Protein Science (1996), 5:538-541. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Crystallization and preliminary diffraction studies of NodL, a rhizobial *O*-acetyl-transferase involved in the host-specific nodulation of legume roots

STEVEN M. DUNN,[1,3] PETER C.E. MOODY,[1] J. ALLAN DOWNIE,[2] AND WILLIAM V. SHAW[1]

[1] Department of Biochemistry, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK
[2] John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

(RECEIVED November 7, 1995; ACCEPTED December 5, 1995)

Abstract: The *NodL* specified *O*-acetyltransferase from the microbial symbiont *Rhizobium leguminosarum* has been overexpressed in *Escherichia coli* and purified using affinity-elution dye chromatography as the key step. The protein has been crystallized at 20 °C in 18% PEG 600, 0.1 M Tris/HCl buffer, pH 8.5, containing 1% dioxane, 0.25% octyl-β-glucoside, and 5 mM coenzyme A using the hanging drop vapor diffusion method. Ambient temperature X-ray diffraction studies reveal the space group to be hexagonal ($P6_322$) with lattice constants $a = b = 77.08$ Å, $c = 160.6$ Å, and $\alpha = \beta = 90°$, $\gamma = 120°$. Crystals that are flash-frozen to 120 K diffract beyond 2.7 Å.

Keywords: affinity chromatography; coenzyme A; lipo-oligosaccharide; *nod* genes; nodulation; X-ray crystallography

The symbiotic association of certain genera of soil bacteria (collectively termed rhizobia) with species of the *Leguminosae* is initiated by an exchange of diffusible chemical signals (Downie, 1994). In response to flavonoids secreted by host plant roots, rhizobial *nod* genes, involved in the biosynthesis and secretion of specific lipo-chitooligosaccharide Nod factors, are transcriptionally activated (Schlaman et al., 1992). On perception of the Nod-metabolite stimulus, host plant roots undergo a controlled program of cortical cell division and morphological change resulting in bacterial infection and root nodule organogenesis, a prerequisite for nitrogen fixation.

All rhizobial Nod factors thus far isolated possess a common core structure; namely a β-1,4-linked *N*-acetyl-D-glucosamine oligosaccharide backbone of three to five sugar residues with an *N*-linked long-chain fatty-acyl substituent on the nonreducing terminus. Host plant specificity is determined by specific substituents on the reducing and/or nonreducing sugar residues, and also by variations in the length and degree of saturation of the acyl chain (Spaink, 1992). One such substituent, present on lipo-oligosaccharides secreted by members of the genus *Rhizobium*, is an *O*-acetyl group at C-6 on the nonreducing terminal residue of the oligosaccharide chain (Spaink et al., 1991; Fig. 1). The enzyme responsible for this addition is the *nodL* gene product, a flavonoid-inducible cytoplasmic protein of 20.1 kDa (subunit molecular weight) shown to possess acetyltransfer activity in vitro (Bloemberg et al., 1994, 1995). Consistent with this property is the observation that NodL shares extensive sequence similarity with two enterobacterial *O*-acetyltransferases, namely, the *lacA* and *cysE* gene products (37% identity across 181 amino acids and 30% identity across 110 amino acids, respectively; Downie, 1989). LacA, for which a precise in vivo biochemical function has yet to be assigned, acetylates various galactosides in vitro (Beckwith, 1987), whereas CysE (L-serine acetyl transferase) is a component of the cysteine synthase complex (Denk & Bock, 1987). The region of most striking similarity in primary structure between NodL, LacA, and CysE is also present in a group of antibiotic acetylating enzymes from bacteria, some of which *O*-acetylate chloramphenicol (Fig. 2). Because the acetyl-acceptor substrates for the above enzymes are structurally dissimilar, it seems probable that the linear consensus sequences are involved in the structural motifs that bind the common acetyl donor, coenzyme A.

Structures of the few protein·CoA complexes that have been determined at high resolution reveal that the mode of CoA binding, as reflected in both gross conformation of the ligand and the relative contributions of hydrophobic and electrostatic interactions, can differ significantly between proteins. For example, CoA bound to mammalian citrate synthase (Remington et al., 1982) adopts a compact, intramolecular hydrogen bonded conformation that is quite distinct from the more extended structure observed for the type III chloramphenicol acetyltransfer-

Reprint requests to: W.V. Shaw, Department of Biochemistry, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK; e-mail: wsw@le.ac.uk.

[3] Present address: Department of Biochemistry and Physiology, IACR-Rothamsted, Harpenden, Herts AL5 2JQ, UK.

Abbreviations: AcCoA, acetyl coenzyme A; CoA, coenzyme A; CAT, chloramphenicol acetyltransferase; IPTG, isopropylthiogalactoside, EDTA, ethylenediaminetetraacetic acid; PEG 600, polyethylene glycol 600; DTT, dithiothreitol.

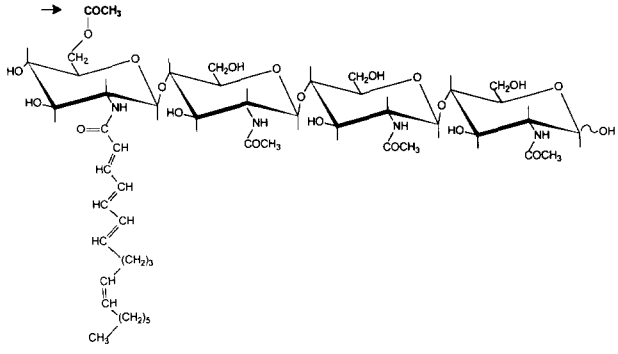

Fig. 1. Structure of the lipo-oligosaccharide signal factor NodRlv-IV (Ac, C18:4; Spaink et al., 1991) excreted by *R. leguminosarum*, illustrating the site of NodL-mediated *O*-acetylation (arrow).

ase binary complex (Leslie et al., 1988). Furthermore, CAT recognizes CoA mainly through a mix of hydrophobic and polar (charged and uncharged) interactions, whereas electrostatic determinants of binding are more prominent in the case of citrate synthase. More recently, the catalytic domain of dihydrolipoyl transacetylase (E2pCD), which is topologically similar to CAT, has been shown to possess both productive and nonproductive binding modes that are characterized by both extended and compact states of bound CoA, respectively (Mattevi et al., 1993).

We have established co-crystallization conditions for the *Rhizobium leguminosarum* NodL acetyltransferase with CoA for several reasons. Firstly, it will be of interest to determine whether the consensus region (Fig. 2) does indeed represent a novel CoA recognition motif. Secondly, although in vitro studies indicate that small chitin oligomers as well as mature lipo-oligosaccharides are acetylated by NodL, kinetic measurements reveal the former to be poor substrates (Bloemberg et al., 1995; S. Dunn, unpubl. data). Structural information should provide an insight into the precise nature of the acetyl acceptor specificity. Finally,

```
rlnod     SIGRHAWIGGGAIILPGVTIGDHAVIGAGSVVTRDVPAGSTAMGNPARVK   182
eclac     TIGNNVWIGSHVVINPGVTIGDNSVIGAGSIVTKDIPPNVVAAGVPCRVI   182
llact     YIEENVWLGAGVIVLPGVRIGKNSVIGAGSLVTKDIPDNVVAFGTPCMVK   184
ecsat     KIREGVMIGAGAKILGNIEVGRGAKIGAGSVVLQPVPPHTTAAGVPARIV   241
atcat     VIGNDVWIGSEAIIMPGITVGDGAVIGTRALVTKDVEPYAIVGGNPAKTI   160
pncat     VIGSDVWIGSEAMIMPGIKIGHGAVIGSRALVAKDVEPYTIVGGNPAKSI   161
savat     EIGNDVWIGRDVTIMPGVKIGDGAIIAAEAVVTKNVAPYSIVGGNPLKFI   169
bserm     VIGNDVWIGQNVTIMPGVIIGDGAIIAANSTVVKSVEPYSIYSGNPAKFI   127

          IG::VWIG  : I:PG: IG: A:IGA ::V:K V P ::::G:P::
```

Fig. 2. Primary sequence alignment spanning the conserved putative CoA-binding region of NodL and related natural product acetyltransferases. The sequences are as follows: NodL from *R. leguminosarum* (rlnod; Surin & Downie, 1988); two galactoside acetyltransferases from *Lactococcus lactis* (llact; Griffin & Gasson, 1994) and *E. coli* (eclac; Hediger et al., 1985); serine acetyltransferase from *E. coli* (ecsat; Denk & Bock, 1987); antibiotic acetylating enzymes from *Agrobacterium tumefaciens* (atcat; Tennigkeit & Matzura, 1991), transposon *Tn2424* (pncat; Parent & Roy, 1992), *Staphylococcus aureus* (savat; Allignet et al., 1993), and *Bacillus sphaericus* (bserm; Monod et al., 1987). The NodL sequence from *R. meliloti* is 80% identical to that of *R. leguminosarum* across this region (Baev & Kondorosi, 1992). Consensus residues that occur in at least six of the eight proteins are highlighted and locations within the region that display equivalence or conservation are indicated by colons.

the question arises as to whether NodL and the plant root cell receptor each bind the signalling molecule by means of similar structural complementarities, albeit with affinities that may be quite different.

The *nodL* gene was isolated from plasmid pIM116 (Surin & Downie, 1988) by *pfu*-mediated PCR-amplification using *Eco*R I/*Hind* III-tagged primers. To facilitate efficient translation, a consensus *Escherichia coli* ribosome binding motif was incorporated into the 5′ primer tag. The major product, following visualization by agarose gel electrophoresis, was purified using the Geneclean kit (Bio 101, Inc.), double digested, and ligated into pKK223-3 (Pharmacia) to yield plasmid pNODL1. DNA sequencing confirmed the absence of any mutations in the *nodL* reading frame. Approximately 30 mg of soluble recombinant NodL was routinely purified from 2 l of *E. coli* (strain JM101) cells containing pNODL1 as follows: four LB-agar plates containing ampicillin (200 μg/ml) were inoculated with a JM101: pNODL1 glycerol stock and incubated at 37 °C for 16 h. Each resultant cell lawn was scraped into 0.5 l of 2YT medium containing 200 μg/ml ampicillin and 0.5 mM IPTG. Flasks were shaken at 250 rpm for 16–18 h and cells were harvested by centrifugation at 4 °C. Cell pellets were resuspended in 7 ml of an ice-cold solution comprising 25 mM Tris/HCl buffer, pH 7.5, 2 mM DTT, 1 mM EDTA and lysed by sonication. Following the removal of cellular debris by centrifugation, the supernatant was heated to 55 °C in a water bath for 2 min, cooled on ice, and then briefly centrifuged to remove precipitated protein. Solid ammonium sulphate was slowly stirred into the ice-cold supernatant to give approximately 42% saturation, and the solution left on ice to equilibrate for 20 min. Precipitated protein was pelleted by centrifugation, resuspended in 40 volumes of 30 mM Tris/HCl, pH 7.5, and loaded onto a Cibacron Blue agarose (Sigma) affinity column (2 × 10 cm) equilibrated in the same buffer at 4 °C. Following the elution of unbound protein, NodL could be affinity-eluted with a pulse of the same buffer containing AcCoA (0.5 mM). The NodL-containing fractions were pooled and applied to a monoQ (10/10) FPLC anion-exchange column equilibrated in 30 mM Tris/HCl, pH 7.5. After extensive washing, NodL was isocratically eluted with equilibration buffer containing 0.1 M NaCl. Leading fractions were assessed for purity by SDS-PAGE, diluted with 2.5 volumes of ice-cold 5 mM Tris/HCl, pH 7.5, and concentrated at 4 °C to greater than 5 mg/ml in Centricon modules (10,000 MW cut-off, Amicon). Protein concentration was estimated using the spectroscopic method of Edelhoch (1967). Storage was at −70 °C in the above buffer and without additives.

Both the independently developed protocol described above and the method of Bloemberg et al. (1994) utilize the selective binding of NodL by immobilized Cibacron blue as the key purification step. The two purification procedures differ in that the elution method reported here employs specific elution by substrate rather than a gradient of increasing ionic strength. In common with Bloemberg and co-workers, we routinely observe a faster migrating, minor contaminant (ca. 2% of the total protein) following SDS-PAGE analysis. N-terminal amino acid sequencing has confirmed this contaminant to be a truncated species of NodL lacking the first 19 residues. Such a fragment is also observed when NodL is overexpressed in, and purified from, *R. leguminosarum* (S. Dunn, unpubl. experiments).

Preliminary crystallization trials were conducted with 4-μl hanging drops at 20 °C according to the vapor diffusion sparse-

matrix sampling method of Jancarik and Kim (1991). Crystals were obtained under several sets of matrix conditions and displayed morphologies ranging from very thin hexagonal wafers to highly disordered needles and polygons. Conditions yielding the latter were subsequently refined through secondary screening and the inclusion of additives. In the absence of CoA, small, ordered hexagonal plates were occasionally observed instead of the large polygons, but refinement of parameters failed to improve reproducibility. The subsequent inclusion of CoA yielded highly reproducible and ordered rod-shaped crystals of hexagonal cross-section at low concentrations (2 mg/ml) of NodL. The crystallization protocol was as follows: stored protein was thawed and diluted to 4.4 mg/ml with 5 mM Tris/HCl buffer, pH 7.5, and briefly centrifuged to remove insoluble material. Hanging drops (12 μl) were prepared by mixing the above solution of NodL with an equal volume of filter-sterilized and degassed crystallization solution (0.1 M Tris/HCl buffer, pH 8.5, 18% PEG 600, 1% dioxane, 0.25% octyl-β-glucoside, and 5 mM CoA) and suspended over 1 ml of the latter in 24-well Linbro boxes. Crystals appeared within 3–4 days and reached maximum size (0.13 × 0.13 × 1.0 mm) after about 5–6 weeks (Fig. 3).

Data were collected from harvested crystals using a Raxis IIc image plate system (Rigaku/Molecular Structure Corp.) mounted on a Rigaku RU200HB rotating anode generator with a copper anode, a nominal focus of 0.3 × 3.0 mm and powered at 50 KV, 100 mA. A graphite monochromator was used to select CuKα (1.5418 Å) radiation. The measurement and reduction of data employed MOSFLM (V5.23; A.G.W. Leslie, LMB Cambridge, UK) and the CCP4 suite (1994). Initial studies at room temperature allowed the space group and cell dimensions to be determined. Data from a 2.0° oscillation exposure was subjected to the auto-indexing routine of Kabsch (1993) that gave a cell consistent with a trigonal or hexagonal space group. Collection of a full data set with multiple redundancy to 3.5 Å allowed the cell to be refined to $a = b = 77.08$ Å, $c = 160.6$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$. Examination of the distribution of intensities allowed the space group to be unambiguously assigned as $P6_322$ (number 182). The R_{merge} for 28,356 observations of 2,730 reflections to 4.0 Å was 9.3%, with an average $I/\text{sig}I$ of 6.7.

The above data are not adequate for high-resolution studies and hence low-temperature techniques were employed. A crystal was transferred to a solution with the same composition as that used for crystallization save for the incorporation of glycerol to a final concentration of 25% (v/v). The crystal was next mounted in a rayon loop and flash-frozen to 120 K. Data to 2.7 Å was clearly seen, and 33° of rotation gave an R_{merge} of 8.4% for 29,018 observations of 7,760 reflections (97.5% complete). However, the data in the outer bin (2.84–2.7 Å) have only 28% of the intensities greater than three times their estimated standard deviations. The use of more powerful X-ray sources is expected to further improve the data. The unit cell alters to $a = b = 75.73$ Å, $c = 160.86$ Å at 120 K, reducing the volume by 3%. Assuming a molecular weight of 20.1 kDa, the V_m value (Mathews, 1968) is 3.3 Da/Å^3, which corresponds to an estimated solvent content of 55%. Structure determination will be attempted initially using both traditional heavy atom methods and replacement of the methionine residues with seleno-methionine.

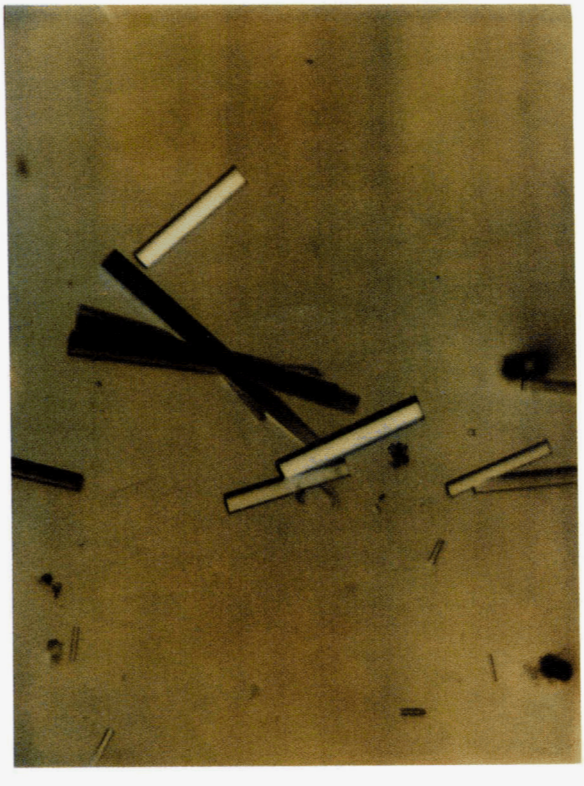

Fig. 3. Crystals of recombinant NodL.

Note added in proof

The structure of LpxA, which shares the [LIV]-G-x(4) hexapeptide signature, is mostly built of parallel β-helix organised as a trimer, the average spacing between parallel β-sheets being 4.8 Å (Raetz & Roderick, 1995). Inspection of the NodL diffraction data reveals the reflection 0 0 34 is by far the strongest, with an amplitude some 20 times the mean value for all the reflections to 4.0 Å. This reflection corresponds to a spacing of 4.73 Å along the 6_3 direction and is consistent with NodL having a similar β-helix trimer structure.

Acknowledgment: This work was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom.

References

Allignet J, Loncle V, Simenel C, Delepierre M, El Sohl N. 1993. Sequence of a staphylococcal gene, *vat*, encoding an acetyltransferase inactivating the A-type compounds of virginiamycin-like antibiotics. *Gene 130*: 91–98.

Baev N, Kondorosi A. 1992. Nucleotide sequence of the *Rhizobium meliloti nodL* gene located in locus n5 of the *nod* regulon. *Plant Mol Biol 18*:843–846.

Beckwith J. 1987. The lactose operon. In: Neidhandt FC, ed. *Escherichia coli & Salmonella typhimurium: Cellular and molecular biology*. Washington, DC: American Society for Microbiology. pp 1444–1452.

Bloemberg GV, Lagas RM, van Leeuwen S, Van der Marel GA, van Boom JH, Lugtenberg BJJ, Spiank HP. 1995. Substrate specificity and kinetic studies of nodulation protein NodL of *Rhizobium leguminosarum*. *Biochemistry 34*:12712–12720.

Bloemberg GV, Thomas-Oates JE, Lugtenberg BJJ, Spaink HP. 1994. Nodulation protein NodL of *Rhizobium leguminosarum O*-acetylates lipooligosaccharides, chitin fragments and *N*-acetylglucosamine in vitro. *Mol Microbiol 11*:793–804.

CCP4. 1994. Collaborative Computational Project, Number 4. The CCP4

suite: Programs for protein crystallography. *Acta Crystallogr D 50*: 760–763.

Denk D, Bock A. 1987. L-cysteine biosynthesis in *Escherichia coli*: Nucleotide sequence and expression of the serine acetyltransferase (*cysE*) gene from the wild-type and a cysteine-excreting mutant. *J Gen Microbiol 133*:515–525.

Downie JA. 1989. The *nodL* gene from *Rhizobium leguminosarum* is homologous to the acetyl transferases encoded by *lacA* and *cysE*. *Mol Microbiol 3*:1649–1651.

Downie JA. 1994. Signalling strategies for nodulation of legumes by rhizobia. *Trends Microbiol 2*:318–324.

Edelhoch E. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*:1948–1954.

Griffin HG, Gasson MJ. 1994. The gene (*lacA*) encoding galactoside acetyltransferase from *Lactococcus lactis*. *Biotech Letts 16*:1125–1130.

Hediger MA, Johnson DF, Nierlich DP, Zabin I. 1985. DNA sequence of the lactose operon: The *lacA* gene and the transcriptional termination region. *Proc Natl Acad Sci USA 82*:6414–6418.

Jancarik J, Kim SH. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J Appl Crystallogr 24*:409–411.

Kabsch W. 1993. Automatic processing of rotation diffraction data from initially unknown symmetry and cell constants. *J Appl Crystallogr 24*:795–800.

Leslie AGW, Moody PCE, Shaw WV. 1988. Structure of chloramphenicol acetyltransferase at 1.75 angstrom resolution. *Proc Natl Acad Sci USA 85*:4133–4137.

Mathews BW. 1968. Solvent content of protein crystals. *J Mol Biol 33*: 491–497.

Mattevi A, Obmolova G, Kalk KH, Teplyakov A, Hol WGJ. 1993. Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p). *Biochemistry 32*:3887–3901.

Monod M, Mohan S, Dubnau D. 1987. Cloning and analysis of *ermG*: A new macrolide-lincosamide-streptogramin B resistance element from *Bacillus sphaericus*. *J Bacteriol 169*:340–350.

Parent R, Roy PH. 1992. The chloramphenicol acetyltransferase gene of Tn*2424*: A new breed of *cat*. *J Bacteriol 174*:2891–2897.

Raetz CRH, Roderick SL. 1995. A left-handed parallel β helix in the structure of UDP-*N*-acetylglucosamine acetyltranferase. *Science 270*:997–1000.

Remington S, Wiegand G, Huber R. 1982. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 angstroms resolution. *J Mol Biol 158*:111–152.

Schlaman HRM, Okker RJH, Lugtenburg BJJ. 1992. Regulation of nodulation gene expression by NodD in rhizobia. *J Bacteriol 174*:5177–5181.

Spaink HP. 1992. Rhizobial lipo-oligosaccharides: Answers and questions. *Plant Mol Biol 20*:977–986.

Spaink HP, Sheeley DM, Van Brussel AAN, Głushka J, York WS, Tak T, Geiger O, Kennedy EP, Reinhold VN, Lugtenburg BJJ. 1991. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature 354*:125–130.

Surin BP, Downie JA. 1988. Characterization of the *Rhizobium leguminosarum* genes *nodLMN* involved in efficient host-specific nodulation. *Mol Microbiol 2*:173–183.

Tennigkeit J, Matzura H. 1991. Nucleotide sequence analysis of a chloramphenicol-resistant determinant from *Agrobacterium tumefaciens* and identification of its gene product. *Gene 98*:113–116.