of the interrelationship of parental nucleosomes and DNA during DNA replication (Seale, manuscript in preparation).

Acknowledgment

I am grateful for the expert technical assistance of Steven R. Lohman, and for the helpful advice and unpublished data of Dr. J. K. Fraser.

References

Bernard, O., and Brent, T. P. (1973), *Biochem. Biophys. Res. Commun. 53,* 1213.
Camerini-Otero, R. D., Sollner-Webb, B., and Felsenfeld, G. (1976), *Cell 8,* 333.
Clark, R. J., and Felsenfeld, G. (1971), *Nature (London), New Biol. 229,* 101.
Fraser, J. K., and Huberman, J. A. (1977), *J. Mol. Biol.* (in press).
Friedman, D. L. (1974), *Biochim. Biophys. Acta 353,* 447.
Friedman, D. L., and Mueller, G. C. (1968), *Biochim. Biophys. Acta 161,* 455.
Friedman, D. L., and Mueller, G. C.(1969), *Biochim. Biophys. Acta 174,* 253.
Germond, J.-E., Bellard, M., Oudet, P., and Chambon, P. (1976), *Nucleic Acids Res. 3,* 3173.
Grisham, J. W., Kaufman, D. G., and Stenstrom, M. L. (1972), *Biochem. Biophys. Res. Commun. 49,* 420.
Hallick, L. M., and Namba, M. (1974), *Biochemistry 13,* 3152.
Hershey, H., Stieber, J., and Mueller, G. C. (1973a), *Eur. J. Biochem. 34,* 383.
Hershey, H., Stieber, J., and Mueller, G. C. (1973b), *Biochim. Biophys. Acta 312,* 509.
Kaufman, D. G., Grisham, J. W., and Stenstrom, M. L. (1972), *Biochim. Biophys. Acta 272,* 212.
Kidwell, W. R., and Mueller, G. C. (1969), *Biochem. Biophys. Res. Commun. 36,* 756.
Krokan, H., Bjørklid, E., and Prydz, H. (1975a), *Biochemistry 14,* 4227.
Krokan, H., Cooke, L., and Prydz, H. (1975b), *Biochemistry 14,* 4233.
Kumar, K. V., and Friedman, D. L. (1972), *Nature (London), New Biol. 239,* 74.
McCarty, K. S., Vollmer, R. T., and McCarty, K. S. (1974), *Anal. Biochem. 61,* 165.
Painter, R. B., and Schaefer, A. W. (1969), *J. Mol. Biol. 45,* 467.
Panyim, S., Bilek, S., and Chalkley, R. (1971), *J. Biol. Chem. 246,* 4206.
Panyim, S., and Chalkley, R. (1969), *Arch. Biochem. Biophys. 277,* 7.
Panyim, S., Jensen, R. H., and Chalkley, R. (1968), *Biochim. Biophys. Acta 160,* 252.
Seale, R. L. (1976), *Proc. Natl. Acad. Sci. U.S.A. 73,* 2270.
Tseng, B. Y., and Goulian, M. (1975), *J. Mol. Biol. 99,* 317.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244,* 4406.

# Structural Investigations of Chromatin Core Protein by Nuclear Magnetic Resonance†

D. M. J. Lilley,*‡ J. F. Pardon, and B. M. Richards

ABSTRACT: A complex derived from chromatin containing one molecule of each of histones H2A, H2B, H3, and H4, termed core protein, was studied by $^{13}C$ and $^{1}H$ nuclear magnetic resonance. $^{13}C$ line widths, when analyzed and compared with those of native and thermally unfolded representative globular proteins, showed that regions of the core protein possess considerable mobility. Studies of $C_\alpha$ and $C_\beta$ line widths, and $C_\alpha$ spin–spin relaxation times, show that this mobility arises from sections of random-coil polypeptide. It is argued that these regions are N-terminal "tails", attached to C-terminal globular polypeptides. The 270-MHz $^{1}H$ nuclear magnetic resonance spectrum shows numerous ring current shifted resonances, indicating that the C-terminal globular domain has a precise tertiary structure. The globular domain most likely forms the histone "core" of the chromatin monomer particle, whilst the basic tails probably wind around the grooves of the double helix, enabling the basic side chains to interact with the DNA phosphate groups. Some biological implications of this model are considered.

Whilst the subunit structure of chromatin is now well established (Woodcock, 1973; Van Holde et al., 1974; Olins and Olins, 1974; Kornberg, 1974; Burgoyne et al., 1974; Noll, 1974a,b; Richards et al., 1976), many details of the substructure of the repeat unit await elucidation. Nuclease resistant monomer core particles from chromatin are composed of two molecules each of histones H2A, H2B, H3, and H4 (Thomas and Kornberg, 1975a), which are enclosed by a segment of DNA double helix comprising 140 base pairs (Pardon et al., 1975). Further structural description requires analysis of both histone–histone and histone–DNA interactions.

The isolation of a complex containing all four histones (Thomas and Kornberg, 1975b; Weintraub et al., 1975), which

† From the Inorganic Chemistry Laboratory, University of Oxford, Oxford OX1 3QR, England (D.M.J.L.), and Searle Research Laboratories, High Wycombe, HP12 4HL, England (J.F.P. and B.M.R.). *Received November 4, 1976.*

‡ Present address: Searle Research Laboratories, High Wycombe, HP12 4HL, England.

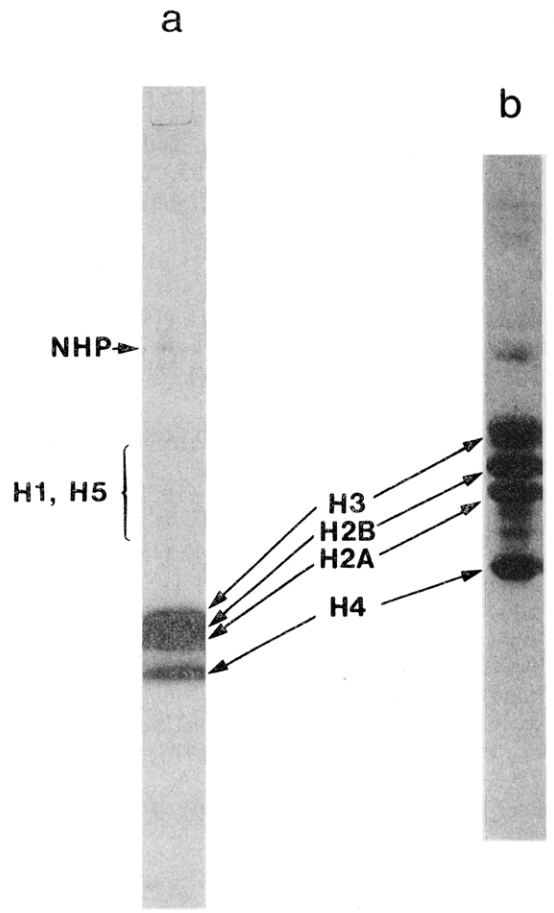


FIGURE 1: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of chicken erythrocyte core protein. (a) Low resolution gel showing entire region from top of stacking gel to bottom of running gel; (b) high resolution gel showing histone region.

is closely related to the protein core of the monomer particle, provides the means to study both interactions. This complex, known as core protein, has a molecular weight (Campbell and Cotter, 1976; Wooley et al., 1977) of 57 000 and is therefore likely to contain one molecule each of histones H2A, H2B, H3, and H4. Analysis of neutron scattering profiles from solutions of the complex suggest that it is approximately disk shaped (Wooley et al., 1977). In the chromatin monomer two such core protein complexes stack with coincident axes to generate the internal core (Pardon et al., 1977; Richards et al., 1977). When isolated and studied at high ionic strength, charge repulsion effects in core protein are screened. For this reason structural features of this complex may be extrapolated to those of the chromatin monomer core in situ.

$^{13}C$ NMR[1] is highly sensitive to the presence of mobile random-coil elements in histone complexes (Lilley et al., 1976; Lilley, 1977). In this paper we have analyzed the spectra of core protein of chicken erythrocytes and calf thymus and show that, despite charge screening, the complexes possess random-coil sections of polypeptide joined to a precisely structured globular domain. We suggest that the random-coil sections are N-terminal tails, which are the sites of primary electrostatic interaction between the histone core and DNA.

## Materials and Methods

*Preparation of Core Protein.* Core protein was prepared from calf thymus erythrocytes using modified published procedures (Weintraub et al., 1975; Wooley et al., 1977). Nuclei, isolated by the methods of Shaw et al. (1974), were disrupted by osmotic shock and the resulting chromatin rendered free of histones H1 and H5 by washes in 0.6 M and then 0.65 M NaCl. Solid NaCl and CHES buffer were worked into the resulting gel to a final concentration of 2 M and 10 mM, respectively, and a final pH of 9.0. After 24 h the gel was centrifuged at 42 000 rpm for 16 h and the supernatant removed from the DNA pellet. This solution of core protein was then dialyzed into the appropriate buffer, containing 2 M NaCl, and concentrated as required by Amicon ultrafiltration. The core protein was characterized by means of polyacrylamide gel electrophoresis, analytical ultracentrifugation, chemical cross-linking (Wooley et al., 1977), and laser light scattering (Campbell and Cotter, 1976).

*Ribonuclease and β-Lactoglobulin.* These were purchased from Sigma London and used without further purification.

*NMR Methods.* $^{13}C$ NMR spectra were recorded at 22.63 MHz on a Bruker WH-90 Fourier transform spectrometer using proton-noise decoupling. Transients ($2–3 \times 10^5$) in 2048 data points were collected for each spectrum, using a 70° (20 μs) pulse, with a 0.2-s recycle time. Samples for $^{13}C$ NMR spectroscopy were studied as 100 mg/mL solutions in buffers containing 30% $D_2O$ for field locking. All free induction decays (FID) were subjected to exponential multiplication. For the normal spectra of histones and other proteins, this corresponds to an artificial line broadening of 3.8 Hz.

Partially relaxed Fourier transform (PRFT) spectra were recorded using the pulse sequence [90°–5 ms–180°–5 ms–FID collection–0.5 s], (Hahn, 1950), whereby resonances underwent $T_2$ relaxation for 10 ms. Exponential multiplication of PRFT FIDs corresponded to a spectral line broadening of 5.6 Hz.

$^1H$ NMR spectra were recorded at 270 MHz on a Bruker HX-270 Fourier transform spectrometer. Transients ($10^3$) were collected in 2048 data points, using a 70° pulse. Core protein solutions (30 mg/mL) were dialyzed into 2 M NaCl–25 mM phosphate in 99.7% $D_2O$, pD 9.4, for $^1H$ NMR.

Chemical shifts for both $^{13}C$ and $^1H$ are quoted relative to external $Me_4Si$.

*Sedimentation Velocity.* Sedimentation velocity runs were performed at 52 000 rev/min using a Beckman Model E ultracentrifuge. Double sector cells were used in an AnD rotor, the boundary being observed by a schlieren optical system. Boundary positions were measured on a microcomparator and sedimentation coefficients calculated from the gradient of a ln $r$ (distance of boundary from rotor center) against time plot, using a correction of 1.6 for the viscosity of 2 M NaCl at 20 °C.

*Gel Electrophoresis.* Sodium dodecyl sulfate–polyacrylamide gel electrophoresis was performed according to Laemmli (1970), using 15% acrylamide running gel and 6% acrylamide stacking gel. Bromophenol blue marker dye was not run off the gel, and the gel was stained in Coomassie brilliant blue R250.

## Results and Discussion

*Purity of Core Protein.* Sodium dodecyl sulfate–polyacrylamide gel electrophoresis, see Figure 1, indicates that core protein extracted by these methods is of high purity. Bands corresponding to only very small amounts of high-molecular-weight nonhistone proteins and histones H1 and H5 were visible. Protein bands were not seen between that for H4 and the bromophenol blue marker dye position, indicating little or

[1] Abbreviations used: NMR, nuclear magnetic resonance; CHES, 2-(*N*-cyclohexyl)ethanesulfonic acid; FID, free induction decays; PRFT, partially relaxed Fourier transform spectra; UV, ultraviolet.

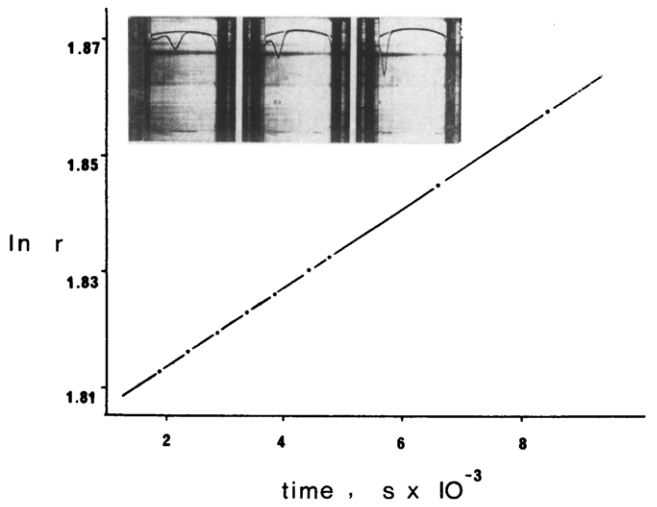


FIGURE 2: Results of a sedimentation velocity ultracentrifuge run on chicken erythrocyte core protein, showing a plot of ln $r$ ($r$ = boundary position) against time. Insert shows schlieren peaks produced by the boundary at 16, 64, and 140 min after the rotor had attained full speed (52 000 rev/min).

no contamination by low-molecular-weight protein. Furthermore, laser light scattering measurements (Campbell and Cotter, 1976) showed that the complex did not dissociate over the concentration range 0.05 to 5 mg/mL, and neutron scattering measurements (Wooley et al., 1977) indicate that there is little or no aggregation over the concentration range 2.5 to 35 mg/mL.

*Hydrodynamic Properties of Core Protein.* The results of a sedimentation velocity experiment on core protein extracted from chicken erythrocytes are given in Figure 2. The ln $r$ against time plot remains linear after 2 h at 52 000 rev/min and a narrow sedimentation boundary is prominent at these times. The small amount of slower sedimenting material trailing behind the peak is probably due to a low H1 and H5 contamination but may also represent a small degree of dissociation of the complex in the ultracentrifuge cell. The gradient measured from the plot using a least-squares linear regression yields an $s_{20,w}$ value of 3.7 S.

*$^{13}C$ NMR Spectra of Core Protein.* The $^{13}C$ NMR spectra of core protein from calf thymus and chicken erythrocytes are shown in Figure 3. The spectra are similar and most differences may be accounted for in terms of the signal-to-noise ratio. Furthermore, the general appearance of the spectra, with narrow resonances superimposed on a broad background, is very similar to that of previously recorded $^{13}C$ NMR spectra from histones (Clark et al., 1974; Lilley et al., 1975, 1976; Lilley, 1977; Tancredi et al., 1976) showing that the general nature of histone structure is independent of the method used for their extraction. Histone $^{13}C$ NMR line widths have been analyzed in some detail elsewhere (Lilley, 1977). Two effects contribute toward line widths, viz. spin–spin relaxation time ($T_2$) and environmental chemical shift nonequivalence. Both effects will produce narrow resonances from carbon atoms present in mobile random-coil sections. Comparison of the line widths of the core protein spectra with those of the monomeric globular protein $\beta$-lactoglobulin, shown in Figure 4, shows that core protein must possess elements of considerable mobility which are not present in a purely globular macromolecule. Globular proteins allow some surface mobility as demonstrated by the small line width of the resonance due to lysine at 39.5 ppm in the spectrum of $\beta$-lactoglobulin (~25 Hz).

The best test for random-coil elements comes from examination of $C_\alpha$ and $C_\beta$ mobility, although it has been demonstrated (Lilley, 1977) that side chain mobility is also greater when the backbone is not anchored to the globular mass. The $C_\alpha$ region (45–60 ppm) is well resolved and several $C_\beta$ resonances may also be distinguished. Comparison of core protein and $\beta$-lactoglobulin $C_\alpha$ regions shows that, whilst the former is composed of narrow resonances, the latter consists of one broad envelope. The $C_\alpha$ region of even quite small globular proteins such as ribonuclease (Lilley, 1977) is still very broad at 22.63 MHz. The presence of $C_\alpha$ line widths of 16–25 Hz in the spectra from the histone core protein (the molecular weight of which is over four times that of ribonuclease) is very good evidence for the resence of random-coil sections. This is substantiated by $C_\beta$ line widths, since rapid rotation about $C_\alpha$–$C_\beta$ bonds when the backbone is rigid does not result in marked resonance narrowing, since this motion is highly anisotropic. Core protein $C_\beta$ resonances are typically of 10–20 Hz line widths compared with >30 Hz in $\beta$-lactoglobulin and ribonuclease. A summary of resonance line widths of core protein and comparison with globular proteins are shown in Table I.

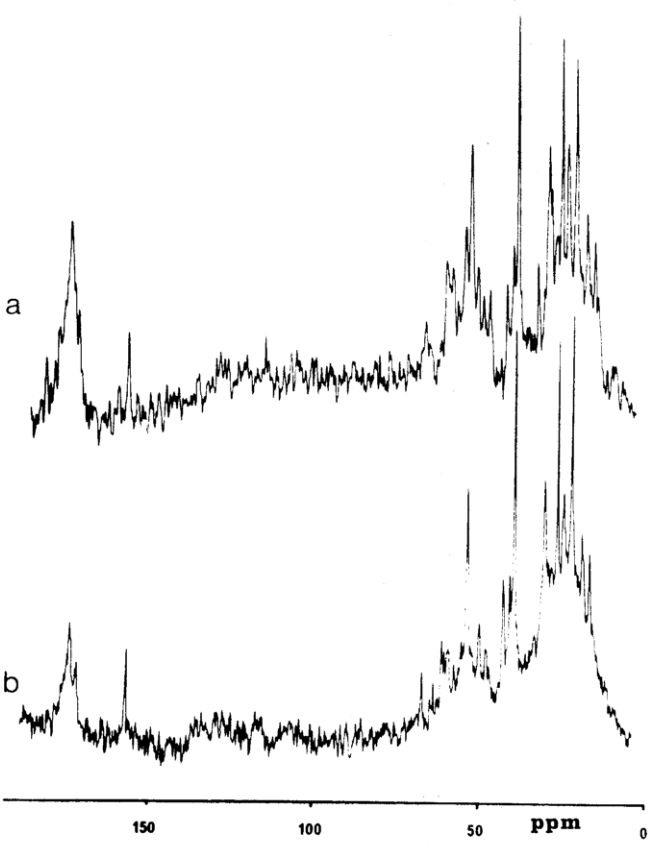


FIGURE 3: $^{13}C$ NMR spectra of 100 mg/mL core protein solutions extracted from (a) calf thymus and (b) chicken erythrocytes, in 2 M NaCl–12.5 mM borate, pH 9.0, in 30% $D_2O$ at 15 °C. The shift scale is relative to external tetramethylsilane.

A further demonstration that the core protein resonance line line widths are typical of random-coil polypeptides is given by the $^{13}C$ NMR spectral line widths of thermally unfolded ribonuclease, also presented in Table I. The line widths of many resonances, including those of the $C_\alpha$ region, are very similar to the narrow resonances of the core protein spectra, especially after allowing for the elevated temperature at which the spectrum was recorded.

We can exclude the alternative explanation for the presence of narrow lines in the core protein spectra, namely, from con-

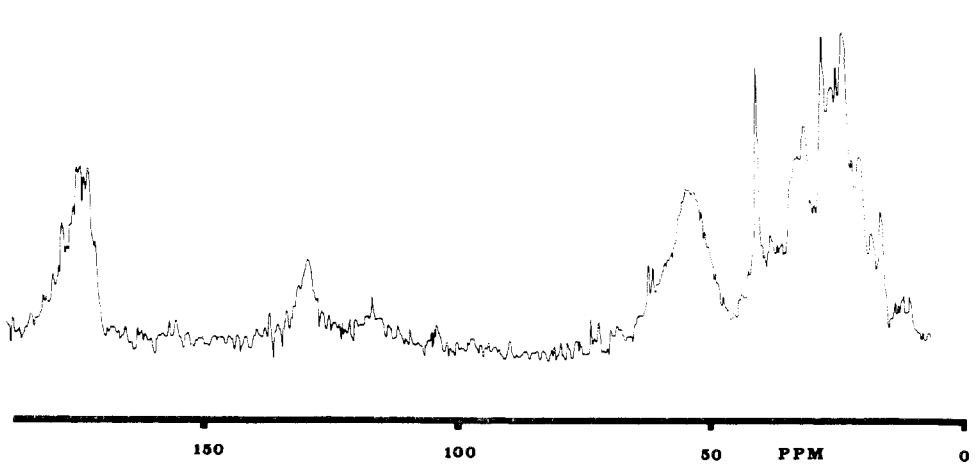


FIGURE 4: $^{13}C$ NMR spectrum of 100 mg/mL solution of $\beta$-lactoglobulin, in 50 mM NaCl-30% $D_2O$.

TABLE I: Some $^{13}C$ NMR Line Widths Observed in Histone Complexes and Comparison with Native and Unfolded Globular Proteins.

| Protein | Resonance (ppm) | Carbon type | Line width[a] (Hz) |
|---|---|---|---|
| Histone, 15 °C | 39.3 | Lys $\epsilon$ | 3 |
| | 26.5 | Lys $\delta$ | 6 |
| | 21.9 | Lys $\gamma$ | 14 |
| | 41.2 | Arg $\delta$ | 4 |
| | 17.8 | Val Me | 14 |
| | 16.8 | Ala $\beta$ | 6 |
| | 67.0 | Thr $\beta$ | 8 |
| | 28.0 | Arg $\beta$ | 21 |
| | 30.1 | Lys $\beta$ | 16 |
| | 53.8 | Lys/Arg $\alpha$ | 16-25 |
| Ribonuclease, 15 °C | 39.5 | Lys $\epsilon$ | 21 |
| | 27.0 | Lys $\delta$ | 26 |
| | 22.0 | Lys $\gamma$ | 36 |
| | 30.0 | Lys $\beta$ | ~45 |
| | 45-60 | $\alpha$ region | Broad envelope |
| Ribonuclease, 73 °C | 39.3 | Lys $\epsilon$ | 3 |
| | 27.0 | Lys $\delta$ | 7 |
| | 22.0 | Lys $\gamma$ | 9 |
| | 18.8 | Thr Me | 7 |
| | 16.8 | Ala $\beta$ | 5 |
| | 42.5 | Gly $\alpha$ | 15 |
| | 49.8 | Ala $\alpha$ | 10 |
| | 54.0 | Lys/Gln/Glu $\alpha$ | 23 |

[a] Observed line widths have been corrected for artificial broadening arising from exponential modification of FIDs.

tamination of a totally globular histone complex by low-molecular-weight mobile proteins. Such proteins would have to be substantially smaller than ribonuclease (mol wt 13 700) or lysozyme (mol wt 14 600) since these give spectra typical of globular proteins under our conditions (Lilley et al., 1976; Lilley, 1977). We have not been able to detect the presence of any protein in our preparations with a molecular weight smaller than histone H4 (11 300) on gels which were overloaded with respect to histones. Furthermore histone $^{13}C$ line widths are reproducible in independently prepared samples, in samples prepared from different tissues and in samples prepared using widely differing extraction techniques. Hence we conclude that the narrow line widths must arise from mobile but integral components within the core protein complex.

A feature of the core protein spectra, when compared with $^{13}C$ NMR spectra of individual histone types (Clark et al., 1974; Lilley et al., 1975), is the greater dominance by resonances assignable wholly or in part to lysine side chains, particularly that of $\epsilon$-lysine at 39.5 ppm. Whilst most histone preparations contain aggregated species (Edwards and Shooter, 1969), the core protein is a relatively discrete complex, and reduction in mobility of surface side chains by intermolecular aggregation is unlikely. Furthermore, for thermodynamic reasons (Tanford, 1954), charged residues are usually found near to or at the surface of globular proteins (Richards, 1963) and there is no reason to expect that the globular component of the histone complex should not follow this general rule. Thus the terminal carbon atoms of the relatively long, charged, side chains of the globular region for the core protein would be expected to produce comparatively narrow resonances in addition to and coincident with the narrow resonances from the charged side chains of the random-coil regions of the protein. The narrow resonance at 34 ppm in the spectrum from calf thymus core protein, assigned to the $C_\gamma$ of glutamic acid side chains, has not been observed in previous histone $^{13}C$ NMR spectra (Clark et al., 1974; Lilley et al., 1975) and the incidence of glutamic side chains in the random-coil "tails" is relatively low. For the same reasons as outlined above for the resonances from lysine side chains, it seems likely that this resonance is also due to surface mobility.

With this in mind an interesting comparison may be made between the spectra of calf and chicken core protein. The latter shows lower intensity of both $\epsilon$-lysine and $\gamma$-glutamate resonances, suggesting the possibility of additional salt linkages in chicken core protein. If this is so, these interactions must be present in a hydrophobic region well protected from the screening effect of the 2 M NaCl. It thus seems possible that a slight degree of unfolding may have occurred during the isolation of the core protein from calf thymus.

*$^{13}C$ Relaxation Times.* Further evidence for the mobility of the polypeptide backbone in the core protein may be obtained by examining the relative relaxation times of $C_\alpha$ carbon atoms. Some analysis of the problem has been presented elsewhere (Lilley, 1977), and a study has been made (Lilley et al., 1976) of relative spin-lattice relaxation times ($T_1$). The dependence of spin-spin relaxation time ($T_2$) upon mobility is essentially monotonic (Doddrell et al., 1972), whereas the interpretation of $T_1$ relaxation times is more complicated, since two roots are obtained in calculating mobilities. Thus it is preferable to compare values of $T_2$ between the core protein and a globular protein of similar molecular weight.

The pulse sequence [90°-5 ms-180°-5 ms-FID collec-

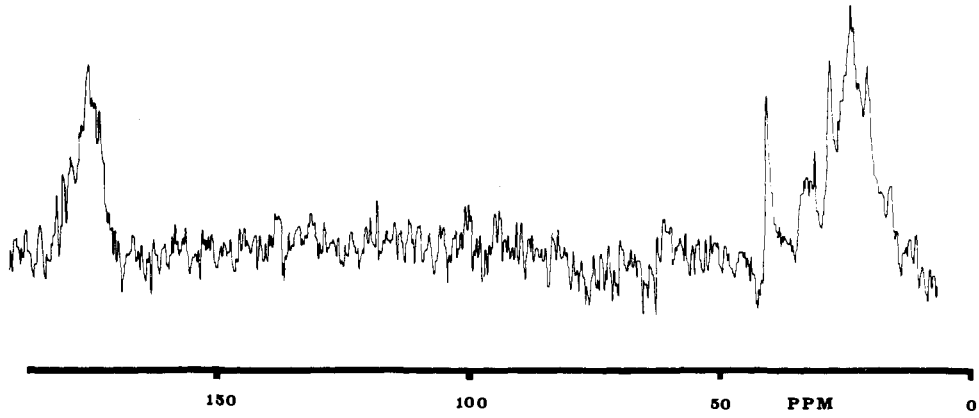


FIGURE 5: Partially $T_2$ relaxed $^{13}C$ NMR spectrum of $\beta$-lactoglobulin, using a total delay between pulses of 10 ms.

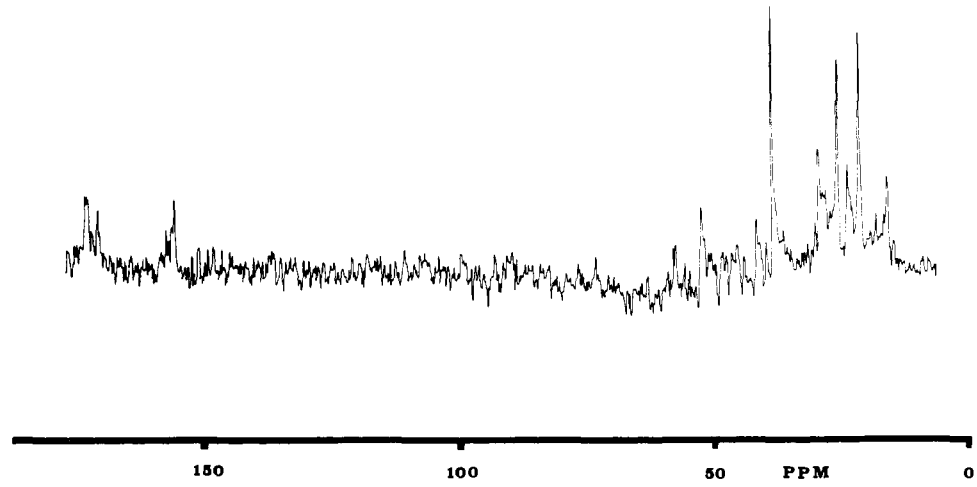


FIGURE 6: Partially $T_2$ relaxed $^{13}C$ NMR spectrum of core protein from chicken erythrocytes, using a total delay between pulses of 10 ms.

tion–0.5 s] (Hahn, 1950) has been used, where all resonances undergo $T_2$ relaxation, with proportional reduction in intensity, for 10 ms. This is a short time relative to the relaxation of the majority of C–H vectors, except those with correlation times of the order of $\tau_R$, the tumbling time of the molecular frame. Figure 5 shows the result of the above pulse sequence upon $\beta$-lactoglobulin. Whilst the partially relaxed spectrum retains the resonances arising from comparatively mobile terminal methylene and methyl groups, the $C_\alpha$ region, 45–60 ppm, has relaxed to the point of being virtually indistinguishable from the noise level. By comparison, the analogous PRFT spectrum of chicken erythrocyte core protein, shown in Figure 6, retains intensity in the $C_\alpha$ region. We conclude that in the core protein a proportion of polypeptide backbone carbon atoms must have mobilities in excess of the tumbling time of the molecular frame, i.e., core protein must possess random-coil sections. Evidence from line widths and relaxation times are therefore consistent.

*$^1H$ NMR Spectrum of Core Protein.* Analysis of the $^{13}C$ NMR spectra of individual histones (Clark et al., 1974; Lilley et al., 1975; Lilley, 1977; Tancredi et al., 1976), together with consideration of primary sequences and theoretical aspects of protein folding (Pardon and Richards, 1973), strongly suggest that N-terminal basic tails will extend from a globular domain composed of the histone C termini. These would be largely stabilized by hydrophobic interactions.

Whilst we have shown (Lilley, 1976) that $^{13}C$ NMR is very useful for examining the *mobile* elements of histone complexes, the spectra allow little insight into the globular domain of the core protein. For this purpose, $^1H$ NMR spectra are required.

The 270-MHz $^1H$ NMR spectrum of calf thymus core protein is shown in Figure 7. This spectrum is quite typical for that of a protein having precise tertiary structure. Numerous ring current shifted resonances are seen, both upfield of the aromatic region from 6.8 ppm, and the methyl region upfield from 1 ppm. Four resonances are reproducibly resolved in the former region, which are probably tyrosine protons within the shift cones of other aromatic side chains. In the absence of assignments for aromatic protons, little precise structural information may be obtained from these shifts, but some general points may be made. In order to produce shifts of the observed magnitude, some aromatic side chains must be within an angstrom of each other (Johnson and Bovey, 1958), i.e., in some form of close interaction. Furthermore the upfield shifts indicate that the interactions are not close stacking with coincident hexagonal axes, but are consistent with either off-centered stacking or inclined axes. Zero length cross-linking by TNM and UV irradiation of chromatin (Martinson and McCarthy, 1975; Martinson et al., 1976) has indicated that aromatic–aromatic interactions are likely to be of importance in both H2B–H2A and H2B–H4 interactions, although the stereochemical requirements for these free radical reactions are difficult to estimate. The methyl peak at 1 ppm shows a removal of intensity to the region 0–1 ppm, and at least three ring current shifted resonances are resolved. We therefore conclude that some phenylalanine and tyrosine residues are in close proximity to isoleucine, leucine, or valine residues.

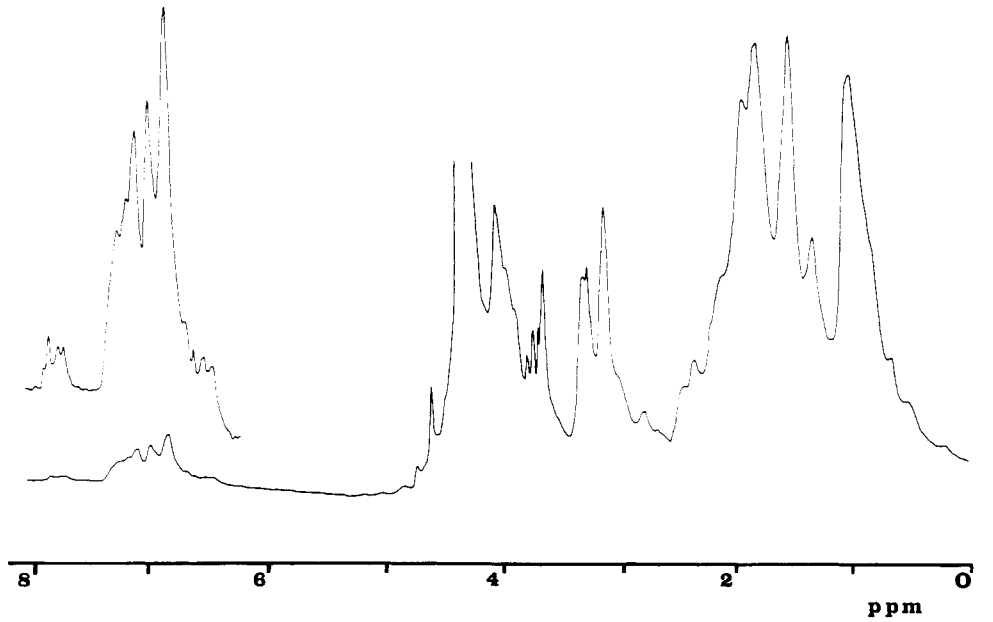


FIGURE 7: The 270-MHz $^1H$ NMR spectrum of core protein from calf thymus. The aromatic region (6–8 ppm) is also shown vertically expanded × 8.

Such interactions would be predicted to occur in a protein interior stabilized by hydrophobic contacts. Further evidence of precise tertiary structure may be gained from examination of histidine $C_2H$ resonances at ~7.9 ppm. Clearly these do not superimpose to give a single narrow peak, indicating that either they are ring current shifted and/or have altered p$K$ values due to their different environments.

Thus the globular domain of core protein is at least in part a precisely structured entity, analogous to a globular protein. Ring current shifted resonances have also been observed (Moss et al., 1976) in the $^1H$ spectrum of the H3–H4 tetramer, indicating that this too has a globular region having a precise three-dimensional structure.

## General Discussion

Analysis of the $^{13}C$ NMR spectrum of histone core protein indicates that part of the complex has considerable localized mobility. The line widths and relaxation times of $C_\alpha$ resonances demonstrate that this mobility arises from the presence of random-coil polypeptide sections. These regions are thought to be N terminal for two reasons. Examination of $^1H$ NMR line widths of aromatic resonances (Bradbury and Crane-Robinson, 1971; Bradbury et al., 1975; Lilley et al., 1975; Pekary et al., 1975a,b) and computer simulation of $^{13}C$ NMR spectra of histones (Clark et al., 1974; Lilley et al., 1975; Tancredi et al., 1976) and their complexes (Lilley, 1977) have in each case indicated the presence of mobile N-terminal "tails." Secondly, tryptic removal of the N termini (Weintraub and Van Lente, 1974; Brandt et al., 1975), itself demonstrating the enhanced accessibility of these regions, perturbs the $^{13}C$ NMR spectra in a manner consistent with the loss of random-coil regions (Lilley, 1977; Lilley et al., 1976; Richards et al., 1976).

Furthermore, it has been shown that the globular region to which the tails attach has a precise tertiary structure, thereby perturbing the proton NMR spectrum in a highly characteristic manner.

Molecular weight determination of trypsin digested core protein (Lilley et al., 1977), together with $C_\alpha$ losses in the $T_2$ pulse sequence experiments, suggest that one-fourth to one-third of the core protein mass is present in the tail regions. Computer simulation of the spectra of the individual histones (Clark et al., 1974; Lilley et al., 1975; Lilley, 1977; Tancredi et al., 1976) supports this estimate. Thus tail lengths of 25–30 amino acids for H4 and 30–40 amino acids for histones H2A, H2B, and H3 are estimated. The signal-to-noise ratio of the core protein spectra prevents a more accurate estimate.

The measured sedimentation coefficient of 3.7 S is in good agreement with other determinations (Weintraub et al., 1975). A molecular weight of 57 000 (Campbell and Cotter, 1976; Wooley et al., 1977) and a radius of gyration of 29.9 Å (Wooley et al., 1977) have also been measured for this complex. Assuming a partial specific volume of 0.745 (Brutlag et al., 1969), it may be calculated that the sedimentation coefficient and the radius of gyration are too low and too high, respectively, for a globular sphere. It is probable that the histone tails are responsible for increasing the radius of gyration expected for a sphere. Furthermore the low sedimentation coefficient may be best explained by assuming that the tails provide an "extra" frictional drag, probably arising from strong hydration. A model for core protein assuming histone tails has also been used in order to explain the observed high angle neutron scattering profile (Wooley et al., 1977).

*Chromatin Monomer Structure.* Early models for chromatin structure (Pardon and Richards, 1973; Van Holde et al., 1974) included the suggestion that the N-terminal regions of the histones interact with the DNA phosphate groups via a groove of the DNA. The random-coil tails observed in isolated core protein do not result from charge repulsion effects since the isolation and observation is made in 2 M NaCl. The more likely reason for their appearance in solution is therefore that they have been released from a distinct nucleoprotein domain. Earlier studies on chromatin have indicated (Simpson, 1970) that histones do not prohibit the entry of reporter molecules to the minor groove of the DNA. Thus it seems likely that the major groove might be the preferred site of interaction for the histone tails (Pardon and Richards, 1973). The accessibility of the histone N termini in chromatin to trypsin (Weintraub and Van Lente, 1974; Brandt et al., 1975) and the differential protection afforded to DNase I attack by the histones (Noll,

1974b; Simpson and Whitlock, 1976) are both in accord with a model of chromatin in which histone tails are an important feature.

In summary, the monomer structure is visualized as follows. The core is composed of the C termini of the four histones, assembled as two comparatively discrete tetrameric disks (Pardon et al., 1977; Richards et al., 1977) of precise tertiary structure. Basic residues of these regions are most likely aligned on the surface so as to follow the "path" of DNA folding about the core. In addition to the charge neutralization thus provided, the primary ionic interactions result from the presence of the basic tails in or around the major groove of the DNA double helix.

The model for the chromatin bead suggests an attractive possibility for its participation during transcription and/or replication, which depends upon the existence of the tails. Since both processes require DNA strand separation, some disruption of the structure is required, and yet there is evidence (Lacy and Axel, 1975; Mathis and Gorovsky, 1976; Weintraub and Groudine, 1976; Foe et al., 1976) showing that active genes have a subunit structure typical of bulk chromatin. We suggest that this is accomplished by means of a loss of the internal hydrophobic interactions whilst retaining the ionic interactions between the DNA and the histone tails. Thus the process becomes favored on enthalpic grounds, since many ionic interactions are not broken, and also on entropic grounds since the process is unimolecular in both directions. A further advantage inherent in the model is a retention of any possible "phasing" of histone position relative to the DNA primary structure. Such a view of possible structure-function relationships in the chromatin monomer is entirely dependent upon the existence of the histone "tails" as discrete entities, and the suggestion bears similarities to a recent model by Weintraub et al. (1976).

Acknowledgments

D. M. J. Lilley wishes to thank Imperial Chemical Industries for an I.C.I. Research Fellowship. We should also like to thank Dr. O. W. Howarth (University of Warwick) for allowing the use of the Bruker WH-90 FT spectrometer and Ms. C. B. Stratford for excellent technical assistance. We should like to thank our colleagues Drs. J. C. Wooley and R. I. Cotter for discussion.

References

Bradbury, E. M., Cary, P. D., Crane-Robinson, C., Rattle, H. W. E., Boublik, M., and Sautiere, P. (1975), *Biochemistry 14,* 1876–1885.

Bradbury, E. M., and Crane-Robinson, C. (1971), Histones and Nucleohistones, Phillips, Ed., New York, N.Y., Plenum Press, pp 85–134.

Brandt, W. F., Böhm, L., and Von Holt, C. (1975), *FEBS Lett.* 51, 88–93.

Brutlag, D., Schlenhuber, C., and Bonner, J. (1969), *Biochemistry 8,* 3214–3218.

Burgoyne, L. A., Hewish, D. R., and Mobbs, J. (1974), *Biochem. J. 143,* 67–72.

Campbell, A. M., and Cotter, R. I. (1976), *FEBS Lett. 70,* 209–211.

Clark, V. M., Lilley, D. M. J., Howarth, O. W., Richards, B. M., and Pardon, J. F. (1974), *Nucleic Acids Res. 1,* 865–880.

Doddrell, D., Glushko, V., and Allerhand, A. (1972), *J. Chem. Phys. 56,* 3683–3689.

Edwards, P. A., and Shooter, K. V. (1969), *Biochem. J. 114,* 227–235.

Foe, V. E., Wilkinson, L. E., and Laird, C. D. (1976), *Cell 9,* 131–146.

Hahn, E. L. (1950), *Phys. Rev. 80,* 380–394.

Johnson, C. E., and Bovey, F. A. (1958), *J. Chem. Phys. 29,* 1012–1014.

Kornberg, R. D. (1974), *Science 184,* 868–871.

Lacy, E., and Axel, R. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 3978–3982.

Laemmli, U. K. (1970), *Nature (London) 227,* 680–685.

Lilley, D. M. J. (1977), in Molecular Spectroscopy, West, A. R., Ed., London, Heyden, pp 96–114.

Lilley, D. M. J., Howarth, O. W., Clark, V. M., Pardon, J. F., and Richards, B. M. (1975), *Biochemistry 14,* 4590–4600.

Lilley, D. M. J., Howarth, O. W., Clark, V. M., Pardon, J. F., and Richards, B. M. (1976), *FEBS Lett. 62,* 7–10.

Lilley, D. M. J., Pardon, J. F., Worcester, D. L., and Richards, B. M. (1977), in preparation.

Martinson, H. G., and McCarthy, B. J. (1975), *Biochemistry 14,* 1073–1078.

Martinson, H. G., Shetlar, M. D., and McCarthy, B. J. (1976), *Biochemistry 15,* 2002–2007.

Mathis, D. J., and Gorovsky, M. A. (1976), *Biochemistry 15,* 750–755.

Moss, T., Cary, P. D., Crane-Robinson, C., and Bradbury, E. M. (1976), *Biochemistry 15,* 2261–2267.

Noll, M. (1974a), *Nature (London) 251,* 249–251.

Noll, M. (1974b), *Nucleic Acids Res. 1,* 1573–1578.

Olins, A. L., and Olins, D. E. (1974), *Science 183,* 330–332.

Pardon, J. F., and Richards, B. M. (1973), *Subunits Biol. Syst., Part B,* 1–70, 305–322.

Pardon, J. F., Worcester, D. L., Wooley, J. C., Cotter, R. I., Lilley, D. M. J., and Richards, B. M. (1977), submitted for publication.

Pardon, J. F., Worcester, D. L., Wooley, J. C., Tatchell, K., Van Holde, K. E., and Richards, B. M. (1975), *Nucleic Acids Res. 2,* 2163–2176.

Pekary, A. E., Li, H. J., Chan, S. I., Hsu, C. J., and Wagner, T. E. (1975a), *Biochemistry 14,* 1177–1184.

Pekary, A. E., Chan, S. I., Hsu, C. J., and Wagner, T. E. (1975b), *Biochemistry 14,* 1184–1189.

Richards, F. M. (1963), *Annu. Rev. Biochem. 32,* 269–300.

Richards, B. M., Cotter, R. I., Lilley, D. M. J., Pardon, J. F., Wooley, J. C., and Worcester, D. L. (1976), in Current Chromosome Research, Jones, K., and Brandham, P. E., Ed., Amsterdam, North-Holland Publishing Co., pp. 7–16.

Richards, B. M., Pardon, J. F., Lilley, D. M. J., Cotter, R. I., Wooley, J. C., and Worcester, D. L. (1977), *Cell Biol., Int. Rep. 1,* 107–116.

Shaw, B. R., Cordon, J. L., Sahasrabuddhe, C. G., and Van Holde, K. E. (1974), *Biochem. Biophys. Res. Commun. 61,* 1193–1198.

Simpson, R. T. (1970), *Biochemistry 9,* 4814–4819.

Simpson, R. T., and Whitlock, Jr., J. P. (1976), *Cell 9,* 347–353.

Tancredi, T., Temussi, P. A., Paslillo, L., Trivellone, E., Crane-Robinson, C., and Bradbury, E. M. (1976), *Eur. J. Biochem. 70,* 403–408.

Tanford, C. (1954), *J. Am. Chem. Soc. 76,* 945–946.

Thomas, J. O., and Kornberg, R. D. (1975a), *FEBS Lett. 58,* 353–358.

Thomas, J. O., and Kornberg, R. D. (1975b), *Proc. Natl. Acad.*

*Sci. U.S.A. 72*, 2626–2630.
Van Holde, K. E., Sahasrabuddhe, C. G., and Shaw, B. R. (1974), *Nucleic Acids Res. 1*, 1579–1586.
Weintraub, H., and Groudine, M. (1976), *Science 193*, 848–856.
Weintraub, H., Palter, K., and Van Lente, F. (1975), *Cell 6*, 85–110.
Weintraub, H., and Van Lente, F. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71*, 4249–4253.
Weintraub, H., Worcel, A., and Alberts, B. (1976), *Cell 9*, 409–417.
Woodcock, C. L. F. (1973), *J. Cell Biol. 59*, 368a.
Wooley, J. C., Pardon, J. F., Richards, B. M., and Worcester, D. L. (1977), in preparation.

# Studies on the Biosynthesis of 16-Membered Macrolide Antibiotics Using Carbon-13 Nuclear Magnetic Resonance Spectroscopy†

Satoshi Ōmura,* Hideo Takeshima, Akira Nakagawa, Jun Miyazawa, Francois Piriou,‡ and Gabor Lukacs‡

ABSTRACT: The origin of the skeletal carbons in the lactone ring of 16-membered macrolide antibiotics has been studied. $^{13}C$-labeled antibiotics, leucomycin and tylosin, have been obtained from the culture broth of *Streptomyces kitasatoensis* 66-14-3 and *Streptomyces fradiae* C-373, respectively, in the presence of appropriate $^{13}C$-labeled precursors, and $^{13}C$ NMR spectra of the antibiotics thus obtained have been measured. It was shown that the aglycone of leucomycin $A_3$ is derived from five acetates, one propionate, one butyrate, and an unknown precursor corresponding to two carbons. The formyl carbon which is characteristic of the basic 16-membered macrolides originates from C-4 of butyrate. On the other hand, the aglycone of tylosin is formed from two acetates, five propionates and one butyrate. Butyric acid and ethylmalonic acid are metabolized to propionyl-CoA or methylmalonyl-CoA through a pathway involving methylmalonyl-CoA mutase, and subsequently incorporated into the lactone ring of tylosin.

Much attention has been paid to the biosynthesis of macrocyclic lactones of 12- or 14-membered macrolides such as methymycin and erythromycin, and the origin of these skeletal carbons has been fully explored (Bentley, 1962; Birch et al., 1964; Grisebach et al., 1961; Vaněk et al., 1961; Kaneda et al., 1962; Friedman et al., 1964).

With regard to the biosynthesis of the lactone ring of 16-membered macrolides, magnamycin A has been intensely studied as a representative unit and several reports have appeared. These investigations suggested that carbon atoms 9–16, 20, and 21 (the numbers refer to those given for leucomycin $A_3$ which is the dihydro product of magnamycin B as shown in Figure 1A) are derived from acetate and carbon atoms 7, 8, and 19 arise directly from a propionate residue, but that precursors other than acetate, which are readily formed from glucose or succinate, are involved in the formation of the portion consisting of carbon atoms 1–6, 17, and 18 as established by using $^{14}C$-labeled precursors (Grisebach et al., 1961; Grisebach and Achenbach, 1962a,b; Achenbach and Grisebach, 1964). In another study, Srinivasan and Srinivasan (1967) concluded that the lactone moiety is synthesized from one propionate residue (carbons 7, 8, and 19) and eight acetate units (all other carbons).

Among known macrolide antibiotics, the basic 16-membered macrolides are characterized by the presence of a formyl group on the lactone ring. Considerable interest has been focused on the discussion as to the origin of the formyl group. Grisebach and Achenbach (1962a) had originally suggested that it arises from succinate, but this possibility was eliminated later (Grisebach and Weber-Schilling, 1968). Srinivasan and Srinivasan (1967) proposed that its origin is acetate.

As seen in the biosynthetic studies of macrolides and ansamycins such as rifamycin S (White et al., 1973) and geldanamycin (Rinehart, personal communication), *propionate* is directly incorporated into the lactone ring. (A substrate written in italics indicates the chemical entity which actually participates in the biological reaction.) Analogous evidence was obtained in the formation of branched-chain fatty acids by biological systems (Gastambide-Odier et al., 1963). Branching methyl groups in these antibiotics and in the branched-chain fatty acids represent C-3 of a propionate residue incorporated as an intact unit.

On the basis of these observations, it seems more reasonable to assume that the formyl group would arise from C-4 of *butyrate* formed from acetate during the biosynthetic process rather than from acetate directly as suggested by Srinivasan and Srinivasan (1967).

In this connection, we have investigated the biosynthetic origin of the lactone ring of leucomycin and tylosin as representative 16-membered macrolides by using $^{13}C$-labeled compounds and $^{13}C$ NMR[1] spectroscopy.

The present work strongly supports the conclusions mentioned above and provides further evidence as to the metabolism of fatty acids related to the macrolide biosynthesis by *Streptomyces* species. Preliminary reports of this work have been published (Ōmura et al., 1975a,c, 1976).

† From the School of Pharmaceutical Sciences, Kitasato University and The Kitasato Institute, Minato-ku, Tokyo 108, Japan. *Received December 7, 1976.*

‡ Centre National de la Recherche Scientifique, Institut de Chimie des Substances Naturelles, 91190 Gif sur Yvette, France.

[1] Abbreviations used are: NMR, nuclear magnetic resonance; CoA, coenzyme A.