorter length (−20 to +6) (19), exhibited editing over 60% (Figure 5). In fact, editing of the two lines containing −20A to T changes exhibited somewhat higher editing than the R54 line containing wild-type sequence. Evidently, either the addition of 11 nt at the 5′ end or 15 nt 3′ to the edited C has affected the editing efficiency of the transgene transcripts in comparison to the transgenes carrying only 27 nt of chloroplast sequence.

Further characterization of transgene transcripts

In order to understand how the increased size of the transcripts carrying 54 nt rather than 27 nt of rpoB sequence has enhanced editing efficiency, it was necessary to determine the exact RNA species produced in vivo. The 5′ end of the transcript is determined by the rrn16 promoter, which initiates transcription downstream of eubacterial-like −35/−10 promoter elements (24). The 3′ end of the transcript is determined by the terminator sequence, Trps16, from the 3′-untranslated region (3′-UTR) of tobacco rps16 (22). The 3′ end of the rps16 gene had not been mapped previously in tobacco; however, it had been determined in white mustard (26). In white mustard, a nuclease has been purified and implicated in cleavage at a recognition sequence creating the 3′ end (27,28). We aligned the Trps16 sequence from tobacco with white mustard and found the tobacco sequence carries a sequence similar to the white mustard nuclease recognition element (Figure 6). We performed 3′ end mapping in the tobacco transgenic plants and found that two RNA species are the products of the transcribed transgene in tobacco. One of the two, 3′ end I, might be a precursor to the more abundant 3′ end II. 3′ End II also matches a region of similar sequence where rps16 in white mustard is cleaved to form the mature 3′ end.

After determining that two RNA species were the product of in vivo transcription of the transgene, we wanted to assess whether transcripts of both sizes were edited to the same extent. Through selective RT–PCR, it was possible to amplify the transcripts with 3′ end I only, unlike the data shown in Figure 5, in which primers were used so that transcripts of both sizes were assayed simultaneously. Of the two RNA species within transgenic plants, the smaller RNAs accumulated more edits in every construct than the longer precursor (Figure 6). This could either result from an inhibition of editing by the extra 3′ sequence, or because the longer transcripts are processed quickly to the smaller size, before significant editing has occurred.
Editing of endogenous transcripts in transgenic plants carrying NTrpoB C473 transgenes

Previously we described editing of chloroplast transcripts in tobacco chloroplast transgenics that were over-expressing a 92 nt maize rpoB gene fragment encompassing the maize chloroplast sequence homologous to NTrpoB C473. Editing of the endogenous site NTrpoB C473 as well as sites NTpsbL C2 and NTrps14 C80, are less edited in these transgenic plants (20). All three sequences were observed to share common elements (Figure 7A) and were therefore described as a ‘cluster’ of editing sites. We analyzed the editing extent of endogenous transcripts of all three cluster members in the five new transgenic lines. All transgenic plants exhibit some reduction, varying from 10 to 30%, of endogenous NTrpoB C473 editing (Figure 7B). In every homoplasmic plant, a reduction of endogenous NTrpoB C473 editing was...
accompanied by a 10–32% reduction of endogenous NTpsbL C2 editing. A reduction in editing extent of cluster member NTpsbL C2 was not observed in R54. Endogenous NTrps14 C80 editing is reduced 12 and 32% in the wild-type sequence-containing R54 and R92 plants, respectively. Reduced Rps14 C80 endogenous editing correlates with plants with large, >20%, reductions in endogenous NTrpoB C473 and NTpsbL C2. Constructs with differences at ~20% had little effect on NTrps14 C80 editing. NTrps14 C80 is less sensitive to over-expression of NTrpoB C473 substrates than is NTpsbL C2 editing. Editing extent of another editing cluster of endogenous sites that share common sequences with each other, NTrps14 C80, which do not carry S1, S2 or S3, are unaffected by expression of transgenes carrying NTpoB C473 (Figure 7C).

A possible reason for differences in extent of inhibition of editing of related endogenous sequences between different rpoB transgenic lines could be different levels of transgene transcripts. The reduction in endogenous transcript editing is thought to be due to competition among transcripts for a limited quantity of editing factors. Transgene transcripts

presumably must be expressed at a sufficient level to engender this competition effect. We assayed the abundance of transgene transcripts in the five lines by semi-quantitative PCR. R92, the transgenic line which exhibited most reduction of editing of cluster members, also exhibited the highest transcript levels (Figure 8). R54, which exhibited no reduction in endogenous editing, had very low transcript levels, presumably due its heteroplasmic state. The other three lines exhibited intermediate amounts of transcripts and less reduction in editing of endogenous transcripts of cluster members than in R92 (Figure 8).

Analysis of chimeric transcripts in vitro

Analyzing the requirements for editing of transcripts incubated in chloroplast extracts in vitro allows more rapid examination of a large number of mutated RNA substrates than is possible with chloroplast transgenic plants, which require considerable labor and months of time for regeneration. We carried out further analysis of the cis-requirements for editing of rpoB using an in vitro editing system. Previously, the RNAs that have been assayed for editing efficiency in vitro have either been synthetic RNAs comprised only of chloroplast sequence and a 3′ KS amplification sequence (6,29,30) or have been surrounded by SK and KS amplification sequences following transcription in vitro by T7 polymerase
intensities correspond to varying transcript abundances. In vivo effects on editing efficiency of wild-type sequences KS and SK sequence surrounding a chloroplast sequence were assayed. Transcripts with the shorter 3' ends were more highly edited than those with the longer 3' ends (Figure 9A), but the difference in editing extent between the long and short transcripts was less than in vivo. When a primer was used that would amplify both long and short transcripts from extract initially incubated with the longer transcript, a higher editing efficiency was obtained. This suggested that the chloroplast extract used for editing in vitro might be capable of 3' end processing.

To assess the processing capacity of the chloroplast extracts used for editing analysis, RNA substrates equivalent to R92 with 3' end I were amplified by PCR and transcribed in vitro. Substrates were then radiolabeled at their 5' ends and incubated in chloroplast extracts to determine whether the longer RNA can be processed in vitro. The RNA was cleaved over time and was nearly fully digested after 2 h (Figure 9). RNA substrates equivalent to R54 and R92 were randomly labeled to show if they were equally cleaved and that no other cleavage products were formed (Figure 9). Only substrates with 3' ends equivalent to in vivo 3' ends I and II were observed. Thus, chloroplast extracts used for assaying editing also exhibit 3' RNA processing activity. Since most transcripts have been processed by 2 h of incubation (Figure 9), most of the transcripts assayed with the primer that amplifies both short and long transcripts are actually the shorter transcripts.

RNA substrates were constructed with the same 5' and 3' ends as in vivo transgenic transcripts to compare editing in vitro versus in vivo. The extent of editing within transcripts with sequence differences versus substrates carrying wild-type sequence correlates with the differences observed in vivo (Figure 10A). To determine whether the presence of KS and SK sequence surrounding a chloroplast sequence affects editing efficiency of wild-type sequences in vitro, we compared a substrate with a KS sequence 3' to the 54 nt rpoB sequence versus substrates carrying the rps16 3' end II (Figure 10B). The insignificant difference between substrates carrying the 54 nt sequence and either the 5' rrn16/MCS 5' end and the rps16 3' end versus the SK and KS 5' and 3' ends led us to utilize the convenient SK/KS sequences for further assessment of editing efficiency of substrates in chloroplast extracts. The 5' and 3' environment of the NTrpoB C473 region is necessarily different in the transgene and in vitro substrates than in actual rpoB transcripts (Figure 10C).

We produced a number of RNA substrates to assess the effect of alterations within the conserved elements on editing efficiency and to identify any other important sequences. DNA templates were created by PCR with bacterial T7 promoter and SK and KS sequences flanking the rpoB region and then transcribed in vitro to produce the substrate RNA. All substrates carried the –31/+22 chloroplast sequence present in the transgenes tested in vivo.

Substrates R1–R11 were created to determine the critical elements within –20/+6 region (Figure 11A). Together these substrates cover the –20/+6 region found to be important in vivo. The editing efficiency of substrates R1–R7 and R9 in tobacco extracts was significantly reduced compared to the wild-type substrates. Of the substrates with reduced editing efficiencies, R4 and R9 had the smallest reduction (20%) in relative editing, and alterations were either not in an important region or not a sufficient sequence change to alter processing by the editing apparatus. Substrate R8 contains the same alterations as R9 but was not reduced in its editing efficiency, suggesting that the –3 and –2 nt contained in R9 are not part of the critical sequence element. Curiously, R10 was a better substrate than R54KS perhaps indicating that the sequence restraints within coding regions result in endogenous sequences that do not always represent the optimal sequences for editing. The increase in editing is only ~30% and may be due to enhancement of RNA structure rather than representing a positive alteration in part of an important sequence element.

Sequence changes in S1, S2 and S3 affect editing efficiency

A single –20 A→U change nearly completely abolished editing, in contrast to the results obtained in vivo with the –31/+22 transgene (Figure 5), but in agreement with the low editing efficiency when only –20/+7 chloroplast sequence is included in the transgene (19). Evidently the presence of the SK sequence 5' to the chloroplast sequence results in greater sensitivity of the transcripts to editing perturbation by alteration of the –20 nt than does the presence of a few nucleotides derived from the rrn16 promoter.

Substrates R2, R4, R7 and R12–R15 were created with differences in S1, S2, S3 to test their importance for editing NTrpoB C473 (Figure 11A). R4 showed small, ~20%, reductions in editing relative to wild type. This can be compared to the 65% reduction in R7 and abolition of editing in R2. Substrates carrying a combination of altered common elements, R12–R15, were created to study the effect of multiple changes. The effects of changes in S1, S2 and S3 are evident in cumulative reductions of editing in substrates with combinations of altered elements compared to wild-type.

Figure 8. Semi-quantitative RT–PCR of transgene transcripts from transplastomic plants. (A) At 22 cycles of RT–PCR, bands of different intensities correspond to varying transcript abundances. (B) After 40 cycles of RT–PCR all bands are at equivalent intensities.
substrates. All substrates with differences in S1 could not be edited by tobacco chloroplast extracts.

Substrates R1–R15 all had sequence changes from purine to pyrimidine and vice versa. Substrates R3–R5 only exhibited minor 20–40% reductions in editing compared to wild type (Figure 11A). To test whether purine to purine and pyrimidine to pyrimidine mutations might disrupt editing more significantly, such mutations were created within the previously defined critical region in substrates R16–R21 (Figure 11C). Again nucleotide changes in S1, as in R16, had the greatest effect on editing. R17–R19 had more severe effects than substrates R3–R5. R18, carrying a change in S2 is a much poorer editing substrate than R4 or R54K5 and confirms that S2 is a critical sequence. Overall, purine to pyrimidine and pyrimidine to pyrimidine base changes had a larger impact on editing than purine to pyrimidine and pyrimidine to purine. In spite of this, R21 had no effect compared to R7 although they both alter S3. S3 is flanked by GA nucleotides and the mutations in R21 happen to create two S3 sequences shifted 1 nt 5’ and 3’ of the endogenous position from the editing site. The sequence of R21 is evidently not sufficiently altered to reduce the editability of the substrate.

When maize rpoB is aligned with tobacco rpoB, ZMrpoB C467 is at the same position as NTrpoB C473, and the surrounding sequence is very similar to that in tobacco. Tobacco chloroplasts can edit a template expressing a +C0/C0+C0+61 region around ZMrpoB C467 (data not shown). Therefore, it is likely that tobacco and maize contain factors that can recognize similar cis-acting elements around NTrpoB C473. The same substrates tested in vitro using tobacco extracts were tested in vitro using maize extracts. Substrates R2–R7, which have sequence alterations in the −20/−5 region, exhibited major reductions in editing efficiency. The same region of sequence is therefore critical for editing in maize as in tobacco. All substrates with differences in S1, S2 and S3 were less efficient editing substrates in maize extracts compared to wild type (Figure 11B). As in the tobacco extracts, changes in S1 and S3 affected editing of the substrate more than S2. In contrast, the single −20 T to A change that severely affected editing in tobacco had little effect on editing in maize extracts. The editing efficiencies of the substrates with mutations within the −20/−5 region were reduced to a greater extent in maize compared to tobacco. Also, substrates R5 and R6 had the largest reductions in editing compared to R2, which had the largest reduction in tobacco. These results indicate that, although the RpoB C467/C473 editing factors in the two species may be similar, some evolutionary...
immediate flanking and adequate region for editing do not

with only C473 were more efficiently edited

in vivo complete in their complement of critical editing factors.

chloroplasts. Chloroplast extracts are therefore likely to be

in vitro and in vitro systems have been particularly useful in studying the cis-acting elements for editing sites due to their relative speed, flexibility of species and cost advantages. Studies in vivo have presently been limited to tobacco because of the technical difficulty of chloroplast transformation. The maize in vitro assay described here should facilitate identification of editing factors in an organism with better genetic resources. This is also the first monocot in vitro editing system and should allow comparisons of editing between dicot and monocot species.

The same substrates have not been assayed previously both in vivo and in vitro to test the biological relevance of the in vitro system. Here we describe the assay of comparable templates for NTrpoB C473 in vivo and in vitro, which allows a direct comparison between systems. Five substrates with sequence differences but the same approximate flanking sequences have been compared to wild-type substrates in vivo and in vitro. All five substrates showed similar in vivo and in vitro editing efficiencies, indicating that data from the in vitro system are applicable to editing in intact plant chloroplasts. Chloroplast extracts are therefore likely to be complete in their complement of critical editing factors.

Transcripts containing −31 to +22 nt surrounding NTrpoB C473 were more efficiently edited in vivo than transcripts with only −20 to +6 nt. Thus, substrates reduced to the immediate flanking and adequate region for editing do not represent the full complement of cis-acting elements that influence editing. Sequence elements that enhance editing are probably present outside of the −20/+6 region around the editing site. Nevertheless, the −20 to +6 region is sufficient to specify editing the proper C target in vivo.

Substrates have been constructed and tested in vivo or in vitro with changes in the nucleotides within the −20/+6 minimal region (Figure 12). Changes in the sequence from −20/−5 had large reductions in the editing efficiency of the corresponding substrate and the critical sequence element is contained within that region. Differences in S1, S2 and S3 consistently affected editing of the templates when assayed in extracts from tobacco and maize. Consistent with our hypothesis that elements conserved between cluster members are important site recognition features, S1, S2 and S3 are critical for RNA substrate editing in both tobacco and maize extracts. A substrate that contains a −20 A to U change is very poorly edited in tobacco but well edited in maize. This suggests that in maize the −20 nt is not critical. The editing factor in maize apparently sufficiently differs from the tobacco factor that it can tolerate a change in the S1 sequence whereas the tobacco factor cannot. In addition, substrates with changes within the −20/−5 region have different reductions in editing efficiency between tobacco and maize. Although editing factors in maize recognize the same region of sequence, they have diverged to the extent that different nucleotides are critical for editing.

Of the 34 editing sites within tobacco chloroplasts 11 share elements similar to S1, S2 and S3, although not all are reduced by over-expression of NTrpoB C473 (Figure 12). One explanation for the lack of a competition effect is the existence of different trans-factors for some or all of these sites. Another possible reason that sites with common cis-acting elements are differently affected by over-expression of NTrpoB C473 is that the same sequestered factor may be

Figure 10. (A) Relative in vitro editing of substrates analogous to transgenic transcripts created with 3′ end II. Editing is expressed as % of the wild-type R54 editing because substrates were assayed by two different extracts, one editing the constructs from 50 to 80% and the other from 22 to 36%. (B) In vitro editing of the R54KS substrate with bacterial sequences SK and KS around the editing site and substrates equivalent to the transgene transcript expressed in transplastomic plants. (C) Diagram of editing templates. Arrows represent the +1 position of transcription initiation. Underlined nucleotides signify the common rpoB sequence around the editing site.

**DISCUSSION**

With the addition of maize extracts described here, assays for chloroplast editing in vitro are available in four species including tobacco, pea and Arabidopsis as well (6,25,31). In vitro systems have been particularly useful in studying the cis-acting elements for editing sites due to their relative speed, flexibility of species and cost advantages. Studies in vivo have presently been limited to tobacco because of the technical difficulty of chloroplast transformation. The maize in vitro assay described here should facilitate identification of editing factors in an organism with better genetic resources. This is also the first monocot in vitro editing system and should allow comparisons of editing between dicot and monocot species.

The same substrates have not been assayed previously both in vivo and in vitro to test the biological relevance of the in vitro system. Here we describe the assay of comparable templates for NTrpoB C473 in vivo and in vitro, which allows a direct comparison between systems. Five substrates with sequence differences but the same approximate flanking sequences have been compared to wild-type substrates in vivo and in vitro. All five substrates showed similar in vivo and in vitro editing efficiencies, indicating that data from the in vitro system are applicable to editing in intact plant chloroplasts. Chloroplast extracts are therefore likely to be complete in their complement of critical editing factors.

Transcripts containing −31 to +22 nt surrounding NTrpoB C473 were more efficiently edited in vivo than transcripts with only −20 to +6 nt. Thus, substrates reduced to the immediate flanking and adequate region for editing do not represent the full complement of cis-acting elements that influence editing. Sequence elements that enhance editing are probably present outside of the −20/+6 region around the editing site. Nevertheless, the −20 to +6 region is sufficient to specify editing the proper C target in vivo.

Substrates have been constructed and tested in vivo or in vitro with changes in the nucleotides within the −20/+6 minimal region (Figure 12). Changes in the sequence from −20/−5 had large reductions in the editing efficiency of the corresponding substrate and the critical sequence element is contained within that region. Differences in S1, S2 and S3 consistently affected editing of the templates when assayed in extracts from tobacco and maize. Consistent with our hypothesis that elements conserved between cluster members are important site recognition features, S1, S2 and S3 are critical for RNA substrate editing in both tobacco and maize extracts. A substrate that contains a −20 A to U change is very poorly edited in tobacco but well edited in maize. This suggests that in maize the −20 nt is not critical. The editing factor in maize apparently sufficiently differs from the tobacco factor that it can tolerate a change in the S1 sequence whereas the tobacco factor cannot. In addition, substrates with changes within the −20/−5 region have different reductions in editing efficiency between tobacco and maize. Although editing factors in maize recognize the same region of sequence, they have diverged to the extent that different nucleotides are critical for editing.

Of the 34 editing sites within tobacco chloroplasts 11 share elements similar to S1, S2 and S3, although not all are reduced by over-expression of NTrpoB C473 (Figure 12). One explanation for the lack of a competition effect is the existence of different trans-factors for some or all of these sites. Another possible reason that sites with common cis-acting elements are differently affected by over-expression of NTrpoB C473 is that the same sequestered factor may be
important for editing of multiple sites, but may not always be limiting. This seems to be the case even among sites whose editing is reduced by over-expression, as there are significant differences in the strength of competition between different members of the RpoB C467 cluster. Some sites may have stronger affinities for a particular trans-factor than other sites. Therefore, all 11 sites may share important cis-acting elements, but the possible relationship of such sites will not be obvious by over-expression of a single RNA editing site in a transgenic plant.

The presence of S1, S2 and S3 in the three genes whose editing is altered in transgenic plants could only be identified by allowing significant gaps, suggesting critical sequence elements could be irregularly spaced from the edited site. This complicates understanding how editing sites are specifically targeted, since any common distance from a cis-acting element to the editing site might differ among sites. Possibly, either a processive mechanism specifies a particular nucleotide for editing that is different between sites from a cis-acting element, or there are elements that require particular spacing that may serve roles in locating the editing components to the particular C to be edited. Any processive mechanism within the NTrpoB C473 cluster would have to differ significantly between sites due to different sequences around the editing site, and the position of 5' and 3' Cs around editing sites. Sensitivity of editing site substrates to altered spacing has been observed in mitochondria, suggesting a potential for a 'molecular ruler' determining the target for editing by its distance from a cis-acting element (32,33). In the NTrpoB C473 cluster, S3 is 4–5 nt away from the edited nucleotide and could be important for editing site targeting. Substrate R21 is not reduced in its editing efficiency and suggests that small perturbations in the precise position of this sequence do not affect editing, at least in rpoB transcripts (Figure 11). In mitochondria, editing sites separated by only a short length of sequence were found to share cis-acting elements (34,35). Chloroplast sites NTatpA C791 and NTatpA C795 contain two S3 elements 5 and 6 nt upstream of the edited sites, respectively, and share the same upstream elements S1 and S2 (Figure 12), but their editing efficiency is unaffected by over-expression of NTrpoB C473.

Some editing sites in plant mitochondria and chloroplasts have been shown to be sensitive to changes in nucleotides immediately adjacent to the editing site (30,33). However, an inserted ZMrpoB C467 editing site from maize sequence, flanked by 32 bases, differs from the tobacco sequence by a 3' adjacent G and is edited as well as the tobacco sequence (5,21). Endogenous editing of NTrpoB C473 and cluster member NTpsbL C2 are reduced when the NTrpoB C473 gene fragment is over-expressed, although the NTpsbL C2 editing site differs in its immediate 5' and 3' nucleotides (20). If NTpsbL C2 shares factors involved in the editing mechanism, then the editing complex for NTrpoB C473 may be insensitive to differences in nucleotides flanking the editing site.

The -20A nucleotide has been determined to be critical in vivo using a 27 nt (−20, +6) template and in vitro using a 54 nt (−31, +22) template surrounded by SK and KS sequences. In vivo, however, a substrate containing a 54 nt region around NTrpoB C473 with the −20U alteration can be more highly edited than a wild-type substrate. A substrate

![Figure 11](image-url)
analogous to the in vivo transgenic transcript is also well edited in vitro compared to wild type. Therefore, the reduction in editing caused by the −20U alteration depends upon the particular sequences 5′ to the −20 nt. Since two independent observations show a critical role for the −20A, it is likely important for editing. Possibly the importance of the region 5′ to −20 relates to the existence of complementary sequence in the mature 3′ end, which could be involved in a secondary structure near the editing site (Figure 12). The extent of the complementarity would differ between the 54 nt region around NTrpoB C473 and the 27 nt region, and would not be present in the in vitro substrate that carries SK and KS flanking sequences. Possibly this region of complementarity can compensate for the −20U mutation in the in vivo R54m1 and R54m2 transcripts. Other unknown RNA secondary structures or cis-elements might also result in the compensation for the −20U alteration.

The presence or absence of critical nucleotides near the C editing target does not solely determine the extent to which an RNA substrate is edited in vivo or in vitro. RNA editing sites are sensitive to not only the amount of local sequence around the edited site, but can also be affected by sequence features significantly distal to the editing site. In vivo and in vitro RNA substrates with 3′ ends containing 44 nt more rps16 sequence, 111 nt away from the edited nucleotide, are edited less efficiently. The endogenous NTrpoB C473 editing site is located far from the 3′ end of its transcript, and possibly the edited machinery adapted to edit this site may have done so without the bulky stem–loops and other elements that normally characterize plastid transcript 3′ ends. Another possibility is that there is sequence in the longer 3′ end that has the ability to sequester editing factors. Indeed, near the end of the longer rps16 3′ end present in the transgene transcripts, there is some sequence similar to the common elements known to be critical for NTrpoB C473 editing (Figure 12). Substrate editing is therefore sensitive to the presence of nearby cis-acting elements, more distant enhancing elements and flanking sequence. These sensitivities could be responsible for the early difficulty in creation of an in vitro system, and suggest why only a limited set of editing sites have been described that are edited in chloroplast extracts in vitro at high efficiency.

ACKNOWLEDGEMENTS

We thank Barbara Eaglesham for maintaining transplastomic plants in tissue culture and Pal Maliga for providing the chloroplast transformation vector. This work was supported by NSF MCB 0344007 and NIH GM 50723 grants to M.R.H and a NIH graduate traineeship to M.L.H. Funding to pay the Open Access publication charges for this article was provided by NSF MCB 0344007.

Conflict of interest statement. None declared.

REFERENCES

1. Handa, H. (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica...


