Site-Specific Mutagenesis of the Nodule-Infected Cell Expression (NICE) Element and the AT-Rich Element ATRE-BS2* of the Sesbania rostrata Leghemoglobin glb3 Promoter

Krzysztof Szczyglowski,a,d Laszlo Szabados,d,1 Susan Y. Fujimoto,a,c David Silver,a,c and Frans J. de Bruijn a,b,c,2

a Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824-1312
b Department of Microbiology, Michigan State University, East Lansing, Michigan 48824
c Genetics Program, Michigan State University, East Lansing, Michigan 48824
d Max-Planck-Institut für Züchtungsforschung, Carl von Linneweg 10, 50829 Cologne 30, Germany

Sesbania rostrata leghemoglobin glb3 (Srglb3) promoter sequences responsible for expression in infected cells of transgenic Lotus corniculatus nodules were delimited to a 78-bp Dral-Hinfl fragment. This region, which is located between coordinates -194 to -116 relative to the start codon of the Srglb3 gene, was named the nodule-infected cell expression (NICE) element. Insertion of the NICE element into the truncated nopaline synthase promoter was found to confer a nodule-specific expression pattern on this normally root-enhanced promoter. Within the NICE element, three distinct motifs ([A]AAAGAT, TTGTCTCTT, and CACCC[T]) were identified; they are highly conserved in the promoter regions of a variety of plant (leg)hemoglobin genes. The NICE element and the adjacent AT-rich element (ATRE-BS2*) were subjected to site-directed mutagenesis. The expression patterns of nine selected Srglb3 promoter fragments carrying mutations in ATRE-BS2* and 19 with mutations in the NICE element were examined. Mutations in ATRE-BS2* had varying effects on Srglb3 promoter activity, ranging from a two- to threefold reduction to a slight stimulation of activity. Mutations in the highly conserved (A)AAAGAT motif of the NICE element reduced Srglb3 promoter activity two- to fourfold, whereas mutations in the TCTT portion of the TTGTCTCTT motif virtually abolished promoter activity, demonstrating the essential nature of these motifs for Srglb3 gene expression. An A-to-T substitution in the CACCC(T) motif of the NICE element also abolished Srglb3 promoter activity, while a C-to-T mutation at position 4 resulted in a threefold reduction of promoter strength. The latter phenotypes resemble the effect of similar mutations in the conserved CACCC motif located in the promoter region of mammalian β-globin genes. The possible analogies between these two systems will be discussed.

INTRODUCTION

Plant hemoglobin genes have been identified and characterized in several legume and non-legume plant species (see Arredondo-Peter and Escamilla, 1991). The function of plant hemoglobin appears to be the same as that of mammalian hemoglobin, namely the binding and/or transport of oxygen (Appleby, 1984; Appleby et al., 1990). In nitrogen-fixing stem and root nodules induced by rhizobia on legume or non-legume host plants, leghemoglobin facilitates oxygen diffusion to the actively respiring, nitrogen-fixing bacteroids within the infected cells. These proteins operate at an intracellular oxygen concentration of ~10 nM or less, which ensures that the oxygen-sensitive nitrogenase enzyme is not irreversibly denatured (Appleby, 1984).

The role of hemoglobin in non-nodulated plants is less clear. It has been postulated that hemoglobin may play a role as an oxygen sensor in the roots or may participate in the transport of oxygen to actively dividing cells (Appleby et al., 1988, 1990). The latter hypothesis has been strengthened by the observation that the expression of non-legume hemoglobin genes in heterologous transgenic plants is mostly confined to a small area of actively dividing and respiring cells in the vicinity of the root tip (Bogusz et al., 1990).

In contrast, leghemoglobin (lb) gene expression in legume plants has been shown to be nodule specific and confined to the infected tissues of the nodule (see de Bruijn et al., 1990;
Nap and Bisseling, 1990). In the case of indeterminate nodules, such as those induced on alfalfa by *Rhizobium meliloti*, *lb* gene expression has been found to be triggered in a single cell layer (interzone II to III), immediately adjacent to the beginning of the nitrogen fixation zone. Moreover, *lb* transcripts appear to be present only in the infected cells of alfalfa nodule tissue (de Billy et al., 1991). It has also been reported that *lb* genes are not expressed in alfalfa nodules that do not contain (intracellular) rhizobia, such as those formed spontaneously on the roots of specific alfalfa lines (Truchet et al., 1989) or induced by auxin transport inhibitors (Hiirsch et al., 1989) or by certain mutant rhizobial strains (Dickstein et al., 1988; de Bruijn et al., 1988, 1990). These observations have suggested that a non-diffusible rhizobial signal, produced by intracellular bacteria or otherwise related to the release of rhizobia from the infection thread, may be required for transcriptional activation of the *lb* genes (see Nap and Bisseling, 1990; de Billy et al., 1991; Dickstein et al., 1991; de Bruijn and Schell, 1992).

In determinate nodules, such as those induced by *Bradyrhizobium japonicum* on soybean, the precise cellular localization of *lb* gene transcripts and *Lb* proteins has not yet been established unambiguously. *lb* transcripts have been found to be most prevalent in infected cells of soybean nodules, although low but significant levels also appear to be present in uninfected cells (Kouchi et al., 1989). The *Lb* apoprotein appears to be located both in infected and uninfected cells of soybean nodules (VandenBosch and Newcomb, 1988).

Chimeric *lb*-reporter gene constructs and transgenic plants have been used to investigate the role of *lb* promoter regions in nodule (cell)-specific expression (see de Bruijn et al., 1990; Sanchez et al., 1991; de Bruijn and Schell, 1992). The promoter regions of the *lbc3* gene from soybean and the *gIb3* gene from the stem-nodulated tropical legume *Sesbania rostrata* (see de Bruijn, 1989) have been studied in the most detail. Both soybean and *Sesbania* *lb* 5' upstream regions have been shown to confer nodule-specific expression on reporter genes in indeterminate nodules formed on transgenic alfalfa (de Bruijn et al., 1989, 1990) and determinate nodules induced on *Lotus corniculatus* plants (Stougaard et al., 1986, 1987, 1990; Szabados et al., 1990). *lb* promoter activity has been shown to be confined to the infected zone of the nodules induced on transgenic *Lotus* plants and appeared to be most prevalent in the infected cells (Szabados et al., 1990; Lauridsen et al., 1993).

cis-Acting elements responsible for nodule-specific expression of both soybean *Ibc3* and *Sesbania* *gb3* genes have been delimited, and common DNA sequence motifs, as well as differences in promoter organization have been described (see de Bruijn et al., 1990; de Bruijn and Schell, 1992). In the case of the soybean *Ibc3* 5' upstream region, a distal strong positive element, a weak positive element, an organ-specific element (OSE), and a negative element (NE) have been identified (Stougaard et al., 1987, 1990). Our previous analysis has revealed the presence of two positive elements in the *Sesbania* *gb3* 5' upstream region. The first element acts like an orientation independent enhancer, while the second, located between coordinates -429 to -48 relative to the *gb3* ATG codon, is responsible for nodule-specific expression (de Bruijn et al., 1989; Szabados et al., 1990).

**cis-Acting factors** have been found to interact with specific regions in the soybean and *Sesbania* *lb* 5' upstream regions (see de Bruijn and Schell, 1992). Two related AT-rich elements (ATRE-BS1 and ATRE-BS2), located immediately 5' of the soybean *Ibc3* OSE region, have been shown to interact specifically with DNA binding proteins from nodules (NAT1 and NAT2) and leaves (LAT1; Jensen et al., 1988; Jacobsen et al., 1990). NAT1 and LAT1 resemble mammalian high-mobility group (HMG) proteins found to be associated with transcriptionally active chromatin (Jacobsen et al., 1990). NAT2 binding activity is barely detectable in soybean roots, and its abundance increases in root nodules until days 10 to 12 after rhizobial infection, at which time the *lb* genes are maximally induced (Jacobsen et al., 1990; Jensen, 1991). A highly related element (ATRE-BS2*) has been identified in the analogous position of the *Sesbania* *gb3* 5' upstream region, which interacts with specific factors from nodules, leaves, and roots (de Bruijn et al., 1988; Metz et al., 1988). Other BS2*-like ATREs are present both upstream and downstream of the BS2* site in the *Sesbania* *gb3* locus and appear to be associated with positive cis-acting elements (de Bruijn et al., 1989; de Bruijn and Schell, 1992; Welters et al., 1993).

ATREs that interact with NAT- and LAT-like DNA binding proteins have also been identified in other late nodulin genes, such as the soybean N23 (Jacobsen et al., 1990) and French bean *glny* (Forde et al., 1990) genes, as well as in the promoter region of a variety of non-nodulin plant genes (see Forde, 1994). The exact function of ATREs remains to be elucidated, although in several cases they have been found to be associated with positive cis-acting elements in plant gene promoters (see Forde, 1994).

Here, we extend our analysis of the cis-acting elements in the *Sesbania* *gb3* gene that are responsible for nodule-specific expression. We show that a 78-bp DNA fragment, which is located between coordinates -194 to -116 of the *Sesbania rostrata* leghemoglobin *gb3* (SrgIb3) 5' upstream region, is essential and sufficient for *lb* promoter expression in infected cells of the nodule (nodule-infected cell expression [NICE] element). We also describe the results of site-specific mutagenesis experiments designed to examine the functional significance of the ATRE-BS2* element and to implicate specific base pairs and highly conserved motifs within the NICE element in nodule-specific expression. Based on the latter results, we draw analogies between *lb* promoter structure and the promoter architecture of mammalian β-globin genes.

**RESULTS**

**Delimitation of the SrgIb3 NICE Element**

Previously, we have shown that the region located immediately upstream of the SrgIb3 transcriptional start site (-429 to -48 relative to the ATG) is essential for nodule-specific
Leghemoglobin Gene Regulation 319

To further delimit the DNA sequences responsible for this mode of expression, a detailed deletion/substitution analysis of the -429 to -48 region was conducted. The modified DNA fragments were fused to the β-glucuronidase (gus) (Escherichia coli uidA) reporter gene and their structure is shown in Figure 1A. The chimeric Srglb3-gus reporter gene constructs were introduced into L. corniculatus plants, as described previously (Szabados et al., 1990). Five to 14 independent transgenic plants were generated for each construct, as listed in Table 1, and nodulated with the Lotus symbiont R. loti NZP2037. The GUS activity of nodule, root, and stem tissues of the transgenic plants was measured, and the results of 5 to 14 independent plants were averaged (Figure 1B; Table 1). GUS expression in stem tissues was at a background level for all constructs examined (data not shown).

Construct LP32, carrying Srglb3 5’ upstream sequences up to coordinate -1914, directed a high level of nodule-specific expression in transgenic Lotus plants, as reported previously (Figure 1B; Szabados et al., 1990). When the -1914 to -431 region was deleted, a four- to fivefold reduction of nodule-specific GUS expression was observed (LP31). Deleting the -1914 to -387 region and substituting the -116 to -51 region (containing the TATA box) with the -56 to +1 region of the nopaline synthase (nos) promoter resulted in a high level (.60 of wild type) of nodule-specific activity (LP154). However, substituting the -161 to -51 region of the -431 to -51 Srglb3 promoter fragment (containing the CAAG and TATA boxes) with the truncated -150 to +1 nos promoter (Ebert et al., 1987) resulted in a virtual loss of nodule-specific GUS expression. At the same time, an increase in root-specific expression was observed (LP45). These deletion/substitution experiments suggested that DNA sequences essential for nodule-specific expression are located between the -161 and -116 Hinfl sites, although it cannot be ruled out that the altered expression pattern observed with construct LP45 (versus LP154) is caused by an unusual interaction between tissue-specific cis-acting sequences of the chimeric Ib and nos promoters.

Deleting the -245 to -194 region of the -431 to -51 Srglb3 promoter fragment resulted in a low (45% of wild type) level light grey (stippled) boxes. The striped boxes denote nos minimal promoter sequences from position -56 to +1 and the striped boxes indicate nos promoter (enhancer) sequences from -150 to -57 (Ebert et al., 1987). The A symbol denotes an internal deletion, and the inverted triangles show the position of a 20-bp BglII linker insertion (see text for details).

(B) GUS expression levels found in transgenic L. corniculatus roots (open boxes) and nodules (stippled boxes) harboring the constructs shown in (A). The GUS levels are indicated in picomoles of 4-methylumbelliferone generated per minute per milligram of protein in the extract and represent an average of GUS activity in tissues of 5 to 14 independent transgenic plants for each construct examined (to minimize the position effect; see de Bruijn et al., 1990; Table 1). The standard error values for each construct are shown in Table 1.
Table 1. GUS Expression Levels and Their Respective Standard Errors in Different Transgenic L. corniculatus Plants

<table>
<thead>
<tr>
<th>Construct</th>
<th>No. of Plants</th>
<th>GUS in Nodules</th>
<th>SE</th>
<th>GUS in Roots</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP32</td>
<td>8</td>
<td>1326</td>
<td>596</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>LP31</td>
<td>12</td>
<td>351</td>
<td>136</td>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>LP154</td>
<td>14</td>
<td>789</td>
<td>258</td>
<td>53</td>
<td>14</td>
</tr>
<tr>
<td>LP45</td>
<td>8</td>
<td>100</td>
<td>13</td>
<td>248</td>
<td>6</td>
</tr>
<tr>
<td>LP213</td>
<td>5</td>
<td>202</td>
<td>108</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>LP91</td>
<td>13</td>
<td>16</td>
<td>2</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td>LP152</td>
<td>8</td>
<td>3943</td>
<td>775</td>
<td>252</td>
<td>20</td>
</tr>
<tr>
<td>LP63</td>
<td>8</td>
<td>17</td>
<td>4</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>LP64</td>
<td>8</td>
<td>199</td>
<td>88</td>
<td>331</td>
<td>15</td>
</tr>
<tr>
<td>LP182</td>
<td>14</td>
<td>22</td>
<td>6</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>LP181</td>
<td>7</td>
<td>32</td>
<td>7</td>
<td>73</td>
<td>3</td>
</tr>
<tr>
<td>LP140</td>
<td>5</td>
<td>16</td>
<td>2</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>LP62</td>
<td>10</td>
<td>20</td>
<td>7</td>
<td>39</td>
<td>6</td>
</tr>
<tr>
<td>LP100</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>LP231</td>
<td>10</td>
<td>67</td>
<td>24</td>
<td>37</td>
<td>9</td>
</tr>
<tr>
<td>LP232</td>
<td>10</td>
<td>44</td>
<td>30</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>LP12</td>
<td>12</td>
<td>616</td>
<td>198</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>LP14-27</td>
<td>16</td>
<td>470</td>
<td>136</td>
<td>57</td>
<td>8</td>
</tr>
<tr>
<td>LP14-23</td>
<td>18</td>
<td>199</td>
<td>58</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>LP14-28</td>
<td>25</td>
<td>404</td>
<td>93</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>LP14-9</td>
<td>16</td>
<td>689</td>
<td>265</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>LP14-17</td>
<td>11</td>
<td>334</td>
<td>222</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>LP14-1</td>
<td>15</td>
<td>791</td>
<td>139</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>LP14-12</td>
<td>19</td>
<td>256</td>
<td>39</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>LP14-10</td>
<td>15</td>
<td>388</td>
<td>143</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>LP14-19</td>
<td>19</td>
<td>576</td>
<td>98</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>LP14-2</td>
<td>16</td>
<td>240</td>
<td>75</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>LP14-5</td>
<td>13</td>
<td>151</td>
<td>41</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>LP14-12</td>
<td>12</td>
<td>212</td>
<td>93</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>LP14-15</td>
<td>16</td>
<td>350</td>
<td>116</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>LP14-30</td>
<td>9</td>
<td>242</td>
<td>82</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>LP14-3</td>
<td>10</td>
<td>1536</td>
<td>794</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>LP14-13</td>
<td>18</td>
<td>299</td>
<td>82</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>LP14-25</td>
<td>15</td>
<td>398</td>
<td>106</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>LP14-7</td>
<td>15</td>
<td>46</td>
<td>19</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>LP14-26</td>
<td>31</td>
<td>20</td>
<td>4</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>LP14-21</td>
<td>10</td>
<td>33</td>
<td>16</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>LP14-24</td>
<td>20</td>
<td>597</td>
<td>153</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>LP14-31</td>
<td>17</td>
<td>17</td>
<td>7</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>LP14-20</td>
<td>14</td>
<td>411</td>
<td>149</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>LP14-4</td>
<td>10</td>
<td>60</td>
<td>19</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>LP14-8</td>
<td>14</td>
<td>456</td>
<td>148</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>LP14-6</td>
<td>22</td>
<td>10</td>
<td>1</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>LP14-22</td>
<td>14</td>
<td>206</td>
<td>55</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>LP14-18</td>
<td>15</td>
<td>569</td>
<td>191</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

*Average GUS activity is expressed in picomoles of 4-methylumbelliferone per minute per milligram of protein.

SE, standard error.

of nodule-specific GUS expression (LP213), suggesting that the corresponding Spst-Dral fragment is not essential for nodule-specific expression. However, deleting the -194 to -194 region of the Stglb3 5'upstream region resulted in a complete loss of GUS activity (LP91). Adding the -150 to -57 nos enhancer element (Ebert et al., 1987) in cis to the -194 to -51 fragment resulted in a very high (300% of wild type) level of nodule-specific expression (LP152), suggesting that the DNA sequences essential for nodule-specific expression are located between coordinates -194 (Dral site) and -116 (HinfI site).

Deleting the -194 to -161 region of the Stglb3 5' upstream region abolished GUS expression completely (LP63), and the addition of the nos enhancer element in cis did not restore a nodule-specific GUS expression pattern (LP64). Insertion of a 20-bp linker DNA fragment into the Ball site (position -161) of the -431 to -51 Stglb3 promoter fragment resulted in abolishment of GUS expression (LP182). Adding the -150 to -57 nos enhancer element to the mutated -431 to -51 fragment did not restore nodule-specific GUS expression but resulted in a slight increase in root-specific expression (LP181).

These results supported the conclusion that the -194 to -116 region is essential for (nodule-specific) Stglb3 promoter activity and showed that the restoration of nodule-specific expression by the nos enhancer element in construct LP152 is not due to the presence of a nodule-specific cis-acting element carried by the nos enhancer.

Cytological (X-galuc) staining of GUS activity in nodules induced on transgenic Lotus plants carrying the constructs shown in Figure 1A was performed. Examples of this analysis are shown in Figures 2A, 2B, and 2D. Staining of hand-cut nodule sections revealed that the -194 to -51 Stglb3 promoter fragment (LP32) directed GUS expression only in the infected zone of the nodule, leaving the nodule parenchyma and cortex layers unstained (data not shown), as observed previously (Szabados et al., 1990). The examination of multiple nodules, containing high-to-low ratios of infected versus uninfected cells, strongly suggested that GUS expression directed by the Stglb3 promoter was confined to the infected cells of the nodule (data not shown; see Szabados et al., 1990). However, the level of resolution obtained by examining these hand sections was not sufficient to draw a definite conclusion about the infected-cell specificity of this promoter. Therefore, 10-μm sections of nodules harboring the LP32 construct were also examined using dark-field microscopy after staining for GUS activity. The results of these experiments are shown in Figure 2A. A section of a complete nodule clearly revealed the presence of small uninfected cells and enlarged infected cells, of which only the infected cells exhibited the blue color indicative of GUS activity (Figure 2A). A higher magnification of these nodule sections strongly supported this correlation (Figure 2B). The smaller uninfected cells of nodules harboring LP32 were black, whereas the larger infected cells displayed blue color.

To rule out that the observed blue color was simply due to the presence of bacteria (bacteroids) rather than to GUS expression, sections of nodules harboring a promoterless gus construct (LP100) were examined after staining for GUS activity. Although the difference between infected (white) and uninfected (dark) cells could clearly be observed in these sections, the infected cells did not display the blue color typically observed with GUS' cells (Figure 2D). Thus, the blue color in infected cells observed in Figures 2A and 2B was not due
Figure 2. Histochemical Staining of Nodule Sections for GUS Activity.

(A) A 10-μm section of a nodule induced on a transgenic *L. corniculatus* plant harboring the LP32 construct (see Figure 1) that was stained for GUS activity (see Methods) and observed under a light microscope.

(B) A higher magnification (x5) of the section shown in (A). The blue color characteristic of GUS activity is confined to the enlarged infected cells of the nodule tissue. No staining can be observed in uninfected cells of the inner nodule nor in the cells of the cortex or the vascular bundle.

(C) A 10-μm section of a transgenic nodule harboring construct LP231 (see Figure 3) that was stained for GUS activity and observed under a light microscope.

(D) A 10-μm section of a transgenic nodule harboring the control construct LP100. The infected cells of the inner nodule do not reveal any blue staining.

(E) A 10-μm section of a transgenic nodule harboring the CaMV 35Sp-gus construct ROK2275 (Szabados et al., 1990) that was stained for GUS activity and observed using dark-field microscopy. Blue staining can be observed in the enlarged infected cells of the inner nodule and a reddish staining in the uninfected cells of the inner nodule and nodule cortex.

(F) The same section shown in (E), but observed using the light microscope. Dark blue staining can be observed in the infected cells of the inner nodule and light blue staining in uninfected and cortical cells.

Bars in (D), (E), and (F) = 100 μm.
to the mere presence of rhizobia, but rather to an infected-cell-specific expression pattern of the Srglb3 promoter. As a second control experiment, the expression pattern of another chimeric gene, consisting of the cauliflower mosaic virus (CaMV) 35S promoter fused to the gus gene (ROK2275; Szabados et al., 1990), was examined. When the 10-μm nodule sections were examined by light microscopy, a high level of GUS activity was detected in the infected (dark blue) cells, whereas a low level of activity was detected in the uninfected cells of the inner nodule and the nodule cortex (light blue-stained cells; Figure 2F). The same sections were also examined using dark-field microscopy. Under these conditions, we observed that the infected cells stained blue, whereas the uninfected cells in the inner nodule and in the nodule cortex stained red-purple (Figure 2E). This unique staining pattern of cells expressing a low level of GUS activity, as was observed under dark-field microscopy, has also been described by Miao et al. (1991). Thus, GUS activity can be detected in uninfected cells under our staining and microscopic conditions. The control experiments shown in Figures 2D to 2F allowed us to conclude from the data shown in Figures 2A and 2B that the Srglb3 5' upstream region (promoter) indeed directs a nodule-infected cell-specific expression pattern in transgenic Lotus plants.

The staining of hand sections of nodules harboring the different deletion/substitution constructs that are shown in Figure 1A revealed a strict correlation with the quantitative tissue specificity data presented in Figure 1B. The constructs directing strictly nodule-specific GUS activity (LP31, LP154, and LP213) showed staining patterns identical to the LP32 construct: cells in the nodule parenchyma, cortex, and vascular bundles were unstained, and the blue color was confined to the infected zone of the nodule. The constructs essentially lacking promoter activity (GUS expression), such as LP91, LP63, LP182, LP181, LP140, LP62, and the promoterless negative control LP100 (Figures 1A and 1B), did not reveal any GUS staining (data not shown). In addition, constructs LP45 and LP64 (Figure 1A), which directed a low level of expression in nodules and an increased level in roots (Figure 1B), showed a low level of GUS staining in the nodule cortex and infected zone (data not shown). Construct LP152 (Figure 1A), which directed a very high level of GUS expression in nodules and an elevated level in roots (Figure 1B), revealed a high level of GUS staining in the infected cells (infected zones). The latter construct also directed a low, but clearly detectable, level of blue color in the cortical cells (data not shown).

These results clearly showed that the −194 (DraI) to −116 (Hinfl) fragment of the Srglb3 promoter (NICE element) is essential for reporter gene expression in the infected zone of the nodule.

The NICE Element Can Impose a Nodule-Infected Cell-Enhanced Expression Pattern on a Heterologous Root-Enhanced Promoter

To further investigate the role of the NICE element in nodule-infected cell-specific expression, the −194 to −116 DraI-Hinfl fragment of the Srglb3 promoter region was inserted, in both orientations, into the SspI site at coordinate −57 of the −150 to +1 region of a truncated nos promoter. The resulting composite fragments were fused to the gus gene, and their structures are shown in Figure 3A (LP231 and LP232). As a control, the wild-type −150 to +1 fragment of the nos promoter, which was also fused to the gus gene, was used (LP62). These chimeric genes were introduced into Lotus plants, and nodule, root, and stem tissues of the resulting transgenic plants were assayed for GUS activity. The results are shown in Figure 3B. The truncated nos promoter (construct LP62) directed a low level of GUS activity in roots and an even lower level of GUS activity in nodules of transgenic Lotus plants. Insertion of the NICE element into the nos promoter resulted in a
threefold (LP231) and twofold (LP232) increase of GUS activity in nodules, respectively, while the GUS activity in roots remained constant (LP231) or was reduced (LP232).

Cytological GUS staining of 10-μm nodule sections revealed that the insertion of the NICE element into the truncated nos promoter element resulted in a nodule-infected cell-enhanced expression pattern, as shown in Figure 2C. We concluded from these results that the NICE element is not only essential for nodule-infected cell expression, but may, in fact, be sufficient for this specific expression pattern.

The NICE Element Contains Highly Conserved DNA Motifs

A previous comparison of the 5' upstream DNA sequences of the soybean and Sesbania lb genes revealed a high degree of homology between them. This is especially evident in a 400-bp region containing the soybean lbc3 OSE/INE and Sesbania glb3 NICE elements (Metz et al., 1988). To identify conserved DNA motifs within the NICE element of the Srglbg3 promoter, the equivalent regions in Sesbania, soybean, pea, alfalfa (Medicago sativa) and M. truncatula lb genes were aligned and their DNA sequences compared. The 5' upstream regions of hemoglobin genes from Trema and Parasponia were also included in the comparative analysis. The results are shown in Figure 4.

Two nodulin gene consensus sequences (AAAGAT and CTCTT) had previously been identified in the 5' upstream regions of late nodulin genes (Sandal et al., 1987). The AAAGAT motif was found in the Srglbg3 NICE element and in lb gene promoters from soybean (lba and lbc1 to lbc3), pea (Pslb), alfalfa (Ms1 and MspIb2), and the non-legume Parasponia (Para). A less conserved AAAGAT-like motif was identified at an analogous position in the promoter region of lb genes from M. truncatula (Mtlb1 and Mtlb2), as well as the hemoglobin gene promoter of the non-nodulated non-legume Trema. An extended conserved motif (TTGTCCTT), which includes the CTCTT consensus sequence identified by Sandal et al. (1987), was identified 2 bp downstream of the AAAGAT motif in all lb

![Figure 4. Conservation of DNA Motifs in the Stglb3 NICE Region and Analogous Regions of (Leg)hemoglobin Genes from Selected Legumes and Non-Legumes.](image-url)

A comparison of the DNA sequences in the 5' upstream regions of (leg)hemoglobin genes from S. rostrata (Metz et al., 1988), soybean (lba, lbc1, lbc2, and lbc3; Stougaard et al., 1987), pea (Pslb; Nap, 1988), M. truncatula (Mtlb1 and Mtlb2; Gallusci et al., 1991), M. sativa (Ms1 and Msplb2; Davidowitz et al., 1991), T. tomentosa (Trema; Bogusz et al., 1990), and P. andersonii (Para; Bogusz et al., 1990) is shown. The coordinates of the first base pairs shown in each line are as follows and were derived from the references cited above: glb3, -200; glb2, -200; lba, -131; lbc1, -126; lbc2, -127; lbc3, -139; Pslb, -198; Mtlb1, -147; Mtlb2, -151; Ms1, -205; Msplb2, -200. The coordinates of the soybean and M. truncatula promoters are relative to the transcription start site, whereas the coordinates of the other loci are relative to the initiator ATG. Regions of DNA sequence conservation are denoted by open or shaded boxes, whereby the shaded boxes denote those motifs discussed in detail in the text. Base pairs that differ from the consensus start site are denoted by a lowercase letter. The horizontal open box in the lbc3 line denotes the extent of the organ-specific element (OSE; Stougaard et al., 1990). The underlined sequences denote the putative CAAT (CCAAG) boxes. The base pairs above the lines were looped out to allow maximal alignment, and their position in the sequence is denoted with a small arrowhead.
promoter regions. Some sequence variation, however, was apparent in the case of the soybean \textit{iba} and \textit{ibc2} promoters.

In addition, a highly conserved CACCC(T) motif was found 17 bp downstream of the TTTGCTCTT motif in the Sesbania and soybean \textit{lb} promoters, whereas the other \textit{lb} and hemoglobin promoters showed sequence divergence in the corresponding regions. Interestingly, an analogous CACCC element has been identified in the 5' upstream region of mammalian \beta-globin genes, as shown in Figure 5A, where it is absolutely essential for gene expression (CACCC; Dierks et al., 1983; Walters and Martin, 1992). Ten base pairs downstream of the CACCC motif, a conserved AAGAG motif, a reverse complement of the CTCTT sequence element, could also be identified; it overlaps the presumptive CAAT (CCAAG) box proposed by Metz et al. (1988).

Thus, the Srglb3 NICE element shares extensive overall DNA homology and specific conserved DNA motifs with the analogous regions of other \textit{lb} promoters, the promoters of hemoglobin genes from non-legume plants, and possibly even cis-acting elements in mammalian \beta-globin promoters, suggesting a functional role of these sequences in gene regulation.

**Site-Specific Mutagenesis of Conserved DNA Motifs and Flanking Sequences in the Srglb3 NICE Element and of the Conserved ATRE-BS2* Element**

To examine the functional significance of the conserved DNA motifs identified in Figure 4, the Srglb3 NICE region was subjected to site-specific mutagenesis. For this analysis, the –214 (Dral) to –148 (NlaIV) region was selected (Figure 1A). This region contains all the conserved motifs discussed previously, except for the overlapping CCAAG and AAGAG motifs. In addition, it carries the ATRE-BS2* element (ATTTTAATTATTAA), shown to interact with multiple trans-acting factors (Metz et al., 1988) and highly conserved in other nodulin and non-nodulin plant gene promoters (see Forde, 1994). Mutant 66-bp (Dral-NlaIV) oligonucleotides were synthesized by using a mixture of 97% wild-type and 3% non-wild-type nucleotides and were cloned in plasmid vector pLK225 (see Methods). Approximately 150 cloned inserts were subjected to DNA sequence analysis, yielding a total of 127 unique mutant oligonucleotides (data not shown). Single and double base-pair changes comprised the most predominant classes of mutant oligonucleotides (38 and 34, respectively).

The mutant oligonucleotides were used to reconstruct full –1914 to –51 Srglb3 5' upstream fragments, which differed from the wild-type 5' upstream region by only the respective single or double mutation. The mutant promoter fragments were fused to the gus reporter gene, and the resulting chimeric genes were introduced into \textit{Lotus corniculatus} plants. GUS activity was measured in different tissues of these transgenic plants, and the results of 9 to 31 independent transformants were averaged.

**Effect of Mutations in the ATRE-BS2* Element on Srglb3 Promoter Activity**

An internal deletion of the –245 to –194 region of the –431 to –51 Srglb3 promoter fragment (which removes the ATRE-BS2* element and flanking DNA sequences) was found to result in an approximately twofold reduction of GUS expression in transgenic \textit{Lotus} plants (LP213 versus LP31; Figure 1). To determine if this reduction was due to the absence of specific ATRE-BS2* sequences, the effects of several single and double mutations in the 3' half of the element on Srglb3 promoter activity were examined. In addition, the effect of a fortuitous...
Leghemoglobin Gene Regulation 325

A

B

Figure 6. Effect of Mutations in the ATRE-BS2* on Srglb3 Promoter Activity.

(A) Structure of the wild-type and mutant constructs used in this study. Construct LP14 is identical to LP32 but was generated using the wild-type synthetic Dral-NIalV fragment. The reconstruction scheme used is as described in Methods for the mutant synthetic fragments. The position of the NICE element is indicated by an open box. The ATRE-BS2* (BS2*) sequence is underlined. The corresponding mutated regions in the LP14-derived constructs are included in a shaded box. The dashes indicate base pairs that are identical to the wild-type ATRE-BS2* sequence. The changed base pairs are as indicated by upper-case letters. The Δ symbol indicates the internal ATRE-BS2* deletion. (B) GUS expression levels in different transgenic L. corniculatus tissues harboring the constructs shown in (A). The specific base pair changes in each construct are denoted by the horizontal arrows next to the GUS activity columns. The results shown for each construct represent the averaged values for 9 to 31 independent transgenic plants (Table 1). For further details see the legend of Figure 1. wt, wild type.

Internal deletion in the ATRE-BS2* element, isolated during the site-specific mutagenesis experiment, was analyzed. The position and extent of the mutations are shown in Figure 6A, and their effect on GUS expression is presented in Figure 6B.

T-to-C mutations in two adjacent T residues of the ATTTTT-AATTATAATTAAA motif (LP14-23 and LP14-28) resulted in a reduction of Srglb3 promoter activity to 30% and 65% of wild-type (LP14) expression, respectively. An A-to-C mutation at the ATTTTTAAAATTATAATTAAA position in ATRE-BS2* (LP14-17) reduced GUS expression to 57% of the wild-type level. However, a TA-to-AT mutation at the ATTTTTAAAATTATAATTAAA position (LP14-9) and a T-to-A mutation at the ATTTTTAAAATTATAATTAAA position (LP14-1) did not influence Srglb3 promoter activity significantly. The T-to-G mutation at position ATTTTTAAAATTATAATTAAA (LP14-12) and an A-to-T mutation at the penultimate position of the ATRE-BS2* element (LP14-10) resulted in a reduction of Srglb3 promoter activity to 35% and 65% of wild-type expression, respectively. In contrast, an A-to-T mutation at the ultimate position of the ATRE-BS2* element (LP14-19) had a negligible effect on Srglb3 promoter activity. Surprisingly, the internal ATRE-BS2* deletion mutant (LP14-27) had only a minor effect (25% reduction of expression) on Srglb3 promoter activity. Finally, in not one case did the introduced mutations change the tissue-specific expression pattern of the chimeric Srglb3-gus genes.

These results suggested that the ATRE-BS2* element affects Srglb3 promoter activity in a quantitative manner but is not essential for nodule-specific expression. The results also showed that the effect of specific, single base pair mutations in the ATRE-BS2* element is similar to or even more extreme than the effect of the large deletion (LP213), as well as the ATRE-BS2* internal deletion (LP14-27). The reason for this apparent discrepancy is unknown, but this phenomenon may be related to the redundancy of ATRE (BS2*-like) elements in the glb3 promoter and their relative location.

Effect of Mutations in the NICE Element on Srglb3 Promoter Activity

The results of the site-specific mutagenesis of the NICE region are shown in Figure 7. A total of 19 specific mutations in the NICE region, starting with the A residue immediately preceding the AAAGAT motif, were selected for the cis analysis. Single or double mutations in the conserved (A)AAAGAT motif of the Srglb3 promoter region (constructs LP14-2, LP14-5, LP14-11, LP14-15, and LP14-30) resulted in a significant (two- to fivefold) decrease in GUS activity. However, the nodule-specific expression pattern directed by these mutant promoter fragments was not altered. Surprisingly, a mutation in the A residue located in between the conserved (A)AAAGAT and TTGTCTCTT motifs (LP14-3) resulted in a 2.5-fold increase in nodule-specific Srglb3 promoter activity.

Mutations in the CTCTT residues of the conserved TTGTCTCTT motif (LP14-7, LP14-26, and LP14-21) resulted in a severe
reduction of GUS activity, corresponding to a virtual abolishment of Srglb3 promoter activity. Mutations in the T residue immediately preceding the CTCTT motif (LP14-13 and LP14-25) had a much less drastic effect, reducing Srglb3 promoter activity to 50 to 60% of wild-type expression. A double mutation in the AT residues immediately 3' of the TTGTCTCTT motif (LP14-24) did not have a discernible effect on Srglb3 promoter activity.

A mutation in the A residue of the conserved CACCC(T) motif (LP14-6) resulted in a complete abolishment of Srglb3 promoter activity. On the other hand, a mutation in the third C residue of the CACCC(T) motif (LP14-22) reduced Srglb3 promoter activity threefold, while mutating the terminal T residue (LP14-18) did not affect the expression measurably.

The effects of mutations in a TCAATG oligonucleotide, which is located two residues 5' of the CACCC(T) motif (Figure 7A), were found to be variable. Changing TCA to ACG (LP14-31) abolished Srglb3 promoter activity, while mutating the central C residue (LP14-20) reduced the promoter activity to only 60% of the wild-type level. In addition, mutating the A of the ATG triplet (LP14-4) resulted in a severe reduction of Srglb3 promoter activity, while mutating the T of this triplet (LP14-8) had only a minor effect on expression. Cytological staining of GUS activity in nodules induced on transgenic *Lotus* plants harboring constructs with mutations in the (A)AAAGAT motif revealed that the (reduced) Srglb3 promoter expression patterns were still nodule infected-tissue specific (data not shown). Likewise, promoter constructs carrying mutations in the CTCTT motif (e.g., LP14-7) still directed a low level of GUS expression in the infected tissue of the nodules. On the other hand, no staining of any tissues could be observed in nodules induced on plants harboring the LP14-26 or LP14-21 constructs, even after 15 hr of incubation in the presence of the 5-bromo-4-chloro-3-indolyl β-D-glucuronide substrate (data not shown).

These results clearly showed that the conserved (A)AAAGAT, TTGTCTCTT, and CACCC(T) motifs in the NICE element identified in Figure 4 play an essential role in Srglb3 promoter activity. They also suggested the importance of a less conserved oligonucleotide, TCAATG, immediately upstream of the CACCC(T) element, for Srglb3 promoter activity. The latter element is located at the Ball site (Figures 1A and 7A) that had

---

**Figure 7.** Effect of Mutations in the NICE Element on Srglb3 Promoter Activity.

(A) Structure of the wild-type and mutant constructs used in this study. Construct LP14 is as described in the legend of Figure 6. The conserved AAAGAT, TTGTCTCTT, and CACCC(T) motifs (see Figure 4) are included in shaded boxes, and the conserved AT and ATG motifs (see Figure 4) are in an open box. The TAATGAT sequence, resembling the yeast GCN4 transcription factor binding site (Barn; She et al., 1993), is underlined.

(B) GUS expression levels in different transgenic *L. corniculatus* tissues harboring the constructs shown in (A). The specific base pair changes in each construct are denoted by the horizontal arrows next to the GUS activity columns. The results shown for each construct represent the averaged value for 9 to 31 independent transgenic plants (Table 1). For further details, see the legends to Figures 1 and 6.
been shown to be important for Srglb3 expression by examining the effect of the insertion of a 20-bp linker sequence (LP182; Figure 1). In addition, this TCAATG oligonucleotide overlaps an octamer motif TAATTGCA (underlined in Figure 7A), which has been shown to be highly conserved in lb promoters of different plant species and shares homology with the binding site of the yeast transcription factor GCN4 (lb155; She et al., 1993).

**DISCUSSION**

In this study, we report the delimitation of a 78-bp region of the S. rostrata glb3 5′ upstream (promoter) region, which is essential for the expression of chimeric Srglb3-gus reporter genes in infected cells of nodules induced on transgenic Lotus plants. This element extends from position −194 to −116 (relative to the Srglb3 ATG) or from coordinates −146 to −68 relative to the major transcriptional start site (Metz et al., 1988). In addition, it appeared to be sufficient to impose a nodule-specific expression pattern on a normally root-enhanced nos promoter element. Based on these results, this element has been designated the NICE element. The NICE element is located in the 380-bp region of the Srglb3 promoter, which had previously been shown to be sufficient for nodule-specific expression (Szabados et al., 1990). Our cytological staining experiments strongly suggested that the expression of the Srglb3 promoter is limited to the infected cells of the nodule and that the NICE element is essential for directing this tissue specificity.

In the case of indeterminate nodules, strong evidence has been presented in support of the hypothesis that lb gene expression is strictly limited to the infected cells of the nodule tissue (de Billy et al., 1991). However, in the case of determinate nodules, the evidence is less clear. Kouchi et al. (1989) showed that in soybean nodules ~3% of the total amount of leghemoglobin is present in uninfected cells. Because fractionated nodule tissues were used in this study, it cannot be ruled out that a low level of contamination of the uninfected cell preparation with infected cells is responsible for this observation (Kouchi et al., 1989).

Moreover, using immunochemical methods, VandenBosch and Newcomb (1988) showed that the leghemoglobin apoprotein is present in uninfected (interstitial) cells of soybean nodules. Unfortunately, in L. corniculatus nodules, no direct measurements of the levels of Lb protein or lb gene mRNA in infected versus uninfected cells have been performed. However, our observations are in agreement with the conclusion reached by Lauridsen et al. (1993) that a chimeric soybean lbc3-gus construct was only expressed in cells of transgenic L. corniculatus nodules invaded by Rhizobium.

In the equivalent region of the soybean lbc3 promoter, two important elements have been identified: an OSE (~139 to −102) relative to the start point of transcription, which is required but not sufficient for nodule-specific expression, and an NE (−102 to −49), which is apparently involved in repression of lbc3 promoter activity (Stougaard et al., 1987, 1990). Together, OSE and NE were found to be able to confer a nodule-specific (enhanced) pattern of gene expression on the CaMV 35S promoter in transgenic L. corniculatus plants (Lauridsen et al., 1991). These results and the observations described here suggested a functional conservation of cis-acting elements in the promoter regions of lb genes responsible for nodule (infected-cell)–specific expression, as proposed previously (de Bruijn and Schell, 1992).

To a certain extent, this is reflected in the observed DNA sequence homologies in this region of various lb promoters (de Bruijn and Schell, 1992; Figure 4). The lbc3 OSE is highly conserved at the equivalent position in the Srglb3 promoter region (32 of 38 identical base pairs; Metz et al., 1988; Szabados et al., 1990; Figure 4). However, the overall homology between the lbc3 NE and the analogous glb3 region is much lower (Metz et al., 1988; Figure 4). The most conserved DNA motifs in the part of the Srglb3 NICE element that share a high degree of homology with the soybean lbc3 OSE include the AAAGAT motif (Sandal et al., 1987; de Bruijn and Schell, 1992) and the TTGTCTCTT motif (de Bruijn and Schell, 1992; Figure 4), which contain the CTCTT sequence proposed by Sandal et al. (1987). Our mutagenesis experiments presented here clearly showed that these conserved DNA sequences (nodulin gene consensus sequences) are important for proper functioning of the Srglb3 promoter. The AAAGAT motif appears to be less important for promoter activity, because mutations in this motif resulted in a reduction of nodule-specific expression. In contrast, the CTCTT portion of the TTGTCTCTT motif is clearly essential for promoter activity, because mutations in this DNA sequence virtually abolish (nodule-specific) expression.

Mutations in the DNA sequence flanking the AAAGAT motif (LP14-2) or the CTCTT motif (LP14-24) do not affect Srglb3 promoter activity (Figure 7), suggesting that the observed phenotypes of the nodulin gene consensus sequence mutations are highly specific. Similar results have recently been described by Ramlov et al. (1993) for the CTCTT and AAAGAT motifs of the soybean lbc3 promoter. Interestingly, the mutation of the A residue immediately preceding the TTGTCTCTT motif (LP14-3) results in a "promoter-up" phenotype (Figure 7), which has not yet been observed in other nodulin promoter systems.

Our insertion (LP181 and LP182; Figure 1) and substitution (Figure 7) mutagenesis experiments also revealed the importance of a DNA sequence (TCAATG), located 7 bp downstream of the TTGTCTCTT motif, for Srglb3 promoter activity. This DNA sequence is only partially conserved in other lb promoters (Figure 4) and has not been implicated in lb promoter activity previously. The equivalent sequence is positioned at the border of the OSE and NE elements in the soybean lbc3 promoter, but its in vivo significance has not yet been described.

The fourth DNA sequence motif, which is clearly essential for Srglb3 promoter activity, is the highly conserved CACCCT motif. A single base pair substitution, changing the DNA sequence from CACCCT to CTCCCT (LP14-6), clearly abolished
promoter activity (Figure 7). This motif, located 8 bp upstream of the putative Srglb3 CAAT (CCAAG) box (Figure 4), is of particular interest, because an analogous motif (CACCC) has been identified either 9 or 23 to 25 bp upstream of the CAAT box in the promoter region of human adult β-globin and β-globin-like genes (Dierks et al., 1983). Mutations introduced into the human β-globin gene CACCC element strongly reduced promoter strength (Walters and Martin, 1992), and a naturally occurring mutation in this element has been associated with decreased β-globin gene activity and the thalassemia syndrome (Walters and Martin, 1992).

Closely linked cis-acting elements to the CACCC element in mammalian β-globin gene promoters, such as the binding site for the erythroid-specific trans-acting factor EryFl, have been identified; they are important for tissue (erythroid)-specific gene expression (Walters and Martin, 1992). The binding site for the EryFl factor includes a consensus motif (WGATWR; W=A or T; R=A or G; Figure 5B), which has also been referred to as the GATA binding site (Evans et al., 1990). The erythroid-specific expression of β-globin genes appears to involve the interaction between different protein–DNA complexes (Walters and Martin, 1992): EryFl bound to the GATA-1 binding site, trans-acting factors bound to the CACCC element, and/or a third trans-acting factor bound to its cognate cis-acting element (AP-1/NFE2 element).

Interestingly, nodulin gene consensus sequences, found in proximity to the CACCC element of the Srglb3 promoter region, resembled GATA-1 binding sites of the mammalian β-globin genes (Figure 5C). For example, the AGATTA sequence, which constitutes a part of the AAAGAT motif, fully matches the WGATWR consensus. In addition, two DNA sequences (AGAGAA and AGGAA), with one mismatch to the WGATWR consensus sequence, are present in a partly overlapping, inverse pattern on the complementary strand in the TTGCTCTCTT motif (Figure 5C). As pointed out above, mutations in these DNA sequences clearly affected Srglb3 promoter activity. Moreover, it is interesting to note that a change of the Srglb3 AGATTA sequence to AGATTC resulted in an approximately threefold increase in promoter activity (LP14-3; Figure 7). This closely resembles the effect (three- to fourfold increase of promoter activity) of a naturally occurring T-to-C mutation in one of the GATA-1 binding sites of the mammalian β-globin promoter (Figure 5A), which leads to the hereditary persistence of fetal hemoglobin (HPFH phenotype; see Evans et al., 1990; Walters and Martin, 1992). It has been suggested that plant and animal globins share a common ancestor, based on protein structure predictions (Landsmann et al., 1988). Whether this protein homology could be extended to structural and/or functional similarities in cis-acting elements of leghemoglobin and β-globin gene promoter elements is speculative at this time. However, studies on the possible relationship between proteins interacting with the NICE element (de Brouijn et al., 1994) and those interacting with similar cis-acting elements in the β-globin gene promoters (e.g., GATA binding proteins; Walters and Martin, 1992) are currently being conducted.

The mutagenesis analysis presented here also involved ATRE-BS2" (ATTTTTAAATTATTTAA; Metz et al., 1988; Figure 6), which is located immediately upstream of the NICE element. This element is highly conserved at the analogous position in the soybean lbc3 and N23 promoters, as well as the French bean glpy 5' upstream region and has been shown to interact specifically with DNA binding proteins from nodules, leaves, and roots (Jensen et al., 1988; Metz et al., 1988; Forde et al., 1990; Jacobsen et al., 1990). ATREs have been found in the promoter regions of a number of other plant genes and, in some cases, have been shown to act as positive regulatory elements (see Forde, 1994). However, the importance of the ATREs in the nodulin gene promoters has been unclear thus far (see de Brouijn and Schell, 1992; Forde, 1994). The mutations in the Srglb3 ATRE-BS2" element examined here have diverse effects on promoter activity, ranging from a threefold reduction (LP14-23 and LP14-12) to a slight enhancement of reporter gene expression (Figure 6). However, while apparently serving as a positive regulatory element, ATRE-BS2" is clearly not essential for nodule-specific expression. The distinct effects on gene expression of mutations in Srglb3 ATRE-BS2" may be related to their differential interaction with tissue-specific DNA binding proteins. This conclusion is supported by results reported by Jacobsen et al. (1990), who observed that oligonucleotides corresponding to the soybean lbc3 BS1 wild-type equivalent or those carrying specific base pair changes therein formed distinct protein–DNA complexes with NAT1 and LAT1 (HMG-like) and NAT2 proteins present in nuclear extracts from nodules and leaves. The differential effect of an internal deletion in ATRE-BS2" versus single base pair mutations in this element may be related to the relative distance from ATRE-BS2" to the start point of transcription and/or to bringing redundant upstream ATREs (Welters et al., 1993) in closer proximity to the transcription start site in the deletion construct. Moreover, specific mutations may be affecting protein binding to a particular face of the helix and thereby alter protein–protein or protein–DNA interactions in the downstream promoter region required for proper Srglb3 gene expression.

It is clear from the promoter analysis presented here and from other published observations on both the Sesbania and soybean lbc 5' upstream regions (see de Brouijn and Schell, 1982; Lauridsen et al., 1993; Ramlov et al., 1993; She et al., 1993) that the promoters of late nodulin genes in legumes are highly complex and contain cis-acting elements required for cell-specific expression and nonspecific enhancement of gene expression, as well as binding sites for various trans-acting factors. What is known to date about the primary structure of the S. rostrata glb3 promoter region is summarized in Figure 8. Although the exact role of several cis-acting elements and binding sites for trans-acting factors remains to be elucidated, it appears from these studies that a relatively short (78-bp) region of the Srglb3 promoter, the NICE element, may be sufficient to confer a low level of infected-cell–specific expression. Distinct conserved DNA motifs in the NICE element...
shown is a schematic diagram highlighting the currently known features (cis-acting elements and binding sites for trans-acting factors) of the Srglb3 promoter region. The positive element denotes the region shown by Szabados et al. (1990) to be involved in enhancement of Srglb3 promoter activity. The BBS1 element (lightly shaded box) represents the binding site for the bacterial binding protein 1 (BBP1) that was identified by Welters et al. (1993). The BBS1 DNA sequence is shown between the brackets above and below the black box. The position and nature of the mutations in these motifs having the most significant effect on Srglb3 promoter activity (see Figure 7) are indicated by open boxes and vertical arrows. For further details, see the legend of Figure 1 and the text.

**Figure 8. Structure of the Srglb3 Promoter Region.**

The observation that the Srglb3 promoter appears to be expressed in the infected versus uninfected cells of the nodule fits in well with our recent hypothesis that DNA binding protein(s), which are derived from the intracellular rhizobial bacteroids and interact with the promoter region of the Srglb3 gene in the nucleus (e.g., the bacterial binding site BBS1; Figure 8), may be involved in infected-cell-specific lb gene expression (Welters et al., 1993). However, the further elucidation of the molecular basis for tissue-specific expression of lb and other late nodulin genes must await the results of the characterization and silencing of the rhizobial and plant genes encoding the trans-acting factors that interact with the nodulin gene promoter regions; further analysis of the cis-acting elements depicted in Figure 8 is also required. In addition, the role of chromatin structure in tissue-specific nodulin gene expression will need to be examined. These studies are now in progress.

**METHODS**

**Construction of Vectors and Gene Fusions**

The construction of pLP32, pLP31, pLP45, pLP63, and pLP64 has been described by Szabados et al. (1990), and the structure of the resulting chimeric Sesbania rostrata leghemoglobin gb3 (Srglb3)-β-glucuronic-dase (gus) reporter gene is shown in Figure 1A. The pLP100 vector was derived from the previously described binary vector pLP17 (Szabados et al., 1990) by replacement of the HindIII-KpnI polylinker of pLP17 with a composite EcoRI-KpnI (EcoRI, Clal, HindIII, Sphl, Pstl, SalI, XbaI, BamHI, SmaI, and KpnI) polylinker. In addition, the two EcoRI sites flanking the nos polyadenylation sequence of pLP17 were removed (F. Ratet and F. J. de Bruijn, unpublished results). To construct pLP140, a 57-bp SspI-KpnI fragment, carrying the -56 (SspI) to +1 region of the nos promoter (Ebert et al., 1987), which was fused to polylinker sequences on the 3’end (including the KpnI site), was inserted into the pLP100 binary vector restricted with Smal and KpnI. Plasmid pLP154 was constructed by replacing the EcoRI-BamHI portion of the pLP140 polylinker with the -387 EcoRI to -116 HindIII fragment of the Srglb3 5’ upstream region. Plasmid pLP213 was derived from pLP31 by deleting the -245 SspI to -194 Dral fragment of the pLP31 5’ upstream region. Plasmid pLP91 was constructed by first replacing the SspI-HindIII-Smal fragment of the pSU19 (Yanisch-Perron et al., 1985) polylinker with the -194 Dral to -51 FokI fragment of the Srglb3 5’ upstream region to generate the intermediate vector pLP90. Subsequently, the HindIII-KpnI polylinker fragment of pLP100 was replaced with a HindIII-KpnI fragment of pLP80, which carries the -194 to -51 Srglb3 sequences. Plasmid pLP152 was derived from pLP91 by fusing the -150 to -57 fragment of the nos promoter (Ecer, Ebert et al., 1987) to the -194 to -51 Srglb3 region at the -194 Dral site. Plasmid pLP152 was derived from pLP91 by insertion of a 20-bp synthetic BglII linker at the unique -161 Ball site in the Srglb3 5’ upstream region. Plasmid pLP181 was derived from pLP182 by fusing the -150 to -57 nos enhancer fragment to the -431 to -51 region at the -431 Ball site. Plasmid pLP262 was constructed by replacing the HindIII-KpnI fragment of the polylinker of pLP100 with the HindIII-KpnI fragment of pLP18 (Szabados et al., 1990), which carries the -150 to +1 nos promoter fragment (Ebert et al., 1987).
the mutagenesis protocol. The DNA sequence of the reconstructed -148 NlaIV fragments was linearized with NsiI, and the resulting termini were rendered blunt ended by using T4 DNA polymerase. After further digestion with KpnI, the resulting DNA fragments were fused to a NlaIV-KpnI fragment, which carries the missing -148 to -51 NlaIV region carrying the mutations was verified. Finally, the EcoRI-KpnI fragment of the Srglb3 (wild-type) promoter. This resulted in the reconstruction of 1.9-kb upstream regions fused to the HindIII-KpnI fragment, which carries the mutant EcoRI-KpnI fragments of the 31 intermediate constructs. This resulted in the reconstruction of 1.9-kb Srglb3 5' upstream regions, carrying the desired mutations, fused to the gus reporter gene.

Site-Specific Mutagenesis and Mutant Promoter Reconstruction

The Dral-NlaIV (-214 to -148) Srglb3 promoter fragment was mutagenized using degenerate oligonucleotides, as described by Hill et al. (1987). Briefly, non-wild-type Dral-NlaIV fragments were prepared via DNA synthesis on an Applied Biosystem (Foster City, CA) DNA synthesizer by including low concentrations (3%) of the three non-wild-type nucleotide precursors at each position during the synthesis. Two restriction sites, 5'-NsiI and KpnI-3', were introduced at the 3' terminus of the synthetic oligonucleotides to facilitate the Srglb3 promoter reconstruction procedure. The products of the DNA synthesis, a single-stranded mixture of degenerated oligonucleotides, were converted to the double-stranded form by mutually primed synthesis starting at the 3' palindromic KpnI sequence. The resulting double-stranded molecules were digested with KpnI and used to replace the Srglb3 Dral-KpnI fragment in the vector pLK225 (see above), thereby simultaneously generating a NlaIV-FokI internal deletion. Thus, a library of recombinant plasmid molecules carrying cloned EcoRI-NlaIV fragments with single and multiple base pair substitutions were obtained. DNA preparations from 150 transformants were subjected to nucleotide sequence analysis using the Sanger method (Sanger et al., 1977) to determine the nature of the randomly introduced mutations. Thirty-one specific mutant fragments were used to reconstruct (mutant) 1.9-kb Srglb3 5' upstream regions fused to the gus reporter gene.

Plasmid DNA from pLK225 derivatives carrying mutant -387 EcoRI to -148 NlaIV fragments was linearized with NsiI, and the resulting termini were rendered blunt ended by using T4 DNA polymerase. After further digestion with KpnI, the resulting DNA fragments were fused to a NlaIV-KpnI fragment, which carries the missing -148 to -51 NlaIV-FokI fragment of the Srglb3 (wild-type) promoter. This resulted in the double-stranded Dral-NlaIV (-214 to -148) Srglb3 promoter fragment was mutagenized using degenerate oligonucleotides, as described by Hill et al. (1987). Briefly, non-wild-type Dral-NlaIV fragments were prepared via DNA synthesis on an Applied Biosystem (Foster City, CA) DNA synthesizer by including low concentrations (3%) of the three non-wild-type nucleotide precursors at each position during the synthesis. Two restriction sites, 5'-NsiI and KpnI-3', were introduced at the 3' terminus of the synthetic oligonucleotides to facilitate the Srglb3 promoter reconstruction procedure. The products of the DNA synthesis, a single-stranded mixture of degenerated oligonucleotides, were converted to the double-stranded form by mutually primed synthesis starting at the 3' palindromic KpnI sequence. The resulting double-stranded molecules were digested with KpnI and used to replace the Srglb3 Dral-KpnI fragment in the vector pLK225 (see above), thereby simultaneously generating a NlaIV-FokI internal deletion. Thus, a library of recombinant plasmid molecules carrying cloned EcoRI-NlaIV fragments with single and multiple base pair substitutions were obtained. DNA preparations from 150 transformants were subjected to nucleotide sequence analysis using the Sanger method (Sanger et al., 1977) to determine the nature of the randomly introduced mutations. Thirty-one specific mutant fragments were used to reconstruct (mutant) 1.9-kb Srglb3 5' upstream regions fused to the gus reporter gene.

Plasmid DNA from pLK225 derivatives carrying mutant -387 EcoRI to -148 NlaIV fragments was linearized with NsiI, and the resulting termini were rendered blunt ended by using T4 DNA polymerase. After further digestion with KpnI, the resulting DNA fragments were fused to a NlaIV-KpnI fragment, which carries the missing -148 to -51 NlaIV-FokI fragment of the Srglb3 (wild-type) promoter. This resulted in the double-stranded Dral-NlaIV (-214 to -148) Srglb3 promoter fragment was mutagenized using degenerate oligonucleotides, as described by Hill et al. (1987). Briefly, non-wild-type Dral-NlaIV fragments were prepared via DNA synthesis on an Applied Biosystem (Foster City, CA) DNA synthesizer by including low concentrations (3%) of the three non-wild-type nucleotide precursors at each position during the synthesis. Two restriction sites, 5'-NsiI and KpnI-3', were introduced at the 3' terminus of the synthetic oligonucleotides to facilitate the Srglb3 promoter reconstruction procedure. The products of the DNA synthesis, a single-stranded mixture of degenerated oligonucleotides, were converted to the double-stranded form by mutually primed synthesis starting at the 3' palindromic KpnI sequence. The resulting double-stranded molecules were digested with KpnI and used to replace the Srglb3 Dral-KpnI fragment in the vector pLK225 (see above), thereby simultaneously generating a NlaIV-FokI internal deletion. Thus, a library of recombinant plasmid molecules carrying cloned EcoRI-NlaIV fragments with single and multiple base pair substitutions were obtained. DNA preparations from 150 transformants were subjected to nucleotide sequence analysis using the Sanger method (Sanger et al., 1977) to determine the nature of the randomly introduced mutations. Thirty-one specific mutant fragments were used to reconstruct (mutant) 1.9-kb Srglb3 5' upstream regions fused to the gus reporter gene.

Plant Transformation

Transgenic Lotus corniculatus cv Rodeo plants were generated as described by Szabados et al. (1990). Binary vectors were transformed into Agrobacterium rhizogenes A4 (Tempe and Casse-Delbart, 1989) by using the freeze–thaw method (Holgen and Willmitzer, 1988). Tissues derived from 5 to 14 independent transgenic plants were tested for each deletion/substitution construct shown in Figures 1 and 3 (see Table 1). Between 9 and 31 plants were analyzed for individual mutated promoter derivatives shown in Figures 6 and 7 (see Table 1). Regenerated plants were transferred to a growth chamber (Conviron, Asheville, NC) (25°C; 8-hr/16-hr photoperiod) and inoculated with a 48-hr culture of Rhizobium loti NZP2037 (Pankhurst et al., 1986). Five weeks after inoculation, stem, root, and nodule tissues were harvested, frozen in liquid nitrogen, and stored until analyzed.

Fluorometric Assay

The quantitative determination of GUS activity in various organs of transgenic L. corniculatus plants was conducted as described by Jefferson et al. (1987). Enzymatic activity was expressed as picomoles of 4-methylumbelliferone produced per minute per milligram of protein in the extract. The latter was determined by using the Bradford assay and BSA as a standard (Bradford, 1976). For the analysis of transgenic plants harboring the constructs shown in Figure 1, a fluorometer (model LS-28; Perkin-Elmer) was used, whereas the other samples were analyzed using a fluorescence spectrophotometer (model F-2000; Hitachi, Tokyo, Japan).

GUS Histochemical Staining

GUS activity in the hand sections was histochemically analyzed as described previously (Jefferson et al., 1987; Szabados et al., 1990). The 10-μm sections were analyzed using the method of De Block and DeBrouwer (1992), except that the nodule tissues were first stained for GUS activity using the 5-bromo-4-chloro-3-indolyl β-glucuronide substrate and subsequently embedded in historesin (Feichter-Jung; Cambridge Instruments, Heidelberg, Germany). The 10-μm sections were made using a standard glass microtome. Stained sections were examined by dark- and bright-field microscopy using an Axioskop microscope (Zeiss, Oberkochen, Germany).

ACKNOWLEDGMENTS

We thank Jeff Schell and the Max Planck Gesellschaft for support during early stages of this project carried out at the Max Planck-Institut für Züchtungsforschung in Cologne (Germany). K.S. was the recipient of a Max Planck Gesellschaft postdoctoral stipend. D.S. acknowledges the U.S. Department of Agriculture Food and Agricultural Sciences National Needs Graduate Fellowship Program for predoctoral support. We also thank Pascal Ratet for his help with constructing some of the vectors, Britta Gronenberg at the Max-Planck-Institut for technical assistance, Uwe Roosbach, Marlene Cameron, and Kurt Steppitz for their help in preparing the figures, and Natasha Raikhel (and coworkers) for advice and use of microscopic facilities. We also thank Jens Stougaard, Erik Østergaard Jensen, Kjeld Marcker, and Brian Forde for helpful discussions and sharing information prior to publication. This work was supported by U.S. Department of Agriculture Grants Nos. 90-37262-5781 and 92-37305-7735, as well as U.S. Department of Energy Grant No. DE-FG02-90ER20021 to F.JdB.
Leghemoglobin Gene Regulation

Received November 1, 1993; accepted January 17, 1994.

REFERENCES


elements in the soybean leghemoglobin lbc3 5' upstream region. EMBO J. 7, 1265–1271.


Site-specific mutagenesis of the nodule-infected cell expression (NICE) element and the AT-rich element ATRE-BS2* of the Sesbania rostrata leghemoglobin glb3 promoter.

K Szczyglowski, L Szabados, S Y Fujimoto, D Silver and F J de Bruijn

*Plant Cell* 1994;6;317-332

DOI 10.1105/tpc.6.3.317

This information is current as of July 9, 2017