FEBS Letters 388 (1996) 66–72 FEBS 17123

NMR investigations of the structural properties of the nodulation protein, NodF, from *Rhizobium leguminosarum* and its homology with *Escherichia coli* acyl carrier protein

Ranajeet Ghose[a], Otto Geiger[b,*], James H. Prestegard[a,**]

[a]*Department of Chemistry, Yale University, New Haven, CT 06520, USA*
[b]*Technische Universität Berlin, Institut für Biotechnologie, FG Technische Biochemie, Seestraße 13, 13353 Berlin, Germany*

Received 28 February 1996; revised version received 23 April 1996

Abstract Heteronuclear NMR methods have been used to elucidate the secondary structure and the general tertiary fold of the protein NodF from *Rhizobium leguminosarum*. A similarity to acyl carrier proteins of the fatty acid synthase system had been suggested by the presence of a phosphopantetheine prosthetic group and a short stretch of sequence homology near the prosthetic group attachment site. NMR results suggest that the structural homology extends well beyond this region. Both proteins have three well-formed helices which fold in a parallel-antiparallel fashion and a prosthetic group attachment site near the beginning of the second helix.

Key words: Acyl carrier protein; Nodulation factor; Fatty acid synthesis; NMR; Secondary structure

1. Introduction

Bacteria of the genus *Rhizobium* are able to interact symbiotically with leguminous plant hosts, leading to the formation of nitrogen-fixing root nodules. Plant flavonoids induce the *nod* genes, which are rhizobial genes essential for nodulation [1]. In *Rhizobium leguminosarum* bv. viciae the *nodABC* and the *nodFEL* operons are involved in the production of lipochitin oligosaccharide signals, which mediate host specificity and which are able to induce nodule primordia [2] and preinfection thread structures [3] on the host plant *Vicia sativa*.

The NodE protein appears to be homologous to a group of β-ketoacyl synthases and NodF shares homology with acyl carrier proteins (ACPs) [4]. From all the *nod* genes, expression of *nodFE* is sufficient for the synthesis of novel multi-unsaturated fatty acids [5–7]. Specifically, the NodFE proteins of *R. leguminosarum* bv. viciae are required for the synthesis of host-specific *trans-2,trans-4,trans-6,cis-11*-octadecatetraenoic acid [6]. This fatty acyl residue has been recognized as an important structural element in conferring host specificity to lipochitin oligosaccharide signals [2].

The homology between the constitutively expressed ACP from *Escherichia coli* (AcpP) [8] and NodF is concentrated around a consensus sequence (-LGXDSL-) thought to be required for the attachment of a 4′-phosphopantetheine prosthetic group to a conserved serine residue (S36 of ACP and S45 of NodF). Like AcpP, the NodF protein can carry this prosthetic group [9]. However, the sequence homology outside the region near the binding site is quite low and this raises the question whether the functional similarities between AcpP and NodF are dictated only by this small stretch of sequence homology or whether there exist other structural similarities between the two proteins, perhaps between their secondary structures and tertiary folds.

ACPs from bacteria and plants have resisted detailed structural work, possibly because of the presence of highly labile structural elements. However, a low resolution NMR structure of AcpP does exist [10]. Here we present a determination of the secondary structure and some initial indications of the general fold of the NodF protein from *R. leguminosarum*, along with an analysis of its structural similarities with *E. coli* AcpP.

2. Materials and methods

2.1. Overexpression and purification of NodF protein

The plasmid pMP2301 was obtained by cloning the *NodF* gene in the expression vector pET9a as described earlier [7] and the strain *E. coli* BL21 (DE3) [11], harboring pMP2301, was used for expression of the NodF protein. In short, this strain was grown on the medium M9 [12] in the presence of kanamycin (50 μg/ml) using ammonium [^{15}N]chloride (1 g/l) as the sole source of nitrogen. At a cell density of 5×10^8 cells per ml, IPTG was added to a final concentration of 0.4 mM. After 4 h of induction, cells were harvested and stored in a freezer at −20°C. Frozen cells (1.3 g) were suspended in 15 ml of buffer A (50 mM Tris-HCl pH 6.8, 0.1 M KCl) and passed twice through a French pressure cell at 20000 lb/in^2. The cell-free extract was slowly stirred at 4°C and isopropanol was added dropwise to a final concentration of 50% (v/v). After 60 min incubation at 4°C, the precipitate obtained was removed by centrifugation at $10000\times g$ for 30 min. The supernatant was dialyzed for 16 h against 2 liters of buffer A and the dialysate was applied to a 30 ml column of DEAE-52 cellulose (Whatman), which had been equilibrated with buffer B (10 mM bis-Tris-HCl pH 6.0, 0.1 M NaCl, 1 mM CHAPS). The column was washed with 100 ml of buffer B. Elution was performed with a linear gradient from 0.1 to 1 M NaCl in buffer B in a total volume of 130 ml, fractions (2.7 ml) were collected, and aliquots were analyzed by PAGE. Only those fractions were combined in which the NodF protein made up more than 95% of the total protein in the fraction, as determined by densitometry. The combined fractions were dialyzed against 2 mM potassium phosphate, pH 6 and lyophilized. From 1 liter of IPTG-induced cell suspension, 1.3 g of wet cells were obtained which contained a total of 158 mg of protein. The dialyzed supernatant after isopropanol precipitation contained 55 mg of protein and the combined fractions from DEAE cellulose chromatography contained 45 mg of purified NodF protein.

In order to characterize the product, electrophoresis on nondenaturing gels [13] was performed. Cell-free extract from *Rhizobium leguminosarum* RBL5560.pMP1255 [9] that had been induced with naringenin, cell-free extracts from *E. coli* BL21(DE3).pMP2301 grown without inducer or in the presence of IPTG, and fractions from in-

*Corresponding author. Fax: (49) (30)-453-6067.

**Corresponding author. Fax: (1) (203)-432-8918.

PII S0014-5793(96)00512-1

duced *E. coli* BL21(DE3).pMP2301 cells after isopropanol precipitation and after DEAE-52 were compared (Fig. 1).

CD spectra of a sample containing 0.2 mg/ml of NodF were obtained at 25.6°C using a AVIV Model 60DS spectrophotometer (Lakewood, NJ). A scan was run from 260 nm down to 185 nm with data collected every 0.5 nm. Fifteen scans were averaged. Quantitative analysis of the results was carried out using the method of Yang et al. [14]. This analysis indicated 34% helix, 2% β-sheet, 16% turn and 47% random coil.

2.2. NMR spectroscopy

NMR samples were 4.5 mM in NodF dissolved in 215 μl of phosphate buffer (14 mM in phosphate, 86% H_2O, 14% D_2O) containing 5 mM DTT (to prevent dimerization of prosthetic group carrying species) and a trace of NaN_3. Two sets of samples were prepared, one at pH 6.0 and another at pH 6.95. Performing experiments at more than one pH value is often useful to avoid accidental degeneracies in chemical shift.

All NMR experiments were performed on a GE Omega 500 spectrometer operating at a ^1H frequency of 500 MHz and equipped with a triple-resonance probe capable of generating magnetic field gradients along all three axes. All spectra were collected at 30°C with ^{15}N decoupling during acquisition using GARP [15].

Amide ^{15}N-^1H chemical shifts were correlated using a gradient-selected, sensitivity-enhanced HSQC [16] with a selective flip-back pulse on water to suppress water and minimize saturation of amide protons which may exchange rapidly with water protons [17]. 512×256 complex points were collected in the direct and indirect dimensions respectively using sweep-widths of 5510.0 Hz (^1H) and 1334.0 (^{15}N) Hz. Sixteen scans per point were collected, this required a total experiment time of about 6 h. Quadrature was obtained in the indirect dimension in a hypercomplex [18] fashion using add-subtract gradients [16,19].

Two sets of TOCSY-HSQC experiments (a modification of the TOCSY-HMQC experiment with a gradient-selected, sensitivity-enhanced HSQC used at the end of the TOCSY part of the experiment, instead of the HMQC) [16,20] were collected: one at pH 6.0 with a mixing time of 60 ms, and another at pH 6.95 with a mixing time of 40 ms. For the former dataset 128, 30 and 512 complex datapoints were acquired with sweep-widths of 5989 Hz, 1333 Hz and 3003 Hz in t_1 (^1H), t_2 (^{15}N) and t_3(^1H) respectively. Acquisition parameters for the latter dataset were similar. The DIPSI-2 [21] mixing sequence was employed to obtain isotropic mixing, with the carrier placed at 6.0 ppm to optimize the Hartmann-Hahn match between amide and α protons. The folded-over peaks belonging to the most downfield -NH protons were unfolded using post-acquisition data processing [22]. Solvent suppression was achieved using a Shinnar-Leroux [23] pulse followed by a SCUBA delay. The residual water signal was further suppressed by the gradients used for coherence selection in the HSQC part of the experiment. Quadrature in the t_1 dimension was obtained in a States-TPPI fashion [24] while for the t_2 dimension quadrature was obtained as in the case of the HSQC experiment described above.

Two sets of NOESY-HSQC (modification of the NOESY-HMQC experiment) [16,25] spectra were also collected, one at pH 6.0 with a mixing time of 100 ms and another at pH 6.95 with a mixing time of 200 ms. The former spectrum was acquired with 193, 32 and 512 complex points in t_1 (^1H), t_2 (^{15}N) and t_3 (^1H) respectively with sweep-widths of 5989 Hz, 1333 Hz and 3200 Hz and acquisition parameters for the latter dataset were similar. Solvent suppression in the former dataset was accomplished in the same way as in the TOCSY-HSQC experiments described above. For the 200 ms dataset, however, no presaturation of the water signal was used, instead the water suppression was achieved using a flip-back pulse on water as in the HSQC experiment described above. All the TOCSY-HSQC and NOESY-HSQC spectra were acquired with four scans per transient and required total experiment times between 38 and 48 h.

In addition to the above basic experiments, several supporting experiments were employed. These follow literature descriptions closely and use acquisition parameters similar to those described above. The HNHA experiment of Vuister and Bax [26] was run on the pH 6.0 sample to obtain the 3-bond α-amide ($^3J_{HNH\alpha}$) coupling constants.

An HSQC-NOESY-HSQC (modification of the HMQC-NOESY-HMQC experiment) [27] was used to obtain NOE information between degenerate amide protons. Deuterium exchange spectra were measured using rapid acquisition HSQC spectra (4–20 min each) on a 3.5 mM NodF sample in 25 mM in potassium phosphate, pH 6.1 containing 5 mM DTT and traces of NaN_3. The sample had been prepared just prior to acquisition. A homonuclear ^1H-^1H NOESY dataset with a mixing time of 200 ms was collected using a Shinnar-Leroux [23] pulse for water suppression.

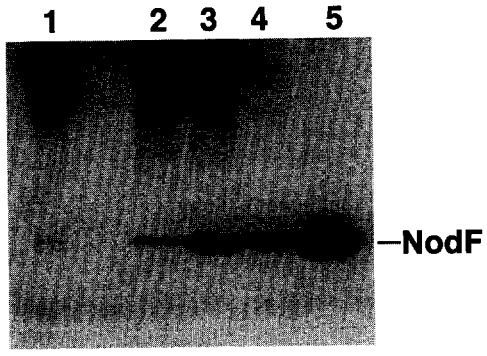

Fig. 1. Purification of NodF Protein. Electrophoresis in nondenaturing gels. The various lanes are: cell-free extract from *Rhizobium leguminosarum* RBL5560.pMP1255 induced with naringenin (lane 1); cell-free extracts from *E. coli* BL21(DE3).pMP2301 grown without inducer (lane 2) or in the presence of IPTG (lane 3); fraction from induced *E. coli* BL21(DE3).pMP2301 after isopropanol precipitation (lane 4) and after DE-52 (lane 5).

All data processing was done using Felix 2.30 from Biosym Technologies Inc. The 3D datasets were typically apodized using either a 90° shifted sinebell or a Kaiser window (window parameter = 10) in the amide proton and ^{15}N dimensions and by a 70° shifted skewed sinebell (skew parameter = 0.40) in the NOESY or TOCSY dimensions. The data in the indirect ^{15}N dimension in the HSQC-NOESY-HSQC and the α proton dimension in the HNHA experiment were apodized using a Kaiser window with the same parameters as above. The high resolution 2D HSQC experiment was processed using a Kaiser window (window parameter = 10) in both dimensions. The exchange data were apodized using 30° shifted skewed sinebell functions (skew = 0.4) in both dimensions. The window sizes in all cases were equal to the number of complex datapoints acquired and the data were zero-filled to double their size (to the nearest multiple of 2) in all cases, prior to Fourier transformation. The data in the homonuclear experiment were zero-filled to 1 K in each dimension and apodized using a 30° shifted skewed sinebell (skew = 0.4) with a window size of 512 points in both dimensions.

3. Results and discussion

3.1. Assignment of NMR spectra

Fig. 2 shows the ^{15}N-^1H HSQC spectrum of NodF at pH 6.0 at 30°C. Since the HSQC spectrum correlates the amide nitrogen to its attached proton, one would expect to see one peak for every backbone amide proton (except for M1), two peaks from each of the sidechain -NH_2 groups of the N and Q residues (NodF has 2 Qs and 6 Ns) and one peak from the -NH group on each of the two R sidechains. The sidechain -NH_2s from the Rs exchange too rapidly with water to be observed [28]. In addition to these, two peaks due to the H^ε protons of W53 and W71 appear shifted downfield of the normal amide proton region. The dispersion in the ^1H and ^{15}N dimensions combined was sufficient to unambiguously resolve and assign nearly all the resonances in this spectrum.

The assignment of the above resonances was based on the standard ^{15}N-directed strategy [29] utilizing both through-bond connectivities to identify spin systems characteristic of various amino acid types (TOCSY-HSQC) and through-space correlation to make sequential assignments (NOESY-HSQC). The assignment procedure began with picking peaks from the 3D ^{15}N-^1H TOCSY-HSQC experiment. This was done using

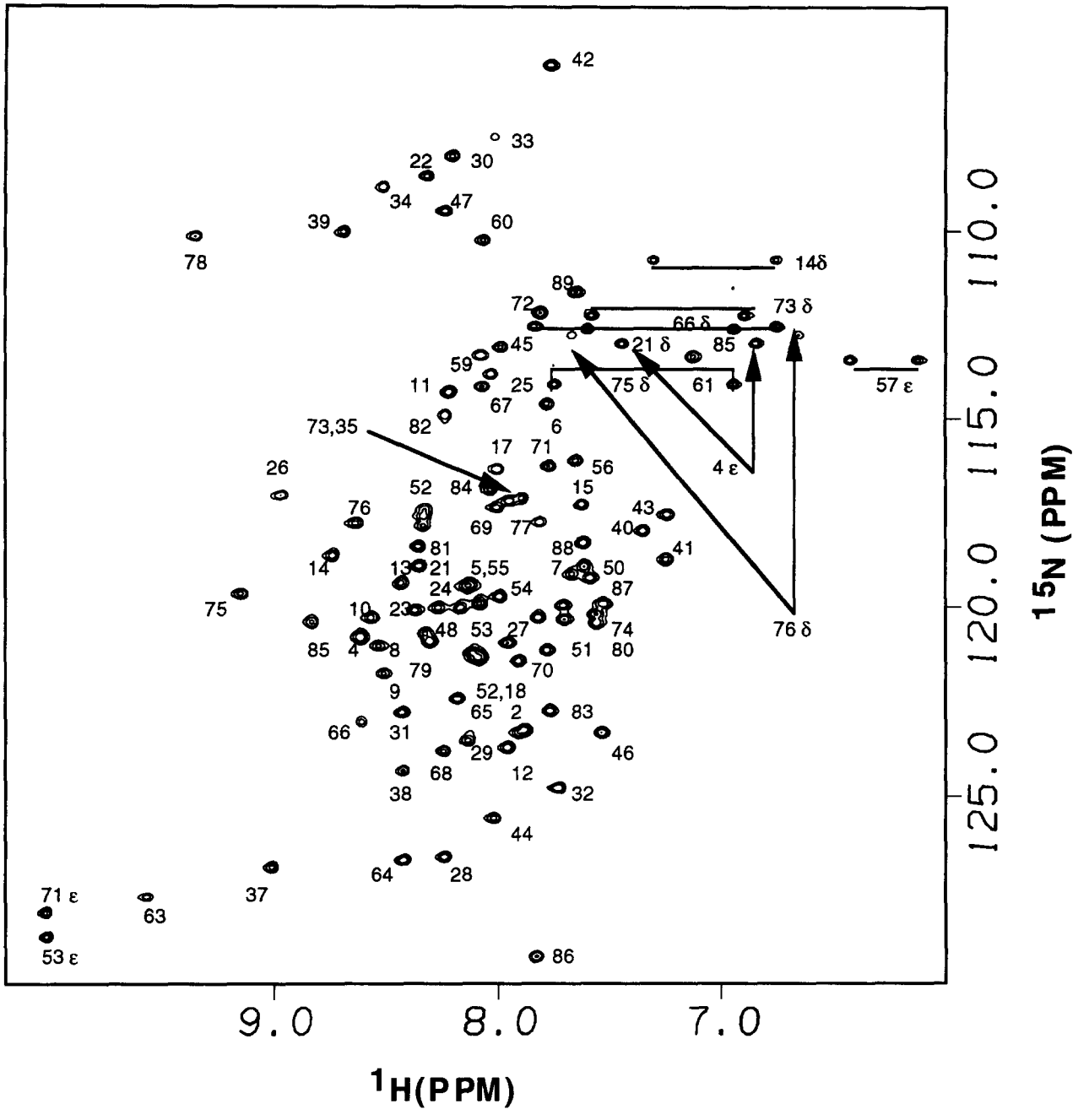

Fig. 2. ^{15}N-^{1}H GSE-HSQC (with water flip-back pulse) on the NodF protein at pH 6.0 at 30°C showing the residue numbers corresponding to the various peaks. The peak due to G86 (labelled 86) is folded over.

the Peakpick program of Chylla [30]. Automatic labelling of peaks belonging to the same spin system was done using a relational database program (DBASE IV), and preliminary assignments to a particular amino acid type were aided by a neural network trained on several assigned datasets from homologous proteins [31]. These preliminary assignments were transferred to appropriate peaks in a 3D NOESY-HSQC dataset for resonances expected to give intra-residue NOE peaks. The subsequent assignment of inter-residue peaks and remaining intra-residue peaks was carried out using a program written in C and was based on an index which reflected the best chemical shift match with assigned TOCSY-HSQC peaks and on the occurrence of multiple connectivities between the same pair of residues. The occurrence of multiple inter-residue connections in these backbone-edited NOESY spectra was taken as evidence of sequential placement of amino acid types. These sequential placements were compared with the NodF sequence to determine the correct sequential assignments. As expected, in most cases $d_{\alpha N(i,i+1)}$ connectivities were seen. In those cases where these were lacking, $d_{\beta N(i,i+1)}$ correlations supported proper sequential assignment. It was possible to assign peaks to almost all the residues using the above procedure. Peaks from the residues A2 and S26 were, however, assigned through a process of elimination. It was not possible to assign the residues D3, N21 and the C-terminus, residues K90 through V92, because of overlap with

Table 1
^{15}N and ^{1}H assignments for the NodF protein at pH 6.0 and 30°C

Residue	$^{15}N^{a}$	NH^{a}	H^{α}	H^{β}	Others
Met-1					
Ala-2	123.3	7.89	4.39*	1.89	
Asp-3					
Gln-4	120.8	8.62	4.03	2.19/2.04	H^{γ} 2.65/2.37
					N^{ε} 113.0
					H^{ε} 7.45/6.85
Leu-5	119.5	8.15	4.13	2.12/1.76	H^{γ} 1.55
					CH_3^{δ} 0.89
Thr-6	114.6	7.78	4.34	3.70	CH_3^{δ} 1.13
Leu-7	119.1	7.67	3.96	1.94/1.57	CH_3^{δ} 0.86
Glu-8	121.0	8.54	4.06	1.96	H^{γ} 2.48/2.34
Ile-9	121.7	8.51	3.75	2.02	H^{γ} 1.13
					CH_3^{γ} 0.82
Ile-10	120.3	8.57	3.47	1.88	H^{γ} 2.02
					CH_3^{γ} 0.87
					CH_3^{δ} 0.76
Ser-11	114.3	8.22	4.28	4.06	
Ala-12	123.7	7.96	4.23	1.65	
Ile-13	119.4	8.44	3.70	2.04	CH_3^{γ} 0.88
					CH_3^{δ} 0.74
Asn-14	118.6	8.74	4.40	2.93/2.81	N^{δ} 110.8
					H^{δ} 7.31/6.76
Lys-15	117.3	7.63	4.06	2.01/1.78	H^{γ} 1.69/1.47
Leu-16	119.2	7.59	4.29	1.86	CH_3^{δ} 0.97
Val-17	116.3	8.01	4.01	2.30	H^{γ} 1.10
Lys-18	121.2	8.11	4.28*	1.90/1.50	H^{γ} 1.67/0.92
Ala-19	123.3	7.89	4.33*	1.50	
Glu-20	119.9	8.09	4.31	2.21/1.95	
Asn-21					
Gly-22	108.5	8.32	4.01/4.75		
Glu-23	120.1	8.37	4.32	2.14/2.02	H^{γ} 2.30
Arg-24	120.0	8.27	4.47	1.87/1.70	H^{δ} 3.26
Thr-25	113.8	8.04	4.39		
Ser-26	117.0	8.98	4.80		
Val-27	120.9	7.96	4.14	2.11	CH_3^{γ} 0.96
Ala-28	126.6	8.25	4.36	1.40	
Leu-29	123.5	8.14	4.25	2.21/1.85	H^{γ} 1.10
					CH_3^{δ} 0.91
Gly-30	108.0	8.21	4.75/3.97		
Glu-31	122.8	8.43	4.38	2.12/1.97	H^{γ} 2.32
Ile-32	124.8	7.74	4.09	2.12	CH_3^{γ} 0.93
Thr-33	107.5	8.02	4.65*		
Thr-34	108.8	8.52	4.75*	3.83	CH_3^{γ} 1.15
Asp-35	117.1	7.90	4.69	2.78/2.42	
Thr-36	120.3	7.82	3.95	4.22	CH_3^{γ} 1.34
Glu-37	126.9	9.01	4.46	2.11/1.33	H^{γ} 2.57/2.38
Leu-38	124.3	8.43	3.91*	1.85/1.51	CH_3^{δ} 0.67
Thr-39	110.0	8.69	4.22	4.36	CH_3^{γ} 1.41
Ser-40	118.0	7.35	4.46	4.12	
Leu-41	118.7	7.25	4.42	1.91/1.71	CH_3^{δ} 0.89
Gly-42	105.6	7.77	4.10/3.88*		
Ile-43	117.5	7.25	4.30	1.83	H^{γ} 1.13
					CH_3^{γ} 0.91
					CH_3^{δ} 0.79
Asp-44	125.6	8.03	4.65	3.02/2.67	
Ser-45	113.1	7.99	4.75*	3.83	
Leu-46	123.3	7.54	4.21	1.84/1.62	CH_3^{δ} 0.91
Gly-47	109.5	8.24	4.08/3.93		
Leu-48	120.7	8.33	4.04*	1.57	CH_3^{δ} 0.88
Ala-49	120.3	7.70	4.05*	1.56	
Asp-50	118.9	7.62	4.65	3.02/2.84	
Val-51	121.1	7.78	3.86	2.48	H^{γ} 1.09
					CH_3^{γ} 0.93
Leu-52	117.8	8.34	4.01	2.90/1.98	CH_3^{δ} 0.76
Trp-53	120.0	8.17	4.60	3.55	H^{δ} 7.42
					N^{ε} 128.7
					H^{ε} 10.01
Asp-54	119.7	8.00	4.32	3.02/2.76	
Leu-55	119.5	8.13	4.13*	1.44	
Glu-56	116.1	7.65	3.84*	1.91	
Gln-57	120.7	8.33	4.03*	1.79/1.58	N^{ε} 113.4
					H^{ε} 6.43/6.21
Leu-58	119.9	7.71	3.98	1.28	CH_3^{δ} 0.76/0.64
Tyr-59	113.3	8.08	4.51	3.20/2.66	
Gly-60	110.2	8.07	4.30/3.95		
Ile-61	113.3	7.13	4.53	1.77	H^{γ} 1.02
					CH_3^{γ} 0.86
Lys-62	121.3	8.09	4.47*	1.36	H^{ε} 3.01
Ile-63	127.7	9.57	4.07	1.87	CH_3^{γ} 0.93
					CH_3^{δ} 0.80
Glu-64	126.7	8.42	4.42	2.06/1.87	H^{γ} 2.20
Met-65	122.4	8.19	4.37	2.15/1.83	H^{γ} 2.28
Asn-66	123.0	8.61	3.98*	2.43	N^{δ} 112.5
					H^{δ} 7.84/6.75
Thr-67	114.1	8.08	4.16*	4.74	CH_3^{γ} 1.22
Ala-68	123.8	8.25	4.75*	1.46	
Asp-69	117.3	8.01	4.52*	2.66	
Ala-70	121.4	7.91	4.14	1.31	
Trp-71	116.0	7.78	4.42	3.85/3.38	H^{δ} 7.26
					N^{ε} 128.1
					H^{ε} 10.02
Ser-72	112.2	7.81	4.27	3.94	
Asn-73	117.1	7.95	4.78*	2.89/2.68	N^{δ} 112.2
					H^{δ} 7.58/6.90
Leu-74	120.2	7.56	4.38*	1.83	H^{γ} 1.06
Asn-75	119.6	9.15	4.94	2.83	N^{δ} 114.1
					H^{δ} 7.75/6.95
Asn-76	117.7	8.64	5.33	2.89/2.62	N^{δ} 112.7
					H^{δ} 7.67/6.66
Ile-77	117.7	7.82	3.58*	2.32	CH_3^{γ} 0.69
Gly-78	110.1	9.35	4.13/3.73		
Asp-79	120.9	8.31	4.49	3.15/2.90	
Val-80	120.4	7.56	3.49*	2.55	CH_3^{γ} 0.86/0.73
Val-81	118.4	8.36	3.38	2.41	CH_3^{γ} 1.10/0.90
Glu-82	114.9	8.24	4.11	2.01/1.41	H^{γ} 2.48/2.14
Ala-83	122.7	7.77	4.13	1.34	
Val-84	116.8	8.04	3.57	2.11	CH_3^{γ} 0.98/0.77
Arg-85	120.4	8.84	3.56*	1.73	H^{γ} 0.76
Gly-86	103.0	7.85	4.74/3.93*		
Leu-87	119.9	7.53	4.38	1.95/1.62	
Leu-88	118.3	7.62	4.24	1.71/1.48	CH_3^{δ} 0.56
Thr-89	111.6	7.65	4.34	4.34	CH_3^{γ} 1.25
Lys-90					
Glu-91					
Val-92					

* indicates α assignments not confirmed by HNHA experiment and assignment solely on the basis of TOCSY-HSQC experiment.
Peaks not listed are either missing or overlapped.

assigned residues or a lack of NOEs for residues in these regions of the protein. The assignments for the ^{1}H and ^{15}N chemical shifts for NodF are shown in Table 1.

3.2. Elements of secondary structure

CD data indicate that NodF is a predominantly α-helical protein. For proteins such as this one would expect to see several characteristic indicators in NMR data. These indicators are summarized in Fig. 3. In terms of sequential NOEs, predominantly weak $d_{\alpha N(i,i+1)}$ connectivities and strong $d_{NN(i+1)}$ connectivities are expected in the helical regions [32]. This pattern was clearly seen in three segments of the protein. These were Q4 through L16, L46 through L58 and I77 through G86. Support for the helical segments in these regions was sought by measuring the $^{3}J_{HNH\alpha}$ coupling constants. Coupling constants smaller than 6 Hz were taken to

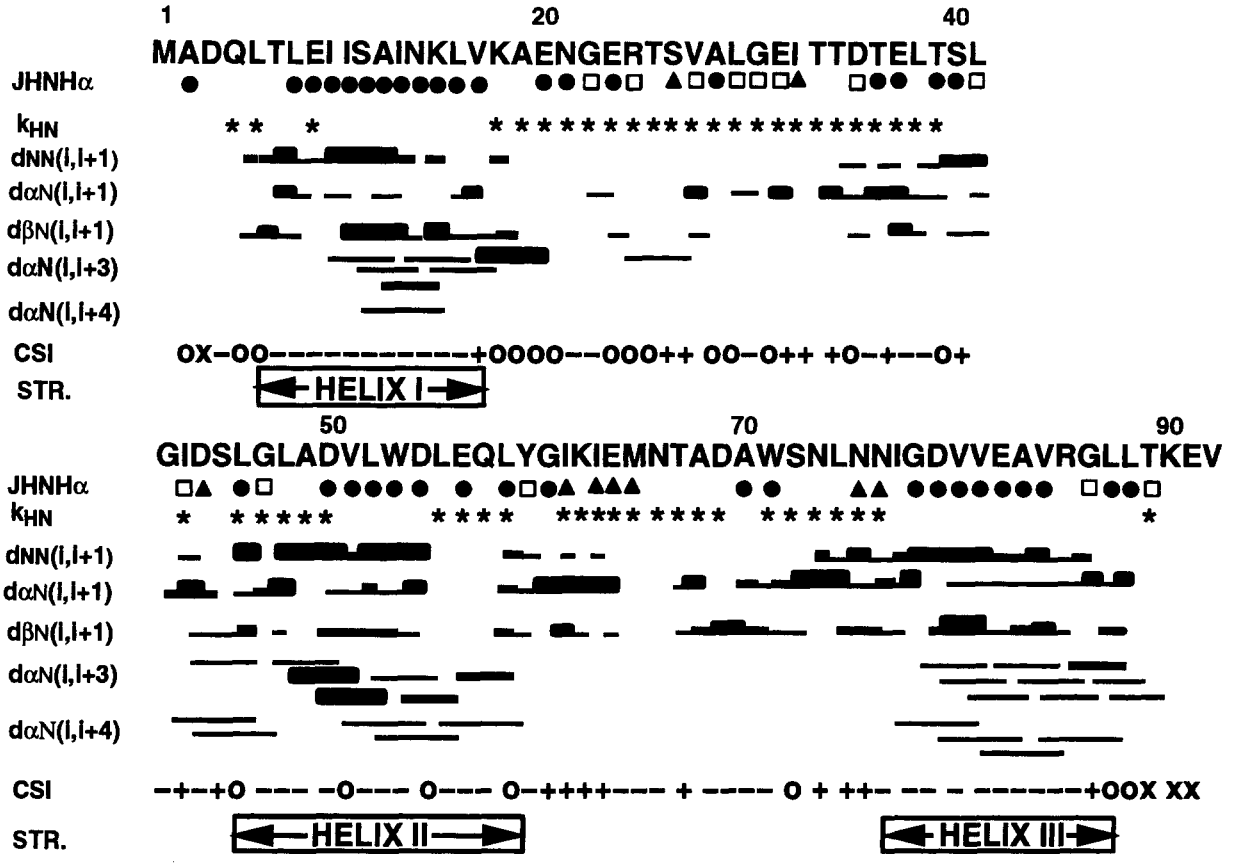

Fig. 3. NOE correlation patterns shown for the NodF protein. The NOEs are classified as strong, medium or weak and the thickness of the vertical bars represents this classification. For the $^3J_{HNh\alpha}$ values <6.5 Hz are represented by filled circles, those of 6.5–8.5 Hz are represented by open squares and those >8.5 Hz are represented by filled triangles. For the CSI (chemical shift indices) the −, + and 0 represent values of −1, +1 and 0 respectively. The xs denote unassigned residues. Rapidly exchanging amide protons are represented by asterisks.

indicate helices [32]. Other evidence which supported the presence of helices in the three regions was an analysis of the α proton chemical shifts in the manner of Wishart et al. [33] and also non-sequential NOE connectivities unique to helical regions ($d_{\alpha N(i,i+3)}$, $d_{\alpha N(i,i+4)}$) [32]. Weak $d_{\alpha N(i,i+3)}$ connectivities were seen in the sequence I9 through N14 (barring A12); L46 through L58 (except L48, D50, W53, D54, E56, Q57) and throughout the sequence G78 through G86. $d_{\alpha N(i,i+4)}$ connections, while rare in the first sequence (seen only from I10), were more abundant in the second and third sequences (seen from V51, L52 and L55; I77, G78, D79 and V81). Insisting on support from at least three of the above-mentioned indications of the presence of helices, we assign helix I to the sequence T6 through L16, helix II to the sequence L46 through L58 and helix III to the sequence I77 through G86. Parts of the regions from K18 to S45 between helix I and helix II and from M65 to N76 between helix II and helix III do not seem to have a definite secondary structure as indicated by lack of a well-defined NOE pattern.

Amide exchange data provide additional insight into the stability of the various secondary structural elements. Protons involved in H-bonds in helical regions of a protein may persist for several days after dissolution in D_2O, whereas those in the unstructured regions exchange rapidly at pH values in the range used here. In NodF, several regions of slow exchange are seen, these seem to be consistent with the position of the three helices. In these regions most of the resonances retain at least 10% of their original intensities for 1–7 h of dissolution in D_2O. However, the resonances corresponding to amide protons of the residues near the two termini of the second helix, seem to exchange rapidly (reduced to below 10% within 15 min of D_2O addition). This indicates that this helix is less stable than the other two. As expected, the amide protons of

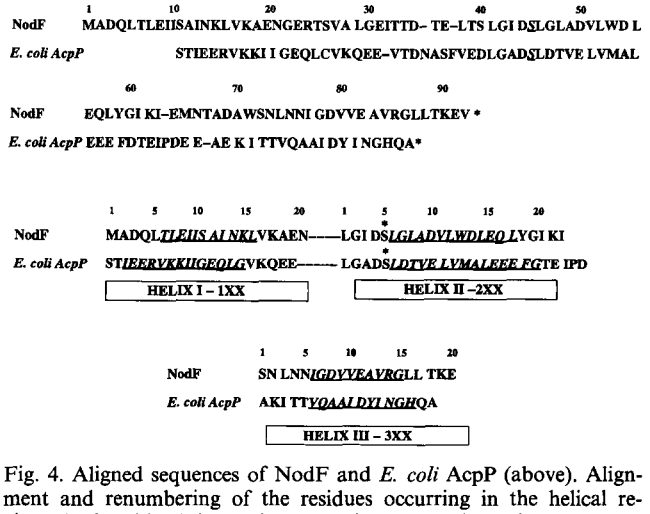

Fig. 4. Aligned sequences of NodF and *E. coli* AcpP (above). Alignment and renumbering of the residues occurring in the helical regions (±5 residues) in NodF and ACP. The residues in helix I are numbered 1XX, those in helix II as 2XX and those in helix III as 3XX (below).

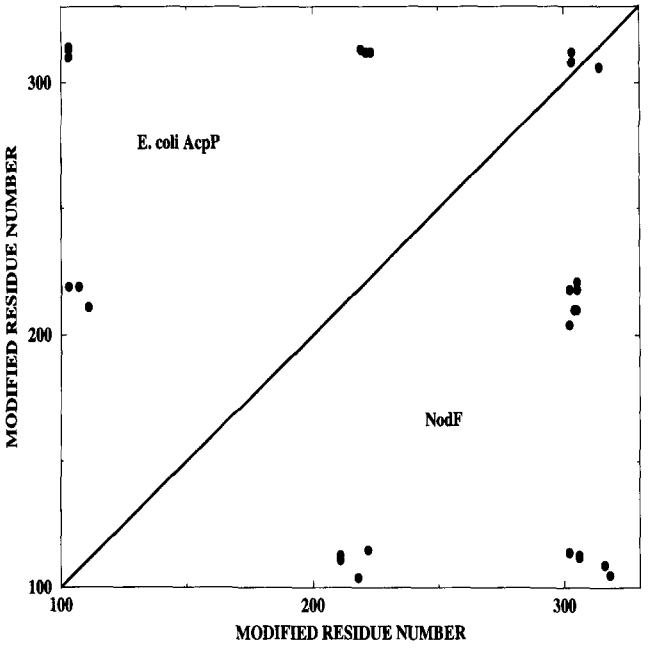

Fig. 5. Correlation plot of the helix-helix contacts in NodF and *E. coli* AcpP. Contacts for NodF are displayed in the bottom right and those for AcpP in the top left.

the residues lying in the unstructured regions between the helices exchange very rapidly (Fig. 3). The amide proton exchange rates in the helical regions of the protein, while significantly lower than those in other regions, are significantly higher than those observed in helical regions in very well structured proteins. This sort of behavior has previously been reported for *E. coli* AcpP [34]. It is also notable that in analogy to AcpP, the N- and C-termini of the second helix, in NodF, seem to be more flexible than the first and third helices, though its central portion is quite stable [34].

The position of the helices in NodF seem to be consistent with that predicted by secondary structure prediction algorithms applied to the NodF sequence. These include the SOPM (Self-Optimized Prediction Method) of Geourjon and Deleage [35] and the HSSP (Homology derived Secondary Structure Prediction) algorithm of Rost, Schneider and Sander [36].

3.3. General fold of NodF

In order to determine the global fold of a protein one relies on a few key NOE contacts between residues far removed in the sequence. Such contacts are usually between pairs of sidechain protons. However, the ^{15}N-directed structure-determination strategy adopted here limits the NOEs primarily to ones involving at least one backbone amide proton or the few contacts with sidechain -NH_2s in glutamine, arginine and asparagine residues. We have therefore attempted to add constraints from the 2D homonuclear NOESY dataset with a mixing time of 200 ms.

It is fortunate that some contacts can be identified between all possible pairs of helices. This indicates a structure which seems to be reasonably well-folded. Contacts are seen between the middle of helix I and the region near the N-termini of helices II and III; between the region near the N-terminus of helix I and the region near the C-termini of helices II and III; between the regions near the N-termini of helices II and III. Several long-range contacts are also seen between the loop between helices I and II (residues T34 through L38) and the middle of helix I, and also with the C-termini of helices II and III. The data are most consistent with an anti orientation of helices I and II and I and III.

3.4. Comparison of the structure with E. coli AcpP

While the sequence homology between NodF and *E. coli* AcpP is low (27%) and largely confined to the region near the binding site for the prosthetic group, it would not be unprecedented to find some additional structural similarities between the two proteins. This proves to be the case for NodF and AcpP. *E. coli* AcpP has three long helices between the residues 3–12, 37–51 and 65–75 and a fourth short helix between residues 56–63 [10,37]. NodF shows two long helices at the C- and N-termini (6–16 and 77–86 for NodF) and a third helix around the mid-region in the sequence (46–58 for NodF). We have not found conclusive evidence for the presence of a fourth helix in NodF corresponding to the fourth short helix (56–63) in AcpP, but this short helix is also missing in some models of AcpP. The binding site for the prosthetic group is near the beginning of the central helix in both proteins (S36 in AcpP, S45 in NodF).

In order to make a more definitive comparison between the general folds of NodF and AcpP, the long-range contacts between the regions around the three helices in the two proteins can be compared. The residues in and ± 5 residues from the ends of the three helices in the two proteins were therefore renumbered. The residues in helix I were numbered 1XX; those in helix II were numbered 2XX and those in helix III were numbered 3XX (Fig. 4). Using this new numbering scheme, a correlation plot (Fig. 5) showing the helix-helix contacts in both AcpP and NodF, was constructed. The correlation plot shows a certain symmetry about the diagonal indicating that AcpP and NodF show similar helix-helix contacts. This confirms a similarity between the general tertiary folds of the two proteins.

Thus, while sequence homology is confined to a short region around the prosthetic group attachment site, structural homology between NodF and AcpP may be more extensive. The long parallel helices play a role in the formation of an acyl chain binding site for AcpP [38]. A similar role may be postulated for NodF.

The fatty acid products formed by AcpP or NodF are however different. While AcpP participates in the synthesis of fatty acids (β-hydroxy myristic, palmitic, stearic, and *cis*-vaccenic acid) needed for essential lipid biosynthesis routes, NodF directs the synthesis of a *trans-2,trans-4,trans-6,cis-11*-octadecatetraenoic acid [6] which forms part of a host-specific lipo-chitin oligosaccharide signal [2]. Surprisingly, recent evidence even suggests that, besides a constitutive AcpP [39] and the inducible NodF, there might be at least two additional ACPs in *Rhizobium*. A third potential ACP was identified by ORF in the complementation unit I of the *fix23* locus in *R. meliloti* [40] and a partially sequenced gene (ORF*) [41] seems to correspond to even a fourth ACP in *Rhizobium*.

Despite the overall similarities between different ACPs, a crucial question remains: is it possible that minor structural differences enable them to direct fatty acids and other β-ketides into certain divergent biosynthetic pathways?

Acknowledgements: R.G. wishes to thank Dr. R.B. Hill for helpful

discussions. This work was supported by Grant GM32243 from The National Institutes of Health and Grant Ge 556/3-1 of the Deutsche Forschungsgemeinschaft.

References

[1] Spaink, H.P. (1995) Annu. Rev. Phytopathol. 33, 345–368.
[2] Spaink, H.P., Sheeley, D.M., van Brussel, A.A.N., Glushka, J., York, W.S., Tak, T., Geiger, O., Kennedy, E.P., Reinhold, V.N. and Lugtenberg, B.J.J. (1991) Nature 354, 125–130.
[3] van Brussel, A.A.N., Bakhuizen, R., van Spronsen, P.C., Spaink, H.P., Tak, T., Lugtenberg, B.J.J. and Kijne, J.W. (1992) Science 257, 70–72.
[4] Shearman, C.A., Rossen, L., Johnston, A.W.B. and Downie, J.A. (1986) EMBO J. 5, 647–652.
[5] Demont, N., Debelle, F., Aurelle, H., Denarie, J. and Prome, J.C. (1993) J. Biol. Chem. 268, 20134–20142.
[6] Geiger, O., Thomas-Oates, J.E., Glushka, J., Spaink, H.P. and Lugtenberg, B.J.J. (1994) J. Biol. Chem. 269, 11090–11097.
[7] Ritsema, T., Geiger, O., von Dillewijn, P., Lugtenberg, B.J.J. and Spaink, H. (1994) J. Bacteriol. 176, 7740–7743.
[8] Rawlings, M. and Cronan, J.E. Jr. (1992) J. Biol. Chem. 267, 5751–5754.
[9] Geiger, O., Spaink, H.P. and Kennedy, E.P. (1991) J. Bacteriol. 173, 2872–2878.
[10] Kim, Y. and Prestegard, J.H. (1990) Proteins Struct. Func. Gen. 8, 377–385.
[11] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Methods Enzymol. 185, 60–89.
[12] Miller, J.H. (1972) Experiments in Molecular Genetics, pp. 431–432, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[13] Jackowski, S. and Rock, C.O. (1983) J. Biol. Chem. 258, 15186–15191.
[14] Yang, J.T., Wu, C.C. and Martinez, H.M. (1986) Methods Enzymol. 130, 208–269.
[15] Shaka, A.J., Barker, P.B. and Freeman, R. (1985) J. Magn. Reson. 64, 547–552.
[16] Kay, L.E., Keifer, P. and Saarinen, T. (1992) J. Am. Chem. Soc. 114, 10663–10665.
[17] Grzesiek, S. and Bax, A. (1993) J. Am. Chem. Soc. 115, 12593–12594.
[18] States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) J. Magn. Reson. 48, 286–292.
[19] Tolman, J.R., Chung, J. and Prestegard, J.H. (1992) J. Magn. Reson. 98, 462–467.
[20] Marion, D., Driscoll, P.C., Kay, L.E., Wignfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) Biochemistry 28, 6150–6156.
[21] Shaka, A.J., Lee, C.J. and Pines, A. (1988) J. Magn. Reson. 77, 274–293.
[22] Clore, G.M., Bax, A. and Gronenborn, A. (1991) J. Biomol. NMR 1, 13–22.
[23] Shinnar, M., Elff, S., Subramanian, H. and Leigh, J.S. (1989) Magn. Reson. Med. 12, 75–80.
[24] Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) J. Magn. Reson. 85, 393–399.
[25] Kay, L.E., Marion, D. and Bax, A. (1989) J. Magn. Reson. 84, 72–84.
[26] Vuister, G.W. and Bax, A. (1993) J. Am. Chem. Soc. 115, 7772–7777.
[27] Vuister, G.W., Clore, G.M., Gronenborn, A.M., Powers, R., Garrett, D.S., Tschudin, R. and Bax, A. (1993) J. Magn. Reson. 101B, 210–213.
[28] Yamazaki, T., Yoshida, M. and Nagayama, K. (1993) Biochemistry 32, 5656–5669.
[29] Clore, G.M. and Gronenborn, A.M. (1991) Prog. NMR Spectr. 23, 43–92.
[30] Chylla, R. Peakpick – Multidimensional NMR peak picking Version 1.0, National Magnetic Resonance Facility, University of Wisconsin, Madison, WI.
[31] Hare, B.J. and Prestegard, J.H. (1994) J. Biomol. NMR 4, 35–46.
[32] Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley and Sons, New York.
[33] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) Biochemistry 31, 1647–1651.
[34] Andrec, M., Hill, R.B. and Prestegard, J.H. (1995) Protein Sci. 4, 983–993.
[35] Geourjon, C. and Deleage, G. (1994) Protein Engineering 7, 157–164.
[36] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[37] Holak, T.A. and Prestegard, J.H. (1986) Biochemistry 25, 5766–5774.
[38] Jones, P.J., Holak, T.A. and Prestegard, J.H. (1987) Biochemistry 26, 3493–3500.
[39] Platt, M.W., Miller, K.J., Lane, W.S. and Kennedy, E.P. (1990) J. Bacteriol. 172, 5440–5444.
[40] Petrovics, G., Putnoky, P., Rheus, B., Kim, J., Thorp, T.A., Noel, K.D., Carlson, R.W. and Kondorosi, A. (1993) Mol. Microbiol. 8, 1083–1094.
[41] Colona-Romano, S., Arnold, W., Schlüter, A., Boistard, P., Pühler, A. and Priefer, U.B. (1990) Mol. Gen. Genet. 223, 138–147.