Nuclear Factors Interact with Conserved A/T-Rich Elements Upstream of a Nodule-Enhanced Glutamine Synthetase Gene from French Bean

Brian G. Forde, Jacqueline Freeman, Jane E. Oliver, and Manuel Pineda
Biochemistry and Physiology Department, AFRC Institute of Arable Crops Research, Rothamsted Experimental Station, Harpenden AL5 2JQ, United Kingdom

The gln-\(\gamma\) gene, encoding the \(\gamma\) subunit of glutamine synthetase in French bean (Phaseolus vulgaris), is strongly induced during nodule development. We have determined the nucleotide sequence of a 1.3-kilobase region at its 5' end and have identified several sequences common to the promoter regions of late nodulin genes from other legume species. The 5'-flanking region was analyzed for sequence-specific interactions with nuclear factors from French bean. A factor from nodules (PNF-1) was identified that binds to multiple sites between \(-860\) and \(-154\), and a related but distinct factor (PRF-1) was detected in extracts from uninfected roots. PNF-1 and PRF-1 bound strongly to a synthetic oligonucleotide containing the sequence of an A/T-rich 21-base pair imperfect repeat found at positions \(-516\) and \(-466\). The same factors also had a high affinity for a protein binding site from a soybean leghemoglobin gene and appeared to be closely related to the soybean nodule factor NAT2, which binds to A/T-rich sequences in the \(lbc_3\) and nodulin 23 genes (Jacobsen et al. (1990). Plant Cell 2, 85–94]. Comparison of NAT2/PNF-1 binding sites from a variety of nodulin genes revealed the conservation of the short consensus core motif TATTTWAT, and evidence was obtained that this sequence is important for protein recognition. Cross-recognition by PNF-1 of a protein binding site in a soybean seed protein gene points to the existence of a ubiquitous family of factors with related binding affinities. Our data suggest that PNF-1 and PRF-1 belong to an evolutionarily conserved group of nuclear factors that interact with specific A/T-rich sequences in a diverse set of plant genes. We consider the possible role of these factors in coregulating the expression of gln-\(\gamma\) and other late nodulin genes.

INTRODUCTION

The development of an effective, nitrogen-fixing symbiosis in root nodules of the Leguminosae depends on a complex series of interactions between the host plant and the invading rhizobia. The plant and the Rhizobium each respond by synthesizing a group of proteins (termed nodulins) that are characteristic of the symbiotic state (Morrison et al., 1988). There are at least 20 to 30 plant-specified nodulins, of which the best characterized and most abundant are the leghemoglobins, whose function is to facilitate \(\text{O}_2\) diffusion to the bacteroids. A second major adaptation is the accumulation in nodules of high levels of glutamine synthetase (GS; EC 6.3.1.2), the plant enzyme catalyzing the first step in the assimilation of the ammonia that is generated by the nitrogen-fixing bacteroids (Atkins, 1987).

In developing root nodules of French bean, the dramatic increase in nodule GS activity that begins 10 days to 12 days after infection (Lara et al., 1983) is mainly the result of the induction of a single GS gene, gln-\(\gamma\), one of five homologous GS genes in the French bean genome (see Forde and Cullimore, 1989). Work with chimeric gene fusions has shown that a 2-kb fragment from the 5'-flanking region of gln-\(\gamma\) is able to direct the appropriate nodule-enhanced pattern of expression in transgenic plants of the forage legume Lotus corniculatus (Forde et al., 1989).

We are interested in the regulatory mechanisms by which gln-\(\gamma\) and other nodulin genes are able to respond to the symbiotic interaction with Rhizobium. Functional analysis of the upstream region of the soybean leghemoglobin \(c_3\) (\(lbc_3\)) gene has revealed both positive and negative regulatory elements and a promoter-proximal sequence required for nodule-specific expression (Stougaard et al., 1987, 1990). A nodule-specific nuclear factor from soybean was found to recognize short A/T-rich sequences in the \(lbc_3\) promoter (Jensen et al., 1988), and related factors were identified in nodules and other tissues of...
Sesbania rostrata and alfalfa (Metz et al., 1988). The same soybean factor also has binding sites in a second late nodulin gene, N23 (Jacobson et al., 1990), but its function as a trans-acting regulatory factor has still not been established.

Although activation of the gln-γ and leghemoglobin genes occurs at about the same stage of nodule development (Lara et al., 1983; Bennett et al., 1989), there is evidence that the regulatory mechanisms may not be identical. In soybean, it has been found that when nodule development is disrupted by growing plants in argon: oxygen mixtures (Hirel et al., 1987), or by use of a nilA mutant of Rhizobium (Studer et al., 1987), there are differential effects on the expression of leghemoglobin and GS genes; in French bean, it has been demonstrated that gln-γ expression is not confined to the root nodule (Bennett et al., 1989).

In the present paper, we report the nucleotide sequence of the 5'-flanking region of the gln-γ gene and the results of an analysis of sequence-specific interactions between this region and nuclear factors from French bean roots and nodules. These studies have provided evidence for the presence of a nuclear factor in nodules and a related factor in roots that bind in a sequence-specific manner to multiple A/T-rich sites in the upstream region of gln-γ. The relationship between these factors and previously reported factors that bind to A/T-rich sequences in nodulin (and non-nodulin) genes is examined and their possible function in regulating gln-γ expression is discussed.

RESULTS

Nucleotide Sequence Analysis of the gln-γ Gene

Figure 1 shows the nucleotide sequence of the 5'-flanking region of gln-γ and the first 400 bp of the transcribed portion of the gene. Over the region of overlap with pcGS-γ1, a full-length cDNA clone for the γ subunit of GS in French bean (Bennett et al., 1989), the sequence of exons 1 and 2 of the gene is identical with the cDNA (starting at position −68 in Figure 1). Previous studies have indicated that there is only a single γ-type GS gene in French bean (Cullimore et al., 1984; Gebhardt et al., 1986), so it is likely that the cDNA and genomic clones correspond to the same gene. Primer extension analysis was used to locate the 5'-end of the gln-γ mRNA in nodules. Figure 2 shows that the major transcriptional start sites in the gln-γ gene are located between positions −73 and −68, about 30 bp downstream of a canonical TATA box sequence. It has not been established whether different transcription start sites are used when the gene is expressed in other tissues of the plant.

From a comparison of the 5'-flanking regions of the soybean N24 gene and the leghemoglobin and N23 gene families, Sandal et al. (1987) identified two short conserved sequences, AAAGAT and CTCTT, each of which is found in one or more copies within 200 bp of the start of transcription in all the genes examined, and they designated these as putative nodulin consensus motifs. The same sequences were subsequently found in two leghemoglobin genes of the stem-nodulating legume Sesbania (Metz et al., 1988). One copy of each motif occurs within the organ-specific element (OSE) of the soybean lbc3 gene, a 37-bp sequence that is involved in directing nodule-specific expression (Stougaard et al., 1987, 1990). When the gln-γ sequence was searched for homologies to the lbc3 promoter, a 15-bp sequence that is related to part of the OSE was identified about 140 bp upstream from the TATA box. Table 1 shows that similar 15-bp sequences are also found at analogous locations in the promoter regions of the soybean N23 and Sesbania glb3 genes. In each case, the OSE-related segment contains a sequence that is identical (lbc3, glb3, and N23), or almost identical (gln-γ), to the AAAGAT motif.

No perfect matches to either of the nodulin consensus motifs are found within the promoter-proximal region of the gln-γ gene, but one copy of each sequence does occur further upstream, and copies of both sequences (mostly in the inverted orientation) are also found downstream of the cap site (Figure 1). However, the two sequences also occur at a similar frequency in the 5'-flanking region of gln-β (Turton et al., 1988), a related GS gene that does not show the same pattern of induction during nodulation (Gebhardt et al., 1986; Bennett et al., 1989; Forde et al., 1989).

The 5'-flanking region of gln-γ contains a number of repeated sequences and sequences with unusual structures (Figure 1). Among these (at −516 and −466) are two imperfect copies of an extremely A/T-rich 21-bp sequence that is identical (Ibc, glbs, and N23), or almost identical (gln-γ), to the AAAGAT motif.

Analysis of DNA-Protein Interactions in the 5'-Flanking Region of gln-γ

We have carried out an in vitro study of interactions between nuclear factors from French bean and sequences in the upstream region of gln-γ using the technique of gel retardation analysis. Figure 3A illustrates in diagrammatic form two overlapping sets of restriction fragments, covering the region from −971 to −21, that were used in these experiments. Each fragment was end labeled with 32P and incubated with crude nuclear extracts from French bean roots and nodules. For each extract and for each fragment, the concentration of herring sperm DNA needed to eliminate nonspecific binding was first established (gen-
Figure 1. 5'-Flanking and N-Terminal Sequences of the gln-γ Gene.

The N-terminal sequence of the γ polypeptide is written below the nucleotide sequence using the single-letter code, and intron sequences are shown in lowercase type. The locations of the intron/exon boundaries were determined by comparison with the sequence of pgGS-γ1, a gln-γ cDNA clone (Bennett et al., 1989). The major transcriptional start sites, as determined by primer extension, are indicated by closed circles. Numbering is from the first base of the ATG translational start codon (designated zero). A putative TATA box sequence and a sequence with homology to part of the OSE of the soybean Ibc γ gene (Stougaard et al., 1987) are boxed. The two 21-bp A/T-rich repeats (see text) are indicated by dashed lines, and other inverted or direct repeats by arrows. Additional sequences referred to in the text are underlined and identified by letters: those labeled a and b match the nodulin consensus motifs identified by Sandal et al. (1987) and those labeled c conform to the TATTTWAT consensus motif (see Figure 6).

erally 0.5 μg to 5 μg per assay). The binding reactions were electrophoresed on acrylamide gels to resolve the DNA-protein complexes from unbound DNA.

Figure 3B presents the results obtained when one set of five fragments was tested with nodule extracts in the gel retardation assay. Three of the fragments (RS2, RS3, and MR3) formed complexes that migrated more slowly than free DNA (lanes d, f, and h) and that were not seen...
Figure 2. Mapping of the 5' End of the gln-γ mRNA by Primer Extension.

The primer extension reactions were carried out as described in Methods using an oligonucleotide specific for the 5'-untranslated region of the gln-γ mRNA and poly(A)* RNA from nodules as template (lane e). As a control, an identical reaction was carried out in parallel on poly(A)* RNA from roots (lane f), which is highly depleted in gln-γ mRNA (Gebhardt et al., 1986; Bennett et al., 1989). Size markers were obtained by carrying out a set of four dideoxy sequencing reactions (lanes a to d) using the same primer and a single-stranded DNA template consisting of the sense strand of the 5' end of the gln-γ gene (see Methods). The major transcriptional start sites, as inferred from the sizes of the major primer extension products, are indicated with arrows, and the location of the putative TATA box sequence is also shown.

in the absence of nuclear protein (lanes c, e, and g). Figure 3C shows the results obtained when the binding reactions were carried out on the second set of restriction fragments. The three fragments (AD1, AD2, and AD3) that span the same region of the gene as RS2, RS3, and MR3 also formed complexes with factors from nuclear extracts (Figure 3C, lanes a, d, and g), but the fourth fragment, AD4, did not (not shown). A distinctive feature associated with AD3 was the formation of a second, less abundant complex of lower mobility (see lane g), which was found to increase in relative abundance at high concentrations of nuclear protein. Both complexes appear to represent sequence-specific interactions because they were each sensitive to competition by a 50-fold molar excess of the unlabeled fragment. Both complexes were also very much more susceptible to competition with poly(dA-dT) than with poly(dG-dC) (data not presented).

These observations suggested that a nodule factor was interacting in a sequence-specific manner with A/T-rich sequences in AD3 and that the formation of additional complexes was due to multiple interactions between this factor and the DNA fragment. To investigate the possibility that the 21-bp A/T-rich repeats (both copies of which are present on AD3, and one each on RS3 and MR3) may be the target sites for this nodule factor, we synthesized a double-stranded oligomer (γ-AT2) containing the sequence of the second of the two repeats (see Methods) and tested its ability to compete for binding in the gel retardation assay. These experiments showed that both of the complexes formed with AD3 could be effectively titrated out in the presence of an excess of the γ-AT2 oligomer (Figure 3C, lanes a and i). In separate competition experiments (not shown), the concentration of γ-AT2 required to obtain 50% inhibition of binding to AD3 was less than one-hundredth the concentration of poly(dA-dT) required to obtain the same effect. The γ-AT2 oligomer also competed effectively for binding to AD1 (lanes b and c), to AD2 (lanes e and f), and to RS2, RS3, and MR3 (not shown). Therefore, it appears that a single type of binding activity was responsible for all of the interactions with the gln-γ gene that were detected in nodule extracts in our experiments. A minimum of four binding sites for this factor would be needed to account for the results obtained in Figure 3. Two of these sites probably correspond to the two copies of the 21-bp A/T-rich repeat, but the exact number and position of the other binding sites have not been determined. The location of the 21-bp A/T-rich repeats close to the ends of RS3 and MR3 may help to explain why binding to these fragments was much weaker than to AD3 (Figure 3B).

Under the conditions used in the present study, no binding of nodule extracts to NR4, MR2, or AD4 was observed. However, Figure 3D shows that NR4 did form a complex with factors from root extracts and that this complex could be competed out by an excess of the unlabeled fragment. Unlike the nodule binding activity, this root factor was not sensitive to the presence of large amounts (up to 1 μg) of poly(dA-dT) nor to competition by a 100-fold molar excess of the γ-AT2 oligomer (not shown).
Table 1. Conserved, OSE-Related Sequences in the Promoter Regions of Four Nodulin Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Conserved Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soybean lbc&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-135</td>
<td>TGAATTTTTATTTTTTTT</td>
</tr>
<tr>
<td>Sesbania glb&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-196</td>
<td>TTAATTTTTATTTTTTTT</td>
</tr>
<tr>
<td>Soybean N23</td>
<td>-177</td>
<td>TTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>French bean gin-&lt;sub&gt;y&lt;/sub&gt;</td>
<td>-245</td>
<td>TTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td>TTTTTTTTTTTTTTTTTTT</td>
</tr>
</tbody>
</table>

* Relative to the translation initiation codon.

Related but Distinct Binding Factors Are Present in Nodules and Roots

To investigate whether the nuclear factors that recognize the A/T-rich binding sites in the gln-<sub>y</sub> gene are specific to nodules, we carried out a gel retardation experiment with the 32P-labeled γ-AT2 oligomer and a crude nuclear extract from uninoculated roots. Figure 4A shows that two complexes were formed when the γ-AT2 oligomer was incubated with the root extract (lane f), both of which had higher mobility than the single complex formed with the nodule extract (PNF-1, lane a). However, only one of the root complexes was effectively competed out by the presence of excess unlabeled γ-AT2 oligomer (PRF-1, lanes f to j). The faster moving root complex, which was not seen in binding experiments with the AD3 fragment (see Figure 4B), was much less susceptible to competition by γ-AT2 and is thought to be due to nonspecific DNA binding activity, although this has not been investigated further. Preliminary experiments with extracts from other tissues indicated that factors related to PNF-1 (Phaseolus nodule factor-1) and PRF-1 (Phaseolus root factor-1) are also present in leaves and in white cotyledons of germinating seeds.

Figure 4B shows the results obtained when we investigated the effects of RNase and proteinase K treatments on the root and nodule factors that bind to AD3. Initial attempts to perform the enzyme treatments at 37°C were unsuccessful because of the temperature sensitivity of both binding factors, particularly PRF-1. When the nuclear extracts were incubated for 10 min at 25°C immediately before the binding assay, the nodule binding activity was retained (lane c), although much of the root binding activity was lost (lane g). Both factors were resistant to a 10-min pretreatment with RNase A (lanes d and h) but both were sensitive to proteinase K (lanes e and i). The proteinase K digestion led to a marked increase in the mobility of the major root and nodule complexes without having any significant quantitative effect on the binding activities when compared with the preincubated controls. A slightly different result was obtained when a range of concentrations of trypsin (up to 30 ng/assay) was substituted for proteinase K. As the concentration of trypsin was increased above a threshold of 2.5 ng/assay, there was a progressive increase in mobility of the PNF-1 and PRF-1 complexes that, in this case, was accompanied by a progressive loss in binding activity (data not shown). At the highest trypsin concentration, the residual binding activity was about 30% of the control.

From these experiments, it is clear that PNF-1 and PRF-1 are at least partially, and probably wholly, proteinaceous. The increase in the electrophoretic mobility of the PNF-1 and PRF-1 complexes after proteinase K digestion may indicate that one of several protein species involved in formation of these complexes is more susceptible than the others to proteolysis but is not required for binding. Alternatively, partial proteolysis of a single DNA-binding protein could have removed an N- or C-terminal segment that is not required for binding. A number of prokaryotic and eukaryotic DNA-binding proteins have been shown to contain a separate DNA binding domain that is linked to the rest of the protein by a protease-sensitive peptide (Geisler and Weber, 1977; Sauer et al., 1979; Supakar et al., 1989).

PNF-1 and PRF-1 also Recognize an A/T-Rich Sequence in a Soybean Leghemoglobin Gene

The similarities already noted between the 21-bp A/T-rich repeats and an A/T-rich site in the soybean lbc<sub>3</sub> gene led us to investigate whether there was cross-recognition between PNF-1 and the lbc<sub>3</sub> binding sites. The target DNA used in these experiments was a double-stranded oligomer (lb-AT1) corresponding to binding site 1 in the lbc<sub>3</sub> gene (Jensen et al., 1988). Of the two A/T-rich binding sites in lbc<sub>3</sub>, this sequence has the higher affinity for the binding protein from soybean nodules (Jensen et al., 1988) but unlike binding site 2 is only distantly related to the gln-<sub>y</sub> repeats. The 32P-labeled γ-AT2 oligomer was incubated with extracts of French bean roots and nodules in the presence of a range of concentrations of the unlabeled lb-AT1 and γ-AT2 oligomers. After electrophoresis and autoradiography, the bands containing the DNA-protein complexes were excised from the gel and the amounts of labeled DNA bound were estimated by scintillation count-
Figure 3. Interaction of Nuclear Factors with the 5'-Flanking Region of *gln-γ*.

(A) Map of the upstream region of *gln-γ* showing the restriction fragments used as probes to study DNA-protein interactions. The locations of four TATTWWAT motifs are indicated by closed circles, and the two 21-bp A/T-rich repeats are indicated by arrows. An asterisk marks the segment with homology to the *ibc* OSE (Table 1), and a right-angled arrow shows the position of the major transcriptional start sites. A, D, H, M, R, and S refer to Accl, Dral, HindIII, MaelII, Rsal, and Spel restriction sites, respectively. The restriction fragments giving the strongest binding to nodule factors are indicated by the black rectangles, those binding weakly by the shaded rectangles, and the fragment that bound to a factor present in roots but not nodules by the rectangle with vertical bars.

(B) Gel retardation assays performed with nodule nuclear extracts and one set of five restriction fragments spanning a 950-bp region upstream of the *gln-γ* gene. Binding reactions included crude nuclear extract from French bean nodules (lanes b, d, f, h, and j) or no extract (lanes a, c, e, g, and i) and 1 ng to 3 ng of the appropriate 32P-labeled DNA fragment. After electrophoresis, the gel was fixed, dried, and autoradiographed overnight (except for lanes e and f, which were exposed for 3 days to visualize the weak binding to RS3). Retarded bands resulting from binding to nodule factors are indicated by arrows. U, unbound DNA fragments.

(C) Gel retardation assays performed with nodule nuclear extracts and a second set of restriction fragments from *gln-γ*. 32P-labeled AD1, AD2, or AD3 were incubated in the presence of nodule extract and a 200-fold (lanes b, e, and h) or a 500-fold (lanes c, f, and i) molar excess of a synthetic double-stranded oligomer (γ-AT2) or in the absence of the oligomer (lanes a, d, and g). The γ-AT2 oligomer contains the sequence of the second of the two 21-bp A/T-rich repeats in the *gln-γ* promoter (see Figure 1 and Methods).

(D) Gel retardation assays performed with root extracts and the *gln-γ* fragment NR4. Binding assays included 0.5 ng of 32P-labeled NR4, root nuclear extract (1.4 μg of protein), and a range of concentrations of unlabeled NR4: 0 ng (lane a), 2.5 ng (lane b), 5 ng (lane c), 10 ng (lane d), 20 ng (lane e), or 50 ng (lane f).
Analysis of Root and Nodule Factors That Bind to A/T-Rich Target Sites in gln-γ.

(A) Binding of root and nodule factors to the γ-AT2 oligomer. Gel retardation assays were carried out on the 32P-labeled γ-AT2 oligomer using nuclear extracts from nodules (lanes a to e) or roots (lanes f to j) and a range of concentrations of unlabeled γ-AT2 oligomer: 0 ng (lanes a and f), 0.6 ng (lanes b and g), 1.5 ng (lanes c and h), 3 ng (lanes d and i), or 6 ng (lanes e and j). PNF-1 is the major retarded complex obtained with nodule extracts; PRF-1 is the major complex obtained with root extracts; ns, nonspecific binding activity.

(B) Sensitivity of PNF-1 and PRF-1 to temperature, ribonuclease, and proteolysis. Binding assays contained 32P-labeled AD3 together with nodule extract (6.6 μg of protein, lanes b to e), or root extract (3.5 μg of protein, lanes f to i), or no protein (lane a). Before addition of the labeled DNA, the binding assays were preincubated for 10 min at 25°C with 3 μg of bovine pancreatic RNase A (lanes d and h), or 3 μg (3 to 6 × 10^5 units) of proteinase K (lanes e and i), or no enzyme (lanes c and g). Lanes b and f: controls without preincubation. Note that the mobility difference between PNF-1 and PRF-1 is less apparent with AD3 than when the smaller γ-AT2 oligomer is the target DNA.

Comparison of A/T-Rich Binding Sites in Nodulin Genes

In an attempt to identify common sequence motifs in the various A/T-rich binding sites that have now been identified in nodulin genes from several legume species, the sequences of the gln-γ A/T-rich repeats have been aligned in Figure 6 with the sequences of A/T-rich binding sites from the soybean lbc3 and N23 genes and the Sesbania glb3 gene. The two gln-γ sequences, which are identical at 18 out of 21 positions, differ at only six and five positions, respectively, from the sequence at binding site 2 of the lbc3 gene (Figure 6). The sequences of the four other binding sites, on the other hand, show very little sequence similarity to the gln-γ repeats and are themselves quite diverse in structure. It is possible, nevertheless, to identify a shorter core motif (boxed in Figure 6) with the consensus sequence TATTWAT (where W = A or T), which is common to all of the binding sites and forms part of an imperfect palindromic sequence of variable length.

When the gln-γ upstream region was scanned in both orientations for sequences conforming to this consensus, we located just two additional matching sequences (at -799 and -380; Figure 1). The former sequence occurs...
Protein binding studies with a soybean lectin gene (lec) led to the identification of two A/T-rich sequences that interact with a factor from immature embryos (Jofuku et al., 1987). Both of these binding sites contain the sequence motif AATTTAAT (in opposite orientations), which differs at only a single position from the nodulin octanucleotide core motif (Figure 6). Jofuku and coworkers reported that the soybean embryo factor did not bind to a 1600-bp fragment from the upstream region of the lbc3 gene, but it is likely that the sensitivity of the gel retardation assay with a fragment of this size would be very low. To determine whether PNF-1 is able to recognize the lbc3 binding sites, we synthesized a double-stranded oligomer (lec-AT2) corresponding to the region from -153 to -132 of the lbc3 gene, the second of the two binding sites (see Jofuku et al., 1987, and Methods). A reciprocal competition experiment was then carried out in which the γ-AT2 and lec-AT2 oligomers were end labeled and mixed with a nodule nuclear extract in the presence of a range of concentra-
Mutational Analysis of the γ-AT2 and lec-AT2 Binding Sites

To investigate in more detail which DNA sequences are important for binding to PNF-1, we carried out a series of competition experiments with mutant oligomers. Figure 8A shows the results obtained when the binding affinities of two mutant forms of the γ-AT2 oligomer were analyzed. One of the mutated oligomers (γ-AT2*) carried four single base pair substitutions while retaining the base composition of the original A/T-rich sequence (see Methods). Three of the substitutions were at the highly conserved first, fourth, and eighth positions of the octanucleotide core motif, which is located in the downstream half of the γ-AT2 sequence (TATTTAAT → AATATAAA); the fourth substitution was in the upstream half of the sequence, within a related octanucleotide motif (TATTTAAT → AATTAATT). These mutations had the effect of reducing the oligomer’s binding affinity for PNF-1 by sevenfold (based on the relative concentrations of the γ-AT2 and γ-AT2* oligomers needed to obtain 20% inhibition of binding to the 32P-labeled γ-AT2 oligomer). In the second mutated version of γ-AT2 (γ-AT2), the entire upstream half of the γ-AT2 sequence was deleted to leave only 10 bp of A/T-rich sequence but with the octanucleotide core motif intact. This deletion had a less marked effect on the binding of PNF-1, reducing its binding affinity for PNF-1 by about threefold (expressed on a molar basis). Taken together, these results indicate that PNF-1 has quite stringent sequence requirements for binding and suggest that the conserved TATTWAT motif is an important component of the binding site.

DISCUSSION

Multiple Protein-Binding Sites in the gln-γ Gene

The 1-kb upstream region of the gln-γ gene, which has now been sequenced, contains all the elements necessary...
Two distinct nodule factors (NAT1 and NAT2) and one leaf factor (LAT1) have recently been shown to bind to several A/T-rich sites in the promoter region of the soybean N23 gene (Jacobsen et al., 1990). NAT1 and LAT1 were found to be related to the high-mobility group I (HMG I) protein of eukaryotic chromatin, but NAT2 differed in having more stringent sequence requirements for binding and in generating a DNA-protein complex with a much lower electrophoretic mobility. Under the standard conditions used in our binding assays, we detected only a single type of binding activity in nodule extracts. However, when herring sperm DNA was used as nonspecific competitor instead of a synthetic polynucleotide mix in binding assays with the γ-AT2 oligomer, we were able to detect a second complex with a very much higher electrophoretic mobility than the PNF-1 complex (data not presented). This second binding factor in French bean nodules appears to correspond to NAT1 because it was soluble in 2% TCA and extremely sensitive to the presence of poly(dA-dT) in the binding assay, properties that are characteristic of the soybean HMG I-like proteins. PNF-1, on the other hand, resembles NAT2 and differs from the HMG I-like factors in that its binding activity is sensitive to heat and to TCA.

PNF-1 Is Closely Related to the Soybean Nodule Factor NAT2

Two distinct nodule factors (NAT1 and NAT2) and one leaf factor (LAT1) have recently been shown to bind to several A/T-rich sites in the promoter region of the soybean N23 gene (Jacobsen et al., 1990). NAT1 and LAT1 were found to be related to the high-mobility group I (HMG I) protein of eukaryotic chromatin, but NAT2 differed in having more stringent sequence requirements for binding and in generating a DNA-protein complex with a much lower electrophoretic mobility. Under the standard conditions used in our binding assays, we detected only a single type of binding activity in nodule extracts. However, when herring sperm DNA was used as nonspecific competitor instead of a synthetic polynucleotide mix in binding assays with the γ-AT2 oligomer, we were able to detect a second complex with a very much higher electrophoretic mobility than the PNF-1 complex (data not presented). This second binding factor in French bean nodules appears to correspond to NAT1 because it was soluble in 2% TCA and extremely sensitive to the presence of poly(dA-dT) in the binding assay, properties that are characteristic of the soybean HMG I-like proteins. PNF-1, on the other hand, resembles NAT2 and differs from the HMG I-like factors in that its binding activity is sensitive to heat and to TCA.

for induction of gene expression in nodules of transgenic L. corniculatus plants (Forde et al., 1989; W.-J. Shen, M.S. Williamson, and B.G. Forde, unpublished results). Gel retardation analysis, using crude nuclear extracts from French bean roots and nodules, has now demonstrated the presence in this region of two types of binding site for nuclear proteins: (1) a sequence between -978 and -861 that binds a factor present in roots but not nodules, and (2) multiple binding sites between -860 and -154 that interact with related but distinct factors from roots and nodules (designated PRF-1 and PNF-1, respectively). PRF-1 and PNF-1 could be distinguished both by the electrophoretic mobility of the complexes that they formed in the gel retardation assay (Figure 4) and by their relative affinities for different A/T-rich sequences (Figure 5), but whether they represent different gene products or post-translationally modified forms of the same protein(s) is unknown.
tions together with the finding that PNF-1 has a high affinity for an NAT2 binding site from the soybean lbc3 gene, we conclude that PNF-1 is probably the homolog in French bean of NAT2.

A/T-Rich Binding Sites in Nodulin Genes Share a Common Sequence Motif

Binding sites for NAT2 and PNF-1 and for related factors in Sesbania and alfalfa (Metz et al., 1988) have now been identified in leghemoglobin genes of both soybean (Jensen et al., 1988) and Sesbania (Metz et al., 1988), in the soybean N23 gene (Jacobsen et al., 1990), and in the gln-γ gene of French bean (this paper). As Figure 6 shows, apart from their high A/T content and partial dyad symmetry, these binding sites show remarkable sequence diversity, indicating a high degree of degeneracy in the sequence requirements for recognition. Nevertheless, a significant degree of sequence specificity is indicated by the finding that PNF-1 has very differing affinities for different A/T-rich binding sites (Figure 7) and by the demonstration that certain single base pair substitutions, even while maintaining the base composition of the target sequence, can have a dramatic effect on the binding of NAT2 (Jacobsen et al., 1990) and PNF-1 (Figure 8). A 16-bp motif, WTWAAATTTTTATTTW, is shared by the soybean lbc3 and Sesbania glb3 binding sites (Metz et al., 1988), but the binding sites subsequently identified in gln-γ (this paper) and in the N23 gene (Jacobsen et al., 1990) conform only poorly to this consensus. However, by aligning the A/T-rich sequences as shown in Figure 6, it was possible to recognize a shorter motif with the consensus sequence TATTTAAT, which is common to all the NAT2/PNF-1 target sites so far reported (Figure 6). Experimental evidence that this motif is important for protein recognition was obtained by analysis of the binding affinities of mutated forms of the γ-AT2 oligomer (Figure 8A) and by the demonstration that a single base pair substitution in the AATTAAATT motif of the lec-AT2 oligomer, which generated a nodulin consensus motif TATTTAAT, greatly increased its affinity for PNF-1 (Figure 8B).

Do A/T-Rich Binding Sites Have a Role in Regulating Nodulin Gene Expression?

The conservation of A/T-rich binding sites in a diverse set of late nodulin genes suggests that they may have an important part to play in regulating nodulin gene expression, but the nature of this role is still uncertain. Studies carried out on a series of 5' deletion mutants of the glin-γ gene found that removal of the region between −597 and −354, which encompasses at least two PNF-1 binding sites, resulted in the loss of detectable gene expression in nodules of transgenic Lotus plants (W.-J. Shen, M.S. Williamson, and B.G. Forde, unpublished results). Similar evidence for an association between the NAT2 binding sites and a weak positive regulatory element in the lbc3 gene has been reported (Jensen et al., 1988). Although recent work has indicated that deletion of two NAT2 binding sites has no dramatic effect on the function of the lbc3 promoter in transgenic plants (Stougaard et al., 1990), the presence of additional binding sites further upstream (Jacobsen et al., 1990) leaves open the possibility that NAT2 is involved in the regulation of this gene. In the soybean N23 gene, a positive regulatory element located between −344 and −293 (Jørgensen et al., 1988) also contains an A/T-rich binding site (Stougaard et al., 1990), and a similar correlation has been noted in the Sesbania glb3 gene (de Bruijn et al., 1989). Thus, there is accumulating, but still circumstantial, evidence to suggest that the A/T-rich binding sites have a role in enhancing gene expression in nodules.

Similarities between NAT2 binding sites and sequence motifs found in the vicinity of scaffold-attached regions of several Drosophila genes (Gasser and Laemmli, 1986) led to the suggestion that NAT2 might be involved in formation of the nuclear scaffold structure (Jacobsen et al., 1990). An alternative possibility is that NAT2/PNF-1 may function by interfering with the binding of the HMG I-like proteins to A/T-rich sequences. Because the binding site requirements for NAT2/PNF-1 are much more stringent than for the HMG I-like factors, this will only affect a certain subset of the HMG I binding sites. The effect that such competition for binding will have will clearly depend on the function of the HMG I-like proteins in plants. In mammals, HMG I may have a role in nucleosome positioning (Strauss and Varshavsky, 1984) or as a component of the nuclear scaffold (Solomon et al., 1986). Thus, the binding of NAT2 or PNF-1 could have an indirect effect on chromatin structure in the vicinity of their target promoters by excluding HMG I-like proteins. A third possibility is that NAT2 and PNF-1 are conventional trans-acting factors, exerting their influence through protein-protein interactions with the transcriptional complex (Mitchell and Tjian, 1989).

PNF-1 and NAT2 May Be Members of a Larger Family of Nuclear Factors

Datta and Cashmore (1989) identified a nuclear protein from pea seedlings (AT-1) that binds to A/T-rich sequences sharing the motif AATATTTTATT in a group of photoregulated genes from pea, tomato, and tobacco. Similar binding sites have been mapped in a variety of seed protein genes that are expressed in developing embryos of soybean (Jofuku et al., 1987; Allen et al., 1989), French bean (Bustos et al., 1989; Riggs et al., 1989), and sunflower (Jordano et al., 1989). In common with the nodulin genes, these A/T-rich binding sites are often present in multiple
copies and are usually located between 500 bp and 1000 bp upstream from the start of transcription. The sequence similarities between these A/T-rich binding sites (which generally include motifs conforming to the consensus WAT₃₋₄A₄₋₅T) and those in the nodulin genes (which conform to the same general consensus) suggest that many of the plant factors that recognize A/T-rich sequences and occur in diverse plant tissues could have related binding requirements. In support of this hypothesis, we have demonstrated that the French bean nodule factor PNF-1 was able to bind to a synthetic target site for a factor from immature soybean embryos (Figure 7).

In a further parallel with the nodulin genes, the A/T-rich binding sites in other plant genes are frequently associated with positive regulatory elements (Bustos et al., 1989; Datta and Cashmore, 1989; Riggs et al., 1989). Most notably, a 55-bp A/T-rich sequence from the β-phaseolin gene, which binds to nuclear proteins from immature French bean cotyledons, was shown to enhance expression from a minimal 35S promoter in transgenic tobacco, not only in seeds but also in other parts of the plant (Bustos et al., 1989). On the other hand, an AT-1 binding site in the tobacco cab-E gene is reported to reside within a negative regulatory element (Datta and Cashmore, 1989).

Thus, a picture is emerging of an evolutionarily conserved and ubiquitous group of nuclear factors having an affinity for certain A/T-rich sequences found in a wide variety of plant genes. Extrapolating from the example of PNF-1 and PRF-1, it is likely that related factors present in different tissues (or even within the same tissue) will have different sequence preferences and will, therefore, be capable of binding to a different (but possibly overlapping) spectrum of A/T-rich sequences. The existence of families of transcription factors with related DNA binding specificities is now well established in metazoans and in yeast (see Mitchell and Tjian, 1989). Further work will be needed to establish whether the group of plant factors that are related by their affinity for similar A/T-rich sequences are also related structurally and functionally, and to determine their role, if any, in the transcriptional control of gene expression in higher plants.

METHODS

Plant Growth

Phaseolus vulgaris (French bean) var Tendergreen was grown from surface-sterilized seed under 16-hr light (20°C)/8-hr dark (16°C) cycles in a sand and gravel mixture (1:2) as described (Cullimore et al., 1983). Seeds were inoculated with Rhizobium leguminosarum biovar phaseoli strain R3622 at the time of sowing and again 2 days later. Nodules were harvested into liquid N₂ 3 to 4 weeks after sowing and stored in liquid N₂ or at -70°C. Roots were obtained from unoinoculated plants grown for 2 weeks using nutrient solution supplemented with nitrate.

Genomic Cloning and Nucleotide Sequencing

Unless otherwise stated, standard recombinant DNA techniques were used (Maniatis et al., 1982). The isolation and restriction mapping of the French bean genomic clones λGSI-56 and λGSI-57 and their respective subclones pN/HB68 and pN/HB57d, which contain the 5'-flanking region of gln-γ, has been reported previously (Forde et al., 1989). Partial sequence analysis of the 2.1-kb insert in pN/HB57d was carried out on both strands by the dideoxy method (Sanger et al., 1977) using T4 DNA polymerase-generated deletions in M13mp18 and M13mp19 (Dale et al., 1989). Sequence data were assembled on a VAX 11/750 computer using DBUTIL (Staden, 1982) and sequence analysis and comparisons were performed on a VAX 11/785 computer using the Wisconsin sequence software package (Devereux et al., 1984).

Primer Extension

The extraction of poly(A)* RNA from roots and nodules and the primer extension reactions were carried out essentially as described previously (Freeman et al., 1990). The primer consisted of a 21mer oligonucleotide (5′-CAGCAGAGACTCTTCACAGAG-3′), which is complementary to a gene-specific region in the 5′-untranslated region of the gln-γ mRNA. Accurate size markers were generated by performing sequencing reactions with the same primer on an M13 subclone of pNH57d carrying the nontranscribed strand of the gln-γ gene. Primer extension reactions were electrophoresed on 6% polyacrylamide sequencing gels in parallel with the size markers and the gels were fixed, dried, and autoradiographed.

DNA Probes

DNA fragments used for gel retardation analysis contained sequences from the 5′-flanking region of gln-γ and were obtained by cleavage of pNH57d with combinations of restriction enzymes (see Figure 3A). The fragments were electrophoresed in 4% acrylamide gels, excised, and electroeluted. The restriction fragments (with the exception of AD4) were then filled in using the Klenow fragment of DNA polymerase I and subcloned in the Smal site of pUC8 or pUC9. After purification of the plasmid DNAs, the inserts were excised by digestion with BamHI and EcoRI and subsequently electroeluted from 1.2% agarose gels. The AD4 fragment was obtained from pNB26, a derivative of pNH57d (Forde et al., 1989), and labeled directly without subcloning. To avoid the denaturation of the DNA that can occur on drying after ethanol precipitation, leading to artifacts in the gel retardation assay (Svaren et al., 1987), pellets obtained after electroelutions were washed with 75% ethanol and drained thoroughly before redissolving directly in 10 mM Tris-HCl, pH 8, 0.1 mM EDTA.

Oligonucleotides were synthesized by phosphoramidite chemistry using a New Brunswick Biosearch 3810 DNA synthesizer. After deprotection, the oligonucleotides were purified by electrophoresis in a 16% acrylamide sequencing gel and recovered from
gel slices by diffusion into 0.5 M NH₄OAc. The oligonucleotides were precipitated in 80% ethanol, washed, dried, and taken up in sterile distilled water. Table 2 gives the sequences of the oligonucleotides that were used for binding studies.

Restriction fragments and double-stranded oligonucleotides were labeled by filling in with α-32P-dATP (3000 Ci/mmol; Amersham International), dCTP, dGTP, and TTP using Klenow, and unincorporated nucleotides were removed by chromatography in Sephadex G-50.

**Nuclear Extracts**

A crude preparation of root or nodule nuclei was obtained using a procedure modified from Willmitzer and Wagner (1981). Plant material (10 g to 40 g, fresh weight) was ground to a coarse powder in liquid N₂ using a pestle and mortar. All subsequent steps were carried out on ice and centrifugations were at 4°C. The powder was transferred to a second mortar containing ice-cold homogenization buffer (7.5 mL/g of tissue) and ground for 3 min to 4 min. Homogenization buffer consisted of 0.25 M sucrose, 10 mM NaCl, 10 mM Mes, pH 6.0, 10 mM 1,10-phenanthroline, 0.15 mM spermine, 0.5 mM spermidine, 20 mM β-mercaptoethanol, 0.2 mM phenylmethylsulfonyl fluoride, 0.01% Triton X-100, the pH being adjusted to 5.3 with HCl. After filtering through four layers of muslin and then through two layers of Miracloth, the homogenate was centrifuged for 15 min at 500 g. The pellet was washed three times in wash buffer (homogenization buffer minus Triton X-100, 2.5 mL/g of starting material), centrifuging for 7 min at 500 g each time.

The procedure for extraction of nuclear proteins was modified from Siebenlist et al. (1984). The crude nuclear pellet was resuspended in 1 mL to 2 mL of NPEB (10 mM Hepes, pH 8.0, 300 mM NaCl, 25% (v/v) glycerol, 0.1 mM EDTA, 0.1 mM EGTA, 0.5 mM spermidine, 0.15 mM spermine, 1 mM phenylmethylsulfonyl fluoride, 7 mM β-mercaptoethanol) and mixed for 2 hr at 4°C by rotation on a Multimixer. Debris was removed by centrifugation at 12,000g for 10 min and by recentrifuging the supernatant for a further 5 min. The crude nuclear extracts were frozen in liquid N₂ in aliquots of 50 μL or 100 μL and stored at −70°C. Protein concentrations in the extracts were estimated by the method of Bradford (1976) using a kit from Bio-Rad Laboratories.

**Gel Retardation Analysis**

Unless otherwise stated, binding reactions consisted of 10 μL of crude nuclear extract (about 3 μg of protein), 1 μL of 20 × mix (300 mM Hepes, pH 8, 20 mM EDTA, 100 mM dithiothreitol), 0.5 μg each of poly(dA·dT) and poly(dG·dC), 2 μg of herring sperm DNA, and 0.2 ng to 0.5 ng (2 fmol to 7 fmol) of 32P-labeled restriction fragment, in a final volume of 20 μL. When 32P-labeled double-stranded oligonucleotides (0.1 ng, 5 fmol) were used as probes, the herring sperm DNA was omitted. Reactions were incubated for 5 min at 24°C before addition of the labeled DNA, and the incubation was continued for 20 min. In competition experiments, the unlabeled competitor DNA was added after the 5-min preincubation and a further 5-min incubation was carried out before addition of the labeled DNA. At the end of the incubation period, 4 μL of 0.1% bromphenol blue was added and one-third of each reaction was loaded on a 5% polyacrylamide gel (80:1 acrylamide:bisacrylamide) that had been pre-electrophoresed at 175 V for 20 min. The electrophoresis buffer was 0.25 × TBE (1 × TBE is 89 mM Tris-borate, pH 8.3, 2 mM EDTA). Samples were electrophoresed at 200 V for 2 hr, the gel temperature being maintained at 13°C to 16°C by water cooling. The gel was fixed in 40% methanol, 10% acetic acid for 45 min, dried, and exposed to x-ray film with two intensifying screens for 16 hr to 72 hr at −70°C.

**ACKNOWLEDGMENTS**

We are grateful to our colleagues in the Biochemistry and Physiology Department for their practical help and advice, particularly Jo Clark and Michele Bianchi, and to Professor Kjeld Markker and his colleagues at the University of Aarhus for helpful discussions. We thank Julian Franklin for providing glasshouse facilities, Martin Cornelius for synthesizing oligonucleotides, and Alison Hopley for carrying out the primer extension reactions. M.P. held a Senior Training Fellowship from the European Economic Community (EEC), and this work was also supported partly by Contract No. BAP/0099/UK of the Biotechnology Action Programme of the EEC.

Received May 7, 1990; revised July 19, 1990.

**REFERENCES**


---

**Table 2. Oligonucleotides Used in the Gel Retardation Experiments**

<table>
<thead>
<tr>
<th>Oligomer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>leb-AT1*</td>
<td>5'-CTGAGATATTTAATTTTAAATTTTATAA-3'</td>
</tr>
<tr>
<td>γ-AT2</td>
<td>5'-CTAGATATTTAATTTTAAATTTTATAA-3'</td>
</tr>
<tr>
<td>γ-AT2*</td>
<td>5'-CTAGATATTTAATTTTAAATTTTATAA-3'</td>
</tr>
<tr>
<td>γ-AT2</td>
<td>5'-CTAGATATTTAATTTTAAATTTTATAA-3'</td>
</tr>
<tr>
<td>γ-AT2</td>
<td>5'-CTAGATATTTAATTTTAAATTTTATAA-3'</td>
</tr>
<tr>
<td>γ-AT2</td>
<td>5'-CTAGATATTTAATTTTAAATTTTATAA-3'</td>
</tr>
<tr>
<td>icf-AT2</td>
<td>5'-GATCCCTAAATTAAAATTTTAAATTTTATAA-3'</td>
</tr>
<tr>
<td>icf-AT2*</td>
<td>5'-GATCCCTAAATTAAAATTTTAAATTTTATAA-3'</td>
</tr>
</tbody>
</table>

* Underlined nucleotides indicate positions at which mutant oligomers differ from the parental type.

* The oligomer containing the sequence of binding site 1 of the soybean Ibc gene (Jensen et al., 1987) was kindly provided by Dr. Erik Jensen, University of Aarhus.


Siebenlist, U., Hennighausen, L., Battey, J., and Leder, P.


Nuclear factors interact with conserved A/T-rich elements upstream of a nodule-enhanced glutamine synthetase gene from French bean.
B G Forde, J Freeman, J E Oliver and M Pineda

Plant Cell 1990;2:925-939
DOI 10.1105/tpc.2.9.925

This information is current as of July 13, 2017