Journal of Supramolecular Structure 11:139–145 (1979)
Biological Recognition and Assembly 65–71

NMR of fd Coat Protein

T. A. Cross and S. J. Opella

Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104

The conformations of the major coat protein of a filamentous bacteriophage can be described by nuclear magnetic resonance spectroscopy of the protein and the virus. The NMR experiments involve detection of the ^{13}C and ^{1}H nuclei of the coat protein. Both the ^{13}C and ^{1}H nuclear magnetic resonance (NMR) spectra show that regions of the polypeptide chain have substantially more motion than a typical globular protein. The fd coat protein was purified by gel chromatography of the SDS solubilized virus. Natural abundance ^{13}C NMR spectra at 38 MHz resolve all of the nonprotonated aromatic carbons from the three phenylalanines, two tyrosines, and one tryptophan of the coat protein. The α carbons of the coat protein show at least two different classes of relaxation behavior, indicative of substantial variation in the motion of the backbone carbons in contrast to the rigidity of the α carbons of globular proteins. The ^{1}H spectrum at 360 MHz shows all of the aromatic carbons and many of the amide protons. Titration of a ^{1}H spectra gives the pKas for the tyrosines.

Key words: fd coat protein, ^{1}H NMR, ^{13}C NMR

In order to understand the function of a protein it is necessary to correlate the structure of the protein at all levels with its dynamic properties. It is generally more straightforward to determine the primary sequence chemically and the secondary and tertiary structure with scattering techniques than to describe the motions present in the molecule. Nuclear magnetic resonance (NMR) spectroscopy is well suited for studying protein dynamics, because it can focus on individual residues even if more than one of a kind is present in the polypeptide and because relaxation behavior of nuclear resonances can often lead to determinations of rotational diffusion rates [1].

The major coat protein of the E coli bacteriophage fd is the subject of our current NMR studies. This protein exists in several biological situations, including as an intrinsic membrane protein and as the nucleoprotein complex of the virion [2]. The protein has great flexibility in its biological function, yet has only 50 amino acids in a single polypeptide chain. This protein may be an example of molecular dynamics being strongly influenced by all levels of structural organization, including primary structure. The direct influence of sequence on properties of the protein is inferred from the amino acids being segregated into acid, hydrophobic, and basic domains [3]. The NMR results can describe the mobility of the domains as well as individual side chains.

Received May 22, 1979; accepted June 12, 1979.

0091-7419/79/1102-0139$01.70

The coat protein has a high α helix content in a lipid- or detergent-bound state as well as in the assembled virion [4, 5]. In the membrane the protein relies on its hydrophobic midsection to span the hydrocarbon bilayer, with the hydrophilic ends associated with the polar phospholipid headgroups. The 19 amino acid hydrophobic stretch makes this a characteristic integral membrane protein. On the other hand, the collection of basic side chains at one end, may interact with the phosphate groups of DNA in the nucleoprotein complex.

The biology and structure of fd are well characterized [2]. The fd virus is a long filament, 9,000 Å by 90 Å, composed of a single strand circle of DNA encased in 2,700 copies of the major coat protein. The only other component is about 5 copies of the gene 3 product at one end of the virus. Upon binding to an E coli F pilus, the DNA and gene 3 protein enter the cell, while the coat protein is stored in the cell membrane. During viral replication the major coat protein is synthesized as a procoat protein with a leader sequence of 23 amino acid residues that serve to make the protein soluble in the cytoplasm. The procoat protein then is inserted into the membrane where it is trimmed, in a separate step, to the mature coat protein [6]. The major coat protein is stored in the membrane asymmetrically with the carboxyl end on the inside of the cell [7]. The virions are assembled when the DNA is extruded through the membrane, where the DNA is packaged inside the coat protein.

While the procoat protein is soluble in both water and membranes, the coat protein itself is completely insoluble in water. It can only be studied in solution in the presence of lipids or detergents, presumably because of its extensive hydrophobic character. In the virus the coat proteins are arranged in such a way that the hydrophobic amino acids do not interfere with solubility in water. The versatility of the coat protein is illustrated by the dramatic transition from water insolubility during storage in the membrane to high solubility in water when assembled in the virion.

The detergent-solubilized coat protein represents a stable conformation of the molecule. It may be the same structure that exists in the cell membrane after synthesis. On the basis of circular dichroism studies of the protein in sodium dodecylsulfate (SDS) micelles and in phospholipid vesicles, the secondary structure is the same for the two preparations [4, 5]. Proteolytic digests of the detergent-solubilized coat protein show that only the hydrophilic terminal sections are exposed, just as expected for an integral membrane protein spanning a bilayer or a micelle [8]. It is an open question as to how extensive the changes in protein structure are upon virus formation.

Because of the wide range of biological structures that contain the coat protein, we are applying a variety of different NMR techniques to the problem. These include solid-state NMR of the DNA and coat protein in the intact virus, and solution ^{13}C and ^{1}H NMR of the detergent-solubilized major coat protein. This paper describes the ^{13}C and ^{1}H experiments that map out the general features of the protein dynamics.

High-resolution ^{1}H NMR is most suitable for monitoring the conformational states of the protein and the chemical environment of individual residues [9]. In some cases the spectral appearance can provide information on molecular dynamics, and, by observing the presence and exchange of non carbon-bound hydrogens such as amide hydrogens, the overall structural integrity and flexibility of protein folding can be characterized.

Natural abundance ^{13}C NMR is a valuable source of information on structure and dynamics of proteins [1, 10, 11]. Relaxation studies are particularly useful because the well-understood nature of $^{13}C-^{1}H$ interactions makes the motional properties available directly. Small globular proteins in solution have been characterized by ^{13}C NMR.

The detergent-solubilized coat protein of fd is shown to have quite different properties than the previous examples of globular proteins.

MATERIALS AND METHODS

The growth of Escherichia coli F^+ (3,300) on a modified SLBH-rich medium infected in late log phase with fd and the purification of the fd from the growth supernatant by polyethylene glycol precipitation, CsCl block gradient centrifugation, and dialysis will be described in detail elsewhere [12]. Gel chromatography with Sephacryl S-200 SF was used to separate the SDS-solubilized protein from the DNA after the virus was broken up. The coat protein collected from column fractions was dialyzed and lyophilized. Samples for NMR were made by dissolving the protein-detergent powder in a minimum volume of ^2H_2O or H_2O buffered with 40 mM borate at pH 9.0, unless mentioned otherwise. The pH readings on a Beckman model 12 meter with a sodium ion-insensitive electrode are uncorrected for deuterium isotope effects.

^{13}C NMR spectra were obtained on a Nicolet NT-150 NMR spectrometer with an Oxford Instruments magnet. Samples were approximately 12 ml in 20 mm sample tubes with a typical protein concentration of 7 mM. All ^{13}C chemical shifts are referred to external TMS through a dioxane standard.

^1H NMR spectra were obtained on a Bruker WH 360 spectrometer. Samples were about 0.5 ml in 5 mm sample tubes with protein concentrations of 3 mM. Spectra were acquired using rapid-scan correlation spectroscopy. Proton chemical shifts were referenced to internal DSS.

RESULTS

The complete natural abundance ^{13}C NMR spectrum of the fd major coat protein in solution is shown in Figure 1. Despite large peaks from the SDS detergent near 20 ppm and 70 ppm, nearly all types of protein carbon resonances are visible, including many aliphatics above 15 ppm. This paper concentrates on the α carbons in the region 50–70 ppm and the aromatic carbons 100–160 ppm. The carbonyls are in the band around 175 ppm. This spectrum was obtained with relatively slow pulsing to ensure equilibrium signal intensities. Continuous square wave-modulated proton decoupling was used, which means that each carbon resonance is a single line and the intensities have contributions from nuclear Overhauser enhancement (NOE).

The α carbon spectral region (50–70 ppm) has an appearance different from that of any protein previously reported. The α carbons of a native globular protein are a rigid part of the protein backbone. The relaxation parameters of broad linewidth, short T_1, and minimal NOE of α carbons correspond to the rotational correlation time of the entire protein, with no evidence of internal motion. This is not the case for the coat protein, with a number of very sharp resonances superimposed on the more numerous broad α carbon lines. The sharp lines have significantly larger T_1s and NOEs than the broad lines. Clearly, heterogeneity in backbone dynamics is being monitored.

The aromatic region of the ^{13}C spectrum is expanded in Figure 2. The spectrum 2a corresponds to that of Figure 1 and contains resonances from all the aromatic carbons of the 3 Phe, 2 Tyr, and 1 Trp of the coat protein. The spectrum 2b, obtained with weak off-resonance-modulated proton decoupling and refocusing of the ^{13}C magnetization to eliminate the signals from carbons with directly bonded hydrogens, has only the non-

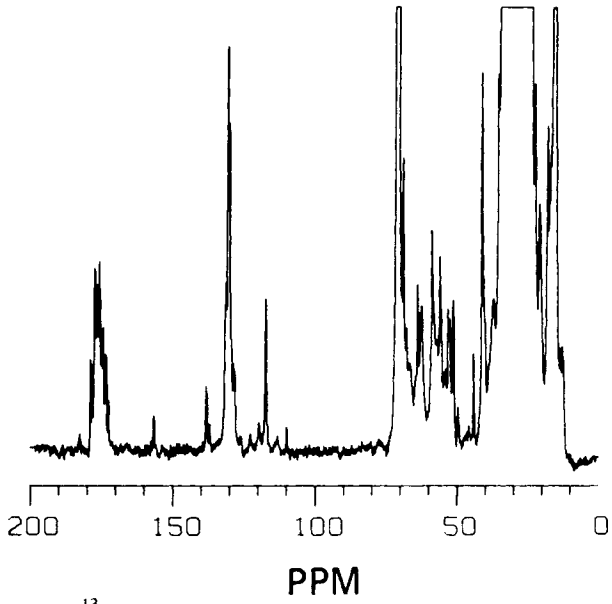

Fig. 1. Natural abundance ^{13}C NMR spectrum of fd major coat protein. 20,000 transients accumulated with 1.5 s recycle delay. Digital resolution of 2.5 Hz with 4 Hz linebroadening added. Referenced to external TMS. ^{13}C frequency of 37.74 MHz.

protonated aromatic carbon resonances. Partial assignments result from the comparison of spectra 2a and 2b.

The C_ζ resonances of the two tyrosines occur at 156.5 ppm at pH 9.0. In the best spectra they have completely separated lines. The 3 phenylalanine C_δ and the $C_{\epsilon 2}$ of the tryptophan come near 138 ppm. The 2 Tyr γ carbons and the δ_2 Trp carbon appear at 129 ppm with the Trp C_γ at 109.5 ppm. These are all narrow resonances, as expected for protein carbons without closeby protons to provide efficient relaxation. And as expected these survive the procedure used for Figure 2b, showing the lines for the nonprotonated carbons.

In a globular protein the other aromatic carbon signals would have broad resonances because of the effect of an attached proton on ^{13}C relaxation. While aromatic group rotation occurs in globular proteins, it is usually limited in extent and is detected through more subtle T_1 and NOE changes rather than drastic line narrowing [11]. A more pronounced effect is seen in the coat protein ^{13}C aromatic spectra. The phenylalanine ring carbons dominate the very sharp signals near 130 ppm. Somewhat broader lines at 118 ppm are from the Tyr C_ϵ. The individual Trp protonated carbons are broader yet and can be seen in the 110–125 ppm region. The finding of such narrow Phe ring carbons and the clear observation of single protonated Trp carbons are remarkable findings. NOE measurements correlate with linewidth measurements, with the sharp lines having high NOE (> 2) and the broad lines a minimal amount (1.1).

The downfield region of the ^1H NMR spectrum of the coat protein is shown in Figure 3. Spectrum 3a is run in H_2O, thus resonances are visible from the amide protons of the entire protein and the carbon-bound hydrogens of aromatic groups. In contrast, spectrum 3c in ^2H_2O after hearing, has no amide resonances but only the aromatic hydrogens. The spectrum 3b is an important intermediate case, where the protein is dissolved in ^2H_2O at 20°C after being lyophilized from H_2O and contains 3–5 amide

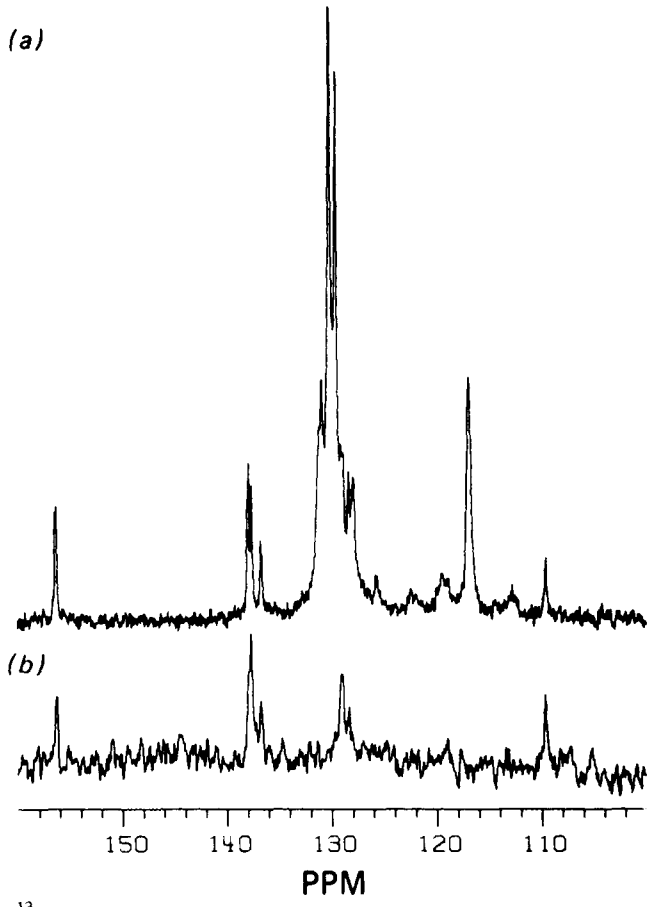

Fig. 2. Aromatic ^{13}C NMR region of fd major coat protein. a) Expansion of Figure 1. b) Non-protonated aromatic carbon spectrum; obtained with off resonance proton decoupling and ^{13}C spin echo formation.

resonances that can be exchanged only by heating. The slow exchange of these few amide protons is evidence for secondary and tertiary structure of the level seen in globular proteins. More sophisticated structural analysis is possible from the individual kinetics of the few amide resonances of spectrum 3b.

The appearance of the aromatic ^1H NMR spectral region is similar to what might be expected from a native protein, especially the upfield two peaks at 6.8 and 6.9 ppm, which can come only from the ϵ protons of the two tyrosines, each resonance being from two equivalent protons on one of the rings. At higher temperatures the doublet structure from coupling to the δ protons is visible, and they titrate upfield with increasing pH. Rapid rotation of the Tyr rings occurs because of the equivalence of the ϵ protons. The corresponding Tyr δ protons are at 7.1 and 7.2 ppm. The phenylalanine ring protons, like the ring carbons, are sharp (7.3 ppm) and show relatively little structure-induced chemical shift nonequivalence. Under some pH and temperature conditions the Trp protons can be discerned among the Phe and Tyr protons. The tyrosine resonances titrate with pKas of 12.5 and 12.6 [12]. The other resonances change little with pH.

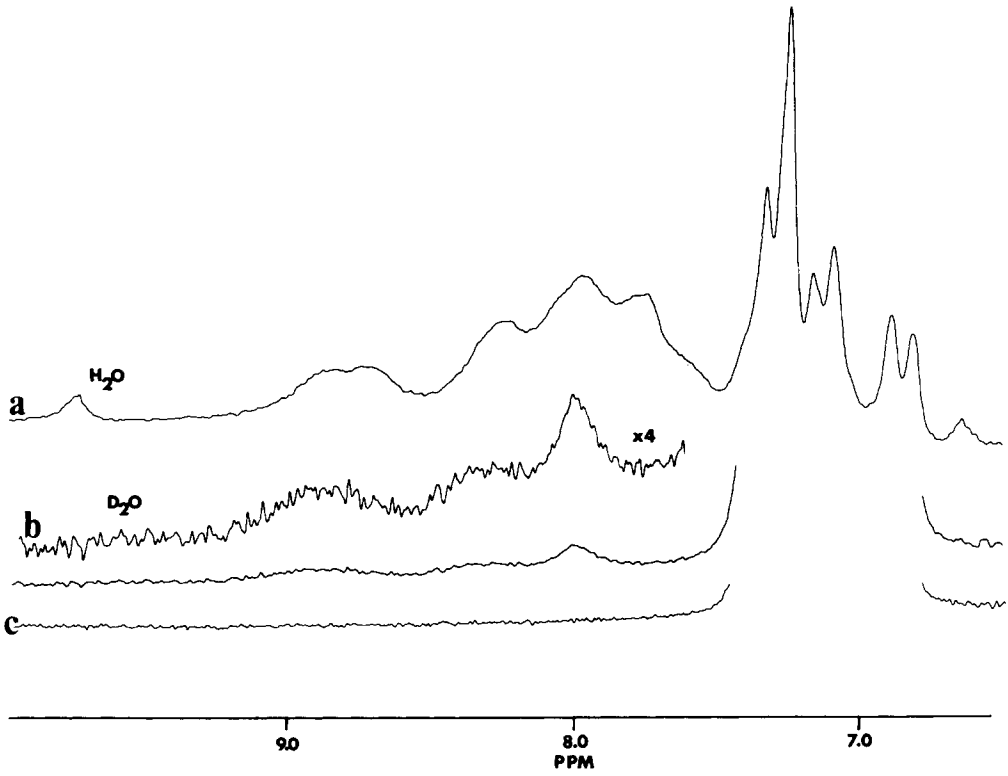

Fig. 3. Downfield ^1H NMR region of fd major coat protein. 64 scans of correlation spectroscopy. ^1H frequency of 360.06 MHz. a) Protein in H_2O, b) protein in ^2H_2O at 20°C with 4X expansion, c) sample b after heating.

DISCUSSION

Detergent-solubilized fd coat protein has a stable native structure that differs significantly from that of a typical globular protein. A high degree of specific secondary and tertiary structure is reflected in the chemical shift dispersion among identical nuclei from different residues. Examples of this are the Tyr ϵ and δ protons, Tyr ζ carbons, and the Phe γ carbons. Denatured proteins have only a single line for such situations [10]. Other evidence for the coat protein being highly structured comes from the slow exchange of some amide protons and the borad lines of the Trp ring carbons as well as most of the α carbons.

The sequence of amino acids 21–26 is Tyr-Ile-Gly-Tyr-Ala-Trp [3]. These are in the hydrophobic central section of the protein. The very high pKas of the two tyrosines indicate that they are buried in the hydrophobic pocket or that they are involved in a highly specific hydrogen or van der Waals bonding arrangement. One residue away is the Trp, which is a rigidly held residue, as seen in the ^{13}C protonated carbon linewidths. These monitors of the hydrophobic domain of the protein show it to be relatively rigid and highly structured. It is quite likely that many of the broad C_α resonances and the slowly exchanged amide hydrogens will be assigned to this region.

The phenylalanine rings are rapidly rotating. Two Phes are at the basic end, and one is in the acidic region of the protein [3]. There is less organized structure in these regions based on the extensive mobility of these large side chains. Chemical evidence also indicates the exposure of the hydrophilic residues to reagents [8].

The results presented here give a qualitative view of the dynamics of a detergent-solubilized protein. There are clearly different regions of backbone mobility, and there are large differences among side chains in rates of motion. The NMR results to date allow a limited amount of interpretation in terms of specific parts of the protein structure. To the extent that groupings of resonance characteristics with primary structure are possible, the central hydrophobic region seems more highly structured than the hydrophilic end regions. More extensive experiments will give a more complete picture of the protein. NMR of the protein in solution and the intact virus can then give an idea of the extent of conformational change associated with virus assembly.

ACKNOWLEDGMENTS

This work is being supported by grants from NIH (GM-24266) and the American Cancer Society (NP-225). The ^{1}H spectra were taken at the Middle Atlantic NMR Facility. The ^{13}C spectra were taken on a spectrometer obtained with a NSF grant. T.A.C. is supported by a Cell and Molecular Biology Training Grant.

REFERENCES

1. Opella SJ, Nelson DJ, Jardetzky O: In Reising HA, Wade CG (eds): "ACS Symposium Series, No. 34. Magnetic Resonance in Colloid and Interface Science." Washington DC: American Chemical Society, 1976, pp 397–417.
2. For a collection of papers on fd see: Denhardt DT, Dressler D, Ray DS (eds): "The Single Stranded DNA Phages." New York, Cold Spring Harbor Laboratory, 1978: 605–626.
3. Nakashima Y, Konigsberg W: J Mol Biol 88:598, 1974.
4. Williams RW, Dunker AK: J Biol Chem 252:6253, 1977.
5. Nozaki Y, Reynolds JA, Tanford C: Biochemistry 17:1239, 1978.
6. Ito K, Mandel G, Wickner W: Proc Natl Acad Sci USA 76:1199, 1979.
7. Wickner W: Proc Natl Acad Sci USA 72:4749, 1975.
8. Woolford JL Jr, Webster RE: J Biol Chem 250:4333, 1975.
9. Wuthrich K: "NMR in Biological Research." New York: Elsevier, 1976.
10. Allerhand A: Accts Chem Res 11:469, 1978.
11. Opella SJ, Nelson DJ, Jardetzky O: J Am Chem Soc 96:7157, 1974.
12. Cross TA, Opella SJ: (in preparation).