ceding paper in this issue).
Neilands, J. B. (1973), in Inorganic Biochemistry, Eichhorn, G., Ed., Amsterdam, Elsevier.
O'Brien, I. G., Cox, G. B., and Gibson, F. (1971), *Biochim. Biophys. Acta 237*, 537.
Porra, R. J., Langman, L., Young, I. G., and Gibson, F. (1972), *Arch. Biochem. Biophys. 153*, 74.
Ramachandran, L. K., and Sastry, L. V. S. (1962), *Biochemistry 1*, 75.
Rosen, H. (1957), *Arch. Biochem. Biophys. 67*, 10.
Sayer, J. M., and Emery, T. F. (1968), *Biochemistry 7*, 184.
Stoops, J. K., Horgan, D. J., Runnegar, M. T. C., de Jersey, J., Webb, E. C., and Zerner, B. (1969), *Biochemistry 8*, 2026.
van der Helm, D., and Poling, M. (1976), *J. Am. Chem. Soc., 98*, 82.

Sulfatides of *Mycobacterium tuberculosis:* The Structure of the Principal Sulfatide (SL-I)†

Mayer B. Goren,* Olga Brokl, Peter Roller,‡ Henry M. Fales,§ and Bhupesh C. Das¶

ABSTRACT: The gross structural features of five families of multiacylated trehalose 2-sulfates elaborated by *Mycobacterium tuberculosis* strain H37Rv are described. The principal sulfatide SL-I is a 2,3,6,6′-tetraacyl-α,α'-D-trehalose 2′-sulfate, whose component carboxylate substituents (and homology) were previously established. In the present study the specific locations of the acyl substituents were assigned. The desulfated glycolipid (SL-I-CF) was methanolyzed on a column of diethylaminoethylcellulose (free base form), affording tri-, di-, and monoacylated trehalose mixtures. The most abundant diacyltrehalose generated was identified as 6,6′-bis(2,4,6,8,10,12,14,16-octamethyl-17-hydroxydotriacontanoyl)trehalose (6,6′-bis(C_{40}-hydroxyphthioceranoyl)trehalose), along with lower and higher homologues. A small amount (about 15%) of the unhydroxylated analogue (phthioceranate) was also recognized. From the monoacylated carbohydrate mixture (chiefly 6-(C_{40}-hydroxyphthioceranoyl)trehalose) surviving trehalose monopalmitate(s) were isolated by preparative gas chromatography of the trimethylsilylated products. Trehalose 2-palmitate was identified as the principal component. Small amounts of the 3 isomer may also be present, but no 6-palmitate was detectable. Gentle acidic solvolysis, which minimizes the possibility of acyl migrations, afforded a different diacyltrehalose, identified by mass spectrometry of the permethylated derivative as principally 2-palmitoyl(stearoyl)-3-phthioceranoyltrehalose. A variant in which hydroxyphthioceranate substitutes at the 3 position was also recognized. The results indicate that the biological acylation processes at the trehalose core are not entirely specific, but instead yield an SL-I family, for the chief member of which a logical structural expression is deduced.

The sulfatides (sulfolipids (Middlebrook et al., 1959)) of *Mycobacterium tuberculosis* comprise some five or more families[1] of multiacylated trehalose sulfates. These contain, in addition to palmitic–stearic acids, several unique multimethyl-branched substituents: phthioceranic and hydroxyphthioceranic acids (see below). In our earlier studies, individual glycolipid sulfates were separated and identified as multiesters of trehalose 2-sulfate. The gross distribution and proportions of the various acyl substituents within the individual sulfatides were determined and reported in part; and the chemical structures of the phthioceranic and hydroxyphthioceranic families of homologues were elucidated (Goren, 1970b; Goren et al., 1971); see Table I.

In the interim, we have learned that the sulfatides may play an important role in the pathogenesis of tuberculosis, probably promoting intracellular survival of the tubercle bacillus by antagonizing phagosomal–lysosomal fusion within phagocytic cells (Goren et al., 1974; Goren, D'Arcy Hart, Young, and Armstrong, in preparation). With these and other biological activities demonstrated (Kato and Goren, 1974), studies to delineate complete structural expressions for the more abundant sulfatides were more vigorously pursued. In the present study we summarize the gross structures of five mycobacterial sulfatides so far characterized and report the structure(s) of the principal sulfatide family (SL-I[2]) of the strain H37Rv.

Materials and Methods

Surface culturing of *M. tuberculosis,* strain H37Rv, lipid extractions, and column chromatographic separation of the various sulfatide families on diethylaminoethylcellulose

† From the Division of Molecular and Cellular Biology, National Jewish Hospital and Research Center, and Department of Microbiology, University of Colorado School of Medicine, Denver, Colorado 80206. *Received December 19, 1975.* This investigation was supported by Grant AI-08401 from the US–Japan Cooperative Medical Science Program, administered by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health, Department of Health, Education and Welfare.

* To whom to address correspondence at National Jewish Hospital.
‡ National Cancer Institute, Bethesda, Maryland.
§ National Heart and Lung Institute, Bethesda, Maryland.
¶ Institut de Chimie des Substances Naturelles, Gif-Sur-Yvette, France.

[1] Designated "families" because of the extensive homology characterizing the acyl substituents and also because of a degree of randomness in the specific location of individual acyl functions, as documented in the present study.

[2] Abbreviations used are: SL-I to -III, the intact sulfated glycolipids, SL-CF for the desulfated analogues; DEAE-cellulose, diethylaminoethylcellulose; Me_3Si, trimethylsilyl; GLC, gas–liquid chromatography; TLC, thin-layer chromatography; mu, mass units; ir, infrared; NMR, nuclear magnetic resonance; E–M, ether–methanol; MS, mass spectroscopy.

(DEAE) have been described previously (Goren, 1970a). In most recent studies we have included about 4 μCi of 1-[^{14}C]-propionate/l. in the Wong's medium for culturing the tubercle bacilli. The label is heavily incorporated into the multi-branched phthioceranic acids in the sulfatides (Goren and Brokl, unpublished data). The biogenesis of the methyl branches, therefore, resembles that demonstrated earlier for the mycocerosic acids (Gastambide-Odier et al., 1963; see also Polgar and Robinson, 1951). The present structural studies were conducted almost exclusively with the neutral glycolipids derived from desulfation of the parent substances in ether (Goren, 1971).

Isolation and Purification of the Sulfatides. The mycobacterial sulfatides are obtained both individually and as mixtures by DEAE chromatography of the crude lipids extracted from *M. tuberculosis* with hexane–decylamine as described earlier (refer especially to Table I of Goren (1970a) which details the elution sequence leading to recovery of SL-I and of some of the minor sulfatides as well). The latter are obtained individually in more homogeneous form by rechromatography on DEAE (as described for SL-II in the preceding reference). Details of additional improvements in elutrients leading to the purification of the more polar SL-III (as an example) will be described in a communication currently in preparation (Goren, Brokl, and Das).

Distribution and Quantitation of Acyl Substituents. The acylated positions on the trehalose cores of the minor sulfatides were determined as described previously for the principal sulfatide SL-I (Goren, 1970b): by permethylation with diazomethane and BF_3-etherate (Neeman et al., 1959), solvolysis, and separation of esters–carboxylic acids from the water-soluble carbohydrate. The former were converted into a uniform mixture of methyl esters with excess diazomethane. As one method of quantitation, the esters were separated in a prolonged silicic acid chromatography to yield palmitate–stearate, phthioceranates, and hydroxyphthioceranates *O*-methyl ethers. Each substituent was then quantitated by infrared spectrophotometry at the 1740–1745 cm^{-1} carbonyl band (Dittmer and Wells, 1969). The H_2O-soluble methylated trehaloses were identified by various chromatographic techniques as described previously (Goren, 1970b).

For preparative GLC of certain degradation products, specifically, Me_3Si derivatives of trehalose monopalmitates, we used a 6-mm diameter column of 3% OV-17 on Chromosorb W(HP), (column 260 °C; argon; 120 ml/min). The desired Me_3Si derivatives were trapped on a small section of OV-17 at room temperature. The trap was eluted with methanol and a small amount of aqueous 1 N HCl was added to regenerate the glycolipid. Excess H_2O was added and the mixture was contacted several times with hexane to extract OV-17. The aqueous–methanol solution containing the glycolipid was millipore-filtered, evaporated to dryness, and reconverted to the Me_3Si derivatives for definitive GLC: (2-mm diameter, 6-ft glass columns of 3% SE-30 on "Varoport 30", 265 °C; nitrogen, 30 ml/min).

Partial Solvolysis of Desulfated SL-I (SL-I-CF). A gentle progressive methanolysis of SL-I-CF, the tetraacyltrehalose obtained by desulfation of NH_4SL-I, is effected on a column of DEAE-cellulose or cellulose–DEAE-cellulose, 2:1, free base form (Goren and Brokl, 1974). SL-I-CF is dissolved and loaded in less than a column void volume of ether–methanol (E–M) 70:30. After 2–4 h contact, the solvolysis products are eluted rapidly with 70:30 E–M and finally with methanol. The various products are separated by chromatography with ether—E–M mixtures on cellulose or on cellulose–DEAE-cellulose (acetate); unaltered SL-I-CF is recycled for additional solvolysis. Five or six 2-h contacts exhaust most of the SL-I-CF.

As a useful alternative, gentle acidic solvolysis generated products different from those produced on the DEAE-cellulose (free base) column. Typically, 50 mg of SL-I-CF in 0.5 ml of $CHCl_3$ was diluted with 250 ml of methanol, and 25 ml of 2 M HCl in CH_3OH was added. The mixture was then left at room temperature or at 37 °C and samples were periodically examined by silica gel TLC in $CHCl_3$–CH_3OH–H_2O–CH_3COOH (70:30:2:2). After sufficient solvolysis (about 50–70 h) the lipids (in ether solution) were washed free of HCl and the products were separated by column chromatography as described in the previous section.

Synthesis of Trehalose Palmitates. Small samples of the various palmitoyltrehaloses were synthesized for ascertaining physical characteristics and behavior in TLC and GLC. Two methods were employed: palmitoylation of 6,6′-ditrityltrehalose on the one hand, and of anhydrous trehalose on the other, with palmitoyl chloride in pyridine. The latter procedure (method II) is simpler and yields the palmitate isomers in the approximate ratios 60:30:7:3 for the 6-, 2-, 3-, and (by inference) 4-palmitates, respectively. These are readily separable and identified both in silica gel TLC and by gas chromatography of the trimethylsilyl derivative, the sequence of elution from an SE-30 column being 3-, 2-, 4-, and 6-palmitoyltrehalose.

For the present study, the 2 and 3 isomers were isolated from the products of 6,6′-ditrityltrehalose acylation, and identified by permethylation analysis, NMR, and by mass spectrometry. Their chromatographic behavior allowed identification of their counterparts from the direct acylation of trehalose. The 6 isomer from the latter reaction was also identified by permethylation and by comparison with an authentic sample kindly provided by R. Toubiana (Gif-Sur-Yvette). Trehalose 2-palmitate, recovered in method II by prolonged multiple chromatographies of mixed palmitates on DEAE-cellulose (acetate) was recrystallized first from dioxane and then from methanol: mp 156.5–158 °C; $[\alpha]^{27}D$ +137 ± 1.2° (c 0.0083 (dioxane)). Anal.[3] Calcd for $C_{28}H_{52}O_{12}$: C, 57.92; H, 9.03. Found: C, 57.89; H, 8.92.

Hexa-O-methyltrehalose. Ditrityltrehalose was permethylated first with $(CH_3)_2SO_4$ and $Ba(OH)_2$ in dimethylformamide, and remethylated exhaustively in benzene with methyl iodide and silver oxide. The product was purified by recrystallization from benzene–trace pyridine (melting point: sinters at 160 °C; melts 230–232 °C). Anal.[3] Calcd for $C_{56}H_{62}O_{11}$: C, 73.8; H, 6.87. Found: C, 74.06; H, 6.88. After detritylation the hexamethyltrehalose was recovered as a viscous syrup which we were unable to crystallize.

Results

The gross distribution and molar proportions of the acyl substituents in the various sulfatides are summarized in Table I. The sulfatides are listed in the order in which they elute from a DEAE-cellulose column (see Materials and Methods); this order is reflected in TLC mobility as well.

Four of the lipids are tetraacyl derivatives; in SL-III (triacylated) the glucose moiety bearing the sulfate is otherwise unsubstituted. Table I also shows that SL-II′ is unique among the sulfatides in a positional isomerism which involves the 4 position on the multiacylated glucose moiety.

[3] Huffman Laboratories, Lakewood, Colorado.

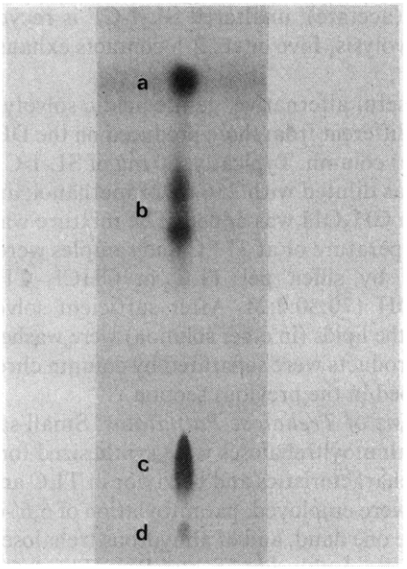


FIGURE 1: Thin-layer chromatogram of mixed glycolipids from DEAE-cellulose-catalyzed methanolysis. Silica gel plate developed 12 cm in $CHCl_3$–CH_3OH–CH_3COOH (100:7:0.7). Components: (a) mixed methyl esters plus unaltered SL-I-CF; (b) two groups of triacyltrehaloses ("mixed" SL-III-CF); (c) diacyltrehaloses ("mixed" SL-IV-CF); (d) trehalose and monoacyltrehaloses ("mixed" SL-V-CF).

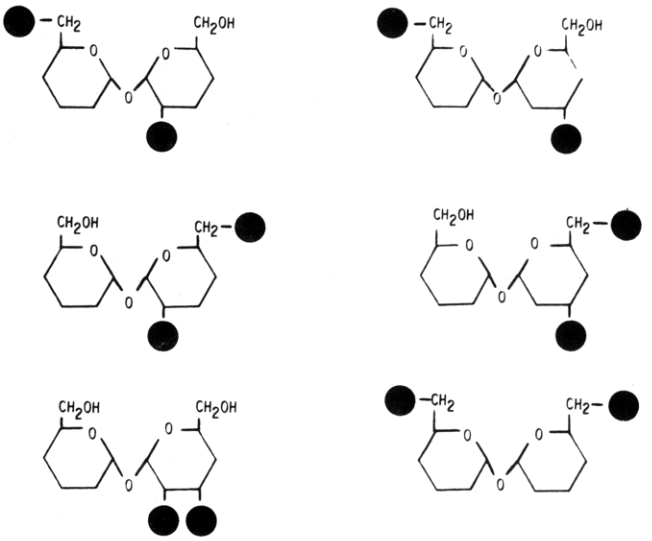


FIGURE 2: Hypothetical diacyltrehalose ("mixed" SL-IV-CF) as derived in a random solvolysis of SL-I-CF. Black circles represent acylated positions (as originally recognized in SL-I). Only the symmetrical 6,6′-diacylated product will not form a tritylated derivative.

SL-I: Structural Studies. Identifiable glycolipid fragments were derived from desulfated SL-I (SL-I-CF) by a partial methanolysis conducted in 70:30 anhydrous ether–methanol solution on a column of diethylaminoethylcellulose (DEAE-cellulose) as the free base (see Materials and Methods).[4] As an example, after four 2-h contacts and chromatographies as described, 300 mg of SL-I-CF afforded, in addition to unaltered starting material, three glycolipids mixtures of *low* ("mixed" SL-III-CF), *intermediate* ("mixed" SL-IV-CF), and of quite *high* polarity ("mixed" SL-V-CF) (29, 65, and 11 mg, respectively). Their chromatographic behavior (see Figure 1) suggested that these might contain three, two, and one acyl substituents, respectively. If the first step of solvolysis of SL-I-CF (a 2,3,6,6′-tetraacyltrehalose) is *random,* four triacyltrehaloses will be produced to comprise "mixed" SL-III-CF. These in turn can give rise to six diacyltrehaloses; i.e., the "mixed" SL-IV-CF as depicted schematically in Figure 2 (the solid circles representing acyl groups). Only the component with both 6 positions acylated is essentially inert toward trityl bromide. Thus tritylation of the "mixed" SL-IV-CF and chromatographic separation and *rejection* of the *tritylated* material afforded about 70% of nonreactive presumed 6,6′-diacyltrehalose. This product is henceforth designated SL-IV-CF as distinguished from "mixed" SL-IV-CF. Its infrared spectrum (Figure 3B) bears a notable resemblance to that of cord factor (Figure 3A) (6,6′-dimycoloyltrehalose (Noll and Bloch, 1955; Goren and Brokl, 1974), especially in the similarity of fine structure in the region 1000–1200 cm^{-1} (arrows). On permethylation and alkaline solvolysis, SL-IV-CF yielded a product identical with synthetic 2,3,4,2′,3′,4′-hexamethyltrehalose (see Materials and Methods). Acidic hydrolysis of this methylated trehalose gave 2,3,4-tri-*O*-methylglucopyranose as the only detectable sugar[5] (identified by gas–liquid chromatography of the Me_3Si derivative). Therefore SL-IV-CF, derived from the DEAE-cellulose-catalyzed solvolysis of SL-I-CF is indeed a 6,6′-diacyltrehalose.

TABLE I: Gross Structures of Five SL Families.

Sulfatide[a]	Trehalose Positions Substituted	Moles Acyl Substituent: Palmitate/ Stearate	Phthio-ceranate[b]	Hydroxy-phthiocer-anate[c]
SL-II′	2,4,6,6′	1	0	3
SL-II	2,3,6,6′	1	0	3
SL-I (principal)	2,3,6,6′	1	1	2
SL-I′	2,3,6,6′	1	2	1
SL-III	2,3,6	1	0	2

[a] Sulfate in 2′ position. [b] Phthioceranate: $C_{16}H_{33}(C(CH_3)HCH_2)_nC(CH_3)COOH$, n = 2–9; 6 principal. [c] Hydroxyphthioceranate: $C_{15}H_{31}C(OH)H(C(CH_3)HCH_2)_nC(CH_3)HCOOH$, n = 2–10; 7 principal.

The 6,6′ Substituents. Silica gel TLC of the methyl esters cleaved from the trehalose core of the permethylated 6,6′ diester indicated the acyl substituents to consist principally of the hydroxyphthioceranic series (now methylated at the hydroxyl group) with small amounts of the other acyl components. The relative amount of *O*-methyl ether (derived from the hydroxyphthioceranic ester) in the total esters was determined by 100-MHz nuclear magnetic resonance (NMR) spectrometry. The contribution of ether methoxyl (chemical shift 3.34 ppm) was some 80–85% of the integral of ester methoxyl (chemical shift 3.66 ppm). Combined gas chromatography–mass spectrometry confirmed this ratio and revealed that the small amounts of methoxyl-free esters were principally

[4] We wish to anticipate the valid question that this mildly alkaline solvolysis may well promote acyl migrations which might invalidate the results. Our data indicate that acyl migrations can occur but do not interfere: mildly acidic solvolysis (vide infra) affords identifiable fragments which complement and corroborate the results of the DEAE-cellulose-catalyzed degradations.

[5] In the more sensitive mass spectrometric examination of permethylated SL-IV-CF, a small amount of unsymmetrical contaminant (with a *tetramethyl glucose moiety*) was detected. Tetra-*O*-methylglucose was not detected in the GLC examination of the hydrolytic products.

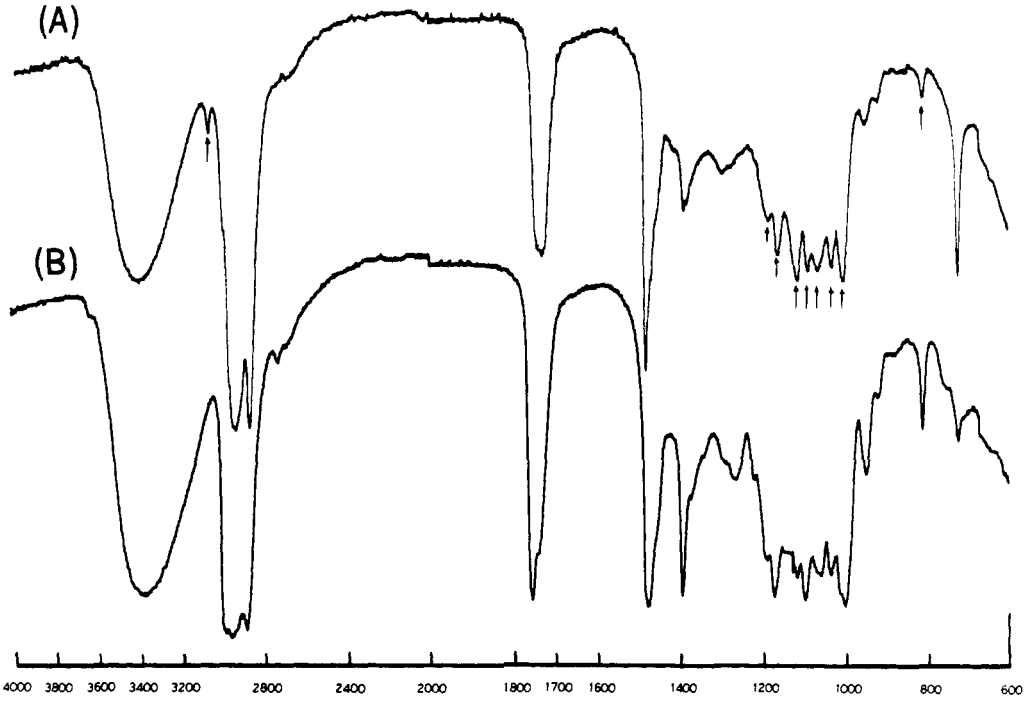


FIGURE 3: Infrared spectra of (A) cord factor (6,6′-dimycoloyltrehalose) and (B) nontritylatable SL-IV-CF derived from DEAE-cellulose solvolysis of SL-I-CF. Arrows in region 997–1175 cm^{-1} point to the fine structure pattern which is highly specific for 6,6′-diacylated trehalose.

phthioceranates (Table I) with $n = 4, 5$. It is clear, however, that the principal permethylated 6,6′-diacyl carbohydrate derived from SL-I is the symmetrical 6,6′-bis(methoxyphthioceranoyl)-2,3,4,2′,3′,4′-hexamethyltrehalose (see Figure 4, i). This diester was synthesized (from hexa-*O*-methyltrehalose and methoxyphthioceranoyl chloride) and by ir and mass spectrometry found to be essentially identical with the naturally derived product.

Mass Spectra of SL-IV-CF Derivatives. Of the numerous mass spectra obtained in this study, we wish to summarize only the most pertinent examples.[6] Figure 4 interprets the fragmentation pattern which is definitive for permethylated SL-IV-CF (both natural and synthesized), consisting of a series of prominent peaks in the higher mass region. According to the interpretations of DeJongh and Biemann (1963), Kochetkov and Chizhov (1966), and Adam et al. (1967), the initial fragmentation of i (Figure 4) would yield the oxonium ion ii, *m/e* 809. By successive losses of methanol (32 mass units) from the ring positions (preceding references) and from the methoxyphthioceranoyl substituent (B. C. Das and M. Goren, unpublished) the ion ii gives rise to the principal series of peaks *m/e* 777, 745, and 713. To the latter, the structure iii is assigned—a composition confirmed by high resolution analysis for $C_{47}H_{85}O_4$ (calculated, 713.645 mu): 713.645 mu. No peaks derivable from contributions of the unmethoxylated *phthioceranate* were evident. Thus i is symmetrical.

An important peak *m/e* 681 (seen in the spectrum) differs from *m/e* 713 by 32 mass units (CH_3OH), and a high resolution analysis shows this interpretation to be correct: $C_{46}H_{81}O_3$ ($809 - 4 \times 32$); calculated: 681.619 mu; found: 681.619 mu. However, we have no evidence how the fragment of this mass is generated; nor can we presently assign for it a meaningful structural expression. Still, this peak characterizes every synthetic permethylated 6-methoxyphthioceranoyltrehalose that we have examined and has until recently been considered definitive for this structure. An interesting variant is discussed later.

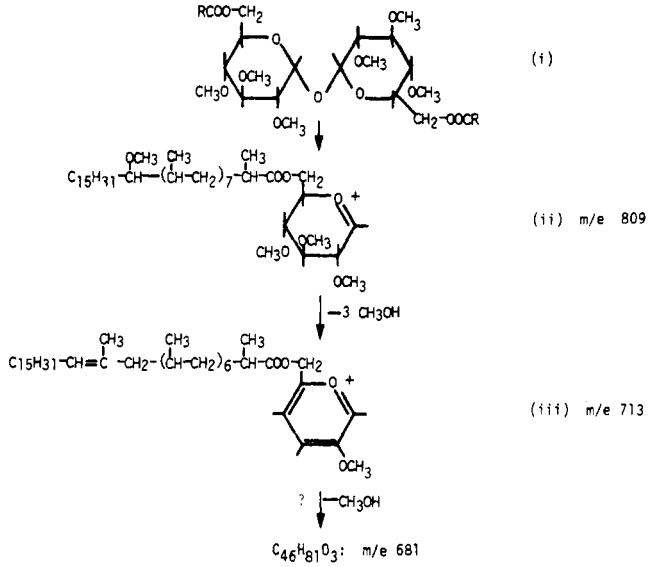


FIGURE 4: Definitive fragmentation pattern of permethylated SL-IV-CF. (i) 6,6′-Bis(methoxyphthioceranoyl)hexamethyltrehalose; R = methoxyphthioceranate as in ii.

All of the preceding are also important peaks in the mass spectrum of permethylated SL-I-CF (i.e., permethylated desulfated *SL-I*) and, therefore, indicate that the glucose moiety *originally bearing the lone acyl function and the sulfate* (the "prime" ring) is substituted at the 6′ position with hydroxyphthioceranate. The mass spectrum of *peracetylated* SL-I-CF (and SL-IV-CF as well) unequivocally verifies this conclusion. The portion of the spectrum contributed by the "prime" glucose moiety of peracetylated SL-I-CF is entirely in accord with

[6] Because of journal space limitations, several spectra discussed herein have been deleted, but the authors will provide these to the interested reader. An extensive analysis of spectra relevant to this structural study is planned as a separate communication (P. Roller, M. B. Goren, and O. Brokl).

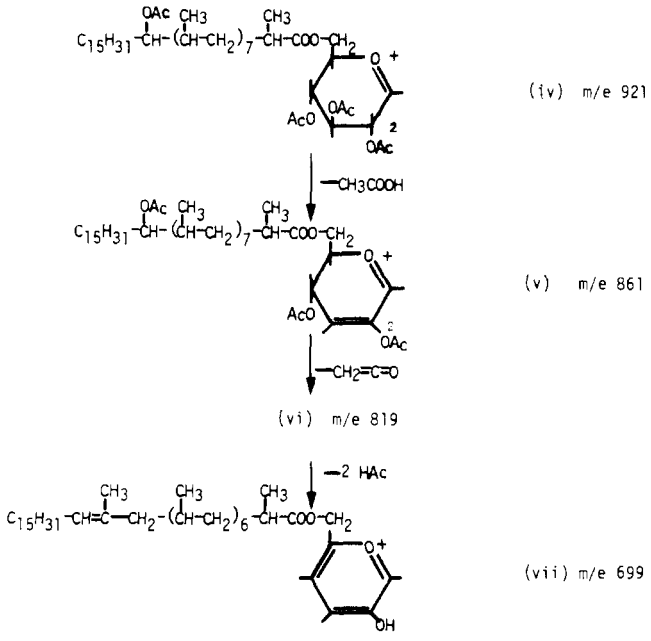


FIGURE 5: Fragmentation of peracetylated SL-I-CF.

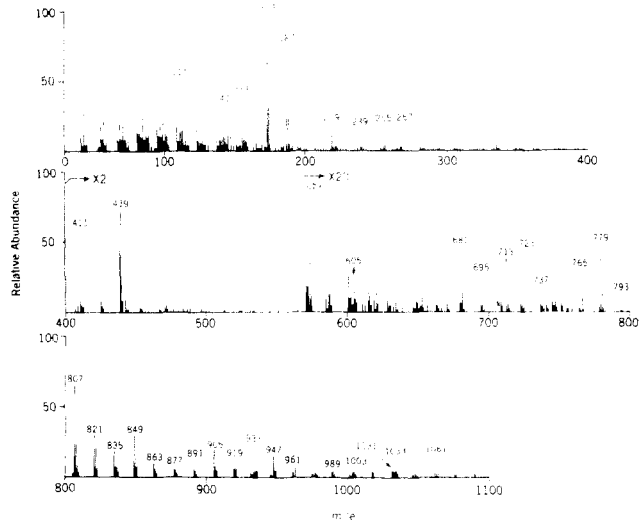


FIGURE 6: Mass spectrum of permethylated SL-IV′-CF (spectrum obtained with JEOL-JMS-01SG-2 instrument; 70-eV ionizing voltage).

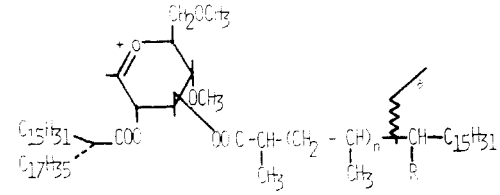


FIGURE 7: Structures assigned to the diacylated oxonium ion derived from permethylated SL-IV′-CF, based on mass spectrum of Figure 6 and analysis in Table II. R = -H for phthioceranate, $-OCH_3$ for methoxyphthioceranate. Fragmentation occurs at a in both series.

the interpretation depicted in Figure 5. Although the primary oxonium ion iv with 4 acetates (*m/e* 921) is not seen,[7] *m/e* 861 [921 minus acetic acid = structure v] is the most prominent peak in the high mass region. In accordance with a sequence seen numerous times in the present studies, e.g., with permethylated trehalose 2-palmitate, with a permethylated 2,3-diacyltrehalose (vide infra) (see Karlsson et al., 1969; Toubiana et al., 1973), fragment v would be expected to expel *ketene* from the 2 position to generate vi (*m/e* 819); elimination of two molecules of acetic acid (from the carbohydrate residue and from the acetoxyphthioceranic acid) then yields the "aromatized" fragment vii (*m/e* 699), a major peak. Comparable fragments attributable to unhydroxylated *phthioceranate* in the 6 positions were not seen.

In summary, the chemical, infrared, NMR, and mass spectrometric data in toto indicate that, in SL-I, the 6 and 6′ positions of the trehalose core are in the main symmetrically substituted, specifically with the hydroxyphthioceranates. However, a small amount of a variant may be present in which the 6 position of the triacylated glucose bears the *unhydroxylated* phthioceranic acid. This is supported by data to be discussed below. Since only two more substituents are present in SL-I, positional assignment to either of the two (phthioceranate or palmitate-stearate) substantially defines the complete structure of the principal sulfatide.

Identification of the Palmitate-Stearate Component of "Mixed" SL-V-CF. Examination of "mixed" SL-V-CF (the penultimate solvolysis product(s) from SL-I-CF) showed it to be almost entirely 6-hydroxyphthioceranoyltrehalose. Other methods failing, we resorted to preparative GLC to isolate any surviving trehalose palmitate and/or homologues. As dictated by experience gained with Me_3Si derivatives of synthetic trehalose palmitates (stearates), the natural analogues from "mixed" SL-V-CF were trapped and recovered (see Materials and Methods). By GLC comparison with the authentic synthetic esters, these products were then identified as almost exclusively 2-palmitoyl(stearoyl)trehalose. Still, the biological acylation leading to SL-I seems to be, to a small extent, random rather than entirely specific: in later experiments, small quantities of presumed 3-palmitate (but not 6-palmitate) have also been recognized. The 3 isomer may be generated by acyl migration: solvolysis of trehalose 2-palmitate on DEAE-cellulose does indeed produce some 3-palmitate (Goren and Brokl, unpublished). However, as discussed immediately following, the principal conclusions were correct. Thus the identification of 2-palmitoyltrehalose as the major palmitate isomer in "mixed" SL-V-CF provides the positional assignment for palmitate (stearate) and inferentially therefore for *phthioceranate*, which must occupy the remaining acylated position: 3. A unique confirmation of this assignment was obtained in an examination of the products derived from gentle acidic solvolysis of SL-I-CF.

Acid-Catalyzed Methanolysis of SL-I-CF: the 2,3-Diacyltrehalose. SL-I-CF was gently solvolyzed in 0.2 M methanolic HCl (see Materials and Methods) since ancillary studies showed that, while the mildly alkaline (DEAE-cellulose) process induced some acyl migration with pure 2- or 3-palmitate (but not with 6-palmitate), only solvolytic cleavage occurred under acidic conditions. In agreement with the preceding results, trehalose 2-palmitate was identified in the SL-V-CF mixture from the acid-catalyzed solvolysis—again, however, with small traces of presumed 3-palmitates. However, the *diacyl*trehalose which was also obtained was different from the SL-IV-CF previously described: in silica gel TLC this product, designated SL-IV′-CF, had significantly lower mobility than the former. Alkaline hydrolysis and TLC of products showed that the acyl components in SL-IV′-CF were

[7] The peak *m/e 921* is prominent in the spectrum of the peracetylated SL-IV-CF (of much lower molecular weight). Like permethylated SL-IV-CF, the peracetylated product is also seen to be symmetrical.

TABLE II: Analysis of Fragments in Mass Spectrum of Permethylated SL IV′-CF.[a]

R	2-Acyl Function	*n*	*m/e*	Composition	*m/e* − (a-H)[b]	Composition[c]
OCH_3	Stearate	7	1061	$C_{67}H_{129}O_8$	807	$C_{50}H_{95}O_7$[d]
OCH_3	Palmitate	7	1033	$C_{65}H_{125}O_8$	779	$C_{48}H_{91}O_7$[d]
H	Stearate	7	1031	$C_{66}H_{127}O_7$	807	$C_{50}H_{95}O_7$[d]
H	Palmitate	7	1003	$C_{64}H_{123}O_7$	779	$C_{48}H_{91}O_7$[d]
H	Stearate	6	989	$C_{63}H_{121}O_7$	765	$C_{47}H_{89}O_7$
H	Palmitate	6	961	$C_{61}H_{117}O_7$	737	$C_{45}H_{85}O_7$
H	Stearate	5	947	$C_{60}H_{115}O_7$	723	$C_{44}H_{83}O_7$
H	Palmitate	5	919	$C_{58}H_{111}O_7$	695	$C_{42}H_{79}O_7$
H	Stearate	4	905	$C_{57}H_{109}O_7$	681	$C_{41}H_{77}O_7$
H	Palmitate	4	877	$C_{55}H_{105}O_7$	653	$C_{39}H_{73}O_7$

[a] Refer to Figure 7. [b] For phthioceranate (a-H) = $C_{16}H_{32}$ (224 mu). For methoxyphthioceranate (a-H) = $C_{17}H_{33}O$ (254 mu). [c] Composition confirmed by high resolution analysis (Ionomet photoplate). [d] The same fragments can be generated from either series.

principally *phthioceranate* and palmitate–stearate along with small amounts of hydroxyphthioceranate.

The mass spectrum of permethylated SL-IV′-CF reveals that it is almost exclusively a 2-palmitoyl/stearoyl-3-acyltrehalose, and it provides definitive evidence that the 3 position is occupied principally by phthioceranate with a smaller contribution by hydroxyphthioceranate. The complete mass spectrum of this product is reproduced in Figure 6. Figure 7 and its associated Table II summarize structures of the principal (phthioceranoyl) and lesser (methoxyphthioceranoyl) variants of the acylated primary oxonium fragment generated from permethylated SL-IV′-CF. The specific fragmentation patterns deduced from Figures 6 and 7 were confirmed by high resolution analysis.

Although most spectra of compounds containing methoxyphthioceranate (Figure 7, R = OCH_3) show extensive C_3 homology (in "*n*"; Goren et al., 1971), the analogue in the spectrum of Figure 6 appears to be almost entirely restricted to the member *n* = 7, as represented by the peaks *m/e* 1033 and 1061. Evidence for lower homologues differing by successive 42 mass units is essentially (and fortuitously) nonexistent. Instead many peaks implicating the participation of at least 4 homologues of the *unhydroxylated* phthioceranate are evident, and these are detailed in Table II.

The principal cleavage of methoxyphthioceranate (Figure 7, at a, which gives a fragment *m/e* 255, was recognized in the earliest studies with the methoxylated branched acids (Goren et al., 1971). In the present form, with the acyl function esterified to the oxonium fragment, the cleavage leaves a residue decreased by 254 mu (confirmed many times by high resolution analysis). In spectra of substances where methoxyphthioceranate is a major component, the peak *m/e* 255 is very large and may even be the base peak. Its unimportant aspect in the present spectrum is convincing evidence that hydroxyphthioceranate is not of major significance in SL-IV′-CF.

From the fragments *m/e* 1033 and 1061, the 255-cleavage generates the fragments *m/e* 779 (with palmitate) and 807 (with stearate). The structural assignments for these two peaks were confirmed by high resolution analysis: for the stearate (calcd for $C_{50}H_{91}O_7$, 807.708 mu) 807.706 mu; for the palmitate (calcd for $C_{48}H_{91}O_7$, 779.677 mu) 779.679 mu. As shown, however, in Table II, identical fragments can also derive from the highest homologue seen in the phthioceranate-substituted series.

The fragmentations considered for the latter involve similar cleavages (Figure 7, at a) from the oxonium ions (R = H) *m/e* 1031, 1003, 989, 961, 947, 919, 905, and 877. These do *not* accommodate contributions from methoxyphthioceranate;

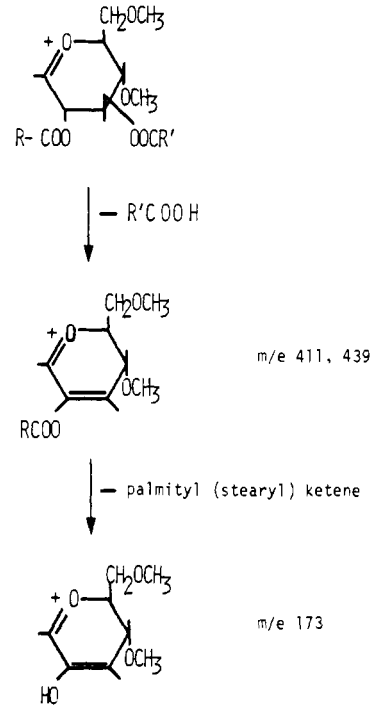


FIGURE 8: Fragmentation pattern of permethylated SL-IV′-CF showing elimination of the 3-acyl function as the carboxylic acid to yield the massive peaks (Figure 6) *m/e* 411 and 439, followed by expulsion of palmityl or stearyl ketene to yield the base peak *m/e* 173. R = palmitate/stearate; R′ = principally phthioceranate and some methoxyphthioceranate (refer to Figure 7).

their prominence and variety, therefore, support the conclusion that *phthioceranate* is the principal branched substituent in SL-IV′-CF. The postulated cleavages at a can, therefore, also generate the fragments *m/e* 779 and 807 (above) as well as the fragments in the series *m/e* 765, 737, 723, 695, and even *681* (of an entirely different composition from the fragment of similar mass discussed with respect to Figure 4). As indicated in Table II, this and other requirements were confirmed by high resolution analysis. Thus the principal contributions to SL-IV′-CF are from phthioceranate, with palmitate and stearate.

The crucial assignment of the 2,3-diacyl structure to SL-IV′-CF is elegantly supported by the remainder of the spectrum. As discussed for peracetylated SL-I-CF (Figure 5), the acyl group at the 3 position is eliminated as the carboxylic acid to leave in this instance the very prominent fragments *m/e* 411 and 439 (see Figure 8). The palmitoyl (stearoyl) substituents

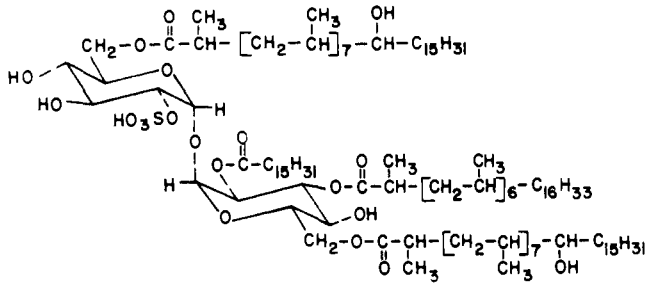


FIGURE 9: Structure of the principal component of SL-I. In a small amount of SL-I the acyl functions in positions 3 and 6 appear to be interchanged. In some preparations of SL-I, stearate appears to predominate over palmitate at the 2 position.

are then expelled as the respective *ketenes* to generate the fragment *m/e* 173. This is the *base peak* in the spectrum of permethylated SL-IV′-CF (and also in that of permethylated 2-palmitoyltrehalose—but not for the 3 or 6 isomer). It may be added that the assignment of the two acyl substituents to the same glucose moiety is of course supported by the contributions from the (unacylated) tetra-*O*-methylglucose, which generates the peaks *m/e* 219 → 187 → 155 (Kochetkov and Chizhov, 1966).

Therefore the diacylated trehalose products derived individually in both the weakly alkaline and in the weakly acidic solvolyses together provide the composite evidence from which the principal structural expression for SL-I is derived (Figure 9): 2-palmitoyl(stearoyl)-3-phthioceranoyl, 6,6′-bis(hydroxyphthioceranoyl)trehalose 2′-sulfate. In some SL-I, the substituents at the 3 and 6 positions of the multiacylated glucose are evidently interchanged. The nature of the 2 substituent and of the 6′ substituent appears to be substantially restricted.

Discussion

The gross structural features of the mycobacterial sulfatides and the exact structures of their acyl substituents were delineated in the earlier studies of this series. Positional assignment for the individual carboxylates of SL-I evolved as a more difficult task. Our early efforts at deducing the total structure from mass spectrometric examination of, for example, permethylated SL-I-CF were entirely unrewarding, and we were compelled to turn to degradative (and ultimately to synthetic) processes to secure lower molecular weight fragments for characterization. Two intensive efforts at achieving a selective and clean cleavage of the desulfated glycolipid at the trehalose glycosidic oxygen were largely unsuccessful, and so these methods were abandoned for the more prosaic solvolytic schemes detailed in the preceding sections. The gentle and quite slow methanolysis catalyzed by DEAE-cellulose (free base) was first recognized when we attempted to chromatograph cord factor (6,6′-dimycoloyltrehalose) on this adsorbent, following a very effective adsorption of contaminant mycolic acids (Goren and Brokl, 1974).

Despite the significant acyl migrations which accompany the DEAE-catalyzed solvolysis of simple trehalose 2- and 3-palmitates, the results of the gentle acid-catalyzed methanolysis of SL-I-CF (conditions wherein the simple trehalose palmitates did not rearrange) confirmed and complemented the results of the DEAE-cellulose catalyzed degradations. The latter results were valid then, probably owing to the greater stability of the 6 substituents in the alkaline methanolysis as compared with that of the secondary acyl functions: whereas 2- and 3-palmitoyltrehaloses are completely solvolyzed in a few hours, the 6-palmitate requires more than 24 h. The stability of the 6 substituent (hydroxyphthioceranate) in the *monoacylated* glucose moiety of SL-I-CF further supports this notion—else it seems unlikely that the symmetrical diacylated SL-IV-CF product would have survived in such (relative) abundance. Moreover if (as in the tri-substituted glucose) the 6 position is already acylated and relatively stable, then a secondary substituent cannot migrate there. We are therefore content that the symmetrical structure assigned to SL-IV-CF derived in the DEAE-cellulose catalyzed solvolysis is secure.

In the gentle acid-catalyzed solvolysis, on the other hand, the 6 substituents are more readily attacked so that an entirely different diacyltrehalose is obtained. Mass spectrometric analysis of this SL-IV′-CF provides the most compelling evidence for the placement at the 2 position of the palmitate–stearate component—a disposition which from this evidence appears to be substantially invariant. It indicates further that the 3 position is principally occupied by phthioceranate, although in some SL-I the branched acid at this position is hydroxylated. This complements the earlier NMR- and GLC–MS-derived observations that some *unhydroxylated* phthioceranate may be found in a 6 position, presumably of the originally triacylated glucose moiety. Accordingly, beyond the earlier recognized structural differences arising out of homology within *all* of the acyl substituents (Goren et al., 1971), the biological acylations themselves which lead to the SL-I family are by no means entirely specific even with respect to type. This may be esthetically unsatisfying; but it is nevertheless consistent with the structural status of a plethora of other kindred substances (see Goren, 1972): the mycolic acids, cord factors, phthienoic acids, mycocerosic acids, and the various mycosides—elaborated in such profusion by the mycobacteria.

Dedication

This communication is dedicated to the memory of a beloved friend and colleague, Werner B. Schaefer, M.D., a co-discoverer of the mycobacterial sulfolipids.

References

Adam, A., Senn, M., Vilkas, E., and Lederer, E. (1967), *Eur. J. Biochem. 2,* 460.

DeJongh, D. C., and Biemann, K. (1963), *J. Am. Chem. Soc. 85,* 2289.

Dittmer, J. C., and Wells, M. A. (1969), *Methods Enzymol. 14,* 491.

Gastambide-Odier, M. J., Delaumeny, J.-M., and Lederer, E. (1963), *Biochim. Biophys. Acta 70,* 670.

Goren, M. B. (1970a), *Biochim. Biophys. Acta 210,* 116.

Goren, M. B. (1970b), *Biochim. Biophys. Acta 210,* 127.

Goren, M. B. (1971), *Lipids 6,* 40.

Goren, M. B. (1972), *Bacteriol. Rev. 36,* 33.

Goren, M. B., and Brokl, O. (1974), *Recent Results Cancer Res. 47,* 251.

Goren, M. B., Brokl, O., Das, B. C., and Lederer, E. (1971), *Biochemistry 10,* 72.

Goren, M. B., Brokl, O., and Schaefer, W. B. (1974), *Infect. Immun. 9,* 142.

Karlsson, K.-A., Samuelson, B. E., and Steen, G. O. (1969), *Biochem. Biophys. Res. Commun. 37,* 22.

Kato, M., and Goren, M. B. (1974), *Infect. Immun. 10,* 733.

Kochetkov, N. K., and Chizhov, O. S. (1966), *Adv. Carbohydr. Chem. 21,* 39.

Middlebrook, G., Coleman, C. M., and Schaefer, W. B.

(1959), *Proc. Natl. Acad. Sci. U.S.A. 45,* 1801.
Neeman, M., Caserio, M., Roberts, J. D., and Johnson, W. S. (1959), *Tetrahedron 6,* 36.
Noll, H., and Bloch, H. (1955), *J. Biol. Chem. 214,* 251.
Polgar, N., and Robinson, R. (1951), *Chem. Ind. (London),* 685.
Toubiana, R., Toubiana, M.-J., Das, B. C., and Richardson, A. C. (1973), *Biochimie 55,* 569.

Methylation of Adenosine in Strongly Alkaline Medium: Preparation and Properties of *O'*-Methyl Derivatives of Adenosine and N^6-Methyladenosine†

Zygmunt Kazimierczuk, Edward Darżynkiewicz, and David Shugar*

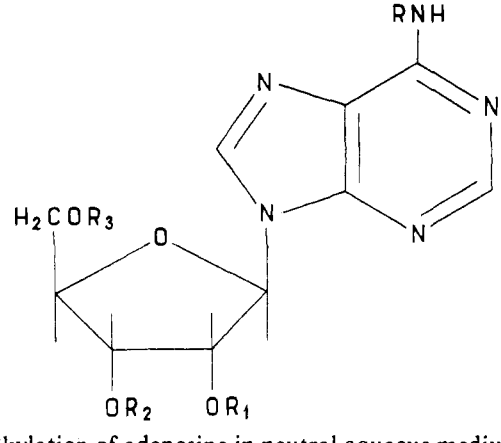
ABSTRACT: In strongly alkaline aqueous medium, 9-substituted adenines, including adenine nucleosides, are relatively resistant to alkylation of the ring nitrogens and the exocyclic amino group. This fact was utilized to obtain the various possible *O'*-methyl derivatives of adenosine by dimethyl sulfate treatment of the latter in alkaline medium, followed by separation of the products on a Dowex OH^- column. In strongly alkaline aqueous dimethyl sulfoxide, several derivatives additionally methylated on the amino group were obtained, including N^6,$O^{2'}$-dimethyladenosine, a component located recently at the 5' terminus of animal cell and viral mRNAs. The latter was also prepared by diazomethane methylation of N^6-methyladenosine in the presence of $SnCl_2$. Alkylation in alkaline medium possesses the advantage that the products are not limited to those involving etherification of cis hydroxyls and is applicable to adenine nucleosides with sugar components other than ribose. The properties of the various *O'*-methyl derivatives, including proton magnetic resonance data, are presented in detail.

A number of procedures have been described for the direct alkylation (etherification) of the cis 2'- and 3'-hydroxyls of ribonucleosides (Martin et al., 1968; Gin and Dekker, 1968). The most effective of these involves catalysis by a Lewis acid, such as $SnCl_2$, of the reaction of a diazoalkane with the nucleoside which, in the case of diazomethane, gives essentially quantitative yields of the 2'-*O*-methyl and 3'-*O*-methyl nucleosides (Robins et al., 1974), the former of which are natural constituents of tRNA and rRNA (Hall, 1971) and mRNA (Perry and Kelly, 1974; Wei et al., 1975; Both et al., 1975).

The foregoing methods are, however, not applicable to nucleosides other than those with cis hydroxyls, or to the preparation of 5'-*O*-alkyl nucleosides, or of the di-*O'*- or tri-*O'*-alkyl derivatives. It was shown elsewhere that the relative resistance to alkylation of the ring N_3 of 1-substituted cytosines in strongly alkaline medium may be utilized to obtain all the possible *O'*-alkyl (methyl and ethyl) derivatives of a cytosine nucleoside (Kuśmierek et al., 1973), including the therapeutically important 1-β-D-arabinofuranosylcytosine (Darżynkiewicz and Shugar, 1974). The reaction is also applicable to cytosine nucleoside 5'-phosphates, and the mechanism involved has been discussed in some detail (Kuśmierek and Shugar, 1973).

In an extension of the foregoing, and in part with a view to the development of a suitable route to the synthesis of the *O'*-alkyl analogues of the important antimetabolite and chemotherapeutic agent *ara*-A[1] (9-β-D-arabinofuranosyladenine), we have now found that the ring nitrogens of adenosine are equally, if not more, resistant to alkylation in strongly basic *aqueous* medium. We have utilized this observation to prepare the various *O'*-methyl derivatives of adenosine in reasonably good yields, and to show that it may be extended to the O'-alkylation of adenine nucleosides in general.

DIAGRAM I: Schematic Representation of the Various *O'*-Methyl and N^6-Methyl Derivatives of Adenosine (e.g., If R = CH_3, R_1 = H, R_2 = CH_3, R_3 = CH_3, the Compound Is N^6,3',5'-Me_3Ado).



Alkylation of adenosine in neutral aqueous medium, or in

† From the Department of Biophysics, Institute of Experimental Physics, University of Warsaw (Z.K.), 02-089 Warszawa, Poland, the Department of Cell Biology and Experimental Therapeutics, Institute of Oncology (E.D.), 02-061 Warszawa, Poland, and the Institute of Biochemistry and Biophysics, Academy of Sciences (D.S.), 02-532 Warszawa, Poland. *Received December 31, 1975.* Supported by a grant from the Polish Academy of Sciences (09.1.3), with partial support from The Wellcome Trust and the World Health Organization.

[1] Abbreviations employed: Ado, adenosine; 2'-MeAdo, 2'-*O*-methyladenosine; N^6,3',5'-Me_3Ado, N^6-methyl-3',5'-di-(*O*-methyladenosine); and similar connotations for other N^6 and *O'*-methyl derivatives (see Diagram I); *ara*-C, 1-β-D-arabinofuranosylcytosine; *ara*-A, 9-β-D-arabinofuranosyladenine; DSS, sodium 4,4-dimethyl-4-silapeptanesulfonate; uv, ultraviolet; TLC, thin-layer chromatography.