MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 5, No. 5, pp. 439-442, 1992

Research Notes

Homology of *Rhizobium meliloti* NodC to Polysaccharide Polymerizing Enzymes

E. M. Atkinson and S. R. Long

Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020 U.S.A.
Received 8 June 1992. Accepted 29 June 1992.

Rhizobium **bacteria form nitrogen-fixing nodules on legume roots. As part of the nodulation process, they secrete Nod factors that are β-1,4-linked oligomers of *N*-acetylglucosamine. These factors depend on nodulation (*nod*) genes, but most aspects of factor synthesis are not yet known. We show here that one gene, *nodC*, shows striking similarity to genes encoding proteins known to be involved in polysaccharide synthesis in yeast and bacteria, specifically chitin and cellulose synthases, as well as a protein with unknown function in *Xenopus* embryos, DG42. This similarity is consistent with a role for the NodC protein in the formation of the β-1,4-linkage in Nod factors.**

The formation of symbiotic root nodules on legumes by *Rhizobium* bacteria requires the action of *Rhizobium* nodulation (*nod*) genes (reviewed in Long 1989). It is now known that extracellular bacterial signal molecules are important for this symbiosis, and production of these factors depends on the presence and expression of the *nod* genes in the bacterium (Faucher *et al.* 1988, Van Brussel *et al.* 1986). The Nod factors are modified oligosaccharides of β-1,4-linked *N*-acetylglucosamine and are thus similar to chitin oligomers (Lerouge *et al.* 1990). The modifications include an *N*-acyl substitution and sometimes a C-6 acetyl on the nonreducing sugar residue, and a C-6 sulfate on the reducing end (Lerouge *et al.* 1990; Schultze *et al.* 1991; Spaink *et al.* 1991; E. M. Atkinson, K. Faull, and S. R. Long, unpublished observations).

The sequences of numerous *nod* genes are known. In some cases, sequence homology has suggested function, and in others direct biochemical assay has demonstrated function. The genes studied so far include those for fatty acid modification, addition of the sulfate and acetyl groups, as well as the synthesis of glucosamine (Baev *et al.* 1991; Roche *et al.* 1991; Schwedock and Long 1990; Spaink *et al.* 1991). However, there is currently no biochemical evidence on the nature of the Nod factor β-1,4-glucan polymerizing activity.

We have found that the *Rhizobium* NodC protein has striking homology to other proteins known to be involved in polysaccharide synthesis in yeast and bacteria, specifically chitin and cellulose synthases, as well as a protein with unknown function in *Xenopus* embryos, DG42 (Sargent and Dawid 1983). Some of these homologies have been reported earlier (Bulawa 1992). The sequence similarity suggests to us that NodC could be the synthetic enzyme catalyzing the β-1,4-linkage in Nod factor production. If a catalytic domain is the basis for the similarity of these proteins, then the presence of this domain in DG42 would be consistent with a role in the synthesis of matrix polysaccharides such as hyaluronic acid. These observations of sequence similarity provide specific hypotheses that can be tested biochemically.

We performed sequence alignments on the translation products of the *nodC* gene from *R. meliloti*, pDG42 from *Xenopus laevis* (Rosa *et al.* 1988), the cellulose synthase gene from *Acetobacter xylinum* (Saxena *et al.* 1990), and the CSD2/CAL1 gene (Valdivieso *et al.* 1991; Bulawa 1992) and the CHS2 gene (Silverman 1989) from *Saccharomyces cerevisiae*. The alignments were done with the University of Wisconsin Genetics Computer Group software, specifically FASTA and BESTFIT (Devereux *et al.* 1984). Multiple sequence alignments were performed by the TULLA program (Subbiah and Harrison 1989). Final alignments were made by a combination of BESTFIT analysis and by hand alignment.

We observed substantial similarity of *nodC* to each of these genes. NodC shows the best match with the DG42 protein of *X. laevis*, with which it displays 26.4% overall identity and 48.8% overall similarity. Ranked in order of decreasing NodC homology are DG42, cellulose synthase, CSD2/CAL1, and CHS2. DG42 and cellulose synthase show extended homology to the amino terminus of NodC. In addition there are four other regions in which all five sequences are well conserved (see Fig. 1). At amino acids 141–143 of NodC there is an acidic region partly conserved in other proteins, while beginning at residue 204 there is a sequence that contains the unusual cysteine cluster that our group and others have previously noted in NodC (Jacobs *et al.* 1985; Long 1991); CHS2 does not have this cluster but other proteins in the family show it. In addition there is another region with some acidic character at amino acids 238–245, and finally a well conserved region at residues 273–283.

NodC was originally identified based on the requirement for the *nodABC* operon in nodulation of alfalfa by *R. meliloti* (Debelle *et al.* 1986; Jacobs *et al.* 1985, Kondorosi

Corresponding author: S. R. Long.
Nucleotide and/or amino acid sequence data are in GenBank, EMBL, and DDBJ as accession number M10923.

et al. 1984). *R. meliloti nodC* mutants exhibit a Nod⁻ phenotype on alfalfa. NodC has been identified in all *Rhizobium* species studied to date, and NodC proteins can functionally complement *nodC* mutations in other *Rhizobium* species. It is now known that the *nodABC* operon is essential in the production of the modified oligosaccharides known as Nod factors (Lerouge *et al.* 1990; Spaink *et al.* 1991; Schultze *et al.* 1991; E. M. Atkinson, K. Faull, and S. R. Long, unpublished observations). All of these molecules are β-1,4-linked *N*-acetylglucosamine. We show here that NodC is similar to several enzymes involved in the synthesis of β-1,4-polysaccharides. The sequence similarity with chitin synthase could be consistent either with NodC being a synthase, or with its simply having a binding domain for UDP-*N*-acetylglucosamine, a probable common precursor for chitin and Nod-factor synthesis. However, the fact that NodC is also homologous to a β-1,4 synthase for at least one other polymer, cellulose, supports the possibility that the protein conservation relates to the synthesis of that particular linkage, rather than for the use of *N*-acetylglucosamine in particular. This would be consistent with a role for NodC in polymerization of the Nod factor backbone.

Chitin is a polymer of β-1,4-linked *N*-acetylglucosamine and, because chitin synthase from yeast has been extensively studied, there are detailed biochemical and genetic data concerning this enzyme. There appear to be at least three chitin synthetic activities in *Saccharomyces* that have been localized genetically: CHS1, CHS2, and CSD2 (also called CAL1) (Valdivieso *et al.* 1991; Bulawa 1992). We previously observed slight homology between CHS2 and NodC (Long 1991), but the identification of CSD2/CAL1 led to the examination of these sequences as a group, which revealed the similarities shown in Figure 1.

Cellulose is a polymer of β-1,4-linked glucose, and cellulose synthase from *A. xylinum* has recently been cloned by two independent groups (Saxena *et al.* 1990; Wong *et al.* 1990). The *bcsA* gene of the cellulose synthase operon matches the cellulose synthase reported by Saxena *et al.* (1990), and it is this protein that has similarities to NodC,

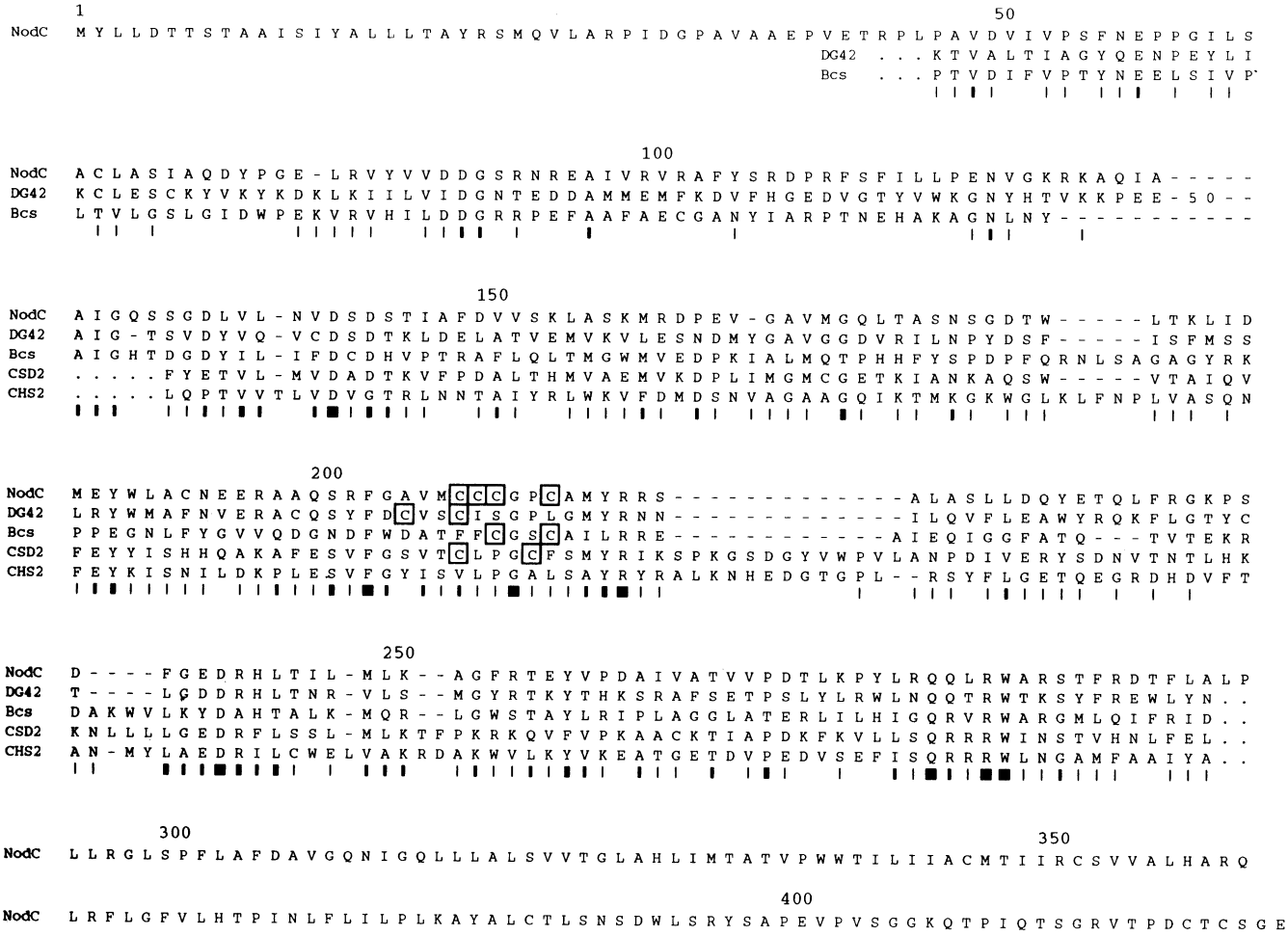

Fig. 1. Multiple sequence alignment of NodC from *Rhizobium meliloti*, DG42 from *Xenopus laevis*, cellulose synthase from *Acetobacter xylinum* (Bcs), and two yeast chitin synthases, CSD2/Cal1 and CHS2. The DG42 and cellulose synthase proteins have homolgy to NodC beginning at amino acid 46 of NodC and extending to amino acid 121. DG42 has a 50 amino acid insert at this point, and then all five peptides show homology from NodC 128 to 291. Note the cluster of cysteines (boxed) near amino acid 207. In the alignment, "|" denotes two out of five matches, "▮" denotes three out of five, "■" denotes four out of five, and "■" denotes complete identity for this amino acid in all five proteins. Sequences shown are the complete NodC sequence, the pDG42 gene product sequence from amino acid 98 to 393 (Sargent and Dawid 1983), the cellulose synthase amino acid sequence from 149 to 384 (Saxena *et al.* 1990), the CSD2/CAL1 sequence from 827 to 1008 (Valdivieso *et al.* 1991, Bulawa 1992), and the CHS2 amino acid sequence from 432 to 617 (Silverman 1989).

CHS2, and CSD2/CAL1. Saxena *et al.* believe this to be the catalytic subunit for cellulose production, based on N-terminal sequencing of synthase purified by product entrapment. Other groups (Mayer *et al.* 1991; Wong *et al.* 1990) have identified the *bcsB* gene of this operon as the catalytic subunit of the synthase. There has been no firm resolution of this discrepancy.

Of all of the sequences examined so far, *X. laevis* DG42 best matches the NodC sequence. The DG42 gene was cloned as a cDNA expressed during *Xenopus* gastrulation (Sargent and Dawid 1983). According to immunolocalization, DG42 accumulates to a peak at the mid-gastrula stage, and decays by the end of neurulation (Rosa *et al.* 1988). At one point during embryogenesis, DG42 makes up about 0.2% of the total poly(A)$^+$ mRNA in the embryo. Bulawa (1992) has proposed that the homology of DG42 to NodC suggests involvement of lipo-oligosaccharides as signals in early vertebrate embryo development. However, we also note that high concentrations of polysaccharides such as hyaluronate are often associated with epithelium-to-mesenchyme transitions, as occurs during gastrulation (Toole *et al.* 1984). Because the homologies of this small family of apparent synthetic enzymes include chitin and cellulose synthases, which produce matrix-type substances, we speculate that the DG42 protein may simply be involved in synthesis of matrix polysaccharides in the developing embryo. This would be consistent with both the homology and localization data.

The homologies presented are intriguing. Each of these proteins is probably a membrane protein, as evidenced by the extended hydrophobic domains C-terminal to the regions of homology. We have previously noted that NodC has four putative transmembrane domains (Jacobs *et al.* 1985). The known requirements for polysaccharide synthesis indicate that a nucleotide-sugar binding domain should be present, and it will be informative to discover if the proteins described here will bind nucleotide sugars. Also, the cluster of conserved cysteines suggests a common domain such as a metal-binding region, since it has been found that cellulose and chitin synthases require divalent cations for their activity (Cabib *et al.* 1983; Wong *et al.* 1990).

There has been one previous proposal for NodC function: John *et al.* (1985, 1988) reported that NodC is in the outer membrane of *Rhizobium* and suggested that its location and inferred topology indicate that it functions as a signal receptor. Based on the homology data presented here, we propose, by contrast, that NodC functions in the synthesis of Nod factors. This may occur in a vectorial fashion through the membrane, as has been shown for some chitin synthases (Cabib *et al.* 1983), or may be cytoplasmic, with factor export occuring independently. NodC may also interact with other *nod* gene products such as NodA (Johnson *et al.* 1989).

Our results suggest that NodC and the other proteins belong to an extended family of β-1,4 polysaccharide synthases from a range of organisms from bacteria to vertebrates. The function of some of the proteins is well known, as in the case of chitin synthase (CHS) in yeast, while some functions can be inferred by genetic analysis, as in the case of *R. meliloti* NodC. While the *Xenopus* DG42 protein has no demonstrated function at this time, the homology now suggests possible tests for function. This proposed family will be likely to gain new members as more genes are discovered in systems that involve polysaccharide synthesis.

ACKNOWLEDGMENTS

This work was supported by DOE contract AS03-82-ER-12084 to S. R. L., and E. M. A. was supported by a Training Grant in Cell and Molecular Biology from the National Institutes of Health to Stanford University. We thank members of our group for useful discussions and computer assistance, and express our thanks as well to J. Dénarié, C. Bulawa, and I. Dawid for discussions and/or communications prior to publication.

LITERATURE CITED

Baev, N., Endre, G., Petrovics, G., Banfalvi, Z., and Kondorosi, A. 1991. Six nodulation genes of *nod* box locus 4 in *Rhizobium meliloti* are involved in nodulation signal production: *nodM* codes for D-glucosamine synthetase. Mol. Gen. Genet. 228:113-124.

Bulawa, C. 1992. CSD2, CSD3, and CSD4, genes required for chitin synthesis in *Saccharomyces cerevisiae*: The CSD2 gene product is related to chitin synthases and to developmentally regulated proteins in *Rhizobium* species and *Xenopus laevis*. Mol. Cell. Biol. 12:1764-1776.

Cabib, E., Bowers, B., and Roberts, R. L. 1983. Vectorial synthesis of a polysaccharide by isolated plasma membranes. Proc. Natl. Acad. Sci. USA 80:3318-3321.

Debelle, F., Rosenberg, C., Vasse, J., Maillet, F., Martinez, E., Denarie, J., and Truchet, G. 1986. Assignment of symbiotic developmental phenotypes to common and specific nodulation (*nod*) genetic loci of *Rhizobium meliloti*. J. Bacteriol. 168:1075-1086.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

Downie, J. A., Marie, C., Scheu, A.-K., Firmin, J. L., Wilson, K. E., Davies, A. E., Cubo, T. M., Mavridou, A., Johnston, A. W. B., and Economou, A. 1991. Genetic and biochemical studies on the nodulation genes of *Rhizobium leguminosarum* bv *viciae*. Pages 134-141 in: Advances in Molecular Genetics of Plant-Microbe Interactions. Vol. 1. H. Henneke and D. P. Verma, eds. Kluwer Publishing, Amsterdam.

Faucher, C., Maillet, F., Vasse, J., Rosenberg, C., Van Brussel, A. A. N., Truchet, G., and Denarie, J. 1988. *Rhizobium meliloti nodH* gene determines production of an alfalfa specific extracellular signal. J. Bacteriol. 170:5489-5499.

Jacobs, T. W., Egelhoff, T. T., and Long, S. R. 1985. Physical and genetic map of a *Rhizobium meliloti* nodulation gene region and nucleotide sequence of *nodC*. J. Bacteriol. 162:469-476.

John, M., Schmidt, J., Wieneke, U., Krussmann, H-D., and Schell, J. 1988. Transmembrane orientation and receptor-like structure of the *Rhizobium meliloti* common nodulation protein NodC. EMBO J. 7:583-588.

John, M., Schmidt, J., Wieneke, U., Kondorosi, E., Kondorosi, A., and Schell, J. 1985. Expression of the nodulation gene *nodC* of *Rhizobium meliloti* in *Escherichia coli*: Role of the *nodC* gene product in nodulation. EMBO J. 4:2425-2430.

Johnson, D., Roth, L. E., and Stacey, G. 1989. Immunogold localization of the NodC and NodA proteins of *Rhizobium meliloti*. J. Bacteriol. 171:4583-4588.

Kondorosi, E., Banfalvi, Z., and Kondorosi, A. 1984. Physical and genetic analysis of a symbiotic region of *Rhizobium meliloti*: Identification of nodulation genes. Mol. Gen. Genet. 193: 445-452.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Prome, J. C., and Denarie, J. 1990. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature 344:781-784.

Long, S. R. 1991. Genetic analysis of *Rhizobium* nodulation. Pages 560-567 in: Biological Nitrogen Fixation. G. Stacey, R. H. Burris, and H. J. Evans, eds. Chapman and Hall, New York.

Long, S. R. 1989. *Rhizobium*-legume nodulation: Life together in the underground. Cell 56:203-214.

Mayer, R., Ross, P., Weinhouse, H., Amikam, D., Volman, G., Ohana, P., Calhoon, R. D., Wong, H. C., Emerick, A. W., and Benziman, M.

1991. Polypeptide composition of bacterial cyclic diguanylic acid-dependent cellulose synthase and the occurrence of immunologically cross-reacting proteins in higher plants. Proc. Natl. Acad. Sci. USA 88:5472-5476.

Roche, P., Debelle, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Denarie, J., and Prome, J.-C. 1991. Molecular basis of symbiotic host-specificity in *Rhizobium meliloti*: *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. Cell 67:1131-1143.

Roche, P., Lerouge, P., Ponthus, C., and Promé. J.-C. 1991. Structural determination of bacterial nodulation factors involved in the *Rhizobium meliloti*-alfalfa symbiosis. J. Biol. Chem. 266:10933-10940.

Rosa, F., Sargent, T. D., Rebbert, M. L., Michaels, G. S., Jamrich, M., Grunz, H., Jonas, E., Winkles, J. A., and Dawid, I. B., 1988. Accumulation and decay of DG42 gene products follow a gradient pattern during *Xenopus* embryogenesis. Devel. Biol. 129:114-123.

Sargent, T. D., and Dawid, I. B. 1983. Differential gene expression in the gastrula of *Xenopus laevis*. Science 222:135-139.

Saxena, I. M., Lin, F. C., and Brown, R. M., Jr. 1990. Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. Plant Mol. Biol. 15:673-683.

Schultze, M., Quiclet-Sire, B., Kondorosi, E., Virelizer, H., Endre, G., Gero, S. D., and Kondorosi, A. 1991. *Rhizobium meliloti* produces a family of sulfated lipooligosaccharides exhibiting different degrees of plant host specificity. Proc. Natl. Acad. Sci. USA 89:192-196.

Schwedock, J. S., and Long, S. R., 1990. ATP sulphurylase activity of the *nodP* and *nodQ* gene products of *Rhizobium meliloti*. Nature 348:644-647.

Silverman, S. 1989. Similar and different domains of chitin synthases 1 and 2 of *S. cerevisiae*: Two isozymes with distinct functions. Yeast 5:459-467.

Spaink, H. C., Sheeley, J. M., VanBrussel, A. A. N., Glushken, J., York, W. S., Tak, T., Geiger, O., Kennedy, E. P., Reinhold, V. N., and Lugtenberg, B. J. J. 1991. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. Nature 354:125-130.

Subbiah, S., and Harrison, S. C. 1989. A method for multiple sequence alignment with gaps. J. Mol. Biol. 209:539-548.

Toole, B. P., Goldbert, R. L., Chi-Rosso, G., Underhill, C. B., and Orkin, R. W. 1984. Hyaluronate-cell interactions. Pages 43-66 in: in The Role of the Extracellular Matrix in Development. Alan R. Liss, New York.

Torok, I., Kondorosi, E., Stepkowki, T., Posfai, J., and Kondorosi, A. 1984. Nucleotide sequence of *Rhizobium meliloti* nodulation genes. Nucleic Acids Res. 12: 9509-952.

Valdivieso, M. H., Mol, P. C., Shaw, J. A., Cabib, E., and Durán, A. 1991. CAL1, a gene required for activity of chitin synthase 3 in *Saccharomyces cerevisiae*. J. Cell Biol. 114:101-109.

VanBrussel, A. A. N., Zaat, S. A. J., Canter-Cremers, H. C., Wijffelman, C. A., Pees, E., Tak, T., and Lugtenberg, B. J. J. 1986. Role of plant exudate and sym plasmid-localized nodulation genes in the synthesis by *Rhizobium leguminosarum* of TSR factor which causes thick and short roots on common vetch. J. Bacteriol. 165:517-522.

Wong, H. C., Fear, A. L., Calhoon, R. D., Eichinger, G. H., Mayer, R., Amikam, D., Benziman, M., Gelfand, D. H., Meade, J. H., Emerick, A. W., Bruner, R., Ben-Bassat, A., and Tal, R. 1990. Genetic organization of the cellulose synthase operon in *Acetobacter xylinum*. Proc. Natl. Acad. Sci. USA 87:8130-8134.