when grown quickly. The identical birefringence (0.22) that they present in (100) and their similarity of unit-cell dimensions, apart from $a$, imply structural resemblances.

Optical data, after Ringertz, on the new orientation, are:

*Uric acid* (biaxial positive): $2V$(calc.) $= 84°$; $B = 0.31$; $\alpha = 1.588$ parallel to $b$; $\beta = 1.739$; $\gamma = 1.898$; $\gamma:c = 45.6°$.

*Uric acid dihydrate* (biaxial negative): $2V$(obs.) $= 40.4°$; $B = 0.22$; $\alpha = 1.508$ parallel to $b$; $\beta = 1.691$; $\gamma = 1.728$ $||$ $c$.

Test refractive-index solutions of $n =$ 1.51, 1.59, and 1.73 distinguish the two forms well, if crystals of sufficient size can be obtained.

Density separation is less satisfactory unless carried out very rapidly (for example, in a gradient column) because of the risk of dehydration. The stability of the hydrate varies widely, being dependent both on temperature and on the local $p$H. The hydrate appears to be more stable at lower temperatures than the anhydrous form, to which it may, however, sometimes change completely within 30 minutes.

Identification is most reliably made by x-ray methods, and a list of the principal powder lines is appended (Table 1). It should be noted that the strong line of the dihydrate at 8.75 Å is well inside the first line (at 6.56 Å) given by the anhydrous acid.

ROBIN SHIRLEY

*Department of Chemical Crystallography, University College, London, England*

### References and Notes

1. K. Lonsdale and P. Mason, *Science*, this issue.
2. W. M. Ord, *The Influence of Colloids upon Crystalline Form and Cohesion* (Stanford, London, 1879).
3. A. Brun, *Arch. Sci. Phys. Nat. Geneva* **7**, 284 (1899).
4. P. Gaubert, *Compt. Rend.* **202**, 1192 (1936).
5. A. N. Winchell, *The Optical Properties of Organic Compounds* (Academic Press, New York, ed. 2, 1954), p. 249.
6. H. Ringertz, *Acta Cryst.* **19**, 286 (1965).
7. These approximate to the orthorhombic axes of R. C. Waters and E. H. Shaw, *Proc. S. Dakota Acad. Sci.* **35**, 83 (1956).
8. Research supported by Science Research Council, London.

9 March 1966

## Immunoglobulin Structure: Variation in the Sequence of Bence Jones Proteins

Abstract. *Analysis of the amino acid sequence of one Bence Jones protein is almost complete. Many points of interchange occur in the amino terminal portion of the molecule relative to partial-sequence data for other proteins. Most, but not all, are compartible with one-step mutations. Such structural variation in immunoglobulin light chains may result from many related genes.*

The hypothesis that "all Bence Jones proteins of the same antigenic type share a fixed portion of their sequence and also have a mutable part," which was first proposed by Putnam *et al.* (*1*) on the basis of comparative peptide maps, has been substantiated by amino acid sequence analysis in several laboratories (*2–4*). Since many of the peptides of Bence Jones proteins are found in normal human $\gamma$-globulin (*5*) and since these proteins are analogous to the "light" polypeptide chains of immunoglobulins (*6*), sequence analysis of Bence Jones proteins has been undertaken to facilitate structural study of normal $\gamma$-globulin. The widespread variation in amino acid sequence of Bence Jones proteins of antigenic type I (type K) that has recently been demonstrated is thought to be related to the biological specificity of the light chains of antibody globulins.

It has been questioned whether the multiple differences in structure represent an accumulation of point mutations or chromosomal rearrangement (*7*). Only by complete amino acid sequence analysis of at least one Bence Jones protein as a standard of reference can the structural differences in fragments of other proteins be interpreted. We have previously presented the sequence of 148 of the assumed 212 amino acid residues of one antigenic type K Bence Jones protein (specimen Ag) including the consecutive sequence of 118 residues in the COOH-terminal half of the molecule (*2*).

In comparison with the partial sequence of one other protein (specimen Roy), only one difference in the COOH-terminal half of the molecule was deemed significant, namely, the interchange of valine and leucine at position 189 in the Roy numbering system (*3*). Since there appeared to be many interchanges in the $NH_2$-terminal portion of these molecules, it has been proposed that Bence Jones proteins have a "variable" region and a "constant" portion (*2, 3, 8*).

We now present the probable sequence of the $NH_2$-terminal portion of Bence Jones protein Ag with the exception of one uncertain area from positions 19 through 32. By reference to the partial sequence of specimen Roy (*3*), to the assumed sequence of portions of specimen Cum (*3*), and to the sequences of small peptides near the disulfide bridges of other specimens reported by Milstein (*4*), a minimum of 18 positions of interchange have been defined for specimen Ag relative to other Bence Jones proteins of the same antigenic type. Altogether, there are 22 positions of interchange when all areas of definitely known sequence are compared for all proteins studied thus far. These involve 32 interchanges and 26 different pairings of amino acids (Table 1). There is no precedent for such variations in sequence in individual specimens of proteins having a similar biosynthetic origin and function in the same species except for instances of polymorphic proteins under control of allelic genes, such as the multiple normal hemoglobins of man and other species.

Figure 1 gives the probable sequence of the $NH_2$-terminal half of Bence Jones protein Ag for positions 1 to 18 and 33 through 106. The procedures for sequence analysis have been described (*2, 8*). By analysis of many chymotryptic peptides, repeated confirmation has been obtained for many of the 148 positions reported (*2*). No evidence for error in any of these positions was found. The continuous sequence of the COOH-terminal half of the molecule (residues 106 to 212) is not shown since it has already been presented (*2*). Since this apparently differs only at position 189 for various proteins, only this position in the COOH-terminal half and the COOH-terminal residue cysteine are indicated in Fig. 1.

Milstein (*4*) has recently confirmed our sequence for 9 of the 19 positions from 115 through 133 and for 13 of the 23 positions from 190 through 212, in the neighborhood of the half-cystine residues. The composition of the undetermined portions of these peptides in his work and in the work of Hilschmann and Craig (*3*) is entirely compatible with the sequence we have reported. Hence, it is quite probable that the only variation in the COOH-terminal half of type K Bence Jones proteins

is the valine-leucine interchange at position 189.

The length of type K Bence Jones proteins is not yet established since the sequence is not complete for any specimen. For comparative purposes (Fig. 1) we have used the numbering system based on the Roy protein (*3*). However, Hilschmann and Craig have indicated this protein may have 213 residues (*3*). Furthermore, in their specimen Cum the composition of the peptide $T_c18_a$, which begins at position 62, requires the presence of two additional amino acid residues between positions 70 and 71 in the Roy numbering system. In our specimen Ag we have established the sequence Phe·Thr (*9*) at this location, which we have numbered as positions 70b and 70c until this question is resolved. If we assume that the unreported gap in sequence in our specimen Ag (positions 19 to 33) is analogous to the same region in Roy, for which only the peptide composition has been reported, then Ag would have 214 residues. This is identical to the number calculated from careful amino acid analysis of this protein. The determination of the complete sequence and the absolute length of several type K proteins is essential for testing various hypotheses of the genetic basis of the sequence variations in these molecules.

From amino acid analysis the empirical formula of Ag is: $Lys_{13}$, $His_3$, $Arg_6$, $Asp_{20}$, $Thr_{20}$, $Ser_{28}$, $Glu_{25}$, $Pro_{11}$, $Gly_{13}$, $Ala_{12}$, $(CyS/2)_5$, $Val_{13}$, $Met_1$, $Ilu_9$, $Leu_{15}$, $Tyr_9$, $Phe_9$, $Try_2$. Of the 214 residues listed, 200 have been identified in the sequence shown in Fig. 1 and in reference *2*, fig. 1. The missing 14 amino acids (which include one half-cystine residue) would exactly fill the gap from positions 19 through 32. Most of these correspond to amino acids found by Milstein (*4*) in the disulfide-bridge peptides that contain half-cystine at position 23. However, all of our data, except that for amino acid analysis, suggest the presence of only two histidines in the molecule.

In answer to a letter on our previous paper (*7*), we have reported, "Further study of the sequence data with reference to Nirenberg's table of codons has led us to conclude that it is unlikely that the sequence differences in Bence Jones proteins are ascribable to frequent short-range inverted duplications . . . as required by the Smithies hypothesis." The additional data in Fig. 1 and Table 1 likewise are not in accord with the Smithies intrachromatid inversion hypothesis (*10*). Since Milstein has later reached the same conclusion from analysis of his data (*4*), the conflicts with the results predicted by the Smithies model need not be considered in detail here.

Talmage (*7*) in an analysis of our earlier data concluded that the eight probable interchanges then identified could all be explained as being due to a single nucleotide substitution. This appeared to be evidence for an accumulation of point mutations. By reference to the table of codons for the genetic code (*11*), it can be verified that 26 of the 32 interchanges (81 percent)—or 20 of the 26 different amino acid pairings (77 percent)—are compatible with a single nucleotide change in the codons for the amino acid pair. Both transitions and transversions would be required. However, six of the pairings would require a change in two nucleotides. Since four of these involve specimen Cum for which the positions have been assigned by inference, verification of these must await sequence analysis of the Cum protein. These positions are (i) serine at position 21 rather than isoleucine, (ii) arginine at position 89 rather than tyrosine, and (iii) leucine at position 90 rather than either aspartic acid or glutamic acid. A fifth instance requiring a two-step mutation may be in doubt, namely the aspartic acid–threonine interchange at position 91 in specimen Ker. Asparagine, which is compatible with a one-point mutation to threonine, is reported as residue 91 for specimen BJ. Since the difficulties of distinguishing asparagine and aspartic acid in sequence analysis are well known, this example remains questionable. Thus, only one definite example of a two-step mutation remains—the alanine-isoleucine interchange between proteins Ker and Rad at position 19.

The case for accumulation of point mutations could be made even stronger. For 20 different amino acids there are 190 possible pairings (or interchanges). For the definitely known codons for amino acids there are 68 possible "one-step" interchanges and thus almost twice as many possible two- and three-step interchanges. We now give examples of 20 amino acid pairings that are compatible with a single nucleotide change in the codons, but there is unequivocal evidence for only one two-step mutation, and this is at position 19 where two single-step interchanges have clearly occurred.

Of course, many positions of interchange other than those listed in Table

Table 1. Definite sequence differences in type K Bence Jones proteins. The position in the polypeptide chain (1, 19, and so forth) is based on the Roy numbering system (*3*). Data for specimens Roy and Cum taken from Hilschmann and Craig (*3*) and for the specimens BJ, Ker, Day, and Rad from Milstein (*4*). Asx and Glx signify that the authors did not report which glutamic and aspartic acid residues are present as the amides. The data given for specimen Cum depend on peptide composition, end group determination, and sequence analogy, but not on actual sequence analysis.

| Protein | 1 | 19 | 21 | 22 | 24 | 46 | 53 | 55 | 72 | 75 | 76 | 89 | 90 | 91 | 92 | 94 | 98 | 102 | 103 | 104 | 105 | 189 | Inv type |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ag | Asp | | | | | Ilu | Asn | Glu | Thr | Gly | Leu | Tyr | Asp | Thr | Leu | Arg | Gln | Leu | Glu | Ilu | Lys | Val | b+ |
| Roy | Asx | | | | | Leu | Lys | Glx | Thr | | | | | | | | | Val | Asx | Phe | Lys | Leu | a+ |
| Cum | Glx | | Ser | | | | | Ala | Lys | Arg | Val | Arg | Leu | | | | | Leu | Glx | Ilu | Arg | Val | b+ |
| BJ | | Val | Ilu | Thr | Gln | | | | | | | Tyr | Glu | Asn | Leu | | | | | | | Leu | a+ |
| Ker | | Ilu | Ilu | Thr | Gln | | | | | | | Tyr | Asp | Asp | Leu | Pro | Pro | | | | | Leu | a+ |
| Day | | Val | | | | | | | | | | | | | | | | | | | | | |
| Rad | | Ala | Leu | Ser | Arg | | | | | | | Tyr | Glu | Thr | Ser | Thr | | | | | | Val | b+ |

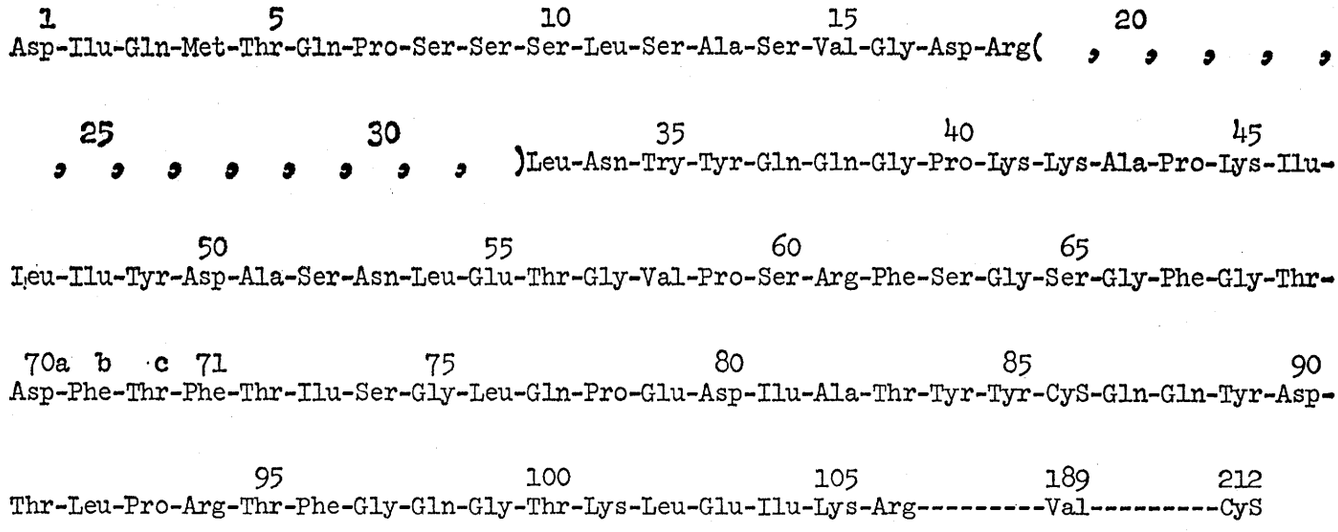


Fig. 1. Partial amino acid sequence of a type K Bence Jones protein (specimen Ag). The Roy numbering system (*3*) is assumed, but two additional amino acid residues are indicated as 70 b and 70 c. The sequence of the carboxyl-terminal portion of the molecule (residues 106 to 212) has been given earlier (*2*). The undetermined portion of the sequence (residues 19 through 32) has been left blank.

1 are probable, as based on the amino acid composition of peptides where the sequence is not fully known. For example, there are 38 positions where the sequence is known in the $NH_2$-terminal half of Ag but only the peptide composition is reported for other proteins. Of these, at least eight or nine represent points of probable interchange not listed in Table 1. On the other hand, no apparent interchanges exist in the 42 positions where the sequence is determined for the COOH-terminal half of Ag but only the peptide composition is known for other proteins.

In accord with conclusions reached from comparative peptide maps (*5*), some proteins will be rather similar in structure, whereas others may differ greatly. For example, there are 25 positions of known sequence in both the Rad and the Ker proteins, yet nine of these represent points of interchange. On the other hand, all of the 96 positions known in the Roy protein are also known in Ag, but only six are points of interchange. At least seven others can be predicted, however, and these include four positions already listed for other proteins in Table 1 (that is, 89, 91, 94, and 98). These results indicate a localization of the areas of variable sequence to restricted segments of the molecule, mainly around the half-cystine residues at positions 23 and 86 and the "switch peptide" constituting residues 102 to 105.

The possible biological significance of a single amino acid interchange is indicated by the strong correlation that is developing between the nature of the Inv genetic factor and the amino acid at position 189. The Inv factor is $a^+$ in three cases where the amino acid at position 189 is leucine (Roy, BJ, Ker), whereas this factor is $b^+$ in Cum and Rad (*3, 4*), where residue 189 is valine. Since the Inv factor in our specimen Ag is $b^+$ and valine is residue 189, the correlation reported by Milstein (*4*) now holds for six cases where the sequence is definitely known (*12*).

Like Milstein (*4*), we have tested various model systems based on the genetic code that would explain sequence variation as being due to crossing over or to a mistake in translation but have not found a unified explanation. For example, according to the hypothesis of Potter, Appella, and Geisser (*13*), some unusual triplets coding only in the variable region could bind to anticodons of transfer RNA which lack complete specificity for their activating enzymes. Thus, different amino acids could be introduced by the same triplet. Generally, this should lead to a single kind of interchange for each amino acid, such as arginine always with lysine, yet arginine is replaced by six different amino acids according to Table 1 (see positions 24, 75, 89, 94, and 105). Likewise, leucine and isoleucine are each involved in five different pairings.

One of the key problems for understanding the nature of antibodies and their biosynthesis is the elucidation of the subtle variations in molecular structure that govern their specificity and provide their heterogeneity. The demonstration that Bence Jones proteins, and by implication the light chains of immunoglobulins, have regions subject to wide variation in amino acid sequence imposes the question whether these arise because of hypermutability of a single structural gene, to chromosomal rearrangement, or to the presence of many genes. Our results are compatible with the presence of many genes for the light chains, each of which could have arisen throughout evolutionary history by a process of duplication and independent mutation. With each step occurring as a one-step mutation, there could be an accumulation of one-point mutations. In time, randomly, a two-point mutation would arise as a second step. These, of course, would be of low frequency as appears to be the case in the results of Table 1. As in the instance of the human-hemoglobin chains, where apparent deletions occur, the length of polypeptide chains might vary. Until sequence analysis is completed for several Bence Jones proteins, this question cannot be resolved.

KOITI TITANI
EDWARD WHITLEY, JR.
FRANK W. PUTNAM
*Division of Biological Sciences, Indiana University, Bloomington*

### References and Notes

1. F. W. Putnam, C. W. Easley, J. W. Helling, *Biochim. Biophys. Acta* **78**, 231 (1963).
2. K. Titani, E. Whitley, Jr., L. Avogardo, F. W. Putnam, *Science* **149**, 1090 (1965).
3. N. Hilschmann and L. C. Craig, *Proc. Nat. Acad. Sci. U.S.* **53**, 1403 (1965).
4. C. Milstein, *Nature* **209**, 370 (1966).
5. F. W. Putnam, *Biochim. Biophys. Acta* **63**, 539 (1962); ——— and C. W. Easley, *J. Biol. Chem.* **240**, 1626 (1965).
6. G. M. Edelman and J. A. Gally, *J. Exp.*

*Med.* **116**, 207 (1962); S. Cohen, *Biochem. J.* **89**, 334 (1963); G. M. Bernier *et al.*, *Biochemistry* **4**, 2072 (1965).
7. D. W. Talmage, *Science* **150**, 1484 (1965); F. W. Putnam and K. Titani, *ibid.*, p. 1484.
8. K. Titani and F. W. Putnam, *ibid.* **147**, 1304 (1965).
9. Abbreviations for amino acid residues: Lys, lysine; His, histidine; Arg, arginine; Asp, aspartic acid; Asn, asparagine; Thr, threonine; Ser, serine; Glu, glutamic acid; Gln, glutamine; Pro, proline; Gly, glycine; Ala, alanine; Val, valine; Met, methionine; Ilu, isoleucine; Leu, leucine; Tyr, tyrosine; Phe, phenylalanine; Try, tryptophan; CyS, half-cystine.
10. O. Smithies, *Nature* **199**, 1231 (1963).
11. R. Brimacombe, J. Trupin, M. Nirenberg, P. Leder, M. Bernfeld, T. Jaouin, *Proc. Nat. Acad. Sci. U.S.* **54**, 954 (1965).
12. The Inv type of protein Ag was determined by Dr. A. G. Steinberg, Dept. of Biology, Western Reserve Univ., Cleveland. In nomenclature, Inv(2) is equivalent to Inv(a) and Inv(b) to Inv(3).
13. M. Potter, E. Appella, S. Geisser, *J. Mol. Biol.* **14**, 361 (1965).
14. We thank Laurel Bailey and Patricia Tindall for amino acid analyses. Supported by grants HE 10116-01 and CA 08497-01 from National Institutes of Health.

28 March 1966

## Sex Attractant of the Pink Bollworm Moth: Isolation, Identification, and Synthesis

Abstract. *The sex attractant produced in extremely small amount by the virgin female pink bollworm moth has been isolated in pure form and identified as 10-propyl*-trans-*5, 9-tridecadienyl acetate. Successful synthesis of this molecule confirms the structure and makes possible its practical use to help control this destructive pest of cotton. The attractant, for which the common name "propylure" is suggested, may be the first naturally occurring compound with propyl branching to be reported.*

The pink bollworm moth, *Pectinophora gossypiella* (Saunders), is one of the most destructive pests in the cotton-growing areas of the world (*1*). Although mention was made in 1957 (*2*) of the presence of a sex attractant in the female insect, no details were given concerning the techniques used for its demonstration. In 1962, Ouye and Butt (*3*) showed that small traps baited with a methylene chloride extract of mating insects lured males, and in 1964, Berger *et al.* (*4*) demonstrated that extracts prepared from the terminal (2 to 3) abdominal segments of virgin female moths elicited excited flight, rapid wing vibrations, and a characteristic upward curving of the abdomen in males. Males readily responded with their characteristic dance to the vapors expelled from a glass pipette contaminated with the abdominal extracts. Field traps baited with a crude methylene chloride extract of the female moths have been used in surveying pink bollworm infestation (*3–7*).

Although some concentration of the sex attractant was accomplished by chromatography prior to 1964 (*4*), and the attractant was thought to be an 18-carbon ester, its isolation in pure form was not realized at that time. We now wish to report the successful isolation, identification, and synthesis of the pure, highly active attractant.

The methylene chloride extractive, prepared from the whole bodies of 850,000 virgin female moths (2 days old), was dissolved in ten volumes of acetone, and the solution was kept overnight at −20°C. The large amount of precipitated white solid was filtered rapidly through a cold Buchner funnel and washed with cold acetone, and the combined mother liquor and washings were freed of solvent at 20 mm-Hg (bath below 40°C). The yellow oily residue was shaken repeatedly with portions of methanol at room temperature and the combined methanol-soluble portions were evaporated to dryness. The resulting oily residue was dissolved in five volumes of acetone, kept overnight at −20°C, and filtered. Evaporation of the acetone filtrate gave an oil that was chromatographed successively on two columns of Florisil (*8*), using a sample-to-adsorbent ratio of 1:30 for the first and 1:200 for the second column. Each column was eluted successively with hexane, 3-, 5-, and 10-percent ethyl ether in hexane (*9*). Only 3-percent ether in hexane removed active material in each case, as shown by exposure of caged male moths to the air from pipettes containing the solution vapors (*4*).

The active fraction was chromatographed on a column of silica gel impregnated with silver nitrate (*10*) and eluted successively with hexane, with 5-, 10-, 25-, and 50-percent ether in hexane, and then with ethyl ether. The active fractions eluted with 25- and 50-percent ether in hexane were combined and rechromatographed on silver nitrate-impregnated silica gel, and eluted successively with 40- and 60-percent benzene in hexane, 5-percent ether in benzene, and finally with ethyl ether. All fractions except that eluted with 40-percent benzene in hexane were active, and these were combined and subjected to preparative gas chromatography (*11*) to give four components with retention times of 8.0, 11.5, 16.5, and 20.0 minutes. Only the component emerging from the column in 11.5 minutes caused a sexual response in caged, male, pink bollworm moths. In this way there was obtained approximately 1.6 mg of the pure attractant as a colorless liquid having no detectable odor. Air expelled from a pipette contaminated with a very dilute solution of the attractant and stored at 5°C for at least 6 weeks was still attractive to caged males in the laboratory.

The infrared spectrum of the attractant showed strong bands at 1755 and 1235 $cm^{-1}$ and a medium band at 1038 $cm^{-1}$, characteristic of a primary acetate group. The spectrum also showed the presence of unsaturation (1660 $cm^{-1}$, weak), a *trans* double bond (965 $cm^{-1}$, medium), and an unbroken chain of at least four methylene groups (723 $cm^{-1}$, broad). The ultraviolet spectrum showed only end absorption, precluding the presence of conjugation. Hydrogenolytic gas chromatography of the attractant by the method of Beroza and Sarmiento (*12*) established the presence of branching in its structure. Additional support for an acetate group was derived from the fragmentation pattern of the mass spectrum, which showed a large base peak when the ratio of mass to charge, $m/e$, was 43 ($CH_3CO$). The mass spectrum showed a molecular weight of 280 for the attractant, and peak matching on a double-focusing high-resolution instrument established the elemental formula as $C_{18}H_{32}O_2$. The above data characterized the attractant as the acetate of a branched-chain, $C_{16}$, primary alcohol with two double bonds (*13*).

Nuclear magnetic resonance spectra were obtained in deuterochloroform by means of a Varian HR-100 spectrometer equipped with a C-1024 time averaging computer and facilities for proton-proton spin decoupling. Time averages of up to 25 scans were used, with the $C^{13}$-labeled satellite of tetra-