Crystal Structure of 1-[3-(Indol-3-yl)propyl]thymine. A Model for Protein–Nucleic Acid Interactions

Donald Voet

Contribution from the Chemistry Department, University of Pennsylvania, Philadelphia, Pennsylvania 19104. *Received July 30, 1979*

Abstract: Ind³-C₃-Thy¹, a model compound for protein–nucleic acid complexes, crystallizes in an extended conformation. Its X-ray crystal structure exhibits hydrogen bonds between neighboring thymine rings and between thymine rings and indole rings on different molecules. In addition, the thymine and indole rings of neighboring molecules associate through normal stacking interactions.

The mutual recognition of proteins and nucleic acids is among the fundamental processes of molecular biology because it is such interactions that are responsible for the transmission and expression of genetic information. Enzymatically mediated reactions involving nucleic acids are often dependent upon accurate recognition of the nucleic acid by the enzyme. Thus, DNA replicases, RNA polymerases, repressor proteins, restriction enzymes, repair enzymes, aminoacyl synthetases, and ribosomes, to mention some of the more obvious examples of enzyme systems utilizing this type of interaction, exhibit a recognition specificity that is often exquisitely precise.

There is a large body of spectroscopic data which indicates that there are specific interactions between nucleic acid bases and protein side groups.[1] Yet very little direct structural information is available concerning these interactions. This is largely because no structure determination of a protein–nucleic acid complex that can provide high-resolution structural data has yet been reported.

In an effort to generate information bearing on the nature of the interactions between nucleic acid bases and protein side groups, we are investigating the X-ray crystal structures of model compounds that contain these components. The present study reports the X-ray crystal structure of 1-[3-(indol-3-yl)propyl]thymine (Ind³-C₃-Thy¹), a compound in which an indole nucleus and a thymine residue are bridged by a propyl group. Thus, this structure constitutes a model for the interactions between the nucleic acid base thymine and the side group of the amino acid tryptophan.

Experimental Procedure

Crystals of Ind³-C₃-Thy¹, which were kindly provided by Professor Nelson Leonard, had been prepared according to the method of Mutai, Gruber, and Leonard.[1] A pyramidally shaped crystal having a 0.65-mm base and a 0.15-mm height was glued to a glass fiber along an edge of the pyramid base (the a axis). The symmetry and systematic absences of reflections in preliminary oscillation, Weissenberg, and precession photographs indicated the space group of the crystal to be $P2_12_12_1$.

All subsequent X-ray measurements were performed using a Picker FACS-I diffractometer equipped with a graphite monochromator and employing Cu Kα radiation (λ 1.5418 Å). The unit cell parameters, as determined by least-squares analysis of the angular positions of 12 independent reflections, are given in Table I.

X-ray intensities were measured using the 2θ scan mode at a scan rate of 1°/min over a basic scan range of 1.75°. Stationary background counts of 20 s were taken at both limits of each scan. The three standard reflections that were monitored after every 50th reflection showed no decay during the course of the data collection process. A total of 1357 unique reflections were measured to the limit 2θ = 125°.

The intensities, I, were corrected for Lorentz–polarization effects. Their standard deviations, $\sigma(I)$, were calculated according to counting statistics,[2] with an additional instrumental instability factor of 0.02. A total of 43 reflections had $I < 2.33\sigma(I)$ and were therefore considered to be unobserved.

The structure was solved by application of the direct methods program MULTAN.[3] All 21 expected nonhydrogen atoms were located in the E map based on the phase set with the highest combined figure of merit.

The structure was refined by full-matrix least-squares methods in which the quantity minimized was $\Sigma w(|F_o| - |F_c|)^2$ and $w = 1/\sigma^2(F_o)$. The atomic scattering factors used were taken from Cromer and Waber.[4] Refinement of the scale factor, the atomic positions, and the isotropic thermal parameters, with later refinement of the anisotropic thermal parameters, reduced $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$ from its initial value of 0.306 to 0.114. A difference Fourier map at this point showed numerous peaks and valleys but few features that could reasonably be associated with hydrogen atoms. Furthermore, a majority of the β_{33} thermal parameters were anomalously large. This suggests that the crystal was disordered in some manner. It was therefore decided to remeasure the diffraction data using a different crystal.

A new crystal of Ind³-C₃-Thy¹, which was an irregularly shaped flake with approximate dimensions 0.40 × 0.30 × 0.07 mm, was glued to a glass fiber along its a axis. The X-ray data were measured and treated as is described above but with a diffractometer scan rate of 0.5°/min due to the small size of the crystal. Of the 1356 unique reflections measured to the limit 2θ = 125°, the 224 with $I < 2.33\sigma(I)$ were rejected as unobserved.

Refinement of the structure was continued using the new data set. Anisotropic refinement of non-H atoms converged with R = 0.118. However, the resulting difference Fourier map was much less noisy than its predecessor and indicated significant electron density at the calculated positions of all of the hydrogen atoms except those of the methyl group. Further refinement of the positions and anisotropic thermal parameters of the nonhydrogen atoms and only the positions of the hydrogen atoms, which were assigned the Wilson plot isotropic temperature factor of 5.35 Å², converged at R = 0.092 based on 1132 observed reflections. The highest peak in the final difference Fourier map had a height of 0.38 e/Å³. The β_{33} anisotropic thermal parameters of some of the non-H atoms were still anomalously large. Thus, the relatively low level of refinement of the structure may be largely attributed to a slight disorder of the crystal structure.

Results

The coordinates of the nonhydrogen atoms of the structure are given in Table II. The molecular structure of Ind³-C₃-Thy¹ as well as the atomic numbering system used in this report are illustrated in Figure 1. The conformation of the propyl group is such that the thymine and indole moieties of the same molecule are out of contact.

Figure 2 illustrates the covalent bond parameters of the structure. The dimensions of the thymine ring closely resemble those found in other thymine-containing structures.[5] The bond lengths and angles of the indole group are within their expected ranges. The root-mean-square deviation of the six atoms of the thymine ring from their least-squares plane is 0.012 Å and that for the nine atoms of the indole group is 0.011 Å. Hence, these groups are each coplanar to within the experimental error of this study.

The hydrogen bonding interactions in the structure of

0002-7863/80/1502-2071$01.00/0

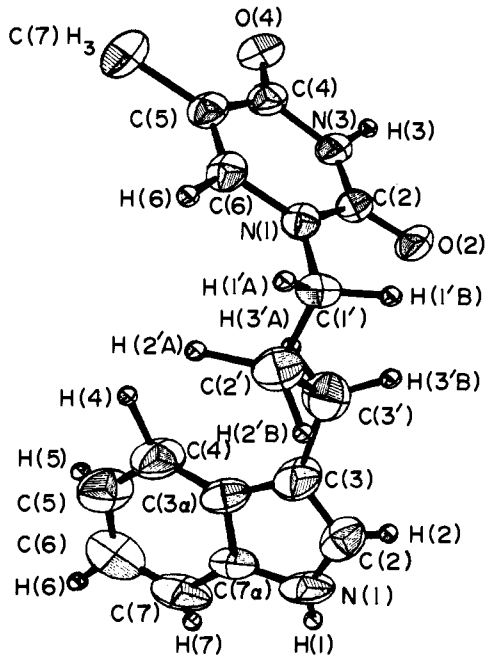

Figure 1. A perspective drawing of Ind[3]-C_3-Thy[1]. Nonhydrogen atoms are represented by thermal ellipsoids at the 50% probability level; hydrogen atoms are represented as spheres at the 25% probability level. The atomic numbering scheme that is used in this report is given.

Table I. Crystal Data for Ind[3]-C_3-Thy[1]

molecular formula	$C_{16}H_{17}N_3O_2$
unit cell dimensions	a = 6.820 (5) Å
	b = 13.875 (6) Å
	c = 15.260 (8) Å
space group	$P2_12_12_1$
Z	4
ρ_{calcd}	1.303 g cm^{-3}

Table II. Coordinates of the Nonhydrogen Atoms and Their Estimated Standard Deviations[a]

atom	x	y	z
IN(1)	1.5618 (11)	1.2670 (7)	0.2407 (8)
IC(2)	1.4563 (15)	1.1942 (8)	0.2089 (7)
IC(3)	1.3060 (12)	1.1778 (7)	0.2629 (7)
IC(4)	1.2157 (16)	1.2673 (8)	0.4056 (8)
IC(5)	1.2689 (22)	1.3380 (10)	0.4594 (7)
IC(6)	1.4316 (26)	1.3941 (9)	0.4432 (11)
IC(7)	1.5418 (20)	1.3748 (9)	0.3733 (12)
IC(7α)	1.4834 (12)	1.3035 (6)	0.3199 (7)
IC(3α)	1.3212 (11)	1.2468 (6)	0.3330 (6)
PC(1')	0.8611 (14)	1.0086 (7)	0.2373 (7)
PC(2')	0.9751 (21)	1.1006 (9)	0.2483 (11)
PC(3')	1.1777 (21)	1.0813 (11)	0.2665 (12)
TN(1)	0.8105 (9)	0.9542 (4)	0.3163 (4)
TC(2)	0.9274 (11)	0.8818 (5)	0.3420 (5)
TO(2)	1.0767 (8)	0.8589 (4)	0.3020 (3)
TN(3)	0.8765 (8)	0.8354 (5)	0.4173 (4)
TC(4)	0.7071 (11)	0.8499 (6)	0.4664 (5)
TO(4)	0.6737 (7)	0.8001 (4)	0.5297 (3)
TC(5)	0.5863 (10)	0.9267 (6)	0.4353 (5)
TC(6)	0.6410 (10)	0.9740 (6)	0.3630 (5)
TC(7)	0.3974 (11)	0.9493 (7)	0.4837 (5)

[a] The atomic coordinates are expressed as fractions of the unit cell edges. The prefixes I, P, and T in atom names refer to the indole, propyl, and thymine groups, respectively.

Ind[3]-C_3-Thy[1] are illustrated in Figure 3. It can be seen there that neighboring thymine residues are associated through a rather distorted N(3)-H···O(4) hydrogen bond (Table III).

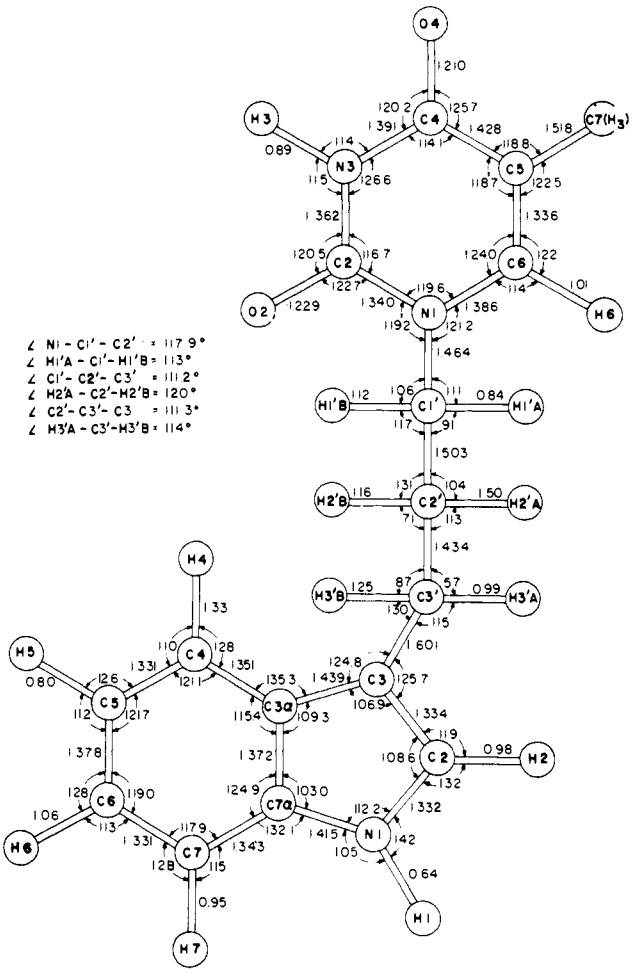

Figure 2. A schematic drawing of Ind[3]-C_3-Thy[1] showing the covalent bond distances (Å) and angles (deg). Standard deviations of these quantities, as estimated from the final cycle of least-squares refinement, average 0.012 Å and 0.9°, respectively, for distances and angles involving nonhydrogen atoms. The corresponding quantities involving hydrogen atoms are 0.06 Å and 4°, respectively.

Table III. Hydrogen Bond Parameters[a]

D-H···A	D···A, Å	H···A, Å	D-H···A, deg
IN(1)-IH(1)···TO(2) (i)	2.851	2.22	170.0
TN(3)-TH(3)···TO(4) (ii)	2.880	2.09	147.0

[a] Roman numerals accompanying atom names refer to atoms related to those in Table II by the following symmetry operations: (i) $3 - x, \frac{1}{2} + y, \frac{1}{2} - z$; (ii) $\frac{1}{2} + x, \frac{3}{2} - y, 1 - z$.

These thymine groups are related by twofold screw symmetry along the a axis and hence form an endless hydrogen-bonded helix about this screw axis. The indole group of another neighboring molecule forms a IN(1)-H···TO(2) hydrogen bond of normal geometry.

The indole residue is stacked on the thymine group of yet another neighboring molecule. The geometry of this association is seen in Figure 4. The dihedral angle between the least-squares planes of these stacked rings is 3°. There are no shorter than van der Waals contacts in this association.

Discussion

The rationale for the present study is that similar molecules tend to interact in a consistent manner and that therefore any interactions seen in the crystal structure of Ind[3]-C_3-Thy[1] are also likely to be important in protein-nucleic acid complexes.

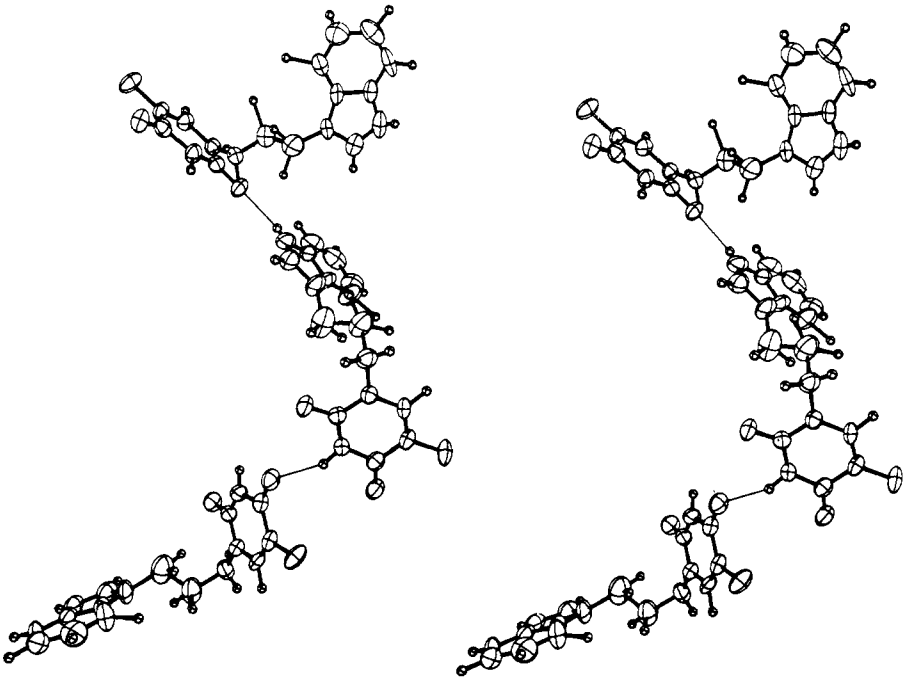

Figure 3. A stereodrawing of three neighboring molecules of Ind³-C₃-Thy¹ illustrating the hydrogen bonding associations present in the crystal structure. Hydrogen bonds are represented by thin lines.

Thus, it is well documented that hydrogen bonding interactions among nucleic acid bases,[5] and to a lesser extent their stacking interactions,[6] occur with only a narrow range of geometries. Likewise, proteins with similar amino acid sequences generally assume similar folding patterns.

Much of the meager crystallographic data reported concerning protein–nucleic acid interactions are based on the high-resolution X-ray structures of nucleotide binding proteins and nucleases. These studies have indicated that the binding sites for nucleic acid bases are often hydrophobic pockets. Aromatic side groups of the proteins are occasionally implicated in the binding of the bases, but no regular pattern is yet apparent, even among related enzymes.

The structures of only two proteins in complex with pyrimidines have been reported at a high enough resolution to reveal molecular detail. These are inhibitor complexes of ribonuclease S[7] and of staphylococcal nuclease.[8] In both of these nucleases the pyrimidine moieties of the inhibitors bind to the enzymes in hydrophobic pockets such that there are no readily classifiable interactions between the bases and the proteins. However, this is not surprising considering the lack of specificity of these nucleases in their requirement for a base.

There are other reported macromolecular structures in which pyrimidines bind to proteins, such as that of the CTP complex of aspartate transcarbamylase[9] or that of tobacco mosaic virus.[10] However, these are all presently of too low a resolution in the region of their binding sites to provide electron density maps that unambiguously reveal the geometries of the protein–nucleic acid interactions.

There have been several crystal structures reported of small molecule complexes of pyrimidines and protein components. In the structures of (glycylglycinato)(cytosine)copper(II)[11] and (glycylglycinato)(cytidine)copper(II) dihydrate[12] there are no specific interactions between the protein and nucleic acid base components. Rather, in both of these structures the base and the dipeptide are liganded to the metal ion. However, in the complex 5-bromocytosine–*N*-tosyl-L-glutamic acid[13] one of the glutamic acid carboxyl oxygen atoms forms a hydrogen bond to each of two symmetry inequivalent cytosine molecules. In contrast, both oxygens of the carboxyl group in the complex cytosine–*N*-benzoylglycine dihydrate[14] form a hydrogen bond with the same cytosine molecule. In the structure of the antiviral agent 5-[*N*-(L-phenylalanyl)amino]uridine,[15] the peptide group of one molecule also doubly hydrogen bonds to the uracil ring of a neighboring molecule. This structure also exhibits normal stacking interactions between the phenyl and uracil rings of yet other molecules. Finally, in the structure of the salt thymin-1-yl acetic acid–tyramine monohydrate[16] the component molecules associate through hydrogen bonding and ionic interactions; there are no ring stacking interactions observed in this structure.

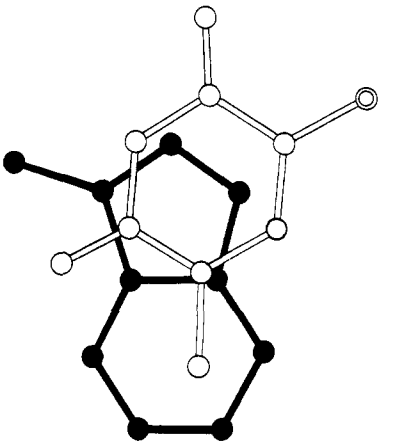

Figure 4. The projection onto the plane of the thymine ring of a stacked indole and thymine pair. The thymine methyl carbon atom, C(7), is represented by a double circle.

None of the foregoing small molecules chemically resemble Ind³-C₃-Thy¹ enough to permit similar patterns of association. Thus, what is required to verify the premise that indole and thymine rings tend to interact in a similar manner are additional structural studies of model compounds containing both of these chemical components.

The stacking association illustrated in Figure 4 corroborates the considerable spectroscopic evidence that the indole ring of tryptophan engages in specific stacking interactions with nucleic acids in general and thymine in particular.[1] This as-

sociation as well as the IN(1)-H···TO(2) hydrogen bond seen in Figure 3 could be utilized by proteins in order to specifically recognize a thymine or a uracil residue. It should be noted that the hydrogen bonding interaction can be made when the uracil or thymine is simultaneously involved in Watson-Crick interaction with adenine. Thus, this study provides a tentative structural basis for the recognition of thymine or uracil by a protein using, in part, specific interactions with a tryptophan residue.

Acknowledgments. I am most grateful to Professor N. Leonard for graciously providing the crystals of Ind[3]-C_3-Thy[1]. I thank Professor K.-I. Tomita for communicating the details of the crystal structure of thymin-1-ylacetic acid-tyramine monohydrate in advance of its publication. This work was supported, in part, by National Institutes of Health Grant GM23966.

Supplementary Material Available: Tables of hydrogen atomic coordinates (Table S1), temperature factors (Table S2), and observed and calculated structure factors (Table S3) (8 pages). Ordering information is given on any current masthead page.

References and Notes

(1) Mutai, K.; Gruber, B. A.; Leonard, N. J. *J. Am. Chem. Soc.* **1975**, *97*, 4095-4104, and references quoted therein.
(2) Stout, G. H.; Jensen, L. H. "X-Ray Structure Determination"; Macmillan: New York, 1968; pp 456-458.
(3) Germain, G.; Main, P.; Woolfson, M. M. *Acta Crystallogr., Sect. A* **1971**, *27*, 368-376.
(4) Cromer, D. T.; Waber, J. T. *Acta Crystallogr.* **1965**, *18*, 104-109.
(5) Voet, D.; Rich, A. *Prog. Nucleic Acid Res. Mol. Biol.* **1970**, *10*, 183-265.
(6) Bugg, C. E.; Thomas, J. M.; Sundaralingham, M.; Rao, S. T. *Biopolymers* **1971**, *10*, 175-219.
(7) Richards, F. M.; Wyckoff, H. W. *Enzymes, 3rd Ed.* **1971**, *4*, 647-806.
(8) Cotton, F. A.; Hazen, E. E., Jr.; Legg, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 2551-2555.
(9) Monaco, H. L.; Crawford, J. L.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *75*, 5276-5280.
(10) Stubbs, G.; Warren, S.; Holmes, K. *Nature (London)* **1977**, *267*, 216-221.
(11) Kistenmacher, T. J.; Szalda, D. J.; Marzilli, L. G. *Acta Crystallogr.* **1975**, *831*, 2416-2422.
(12) Szalda, D. J.; Kistenmacher, T. J. *Acta Crystallogr., Sect. B* **1977**, *33*, 865-869.
(13) Ohki, M.; Takenaka, A.; Shimanouchi, H.; Sasada, Y. *Chem. Lett.* **1974**, 551-554.
(14) Tamura, C.; Hata, T.; Sato, S.; Sakurai, N. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 3254-3261.
(15) Berman, H. M.; Hamilton, W. C.; Rousseau, R. J. *Biochemistry* **1973**, *12*, 1809-1814.
(16) Ogawa, K.; Tago, K.; Ishida, T.; Tomita, K.-I., submitted for publication.

Biosynthesis of the *Cephalotaxus* Alkaloids. Investigations of the Biosynthesis of Deoxyharringtonine, Isoharringtonine, and Harringtonine[1]

Amy Gitterman, Ronald J. Parry,*[2] Richard F. Dufresne, Daniel D. Sternbach, and Michael D. Cabelli

Contribution from the Department of Chemistry, Rice University, Houston, Texas 77001. *Received August 17, 1979*

Abstract: Precursor incorporation experiments have been used to investigate the biosynthesis of the antitumor *Cephalotaxus* alkaloids deoxyharringtonine (**2**), isoharringtonine (**3**), and harringtonine (**4**). It has been established that the acyl portion of deoxyharringtonine is derived from leucine via a pathway that resembles the conversion of valine into leucine in microbial systems. Incorporation experiments with DL-[9-^{14}C]deoxyharringtonic acid (**6**) have also shown that **6** is the precursor of isoharringtonic acid (**12**) and harringtonic acid (**14**) in vivo. Doubly labeled deoxyharringtonine has been synthesized and its ability to serve as an intact precursor of harringtonine has been examined. The results of this experiment indicate, but do not prove, that deoxyharringtonine is directly converted to harringtonine without prior deacylation.

Introduction

Plants of the genus *Cephalotaxus* (Cephalotaxaceae) have been found to contain a number of alkaloids that are esters of the parent alkaloid cephalotaxine (**1**) (Figure 1). These ester alkaloids include deoxyharringtonine (**2**), isoharringtonine (**3**), harringtonine (**4**), and homoharringtonine (**5**).[3] Deoxyharringtonine and its congeners have been shown to exhibit significant antitumor activity against experimental P388 leukemia and L-1210 leukemia in mice.[3] Recent investigations in the People's Republic of China have shown that harringtonine and homoharringtonine are effective in the treatment of human cancers.[4]

The biosynthesis of the ester alkaloids **2**-**5** poses two interesting problems. The first concerns the origin of the unusual diacids that are attached to cephalotaxine in these alkaloids. The second concerns the stages at which the individual diacids are linked to cephalotaxine. Experiments bearing upon each of these problems are outlined in this paper.

Results and Discussion

The key to the elucidation of the biosynthesis of the acyl portions of the antileukemic *Cephalotaxus* alkaloids was provided by the recognition that the diacid **6** linked to cephalotaxine in deoxyharringtonine, henceforth referred to as deoxyharringtonic acid, bears a close resemblance to a diacid intermediate involved in the biosynthesis of leucine from valine in microorganisms.[5] On the basis of this resemblance, we arrived at the hypothesis for deoxyharringtonic acid biosynthesis shown in Scheme I.

The hypothesis predicts that 3-carboxy-3-hydroxy-5-methylhexanoic acid (**8**) should be an intermediate in the biosynthesis of **6** and that carbon atoms 3-8 of **8** should be derived from L-leucine (**7**). Since the presence of diacid **8** or its derivatives in *Cephalotaxus* plants had never been reported, its presence was sought by isotopic trapping. A synthetic sample of racemic **8** was prepared from 4-methyl-2-pentanone using methods developed for the synthesis of deoxyharringtonic

0002-7863/80/1502-2074$01.00/0