A PASTEUR PROBE

A Proposed Experiment

Prepared by Joshua Lederberg

Technical Report No. IRL-1016 March 10, 1965

Prepared under

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Grant NsG 81-60

"Cytochemical Studies of Planetary Microorganisms
Explorations in Exobiology"

Principal Investigator: J. Lederberg

Program Director: E. Levinthal

Instrumentation Research Laboratory, Department of Genetics

STANFORD UNIVERSITY SCHOOL OF MEDICINE

Palo Alto, California

The significance of optical activity for the recognition of life is too well known to require further amplification. As documented in the attached paper by Halpern and Westley, a method is available whereby important metabolites like amino acids can be scanned for optical activity with very high sensitivity, the detection of 100 nanograms being rather easily accomplished (not unreasonably sensitivities down to 1 nanogram should be achievable within the general state of the art). This method depends on the coupling of an optically active reagent, such as L-N-trifluoroacetyl-prolyl chloride, to the amino acid ester. If the amino acid is racemic, two diastereoisomers, the L-D and the L-L dipeptides will be formed, and these often prove to be readily resolvable by gas chromatography. The reactions involved are quite straightforward, run smoothly and quantitatively, and can be automated quite readily. The same approach should be easily generalized to other optically active species, organic acids generally, as well as alcohols and amines, and is being explored accordingly, especially for applications to carbohydrates. Besides their abundance and multifarious functions in the cell, carbohydrates have the advantage that methods are available whereby they can be degraded to a unique asymmetric compound. This would make it possible to test the whole genus of carbohydrates for net optical activity without needing to specify exactly which sugar is in question.

The work with amino acids does point to this limitation, namely that if a wide variety of organic molecules are present in the sample, the gas chromatograph would not be easily interpretable, since any two peaks might be related diastereoisomers, or totally unrelated molecules. This difficulty could be circumvented in principle in several ways:

a. a two-stage separation, the first without the introduction of an optically active probe - for example, trifluoroacetylation; then each fraction would be tested by resolvability, using D and L reagents separately and together prior to the second stage. This is clumsy and may be difficult to implement without racemization.

b. a single stage reaction run in parallel with D- and L- reagents. This may suffer from calibration problems circumvented by (c).

c. a single stage reaction with ratio-detection of D- and L-complexes. For example, suppose we prepared the enantiomeric reagents with differential labels, for sake of argument say levo-tritium and dextro-C^{14}. After coupling

to the mixed target material, the product is then chromatographed. For each symmetrical target molecule, the $^{3}H/^{14}C$ ratio will remain uniform in a single peak. However, if an optically active asymmetric molecule is encountered, and gives rise to resolvable diastereoisomers, the tritium will be concentrated in one peak, the C^{14} in another; that is to say there will be a swing in the ratio of the labels to one another. If the target molecule is racemic, two peaks will also be formed (one containing L-D plus D-L; the other L-L and D-D) but the label ratio in each peak will remain constant. Therefore, even when a wide variety of substances may be present in the sample, the ratio recording of the chromatograph output will be influenced only by optically active species. Overlapping peaks will interfere only insofar as they attenuate the shift in ratio by diluting the difference in label.

This approach therefore requires the fewest assumptions about the specific molecules being sought; naturally, there are technical considerations on the choice of a variety of reagents and columns best suited for different classes of substances.

Tritium and C^{14} were mentioned as differential labels only for purposes of illustration, though they might well serve for certain purposes. With halogenated reagents, only one radioactive isotope may be needed, electron capture and other methods giving excellent detection of total material. Other ultrasensitive techniques, such as neutron capture methods, alpha-particle backscattering, and so on, also suggest themselves. The greatest utility might be found from mass-spectrometric detection, e.g., with O^{18} labelling of the trifluoracetyl group and thermal cracking of the chromatograph effluent; with careful choice of materials, monitoring the m/m+2 ratio would give a very fast, highly sensitive recording for optical activity; the same instrument without cracking could give the full mass spectrum of just the interesting fraction, i.e., data from which to deduce the chemical nature of the optically active species.

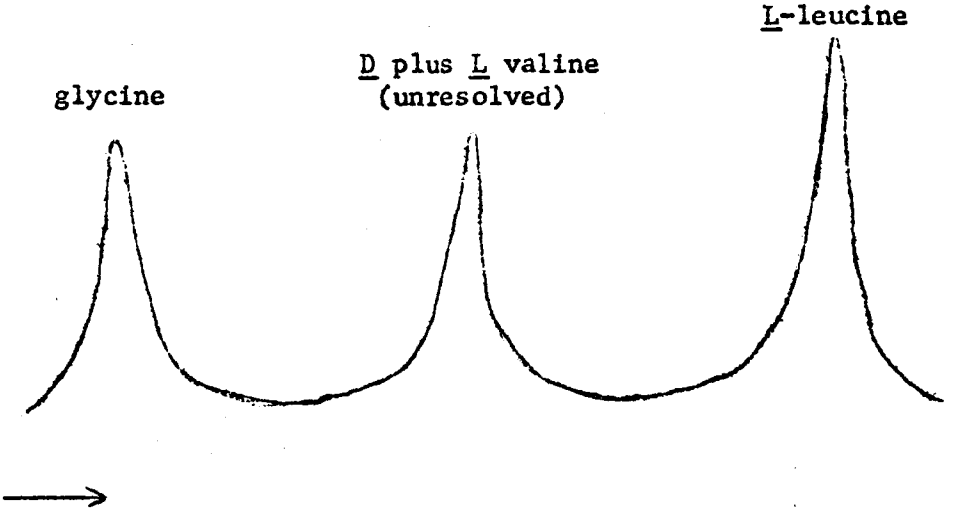
The work already done could be incorporated directly into a useful life-detecting experiment, namely for the properties of amino acids partly separated by another sub-system. We propose to continue our investigations on refinements along the lines indicated that would yield a system giving a general approach to the detection and identification of minute amounts of optically active materials.

Samples

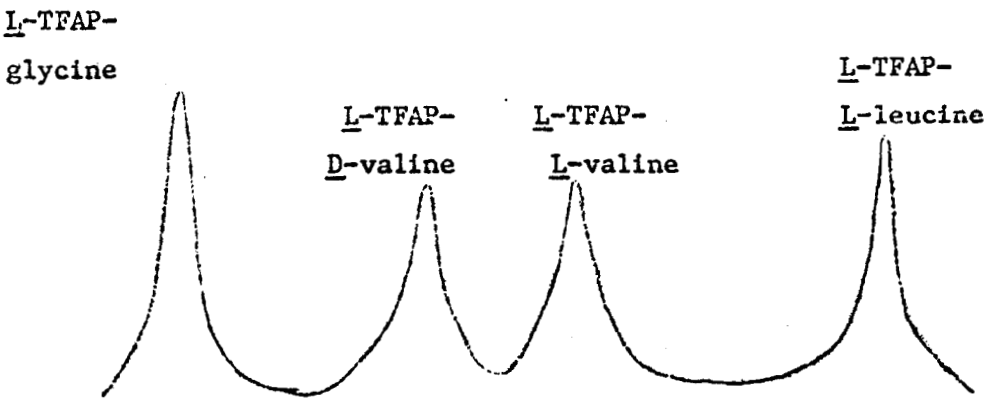
Our present thinking encompasses the utility of soil samples of the order of 1 to 10 grams, the collection of which would involve a subsystem commensurate with the complexity of the analyzer (i.e., gas chromatograph, mass spectrometer and thereon). The reagents and their handling should be manageable within a kilogram and a power requirement very small compared to the analyzers. There should be no problems of sterilization of these reagents. If means of scaling down these analyzers are found, samples of the order of 1 - 10 mg might plausibly be expected to yield an interesting result.

Collection of volatiles from the atmosphere also deserves consideration, perhaps with the help of morning dewfalls. It should be pointed out that on the permafrost model, most of the "volatile" material of Mars will be distilled or leached out and fossilized at some depth beneath the surface in equatorial regions. Subarctic zones roughly at the times of waves of darkening have the best chance of surface exposure, but digging at the equator should give a similar result - as would shaded crevices. Microorganisms might be expected to be most abundant between the surface and the permafrost and especially in the vicinity of such crevices.

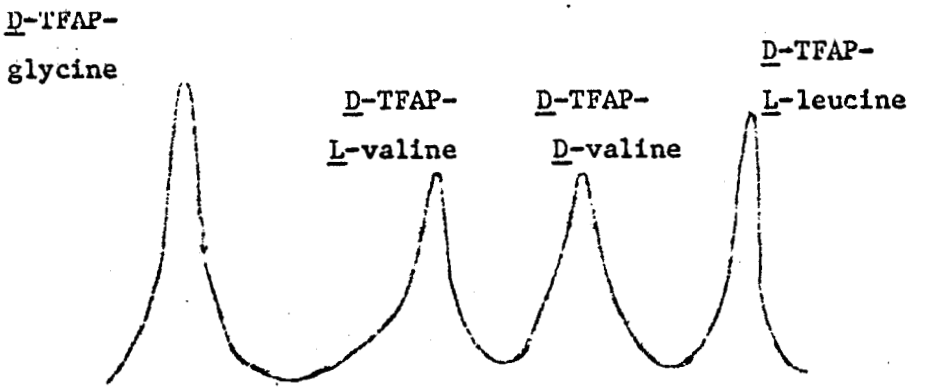
Consider a mixture of glycine (symmetric), DL-valine (racemic) and L-leucine (optically active) in equal amounts of each isomer. The following chromatograms would be realized. (TFA stands for trifluoracetyl; TFAP stands for trifluoracetyl-prolyl.)



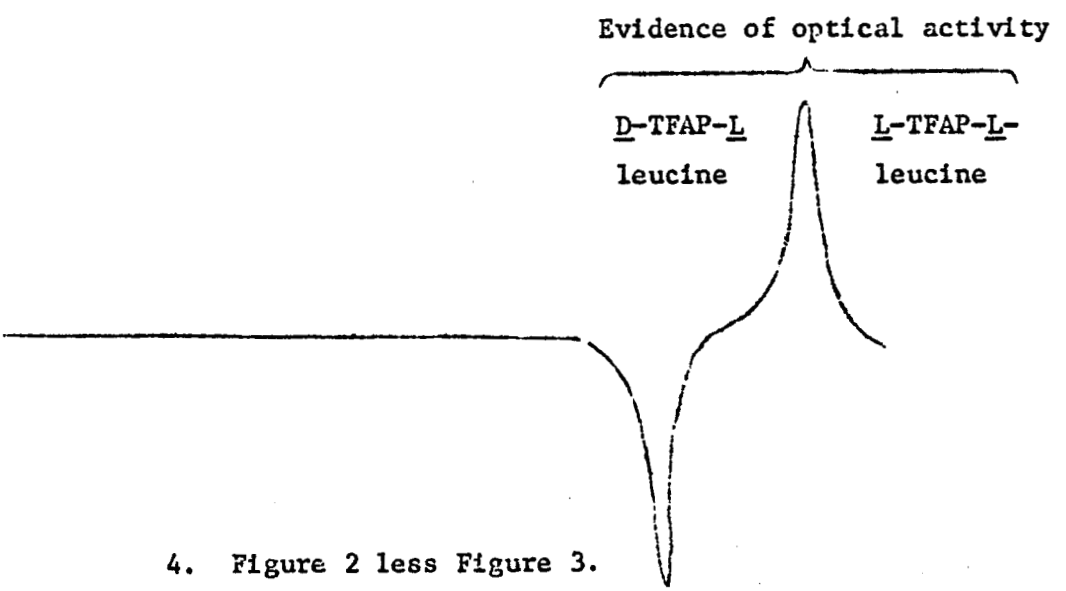
1. N-TFA amino acids



2. L-TFAP peptides



3. D-TFAP peptides



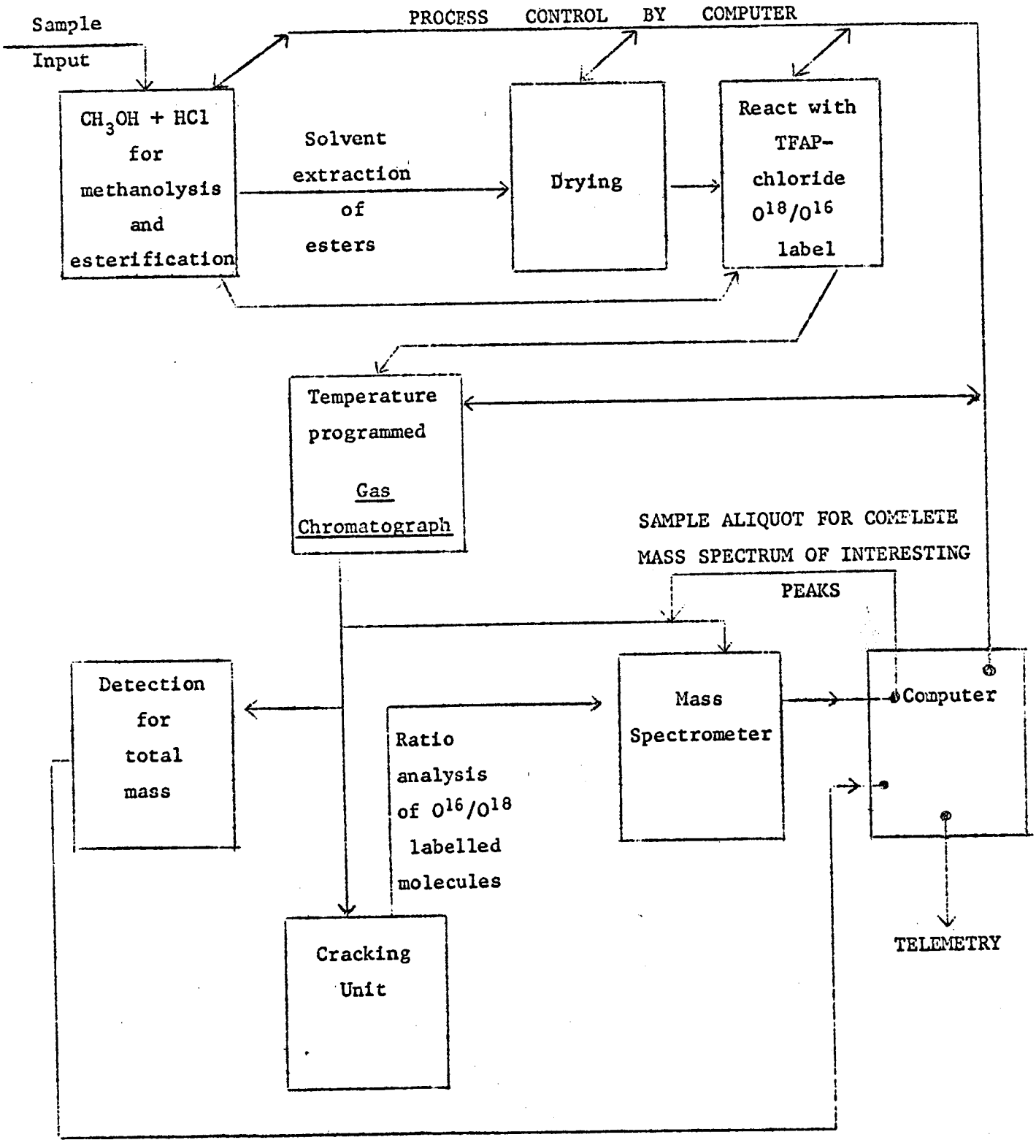
4. Figure 2 less Figure 3.

Figure 4 could be obtained either by (b) running parallel columns precisely calibrated and controlled with respect to the reaction with TFAP or (c) differential labelling of the TFAP used in a single reaction.

FLOW DIAGRAM

A Complete Subsystem for the Automated Biological Laboratory.

This would have the most flexibility, sensitivity and richness of information.



Report to the National Aeronautics and Space Administration

"Cytochemical Studies of Planetary Microorganisms - Explorations in Exobiology"

NsG 81-60

Status Report Covering Period April 1, 1966 to October 1, 1966

Instrumentation Research Laboratory, Department of Genetics

Stanford University, School of Medicine

Palo Alto, California

Joshua Lederberg

Principal Investigator

Elliott C. Levinthal

Director

Instrumentation Research Laboratory

A. INTRODUCTION

This Status Report covers the activities of the Instrumentation Research Laboratory from April 1, 1966, to October 1, 1966. Major technical efforts are described in separate technical reports and papers. The status report refers to these and summarizes continuing projects.

We are now located in the new laboratory facilities provided by NASA Grant NsG-(F)-2. Work under grant NsG 81-60 includes areas of research that are closely related to efforts being carried out in the Department of Genetics under other grants or contracts. This includes Air Force Contract AF 49(638)1599 for "Molecular Biology Applications of Mass Spectroscopy," National Institute of Neurological Diseases and Blindness Grant NB-04270 entitled "Molecular Neurobiology" and for work carried out by the Advanced Computer for Medical Research (ACME) program supported by the National Institute of Health, Division of Research Facilities and Resources under Grant FR00311-01. There is collaboration with the work in the Computer Science Department on artificial intelligence carried out under support of the Advanced Research Projects Agency SD 183. In addition, work is being done on "Genetic Studies of Mammalian Cells," National Institute of Health under Grant CA04681-08. The relationship of the work carried out under NASA grant to these other activities continues to prove of great mutual benefit in all cases.

The general project areas of the Resume are:

I. Fluorometry
II. Gas Chromatograph and Optical Resolution
III. Mass Spectrometry
IV. Computer Managed Instrumentation
V. UV Microspectrometry

During the six month period described above, two papers were presented to the American Astronautical Society and three papers submitted to journals for publication. A listing of these reports is included in this status report. Information covering personnel changes is presented.

B. PROGRAM RESUME

I. Fluorometry

a. Detection and Identification of Bacteria Using a Fluorometric Aminopeptidase Assay

In the previous status report, it was shown that amino acid β-naphthylamides (BNA) could be used as fluorogenic substrates in a very sensitive assay for aminopeptidase activity in soil. This work has now been extended to an investigation of the relative specificity of a number of pure strains of bacteria towards the eighteen different amino acid BNAs.

In the initial series of experiments, bacteria, kindly provided by Dr. E. M. Lederberg, were grown for a number of days on Nutrient Spizizen Agar plates. Bacterial colonies were then collected from the plates and suspended in Spizizen solution. Eighteen equal aliquots (40λ) of this suspension were incubated with each of the substrate solutions (2.0 ml. 10^{-4} M. amino acid BNA in 0.1 M. tris acetate, pH 8) for 4 hours at 37°C. Enzymatic activity was measured by observing the fluorescence of the released β-naphthylamine using an Aminco-Bowman fluoromicrophotometer setup with a Corning 7-37 as primary filter and a Wratten 47B as secondary filter. The results are summarized in Table 1 with the BNAs listed in order of attack with the relative hydrolysis expressed as a percentage of the attack on the primary amino acid BNA. The bacteria have been grouped according to their primary substrate.

Table 1. Aminopeptidase specificity of bacteria grown on Nutrient Agar Spizizen (NSA) plates

Species	Culture	Order of attack (% primary amino acid BNA) 1	2	3	4	5
B. subtilis	SB 19	Gly	Arg (37)	Hypro (36)	Ala (27)	Lys (7)
B. subtilis	SB 763	Gly	Arg (32)	Lys (11)	Phe (7)	
B. subtilis, W23	SB 623	Gly	Arg (48)	Hypro (27)	Pro (21)	Hist (18)
B. subtilis	SB 719	Gly	Arg (94)	Leu (94)	Phe (44)	
B. subtilis	SB 720	Gly	Arg (72)	Leu (62)	Val (32)	Asp (32)
B. subtilis var. niger	SB 718	Gly	Arg (68)	Lys (36)	Phe (27)	Leu (27)
B. aterrismus	UW353 M8	Gly	Arg (31)	Lys (8)	Phe (7)	
B. licheniformis	SB 623	Gly	Lys (60)	Phe (47)		
B. megatherium	SB 710	Arg	Phe (71)	Met (71)	Leu (58)	Lys (23)
B. laterosporus	SB 701	Arg	Phe (62)	Leu (44)	Met (16)	Lys (16)
B. rotans	SB 724	Arg	Leu (66)	Phe (20)	Lys (20)	
B. coagulans	SB 722	Arg	Phe (52)	Tyr (32)	Typ (26)	Leu (21)
B. circulans	SB 705	Arg	Hist(50)	Pro (25)	Lys (12)	Hypro(12)
B. lentus	SB 708	Arg	Lys (19)	Phe (17)	Leu (10)	Met (6)
B. globigii	SB 512	Arg	Gly (97)	Phe (35)	Leu (27)	Lys (26)
B. cereus var. terminalis	SB 514	Leu	Isoleu (8)	Met (5)		
B. cereus	SB 521	Leu	Met (10)	Phe (6)	Lys (4)	
B. polymyxa	SB 515	Leu	Met (4)	Arg (3)	Lys (2)	
B. mycoides	SB 511	Leu	Isoleu (1)			
E. coli K12	W1177	Ala	Leu (34)	Lys (13)	Arg (9)	Met (7)
E. coli K12	W3693	Ala	Lys (16)	Arg (13)	Gly (9)	Leu (8)
B. pumilis	SB 715	Phe	Arg (20)	Lys (12)	Gly (10)	Leu (4)
B. brevis	SB 900	Pro	Arg (9)	Gly (5)	Phe (5)	Asp (4)
B. sphaericus	SB 717	Glu	Ala (50)	Leu (45)	Pro (45)	Val (4)

In order to test whether the differences in specificities among the bacteria were due to differences in permeability, a number of bacteria were lysed by sonication, and after spinning down any intact cells, were assayed as before. There were negligible differences in the pattern of attack between the cell-free system and the intact bacteria. The variety of attack exhibited by the bacteria appeared to be due to the balance and specificity of a number of aminopeptidases in the cell. To examine the effect of growth medium on the balance of these enzymes a second set of experiments were carried out in which the bacteria were grown on minimal plates containing just Spizizens, glucose and agar. Not all strains grew under these conditions, but thirteen strains grew sufficiently to be assayed as described previously and the results are summarized in Table 2. In this second series of assays, only the eight most commonly attacked substrates were tested in order to simplify the procedure. The eight substrates were the BNAs of glycine, alanine, proline, leucine, methionine, phenylalanine, arginine, and lysine. In order to demonstrate the different effect that the change of growth media had on the bacteria, they are listed in the same groups as in Table 1 with the primary substrate after growth on NSA plates in parentheses.

By comparing the results of these two series of experiments it can be seen that the aminopeptidase specificity of bacteria is affected to different extents by the growth media. E. coli and B. coagulans, B. globigii, B. cereus and B. polymyxa attacked the same primary substrates, but the remaining strains had completely different patterns of attack when grown on the two different media. This suggested using two growth media as an additional tool for identification. In order to investigate the causes for the change in specificity with change of growth media, B. mycoides was examined in detail. When grown on NSA medium, B. mycoides exhibited almost exclusive attack on leucine BNA, whereas, when grown on minimal plates the bacteria was much less specific attacking almost equally lysine, phenyl-alanine, leucine and arginine BNAs. In order to test whether enzyme

Table 2. Aminopeptidase specificity of bacteria grown on Minimal plates

Species	Culture	Order of attack (% primary amino acid BNA) 1	2	3	4	5
B. subtilis (Gly)	SB 19	Arg	Lys (68)	Pro (55)	Leu (46)	Gly (30)
B. aterrimus (Gly)	UW353M8	Leu	Phe (69)	Met (55)	Pro (17)	
B. licheniformis (Gly)	SB 623	Leu	Arg (50)			
B. rotans (Arg)	SB 724	Leu	Met (90)	Ala (77)	Phe (67)	Lys (60)
B. coagulans (Arg)	SB 722	Arg	Leu (22)	Lys (16)		
B. circulans (Arg)	SB 705	Phe (100)	Arg (100)	Leu (55)	Lys (52)	Met (41)
B. globigii (Arg)	SB 512	Arg	Leu (77)	Lys (70)		
B. cereus var. terminalis (Leu)	SB 514	Leu	Phe (16)	Arg (15)	Lys (13)	Met (10)
B. cereus (Leu)	SB 521	Leu	Arg (10)	Met (7)	Phe (4)	
B. polymyxa (Leu)	SB 515	Leu	Phe (22)	Lys (16)	Arg (14)	Met (14)
B. mycoides (Leu)	SB 511	Lys (100)	Phe (100)	Leu (78)	Arg (55)	Met (38)
E. coli K12 (Ala)	W3693	Ala	Lys (26)	Leu (21)	Gly (11)	Arg (9)
B. brevis (Pro)	SB 900	Pro	Arg (12)	Gly (4)		

repression was playing a part in this change, the bacteria was grown on (1) minimal + 1 mg.% lysine, (2) minimal + 1 mg.% phenylalanine, (3) minimal + 1 mg.% arginine, and (4) minimal + 1 mg.% leucine.

It appears from these results that there are two major aminopeptidases in B. mycoides, one of which is specific for leucine BNA and a second which attacks phenylalanine, lysine, arginine and probably methionine BNAs. In experiments (1), (2) and (3), the addition of one of the three amino acids produced by the action of the second enzyme causes a repression of this enzyme during growth resulting in a high leucine aminopeptidase specificity, whereas when leucine is added to the growth medium (experiment 4), the bacteria again shows a broad pattern of activity although it should be noted that the leucine aminopeptidase activity is not repressed as severely as the second enzyme in experiments (1) through (3). Presumably in the NSA medium, all amino acids are present and therefore all the peptidases are suppressed to some extent but the leucine aminopeptidase is least affected. The remarkable specificity of the leucine aminopeptidase in this bacteria prompted us to compare it with leucine aminopeptidase from swine kidney (Worthington Biochemical Corporation). This latter enzyme was assayed in the same way as the bacteria and gave the following order of attack: methionine BNA (100%), leucine BNA (92%), tryptophan BNA (22%), and phenylalanine BNA (8%). Three other bacteria with highly specific enzymes are B. brevis (proline aminopeptidase), E. coli K12 (alanine aminopeptidase) and B. pumilis (phenylalanine aminopeptidase).

In conclusion, this technique appears to be useful both as an additional means of identifying bacterial strains and also examining the specificity and balance of different aminopeptidases within a pure strain. However, it has been clearly demonstrated that the aminopeptidase activity is very sensitive to growth medium and therefore if this is to become a reliable taxonomic procedure, bacteria must be grown for equal times on carefully specified synthetic media in order to guarantee reproducibility. With a view to using this technique for an exobiological assay, we have shown that these compounds as hydrochlorides are stable to heat sterilization (24 hours at 130^{o}).

Table 3. Effect of changes in growth media on aminopeptidase specificity of B. mycoides (SB 511)

Growth medium	Order of attack 1	2	3	4	5
NSA	Leu	Isoleu (1)			
Minimal	Phe (100)	Lys (100)	Leu (78)	Arg (55)	Met (38)
(1) Minimal + 1 mg.% lysine	Leu	Phe (14)	Met (12)	Lys (10)	Arg (5)
(2) Minimal + 1 mg.% phenylalanine	Leu	Met (11)	Phe (6)	Lys (5)	Arg (3)
(3) Minimal + 1 mg.% arginine	Leu	Phe (18)	Lys (14)	Arg (13)	Met (12)
(4) Minimal + 1 mg.% leucine	Phe	Met (93)	Lys (76)	Leu (62)	Arg (58)

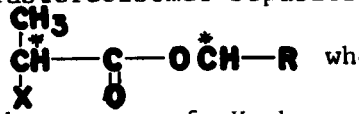
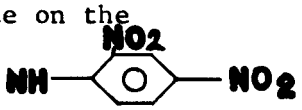
b. Nanosecond Flash Fluorometry (Phosphorimetry)

The instrumentation development has been completed. Further research using the fluorimeter is being carried out by Dr. Stryer of the Biochemistry Department under other grants.

The nanosecond fluorimeter which was developed suitable for measuring the emission kinetics of chromophores which have excited state lifetimes longer than 1 nsec is described in a paper by L. Hundley, T. Coburn, E. Garwin, and L. Stryer, "A Nanosecond Fluorimeter," Review of Scientific Instruments (to be published).

The acquisition and processing of the nanosecond data were accomplished in entirety by computers. The pulsed light source, an oxygen spark-gap lamp operated in a relaxation mode, is simple to construct and use. The lamp exhibits high intensity over a broad spectral range as well as a short pulse duration. There were 6×10^{11} photons per pulse at a repetition rate of 2.1 kHz. The spectrum of the lamp was approximately flat between 200 and 600 m. The rise and fall times of the light pulse, as measured with a photodiode, were 0.7 nsec and 1.1 nsec. The accompanying current pulse had rise and fall times shorter than 0.43 nsec. A 1P21 photomultiplier tube was used as the detector. The output of a 1P21 was connected to a sampling oscilloscope that was triggered by a synchronous current pulse from the light source. The analog output of the sampling oscilloscope was digitized, stored, and averaged on a LINC computer, and then transferred to magnetic tape for processing on an IBM 7090. The nanosecond data were analyzed in terms of a convolution integral which took into account the shape of the exciting light pulse as seen by the detection system. The use of the apparatus is illustrated by two examples: 1,4-bis-2-(4-methyl-5-phenyloxazolyl) benzene, which has a lifetime of 1.5 nsec, and 1-anilino-8-naphthalene surfonate bound to serum albumin, which has a lifetime of 17.1 nsec.

II. Gas Chromatography and Optical Resolution

In a continuance of our work on the optical resolution of amino acids by GLC, we have applied this technique to a study of the racemization in peptide synthesis[1] and to the optical analysis of cyclic amines.[2] While many useful observations on the best conditions for diastereoisomer separation have been made[3] most of our GLC analytical procedures have been obtained by "a hit or miss" approach. We are now starting an investigation aimed at elucidating the mechanism of GLC diastereoisomer separation. As a starting point we are looking at the system  where we change X. Preliminary results indicate that the nature of X has a great influence on the separation. Thus when X = Cl, Br, N = C = O, N = C = S or  no separation can be obtained. If X = NHCHO, $NHCOCF_3$ or $NHCOCH_3$, the diastereoisomers are readily resolved.

While awaiting the completion of an enriching device, the GLC (gas liquid chromatograph) was coupled to the EAI (Electronics Associates Inc.) QUAD 300 quadruple mass spectrometer in a non-enriching configuration.

The basic constraints in design of the interconnection were manifested in the requirements that a volume rate of flow equal to 4×10^{-1} cm^3/sec issue from the terminal end of the GLC volume at a pressure of approximately 1×10^{-3} Torr, that the pressure in the MS (mass spectrometer) not exceed 1×10^{-5} Torr, and that the temporal phase difference in the responses of the GLC and the MS be less than 1 second.

The material rate of flow $\dot{Q}$ through any section of the system is related to the pressure P and the volume rate of flow R by

$$\dot{Q} = PR,$$

---

[1] B. Halpern, L. Chew, and J. W. Westley, "Investigation of Racemization during Peptide Bond Formation by GLC of Diastereoisomeric t-BOC-Amino Acid Amides," (submitted to Anal. Chem.).

[2] B. L. Karger, R. L. Stern, W. Keane, B. Halpern and J. W. Westley, "GLC Separation of Diastereoisomeric Amides of Racemic Cyclic Amines," J. Anal. Chem., 39, p. 228 (1967).

[3] B. Halpern, J. W. Westley, and B. Weinstein, "Investigation of Steric Effects about the Amide Bond by Gas-Liquid Chromatography," Nature, 210, p. 837 (1966).

where, for our purposes, P will be expressed in Torr, R in cm^3/sec, and consequently $\dot{Q}$ in Torr cm^3/sec. The temperature is assumed constant throughout the system. At the terminal end of the GLC column, then we have 4×10^2 Torr cm^3/sec. The pumping speed of the MS vacuum system is approximately 2×10^5 cm^3/sec. Thus, the MS will pump a $\dot{Q}$ of 2 Torr cm^3/sec at 1×10^{-5} Torr. We see then that only one part in 200 of the effluent from the column can be admitted to the MS. The remainder of the effluent must be rejected. Our method of rejection involved the differential pumping scheme illustrated schematically in Figure 1.

The gas flow design parameters characterizing the various lines in the system are summarized in Table 4 wherein the numbers heading the columns are keyed to Figure 1.

Table 4. Initial design parameters on which the development of the system was based

	1	2	3	4	5	6	7
R $\left[\frac{cm^3}{sec}\right]$	$4x10^{-1}$	$2x10^{-1}$	$2x10^{-1}$	$5x10^{2}$	$5x10^{2}$	5	$2x10^{5}$
P [Torr]	$1x10^{3}$	$1x10^{3}$	$1x10^{3}$	$4x10^{-1}$	$4x10^{-1}$	$4x10^{-1}$	$1x10^{-5}$
$\dot{Q}$ $\left[\frac{Torr\ cm^3}{sec}\right]$	$4x10^{2}$	$2x10^{2}$	$2x10^{2}$	$2x10^{2}$	$2x10^{2}$	2	2

The tee between the GLC column and the valve V1 was constructed of 304 stainless steel sections of 0.010" ID and 0.020" ID each of not more than 10 cm in length. Gas flow transit times through the tubes were thereby kept short and roughly equal.

In order to get 5×10^2 cm^3/sec out of line No. 5, valves V1 and V2 were close coupled by less than 3 cm of 1/8" ID stainless steel tubing and line No. 5 expanded from a nominal 1 cm ID stainless steel section to a 3/4" ID feeder into the mechanical pump.

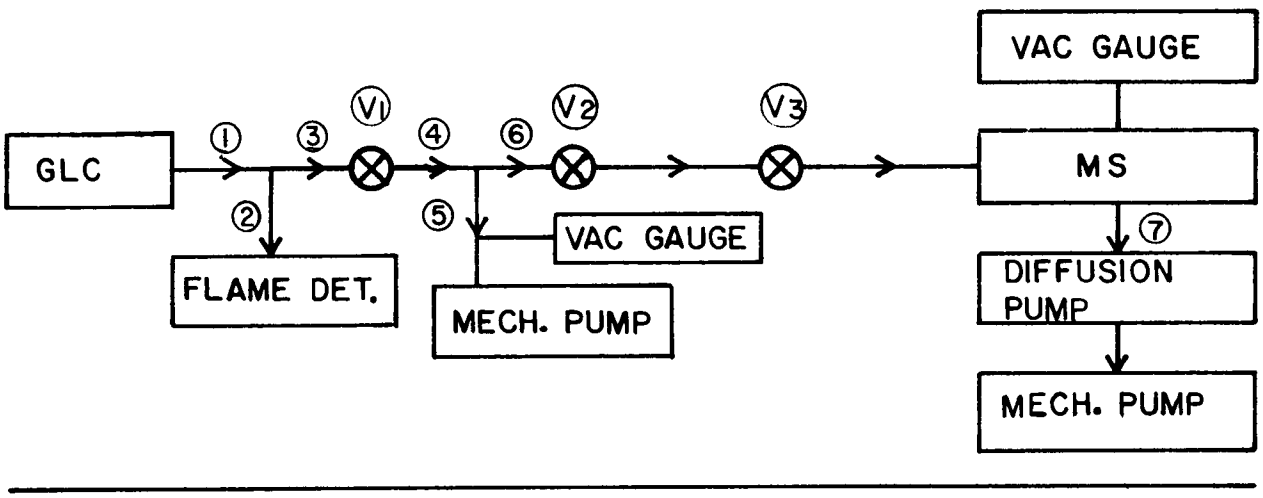


Figure 1

Schematic illustration of non-enriching coupling scheme for connecting GLC to MS.

Nupro stainless steel 25A "very fine metering valves" were employed for V1 and V2. A 1/8" ID glass-to-metal seal on the exit end of V2 and a 3/8" ID glass-to-metal seal on the inlet end of V3 were interconnected by a section of glass tubing. V3 was a 3/8" stainless steel bellows sealed Hope valve with a Viton O-Ring sealing gasket. The valve was close coupled to the inlet flange of the MS.

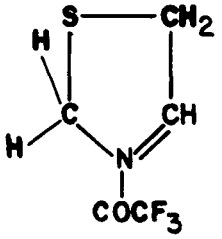
The entire section of lines and valves from the GLC to the MS was kept at a uniform temperature by an oven enclosure.

The system performed close to expectations. It was found that the 1 degree tapered needle in V2 would not withdraw sufficiently far to enable the MS pressure to rise about 3×10^{-6} Torr, and thus this valve was generally run wide open. An increase of MS pressure could be achieved by reducing GLC effluent peak mixing and interfering with the relative time responses of the GLC and the MS.

With the system described, prototype Pasteur probe experiments have been carried out and as little as 10^{-5} gm of optically active amino acids could be detected. We are now improving the valving system in order to include a Biemann enrichment device which should increase the sensitivity by two orders of magnitude.

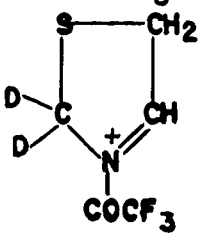
Figures 2 to 5 illustrate the functioning of the Pasteur probe experiment. Figures 2 and 4 are the GLC effluents as a function of time (GLC carried out at 185^{o}, using a 0.5% EGA column (5' x 1/8") on chromosorb W). Figures 3 and 5 result from repeated sweeps of the mass spectrometer over the range m/e = 180 to 190. The peaks of interest are m/e = 184 and 186 due to the fragments

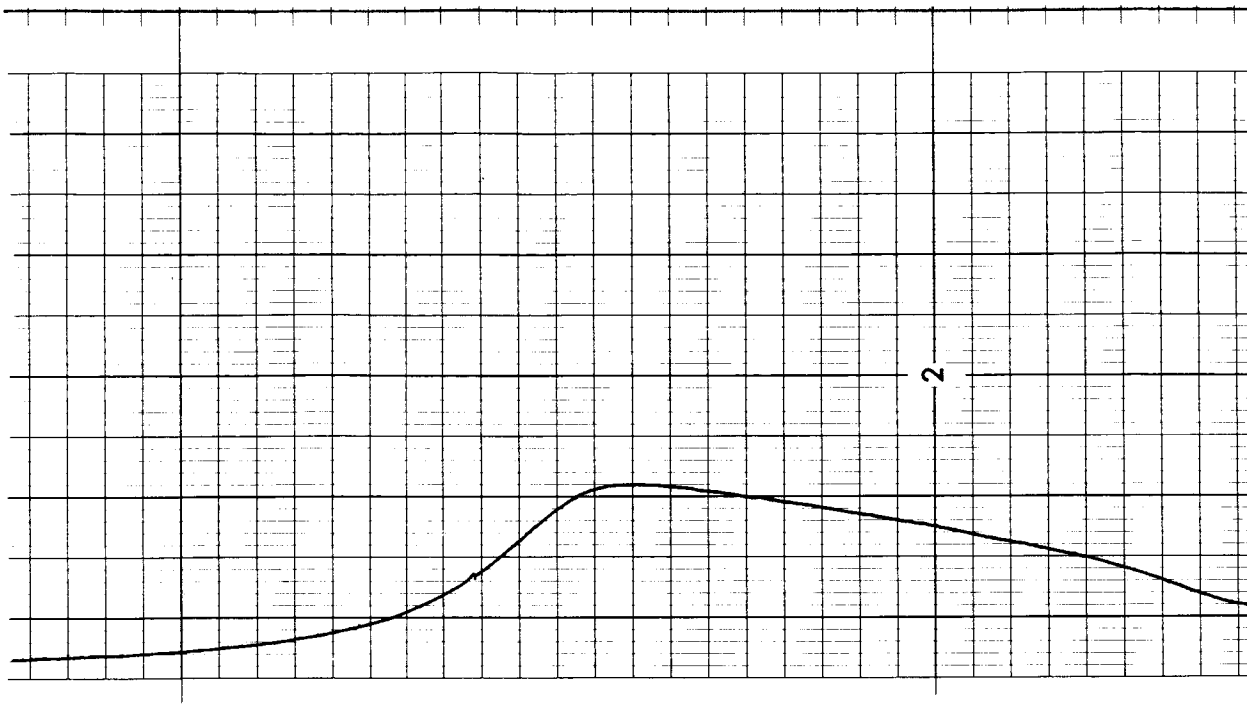
m/e = 184

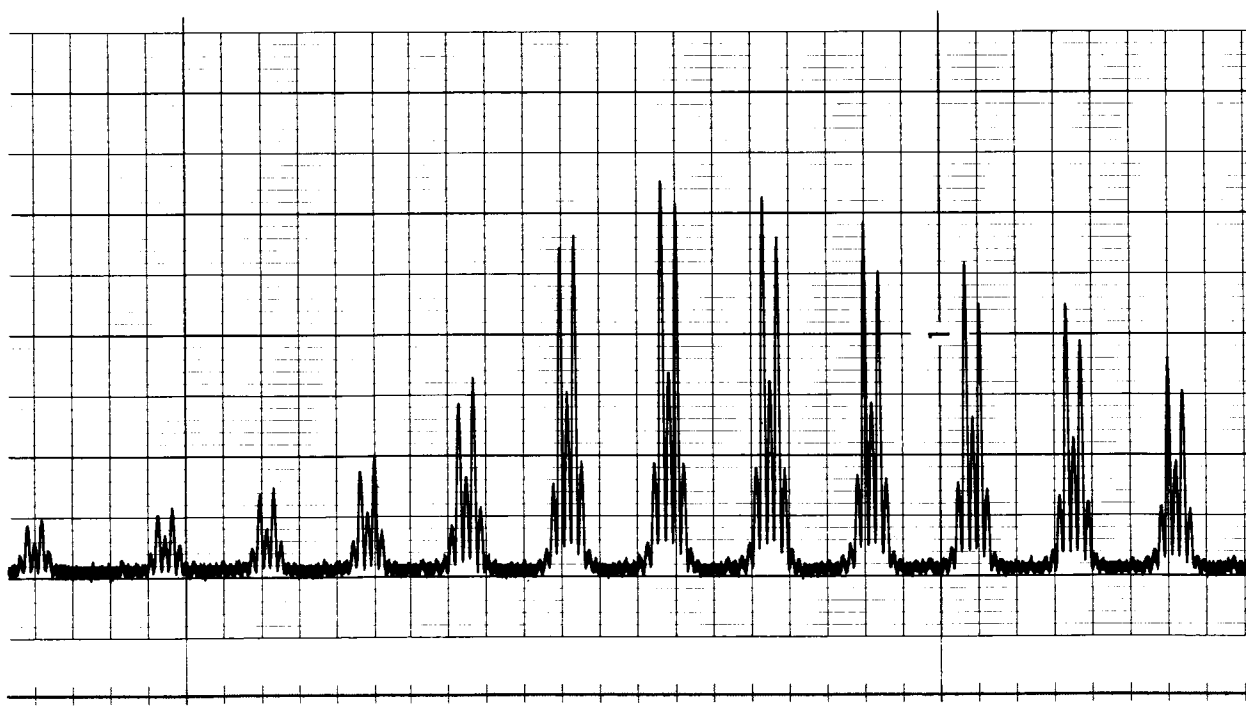


and

m/e = 186




2


1

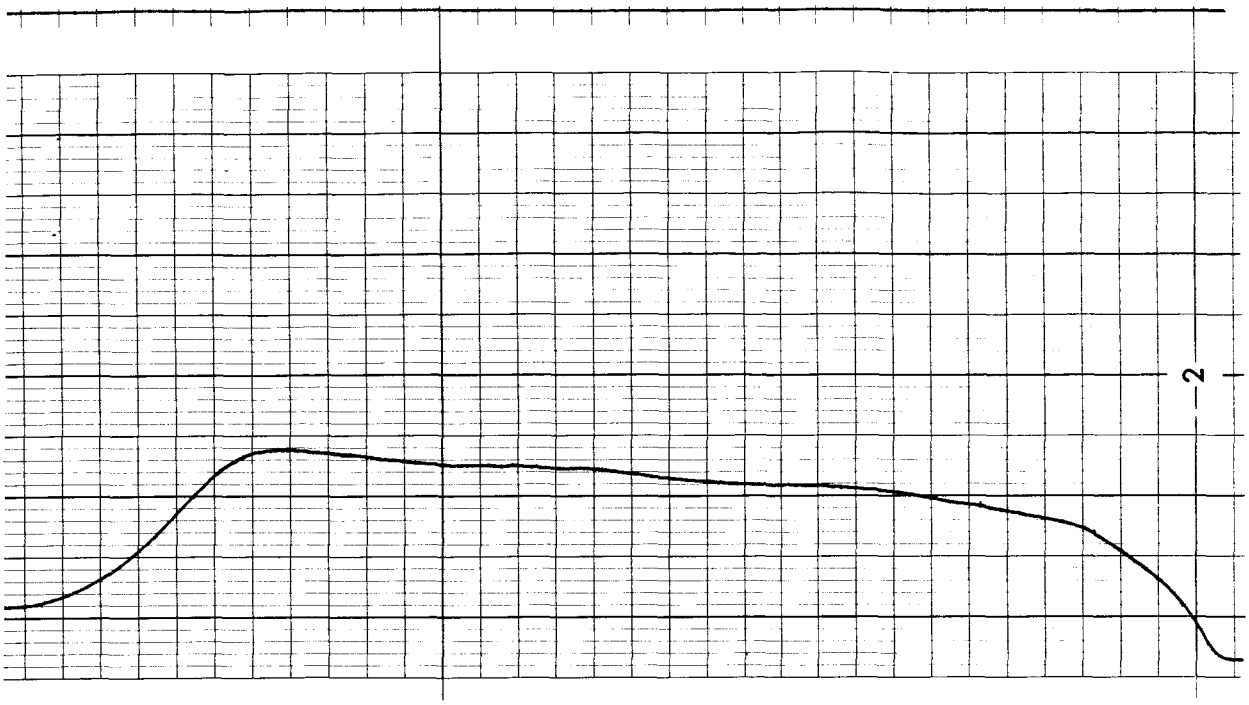


FIGURE 2

GLC of DL-alanine derivative

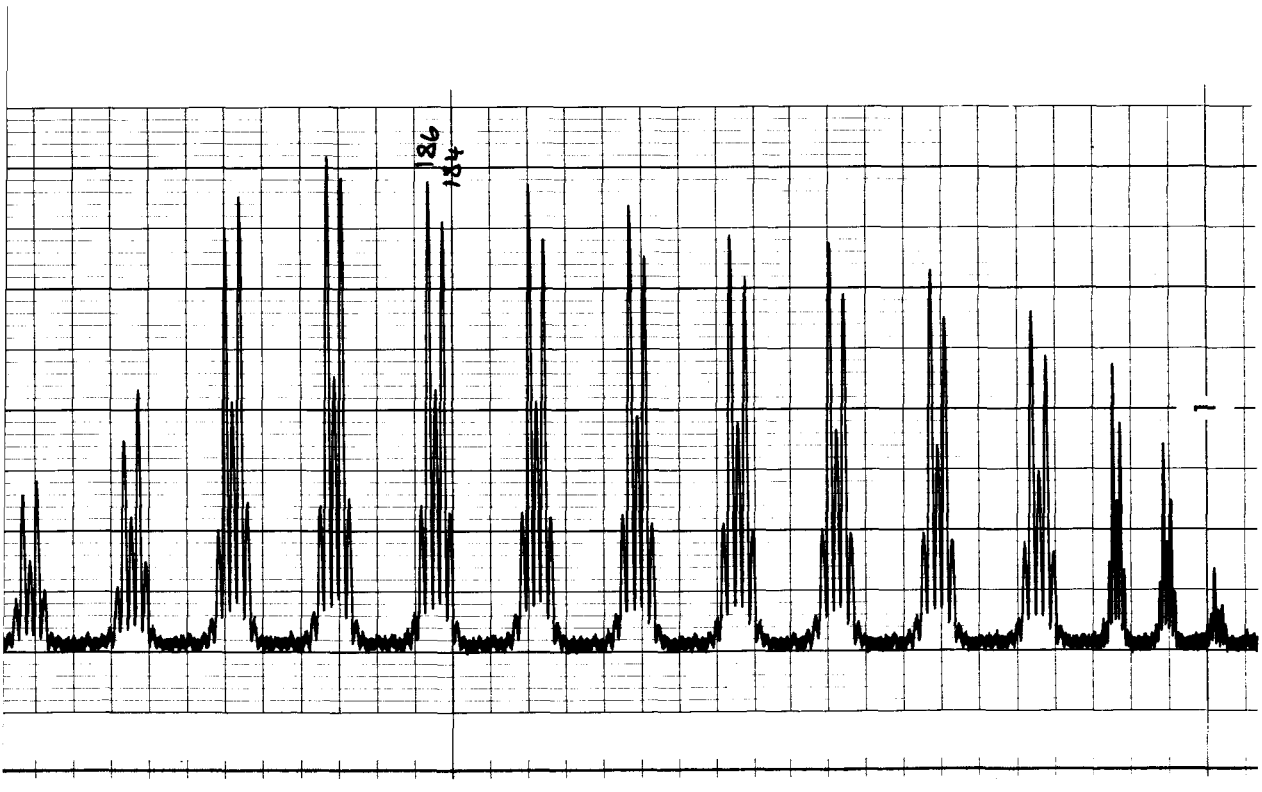
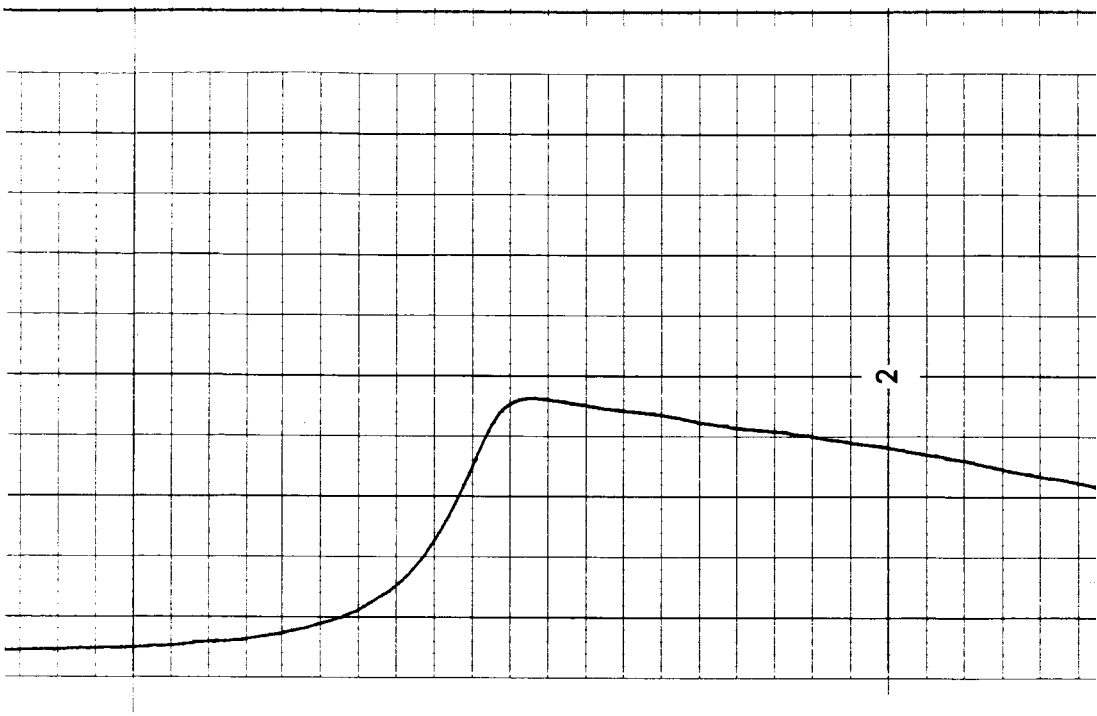


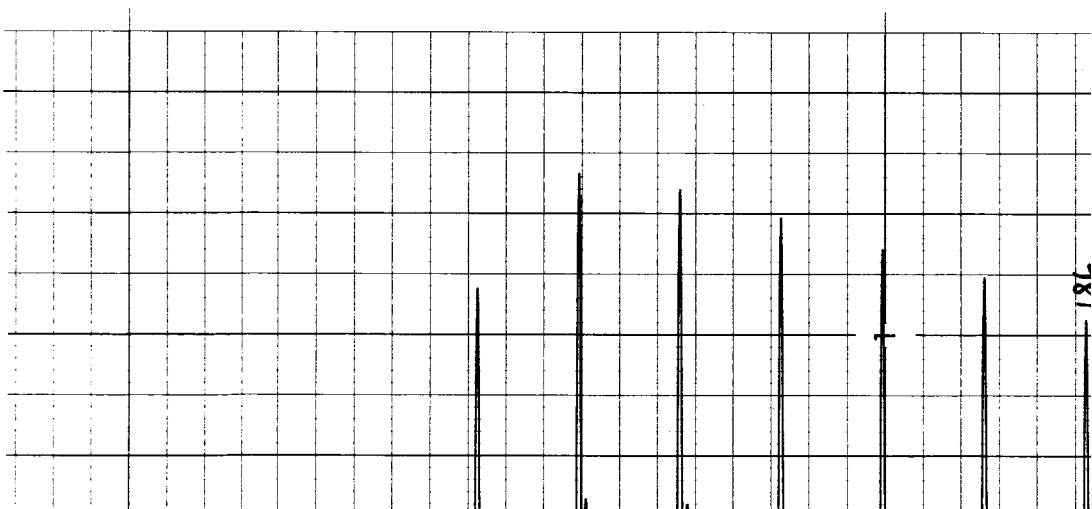
FIGURE 3

Mass spectra scan (m/e = 180 to 190)

DL-alanine derivative



GLC c



Mass s

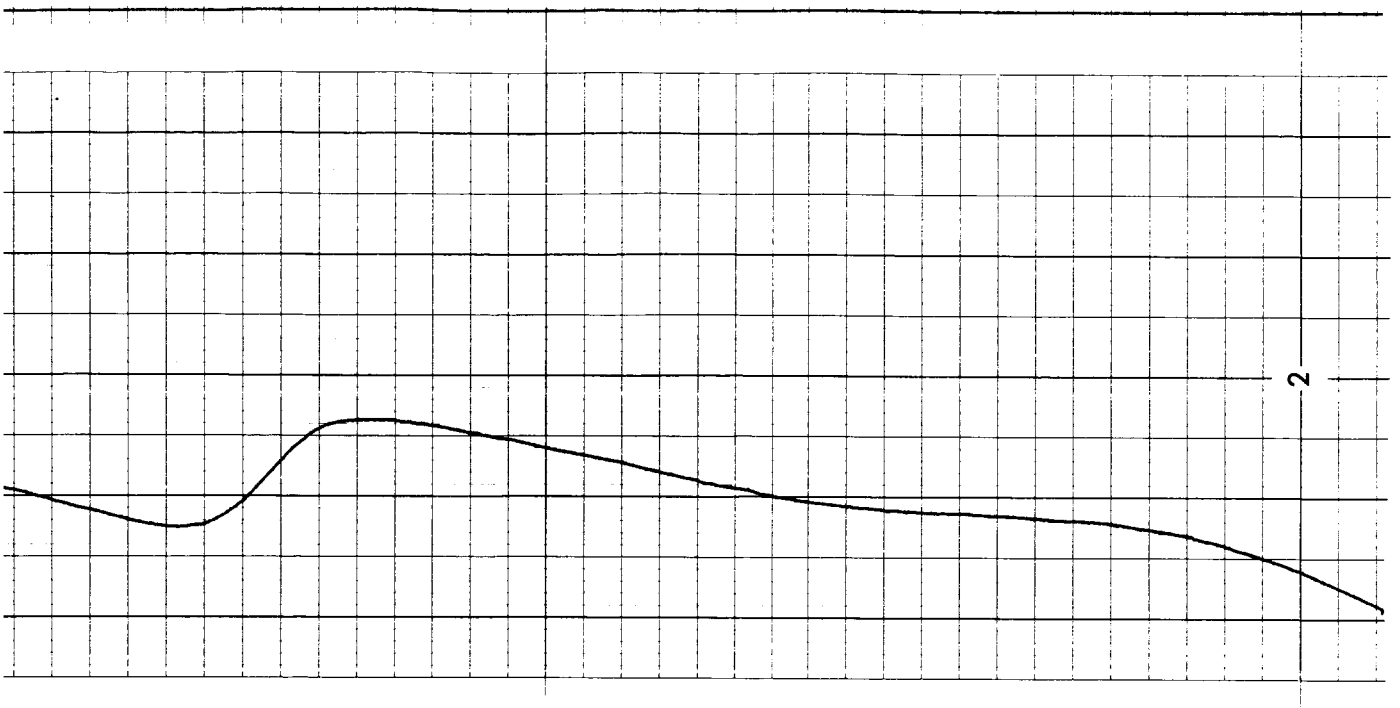


FIGURE 4

f L-alanine derivative

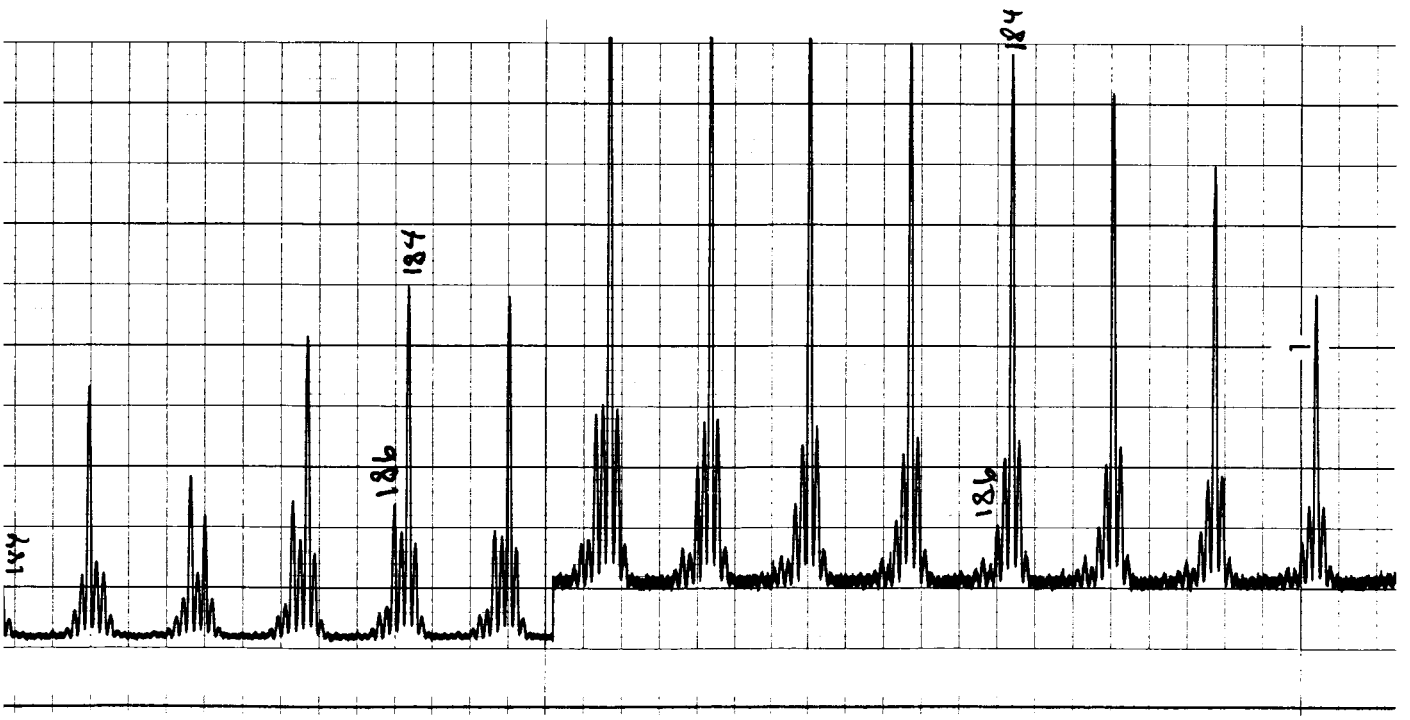


FIGURE 5

ectra scan (m/e = 180 to 190)

L-alanine derivative

In Figure 3 for a DL-aminobutyric acid derivative, the ratio $\frac{m/e\ 184}{m/e\ 186}$ remains uniform, except for differences that can be accounted for by changes in the envelope of the GLC output with time, saturation effects, and the particular sweep parameters chosen. For the optically active amino acid (Figure 5 for L-alanine derivative) we note that the deuterated fragment is concentrated in one peak giving a dramatic and readily observable change from a value large compared to unity through unity at the cross over of the two GLC peaks to a value small compared to one.

The GLC MS system (a Varian aerograph gas chromatograph coupled to an EAI QUAD 300 laboratory mass spectrometer) has also been used extensively to study the mass spectra of several sets of diastereoisomers, derived from amides, esters, and peptide esters. Our experiment shows that there are no significant differences in the mass spectra of the DL and LL diastereoisomers.

III. Mass Spectrometry

a. Analysis of Natural Products

The Atlas CH-4 Mass Spectrometer in Professor Djerassi's laboratory in the Department of Chemistry has yielded the results reported in the following papers:

Djerassi, C.; MacLeod, J. K.: Mass Spectrometry in Structural and Stereochemical Problems CIX. The Non-specificity of Hydrogen Rearrangements in Aryl Alkyl Ethers. J. Am. Chem. Soc., 88, 1840 (1966).

Djerassi, C.; Duffield, A. M.; Sample, S. D.: Mass Spectrometry in Structural and Stereochemical Problems CX. The Course of the Electron Impact-Induced Dehydrochlorination of Primary Alkyl Chlorides. Chem. Comm., 193 (1966).

Djerassi, C.; MacLeod, J. K.: Mass Spectrometry in Structural and Stereochemical Problems CVIII. Deuterium Isotope Effects in Site-Specific Mass Spectrometric Rearrangement Processes. Tetrahedron Letters, 19, 2183 (1966).

Djerassi, C.; Sample, S.: Mass Spectrometry in Structural and Stereochemical Problems CIV. The Nature of the Cyclic Transition State in Hydrogen Rearrangements of Aliphatic Sulfides. J. Am. Chem. Soc., 88, 1937 (1966).

Djerassi, C.; Brown, P.: Mass Spectrometry in Structural and Stereochemical Problems CVI. Occurrence of Alkyl and Aryl Rearrangements in the Fragmentation of Some Organic Carbonates. J. Am. Chem. Soc., 88, 2469 (1966).

Djerassi, C.; Thomson, J. B.; Brown, P.: Mass Spectrometry in Structural and Stereochemical Problems CXIV. Electron Impact-Induced Rearrangement of Thiocarbonates, Carbamates and Thiocarbamates. J. Am. Chem. Soc., 88, 4049 (1966).

b. Mass Spectral Microanalysis of Solids

A study is in progress to determine the suitability of laser induced vaporization as a mechanism for enabling spatial resolution in the mass spectral analysis of organic solids.

Initial data has just been obtained, as of the time of preparation of this report, from a system consisting of a pulsed ruby laser coupled to a Bendix TOF (Time-of-Flight) mass spectrometer. The choice of the particular laser employed for this preliminary work was determined primarily by considerations of availability rather than special suitability.

The laser is an early model procured by the laboratory for a prior requirement. The manufacturer's specifications list a peak power (a quantity not specifically defined) of 1×10^{3} watt, pulse width of 0.5×10^{-3} sec, beam divergence of 0.5×10^{-3} radian, and a laser light wavelength of 6943A.

Certain checks on the performance of the laser have been made. A rat's nest calorimeter has been constructed and employed for the determination of the energy content of the pulsed ruby output. Laser output measurements were made at ranges of 12 cm and 365 cm. The latter were free of appreciable flash tube contribution. The laser was found to have a threshold of 190 joules input and to give a maximum output of 1.0×10^{-2} joules at the maximum available input of 440 joules.

Beam divergence has been found near threshold to be in reasonable accord with the manufacturer's specification. However, at maximum output there is as yet incomplete data suggesting that the beam spread is roughly an order of magnitude higher.

The ruby laser is mounted outside the vacuum chamber. The laser is aimed upon the target by use of a He-Ne 1 milliwatt CW gas laser normally reflected off the ruby. That portion of the ruby pulse fired upon the target is reflected off an approximately 50-50 beam splitter, condensed by a 10 cm focal length lens and introduced to the vacuum environment of the mass spectrometer through a glass window.

If we assume that roughly 1/3 of the ruby radiation is incident upon the sample, then we estimate that at 3×10^{-3} joules of pulsed laser light is focused to 5×10^{-3} cm^2 at the target, corresponding to 0.6 joule/cm^2.

The Bendix TOF mass spectrometer produces a complete spectrum every 10^{-4} seconds. The present mode of data retrieval involves photographing the output displayed on an oscilloscope. We are instrumented to commence observation on any given Bendix spectrum following firing of the laser flash tube and to display any desired number of successive spectra vertically displaced from one another on the scope face.

Under conditions of maximum laser output energy and optimum optical alignment, we have produced laser induced output from a crystalline sample of N-dinitrophenyl (DNP)-L-isoleucine.

Observations have been conducted over the mass range 16 through 47 amu. Laser induced peaks are observed at 27, 29, 30, 39, 40, 41, and 42. Background peaks at 16, 17, 18, 28, and 32 would for the moment mask laser induced peaks at these masses. The peaks persist for about 0.7×10^{-3} sec after firing of the flash tube. The preliminary nature of this work is to be stressed.

c. Computer Manipulation of Chemical Hypotheses

DENDRAL is the name of a notation for chemical structures suitable for computer manipulation. We have now implemented BERZELIUS, a computer program built around the DENDRAL notation. This is written in the LISP language, and was built on the Q-32 time-sharing system of System Development Corporation at Santa Monica, California. Besides its data structures, BERZELIUS requires over 30,000 words of core memory. It would have been impossible to write such a complex program without the near real-time interaction with the computer allowed by the time-sharing system. On account of the imminent closing out of the Q-32, we must now devote a few months to translating the program to the next available computer system, a PDP-6 recently installed at Stanford.

At the present time, BERZELIUS is operating with the following features. These apply to acyclic molecules only at the present time but blocks of program for dealing with cyclics have also been demonstrated.

1. Translation of arbitrary structural representations to canonical DENDRAL.

2. Translation from connection tables and other forms to DENDRAL.

3. A generator to produce all possible unique structural isomers of a given molecular composition.

4. A filter table with specified subgraphs. Prospective structures are eliminated if they contain such unwanted functional groups, for example, .CH=CHOH (enols) or .CH..OH OH (gem-diols). The generator incorporates this filter efficiently, not merely a retrospective scan over all topologically possible structures. The filter thus embodies chemical common sense in the context of the program.

5. A cumulative dictionary of radicals representing the solution to previously solved subproblems. Thus the dictionary already 'knows' that -C2H5 has only one isomer, ethyl, while -C3H7 has two, propyl and isopropyl. This feature saves an enormous amount of repetitive effort as the program learns more chemistry.

6. A zero-order dissector. This routine cuts a given molecule in all possible ways, one bond at a time, and returns a list of the mass numbers of the radicals so obtained. This is a so called zero-order mass spectrum.

7. A zero-order mass spectrum analyzer. This accepts a list of mass numbers and a molecular composition. It returns a list of structures that would generate no radicals other than those in the spectrum. It also reports any mass numbers not accounted for by the structure given.

8. A plausibility feature in the generator departs from canonical DENDRAL sequence, rearranging these heuristically. It is already useful and is under further study to optimize the sequence in which structures (hypotheses) are generated for test.

The program has been repeatedly demonstrated and shown to outdo human competition in narrowly defined kinds of problems -- especially isomer elaboration and the detection of isomorphic representations of structures. It is beginning to simulate human inductive behavior.

IV. Computer Managed Instrumentation

a. Cooperative ACME Program

The NIH has awarded a substantial facilities grant to Dr. Lederberg on behalf of the Stanford Medical School for an Advanced Computer for Medical Research, acronymmed ACME. The principal functions of this facility will be time-shared access of remote terminals throughout the school to the computer, and the multiplexing of laboratory data from several remote locations into the system, and for further computerized management of the experiments. On the premise that this is an excellent simulation of a computer-managed automated biological laboratory for Mars, the Instrumentation Laboratory is deeply involved in this program, and is cooperating in system support for it. Other investigators remain responsible for the details of the experiments at their terminals.

During this period, a configuration has been designed consisting of an IBM 360 Model 50 computer with 64K of fast core and 1000K bytes of 8 μs 'bulk core' which will be the heart of a core-sharing multiterminal system. (The processor is time shared among the programs resident in core and is available to automatic cycle-stealing for data transmission.) In view of wide interest in the design, we include in this section of our status report a report by Mr. Wiederhold detailing the system.

The system also includes an IBM 1800 data acquisition computer to handle the laboratory data lines. In addition, studies are under way to specify special purpose hardware for an augmented data interface to the Model 50.

Delivery is anticipated in mid-November, the NIH funding having finally been scheduled to take effect beginning October 1, 1966.

A Summary of the ACME System

by

Gio Wiederhold

(Presentation given at the ONR Computer and Psychobiology Conference May 17, 1966 at the U. S. Navy Postgraduate School, Monterey, California)

The ACME computer system at the Stanford Medical School is designed to provide powerful computing to research laboratories in the Medical School. The type of computation service planned is regular batch processing (mainly at night) and real time interactive capability for on line experiments. The remainder of this description will concern itself with the proposed implementation of this latter facility.

In order to accommodate all the laboratories in the Medical School with their widely varying data rates, the system is designed to share the available computer time. The amount of time allotted to a user, however, is not one fixed unit per period, i.e., 200 ms every 10 seconds, but rather the time required by him to acquire and process one data point.

The data can be obtained in a number of different ways (see appendix CN-2, SK-2).

Data may be typed on a typewriter station. All textual and programming information is entered this way.

Slow to medium speed (up to 1000 samples/second) analog signals can be entered via a subsidiary computer which can either scan input voltages at the 1000/second rate or respond to separate interrupt signals given by the experiment or experimenter. The precision of the conversion can be up to 14 bits (.02%) precise. This computer will convert, reformat and preprocess these data and then transmit them to the main computer.

Digital input up to 16 bits wide can be handled in a manner similar to the analog input with the subsidiary computer, in order to serve users that have their own digitizing equipment or whose experiments actually generate digital output.

Users of the system with much higher rate demands, especially those that currently have small computers installed, will be able to connect directly to the main computer via a high speed (up to 125,000 samples per second) parallel data link. In this case they may program the interactive aspects of their experiments themselves, but have the resultant data processed on a larger machine.

The IBM 360 Model 50 which we propose to use as the main computer has been chosen with a configuration that will support this type of multiple user activity. Its main (core) memory size is one million bytes or characters, or 250,000 words or values. The backup storage is an IBM pie file, which stores data on strips of magnetic tape, 2000 of them, which are all individually retrievable within 0.6 seconds. It has a total capacity of 400 million bytes. The subsidiary computer proposed is an IBM 1800 process control computer, connected via a special direct channel to the main computer.

Results may be listed on the system printer, of course, but the emphasis will be on typing out the results on the typewriters in the laboratories.

Data may also be returned in analog or digital form via the 1800 at rates comparable to those of the input capability; generally within a few seconds after the results have been produced inside the Model 50. The digital lines may be used to drive plotters producing graphical summaries of the experiments. Display equipment of various types can be installed in the laboratories and be driven from the analog or digital 1800 outputs.

The small computers can also be used to distribute output via their typewriter, display tubes and plotters, or they may use the results to automatically control the continuation of the experiments.

It is quite obvious that a system of the described scope is not supported by any computer manufacturer. We are therefore designing and writing a simple but complete support package including an interactive compiler, a supervisory system, input-output procedures and data acquisition and distribution routines. The system design is such that continuous guidance is provided to the user via the typewriter (see ACME note RC-1 and appendix RU-1).

To make such an approach at all economically possible, we are programming using an IBM compiler (FORTRAN H) to allow us to write the system within the allotted time span. This should also make it possible to later share the results of our software development work with others. The language that our system will compile and which we hope will be acceptable to our non-computer specialist users is a subset of PL-1, a FORTRAN like procedural language, defined by IBM for use on 360 systems. Thus we hope that procedures checked and proven on our interactive compiler can be filed and used on ours and other computer libraries under standard systems (see ACME note on PL-1).

The input-output system will include file handling and retrieval facilities where all data filed will be automatically labeled with all pertinent information to optimize the usefulness of the collected information. No elaborate buffering procedures will be used to minimize a user's inactive time, rather the time that he cannot use for computation will be turned over to the next user in the queue.

Since for some experiments the system reaction time will be quite critical, we will have to limit the number of these users in a given period. An attempt to utilize every available computer cycle for this work can only result in system overloading and failure. However, a number of users with non-critical problems, routine processing or information retrieval can balance the system to achieve reasonable total utilization, and we plan also to provide facilities for this type of use as soon as the more critical uses are satisfied.

The project currently uses several different computers around Stanford to check out parts of the system and is preparing to do a simulation study of the queuing algorithm. A small technical group is building prototype interconnection equipment for the various data interfaces.

The current planning work is sponsored by a Macy foundation planning grant and further funding has been requested from NIH. Much credit for ideas and procedures goes to other computer installations and other people, notably project MAC at M.I.T., MEDLAB at the Latter-Day Saints Hospital in Salt Lake City, The University of California at Berkeley Computation Center and ARPA project, U. C. San Francisco Medical School, U.C.L.A. Health Sciences, etc., and of course the Computation Center and the Computer Science Department of Stanford itself.

Additional ACME notes referred to in the text can be obtained through Gio Wiederhold, Associate Director for the ACME facility.

b. Mass Spectrometry

The Bendix Time-of-Flight mass spectrometer is now tied directly to the LINC computer. A program has been completed for determining peak positions in the spectra produced by the Bendix TOF.

The raw data used in this work were digitized analog signals, passed through a log amplifier, each containing the spectral information from one sampling pass (about 5 minutes) of a Bendix TOF spectrum. For these signals the position of m_i, an integer mass position i, is given by $m_i = ct^2$, where t is the flight time in the TOF spectrometer. A portion of such a spectrum (to about mass 350) is shown in Figure 6. A transformation is applied so that as near as possible $m_i = c't$ (from here on t is measured in digital parameters, normalized so that the time between each pair of adjacent data points equals 1). C' has been set so that as nearly as possible there are 25 such digital data points for Δ m/e = 1. We will call this processed information a "linearized" spectrum, $x(t)$. An example of this "linear" spectrum is shown as the upper trace of Figure 7.

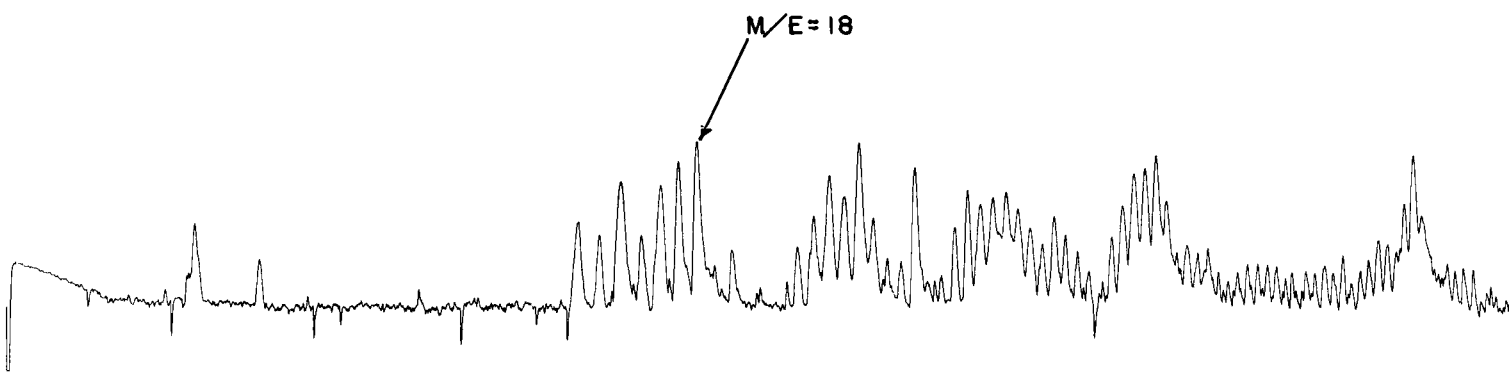
Plotting of this $x(t)$ has been useful in aiding operator identification of mass peaks. The signal has the following properties:

a. Integer mass numbers give signals of varying amplitudes at each position out to about mass 50 to 100. Maxima amplitudes are about 14 mass peaks apart.

b. At higher masses maxima continue to appear at approximately modulo 14 with very little or no signal between peak groups until the molecular ion is reached.

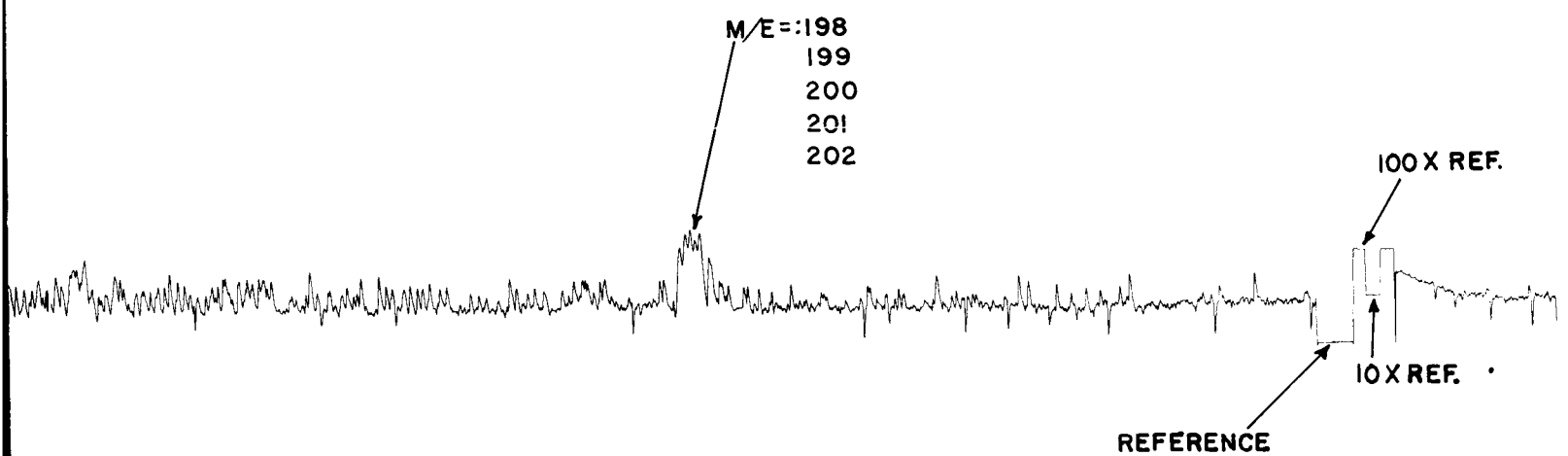
c. The peaks tend in general to be smaller and the groups tend to have fewer peaks as the mass increases.

d. At larger mass numbers, often well before the molecular ion, the resolution between peaks often is very poor, causing the peaks of a group to run together to make what appears to be one poorly defined hump. (This is a condition that should not have to be contended with; a mass spectrometer should have better resolution. However, it may be argued that any mass spectrometer will have similar characteristics beyond some high mass.



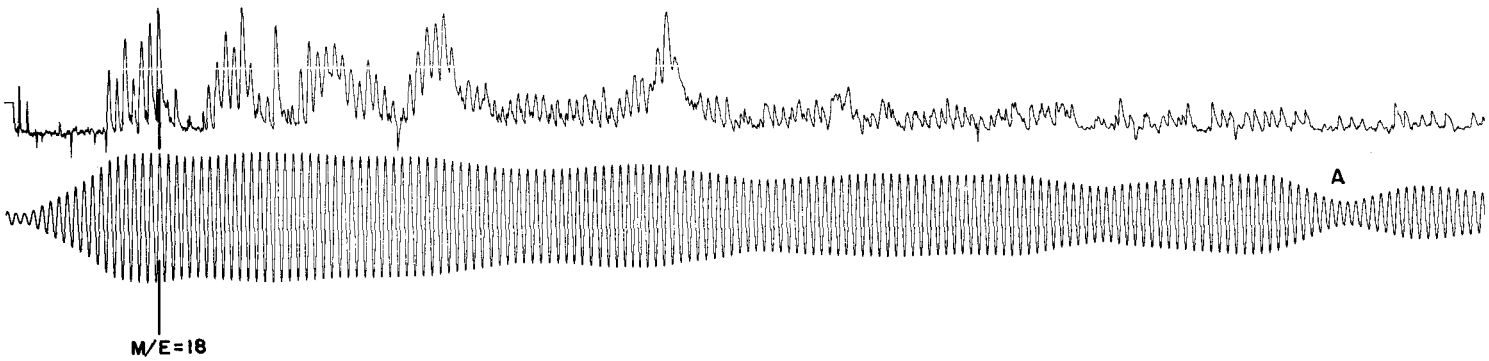
FIG

A digital plot of the di
The steps at the far rig
This spectrum was made o
rate was such that it ta
to m/e 200.



URE 6

gitized spectrum of melamine.
t are calibration levels.
the Bendix TOF. The scan
es about 175 seconds to scan



The upper trace is
The lower trace is

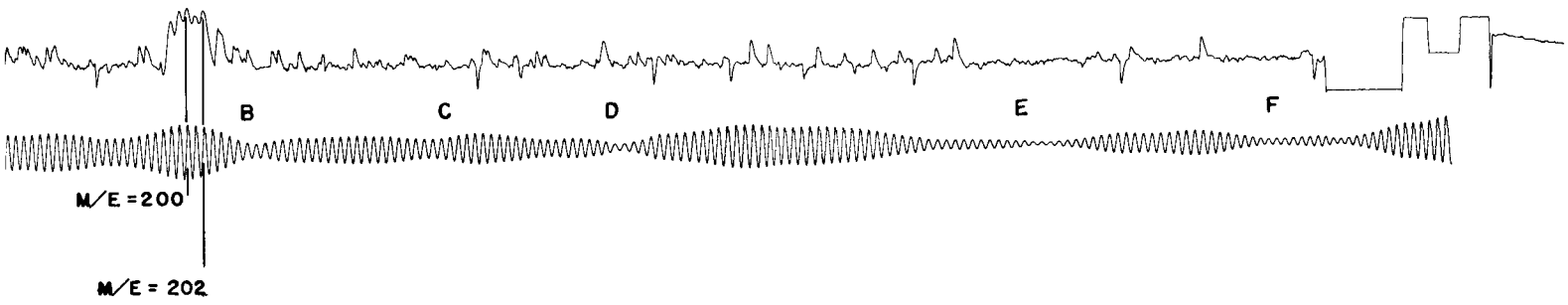


FIGURE 7

a digital plot of x(t).
$R_2(\tau_2)$.

The resolution of the Bendix is less than that needed for many of the samples submitted to it at this laboratory.)

e. The mass number, as indicated by the absolute position is accurate to only two or three integer mass numbers.

f. The linearity of mass numbers to position is accurate to about 10 percent over a mass range of 10 to 20 amu.

g. Due to the use of a logrithmic amplifier, there is a high gain for small signal amplitudes and thus very small signals are observed at almost every mass position.

h. Again due to the logrithmic amplifier, the apparent noise is very small for large peaks and very large for small peaks. As could be expected for very small peaks, the noise is very much larger than the peak.

While these observations are from experiments with the Bendix TOF spectra, they are typical of the spectrum output of the Atlas CH-4 or the low resolution spectra of the MS-9. Hence, it is felt that the study is of general applicability.

Efforts were made to write computer algorithms to interpret the spectra much as a human would be locating easily observed peaks, measuring to others, counting where possible and bridging gaps in the peaks by knowledge of the observable peaks and the known spacing expected between peaks. Such efforts lead to little success. The algorithms either showed far less skill than a human or as they were made more powerful they became unstable and generated false information. One of the most troubling details is an algorithm that will recognize a peak and ignore noise.

Use of correlation techniques was then tried. Correlations may be made on any signal. Two useful ones in the case of periodic signals are the cross-correlation and autocorrelation. The autocorrelation has the property that it will detect the period of a periodic signal in the presence of large amounts of random noise. If the period is known, the crosscorrelation of the noisy signal with a model signal containing a periodic signal of the known signal frequency, and a known reference phase, will result in an

output that will indicate the phase (of the particular frequency) of the original noisy signal.

Our problem was to locate where the peak was, regardless of its amplitude, and to do this reliably in the presence of noise. Hence, crosscorrelation was used.

The other function used in the crosscorrelation was a set of ten "masking" peaks; (5 were used originally, but 10 gave better results). These 10 peaks were an empirical copy of peaks in a typical spectrum, noise free, and with a period that equaled, as nearly as possible, that of $x(t)$. The mean of this masking signal was set to zero. The masking signal is analogous to the function $y(t)$ discussed below.

The classical crosscorrelation function is as follows:[4]

$$R_{xy}(\tau) = \lim_{T\to\infty} \int_{-T}^{+T} x(t + \tau)\, y(t)\, dt$$

This function has several interesting properties. First at a single value, (τ = constant), it may be compared with the Fourier function:

$$F_x(y) = \int_{-\infty}^{\infty} x(t) e^{-2\pi jyt}\, dt$$

This function is well known to rise to maximum as the period of $x(t)$ equals the period of $EXP(-2\pi jyt)$. Further if only the real part of the Fourier function is taken:

$$F_{x_n} = \frac{1}{T} \int_{-T}^{T} x(t) \cos \frac{n\pi t}{2T}\, dt$$

---

[4]Wilbur B. Davenport, Jr., and William L. Root, Random Signals and Noise, McGraw-Hill Book Co., p. 70 (1958).

it may be recognized as the equation for the coefficients of the Fourier series. This function is known to have maximums as the periods of $x(t)$ and $\cos^{n\pi t}/2T$ become equal, and in phase. Thus it could be expected that $R_{xy}(\tau)$ also will have maximums when the period and phase of $x(t + \tau)$ equals that of $y(t)$.

In our case we select $y(t)$ to have as near the same period of $x(t)$ as possible. Thus $R_{xy}(\tau)$ is an indication of the phase or peak position of $x(t)$.

To show that the crosscorrelation functions have a signal-to-noise enhancement property, consider the following:

$x(t)$, the signal, may be considered as noise free signal $X(t)$ and noise $x^*(t)$. If $X(t)$ and $x^*(t)$ are independent, then

$$R_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X(t + \tau)\, y(t)\, dt$$

$$+ \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^*(t + \tau)\, y(t)\, dt$$

If $x^*(t)$ is random with mean zero (which is never completely true in real cases, but it usually has sufficient elements of randomness such that the analysis with that assumption are of guiding value) the second term goes to zero.

Of course in actual practice, we cannot let T go to infinity since $x(t)$ is bound by the limits of the data base. Another constraint on T is the computer time required to evaluate the function. Normally the computer time required would be linear with T. However, the methods described by Gentlemen and Sande[5] shorten computer time required for very large T.

---

[5] W. M. Gentlemen and G. Sande, "Fast Fourier Transfors for Fun and Profit," AFIPS Conference Proceedings, 29, FJCC (1966).

T has been selected such that -T to +T covers 10 peak positions, with 250 sampling points. There is no pretense that this is optimum. At this point it works and is manageable in our LINC computer.

The actual function evaluated by our algorithm is a two step crosscorrelation (although in the actual logic they are performed in an interleaved manner). First:

$$R_1(\tau_1) = k_1 \sum_{i=1}^{250} x(\tau_1 + i)\, y(i)$$

This function, if plotted, produces a spectra-like plot. Noise is suppressed. The peaks are quite well shaped with most of them having unmistakable maximum points. However, the amplitude information is "smeared." Each peak in the original spectrum contributes to peaks in this function. Hence, areas that had very little or no signal now begin to show peaks. This is the "bridging" action that allows peak counting to be attempted by a computer algorithm. The signal level, which is dependent upon the strength of the periodic function in the original spectra tends to vary greatly, rising to large values where there was a group of peaks in the original spectra.

There still is enough noise to be troublesome. However, the principal defect seems that where large data gaps are bridged the phase does exhibit drift. The solution has been to severely clip R_1 so that it has a reasonably constant amplitude and form the new function:

$$R_2(\tau_2) = k_2 \sum_{j=1}^{250} R_1(\tau_2 + j)\, y(j) \quad \text{where} \quad y(j) = y(i)$$

This has demonstrated good phase lock giving reliable results up to the mass range of 200 to 300, depending upon the quality of the original spectra. This is quite comparable with human interpretation of the data. A plot of $R_2(\tau_2)$ is shown as the lower trace in Figure 7.

Figure 6 is a plot produced upon the Calcomp digital plotter of the digitized analog spectrum. The Bendix TOF was set to scan at a rate of "5", taking about 175 seconds from the beginning to m/e 200. Time is linear on this plot, $m/e = kt^2$. Peaks 18 and those of the mercury isotopes are indicated for comparison with those in Figure 7. The steps at the far right of the spectrum are calibration levels of the logarithmitic transformation, the steps indicating one decade. This plot has no particular value to the described computer operation, only verifying that the experiment went without any gross error. The plot may be omitted at will.

The upper portion of Figure 7 is a Calcomp digital plot of the transformed spectrum. Now the x axis is a reasonable linear function of m/e. Note that the peaks about 18 have been compressed and those at 200 have been expanded from the plot of Figure 6. This is the signal previously termed x(t).

The lower portion of Figure 7 is a Calcomp plot of the double correlation described. The amplitude of this signal depends upon the amplitude of the periodic signal in x(t) and upon the closeness of the match of the period times of y(t) and x(t). As long as this correlation is above a certain amplitude there is a high degree of confidence that it is in phase lock with x(t) and does point to peak locations. The level at A and C is about the minimum level that will give such confidence. At points B, D, E and F, the phase lock has completely broken, hence all information from the correlation to the right of any of those points is meaningless.

It is suspected that the loss of phase lock at these points is due to mismatch of the periodicity of x(t) and y(t). We propose to do a frequency check at these points on x(t) and use that to improve y(t). Then that section of the spectrum may be reinvestigated. A Fourier or autocorrelation technique may be used. However, the LINC programming required for this experimental work is quite lengthy. It is anticipated that this may be done at an early date upon the ACME IBM 360 Model 50 computer at a great saving in programer time.

Again the plots of Figure 7 are optional. The program may proceed directly to produce the typeout of Listing A and Figure 8.

Listing A, shown in Table 5,is the actual digital output of the analysis. The largest peak has been normalized to 100. (There just happen to be two large peaks in the spectrum that are equal to within 1 percent.) Peaks 200 and 202 are just over the threshold of 1 percent. Peak 199 apparently was just under 1 percent and the threshold logic discarded it. This threshold is arbitrary and may be set at any level.

Figure 8 is again a Calcomp output. This is a conventional bar graph plot of the spectrum.

It should be noted that there was no operator transcribing of data points in this whole sequence. All control exercised was by program call and program parameters to the LINC, plus the manual controls of indexing the Calcomp plotter to starting points.

Further studies are being carried out on the signal-to-noise enhancement achieved by this method.

c. Investigation into Denaturation Characteristics of Specific DNA

At elevated temperatures DNA molecules "melt" from ordered bihelical to a random "denatured" coil form. The melting is caused by disruption of hydrogen bonds in the helix, and is accompanied by changes in such physical properties as intrinsic viscosity, light scattering, buoyant density, sedimentation velocity, optical rotatory dispersion, and optical absorbance. The change in absorbance is one of the most striking and easiest to measure. It takes place over a temperature range of several degrees. The midpoint, T_m, is strongly dependent on the ionic strength of the solution and on the G-C content with T_m rising with ionic strength and with G-C content.[6]

---

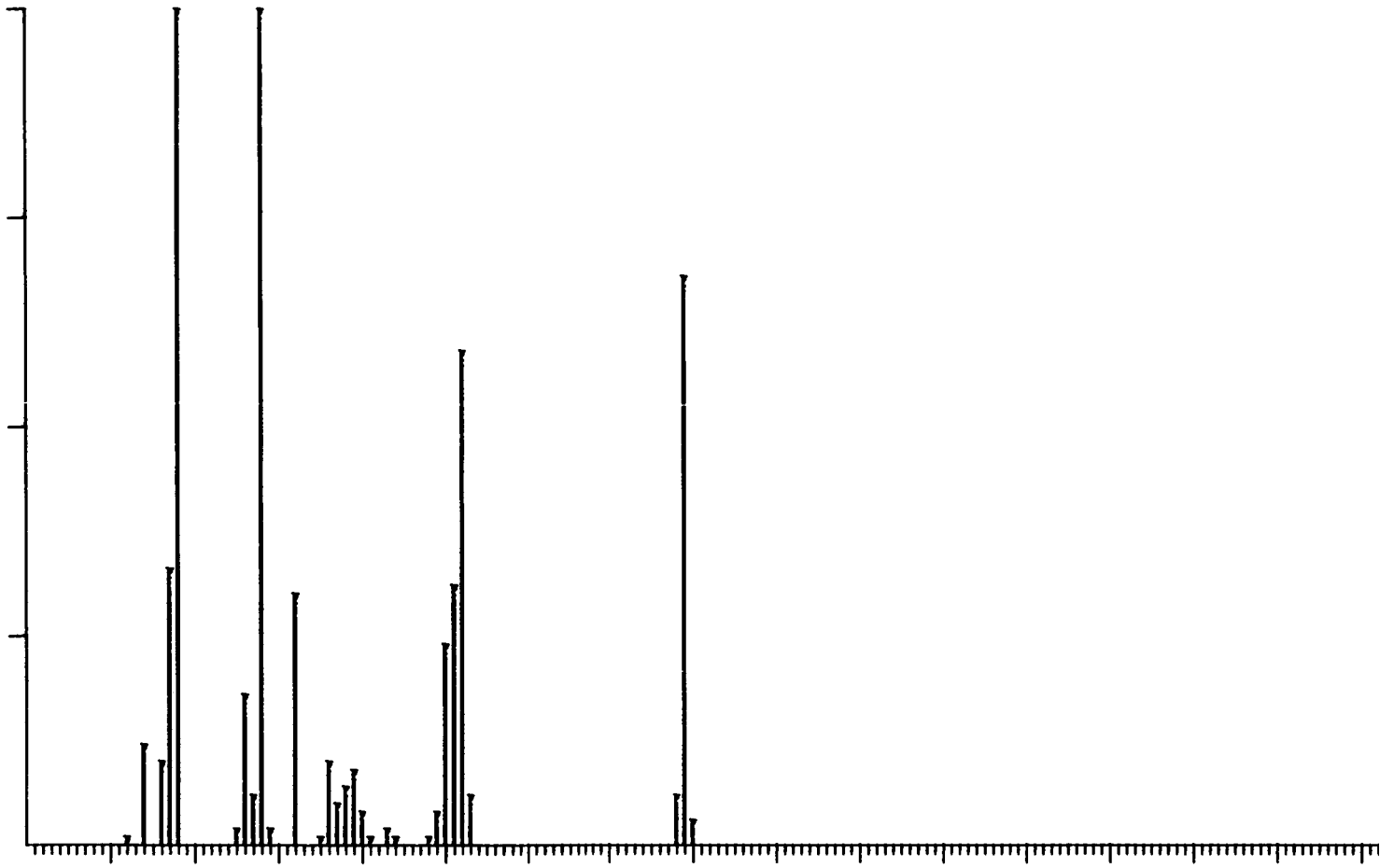
[6] P. Doty, Harvey Lectures, 55, p. 103 (1961).

TABLE 5

MASS	PERCENT OF MAX PEAK									
0	0	0	0	0	0	0	0	0	0	0
10	0	0	1	0	12	0	10	33	100	0
20	0	0	0	0	0	2	18	6	100	2
30	0	0	30	0	0	1	10	5	7	9
40	4	1	0	2	1	0	0	0	1	4
50	24	31	59	6	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	6	68
80	3	0	0	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0	0	0
110	0	0	0	0	0	0	0	0	0	0
120	0	0	0	0	0	0	0	0	0	0
130	0	0	0	0	0	0	0	0	0	0
140	0	0	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0	0	0
160	0	0	0	0	0	0	0	0	0	0
170	0	0	0	0	0	0	0	0	0	0
180	0	0	0	0	0	0	0	0	0	0
190	0	0	0	0	0	0	0	0	0	0
200	1	0	1	0	0	0	0	0	0	0
210	0	0	0	0	0	0	0	0	0	0
220	0	0	0	0	0	0	0	0	0	0
230	0	0	0	0	0	0	0	0	0	0
240	0	0	0	0	0	0	0	0	0	0
250	0	0	0	0	0	0	0	0	0	0
260	0	0	0	0	0	0	0	0	0	0
270	0	0	0	0	0	0	0	0	0	0
280	0	0	0	0	0	0	0	0	0	0
290	0	0	0	0	0	0	0	0	0	0
300	0	0	0	0	0	0	0	0	0	0
310	0	0	0	0	0	0	0	0	0	0
320	0	0	0	0	0	0	0	0	0	0
330	0	0	0	0	0	0	0	0	0	0
340	0	0	0	0	0	0	0	0	0	0
350	0									

<PROG>>

Listing A



FIGURE

8

It has been noticed by various authors that individual genetic markers appear to melt over a very narrow temperature range as evidenced by the fact that almost all of the transforming ability of the marker is often lost over a range of a fraction of a degree.[7,8] In addition, early theoretical treatment of the melting phenomenon indicated that the melting range for a particular DNA might be extremely narrow.[9] This gave rise to the expectation that it might be possible to detect melting of individual molecular species of DNA in a heterogeneous sample by sufficiently accurate measurement of the melting curve. If so, these observations should make it possible to characterize the DNA more precisely and eventually expedite isolation of individual DNA fragments. The instrumental precision required would be high, meaning that computer control of the type available from the ACME system would be desirable. Initial experiments on B. subtilis DNA using manual temperature programming and measuring absorbance changes with the Zeiss PM QII Spectrophotometer indicated that the number of DNA fragments in the preparation was probably too high for effective resolution by such techniques. In order to obtain further information an intensive search of relevant literature was undertaken.

Literature Survey

Early theoretical treatments of DNA melting were based on the assumption that all base pair bonds are equivalent. In fact, the A-T bonds are considerably weaker than G-C bonds, as evidenced by the dependence of melting temperature on G-C content mentioned above. This means that a DNA made up of sections of varying G-C content might be expected to melt in sequence, with those sections having the lowest G-C content melting first. There are many difficulties in analyzing the melting mathematically, but a relatively complete treatment has recently been given of the case where the molecules

---

[7] J. Marmur and D. Lane, Proc. Natl. Acad. Sci., 46, p. 453 (1960).

[8] W. Ginoza and B. H. Zimm, Proc. Natl. Acad. Sci., 47, p. 633 (1961).

[9] B. H. Zimm, J. Chem. Phys., 33, p. 1349 (1960).

are quite short, so that bonds can be expected to break from the ends in zipper fashion, rather than within the molecule.[10] In this case they concluded that melting of most individual molecular species would cover a temperature range of a few degrees or so, making it extremely difficult to isolate the effect of a single species in a heterogeneous sample with a total melting range only slightly greater. They also made a semi-quantitative analysis of longer molecules which might have alternating helix and random coil regions, and concluded that in this case the transition is split into two parts. The first part is similar to that treated above, but ends with the strands still held together by regions of high G-C content. The second covers the splitting apart of these two strands, and should occur over a very narrow temperature range for an individual molecular species. Different species would have different temperatures of strand separation, depending on the G-C content of their most heat resistant sections and thus the total temperature spread for strand separation of a heterogeneous sample would also be relatively wide.

Later papers in this series attempted to extend the treatment to higher molecular weight species on the assumption that the molecule was built up of various sized blocks of A-T or G-C pairs. The primary conclusion was that fluctuations in G-C content occurring over long regions were critical in determining transition behavior.[11]

A number of experimental papers by various authors substantiate the theoretical conclusions discussed above.

1. Two Phase Transition

It is almost certain that there are at least two steps in the transition. This is illustrated by the absorbance of DNA when cooled after being heated to temperatures sufficient to cause increased absorbance. There is a substantial change in the low temperature absorbance in such cases only if

---

[10] D. M. Crothers, N. R. Kallenbach, and B. H. Zimm, J. Mol. Biol., 11, p. 802 (1965).

[11] D. M. Crothers and N. R. Kallenbach, J. Chem. Phys., 45, p. 917 (1966).

the temperature has been raised to a point considerably higher than T_m. A curve can be drawn showing the increase in low temperature absorbance of heat treated samples. The T_m of this curve, where again T_m refers to the temperature to which the sample must be raised to cause half of the total increase in absorbance possible in the process, may be several degrees above the T_m as usually measured.[12] A similar situation is noticed in the T_m for loss of genetic markers, which is invariably higher than the T_m usually measured.

A logical explanation for such results is that as postulated above the usual T_m corresponds to breakage of most of the individual bonds or groups of bonds in the DNA chain, while leaving the two strands joined at some point. The native (or near native) molecule can still be reformed quickly in the cooling process by "zippering up" starting at the point or points of attachment. The second T_m corresponds to strand separation. The separate strands would obviously be much less readily renatured.

Further corroboration for the theory can be found in the behavior of DNA which has been crosslinked by treatment with nitrous acid or nitrogen mustard. This DNA regains its original properties, including transforming ability at low temperature, even after being heated sufficiently to denature non-crosslinked DNA completely.[13,14] The latter authors suggest that the use of the word denaturation be reserved for those treatments providing a random coil form which persists even when conditions are changed so that the helix form is thermodynamically stable. They propose that melting be used to refer to quickly reversible changes in the helix.

---

[12] J. Marmur and D. Lane, Proc. Natl. Acad. Sci., 46, p. 453 (1960).

[13] E. F. Becker, B. K. Zimmerman, and E. P. Geiduschek, J. Mol. Biol., 8, p. 377 (1964).

[14] K. W. Kohn, C. L. Spears and P. Doty, J. Mol. Biol., 19, p. 266 (1966).

2. Wide Melting Curve for Individual Molecules

Melting curves have been made for several phage DNA preparations in which almost all of the molecules contained all the DNA from a given phage particle. The melting curve of homogeneous T2 DNA for instance is only slightly narrower than that of heterogeneous sea urchin DNA measured under the same conditions.[15] The curve for homogeneous lambda DNA in 10 percent formaldehyde is about 10^{o} wide from the 10 percent melted to 90 percent melted points.[16] Even more conclusive is a series of electron micrographs made on the latter samples. The formaldehyde hydrogen bonded to the melted portions apparently preventing renaturation, since after partially melted samples were cooled absorbance remained the same as at the high temperature. Thus the high temperature conditions were presumably maintained during the preparation of the micrograph samples. The micrographs of samples subjected to temperatures between 48^{o} and 53^{o} (the latter is about the T_m under these conditions) showed that the long strands of DNA had developed regions of disordered structure, ranging from short segments ("sites") of the order of 60 base pairs long, to much longer "multiple sites." The positions of these segments were not exactly the same on different molecules, even when heated to the same temperature, but they were overwhelmingly concentrated on one part of the molecule. Most of the initial sites were near the middle, with two other higher temperature maxima appearing on the same side of the central zone. The average number of sites increased with temperature in good agreement with the change in absorbance. Obviously, the entire molecule does not melt over a very narrow temperature range.

3. Sharp Strand Separation for Individual Molecules

As mentioned above, individual genetic markers often lose their transforming ability over a relatively narrow temperature range. The narrow range appears to be connected with short molecules, with a fairly wide range when the

---

[15] E. P. Geiduschek, J. Mol. Biol., 4, p. 467 (1962).

[16] R. B. Inman, J. Mol. Biol., 18, p. 464 (1966).

molecules are relatively long.[17] In addition, the T_m of the denaturation of a particular marker is reduced as the molecular weight is reduced. A logical explanation for this behavior is that short molecules containing a particular genetic marker are all apt to have the same region of high G-C content, and thus would separate their strands at the same temperature. Longer molecules would have a high probability of having regions of higher G-C content than the marker itself, leading to higher T_m, and, if randomly broken from the original extremely long unit, would also have a greater variation in the number and composition of their G-C sites, leading to a fairly wide denaturation temperature range.

One disturbing aspect of the behavior of short DNA molecules should be noted. Their relative infectivity is quite low. In the case of the fractions used to study relative performance of highly sheared molecules (molecular weight about 10^6) the transforming ability was only about 0.4 percent that of unsheared molecules.

4. Conclusions of Literature Survey

On the basis of the preceding evidence, it appears impossible to achieve the initial objectives of this program. The only part of the denaturation process which appears to be completed over a very narrow temperature range is the strand separation of very short particles. The creation of such short particles not only reduces the transforming activity markedly but results in a heterogeneous population of molecules. If it is assumed that the original DNA particle has a molecular weight of the order of 3×10^9, then shearing to a molecular weight of 1×10^6 would result in a population with 3000 different species even if all molecules were sheared at exactly the same points. This is almost certainly not the case, so there are probably many thousands of different molecular species in a sheared preparation. Even with only 3000 species, each with a melting range of 0.2^o or more, and with a total melting range for the entire population of 10^o,

---

[17] W. R. Guild, J. Mol. Biol., 6, p. 214 (1962).

at least 60 different species could be expected to be melting at any temperature near the center of the range, making it extremely difficult to observe the effects of a single one.

Alternate Approaches

Since many molecules are melting at a time it is impossible to separate a single species from the rest using melting temperature as a criterion. However, it may be possible to fractionate groups of molecules in which the concentration of a single species is increased on the order of 50 fold if the molecules melting at a given temperature can be isolated.

One way to fractionate is by cesium chloride density gradient centrifugation, since the density of the melted strands is greater than that of native DNA. Thus a DNA sample could be centrifuged after exposure to higher and higher temperatures, and the heavy layers separated to give a series of fractions. Unfortunately, the procedure is time consuming, and the amount of treated DNA which can be obtained in a single experiment is small.

Native and denatured DNA have been separated on methylated albumin columns[18] with the native DNA appearing earliest in a stepwise elution with NaCl. The denatured DNA can then be renatured if desired. A series of such runs after treatment at successively higher temperatures should give denatured DNA fractions of varying composition. Unfortunately, some of the native DNA is also recovered in the denatured fraction.

A second chromatographic method utilizes hydroxyapatite columns, which retain double stranded DNA while permitting the single stranded to elute in 0.08 M phosphate.[19] In this case the column itself can be heated while elution is proceeding. If there is sufficiently accurate temperature control

---

[18] M. Roger, C. D. Beckmann, and R. D. Hotchkiss, J. Mol. Biol., 18, p. 156 (1966).

[19] Y. Miyazama and C. A. Thomas, J. Mol. Biol, 11, p. 223 (1965).

the denatured DNA can be eluted in order to its denaturation temperature. This procedure has been used to provide separation of denatured fractions of differing G-C content from low molecular weight DNA.

Native DNA has also been chromatographed on magnesium polymethacrylate columns, with some fractionation with respect to G-C content. This fractionation was not extensive, but if it could be increased it might be possible to use this technique to separate species without denaturation.

There are a few cases where it is possible to concentrate a single portion of the bacterial DNA using phage capable of localized transduction. Such phage contains a section of bacterial DNA which can be increased in its relative concentration by a factor of 200 or so.[20] The chromatographic approaches can concentrate an individual species somewhat, but cannot provide anything close to this specificity. The obvious disadvantage of the use of biological concentration is the difficulty of finding a suitably cooperative phage for a particular section of the DNA of a particular bacteria.

Conclusions and Proposed Program

Probably the most promising chromatographic substrate is hydroxyapatite since the separation is direct and simple, and the column is relatively inert. This will be explored in the near future with computer control of the temperature and readout of the absorbance of individual fractions planned. The other chromatographic substrates will be considered if the hydroxyapatite proves to be deficient in some respect. In addition, the possibility of utilizing phage to separate a particular section of DNA more efficiently will be further considered.

V. UV Microspectrophotometry

Our efforts on video scanning techniques are continuing with emphasis in their application to the data readout of a Model E (Beckman) ultracentrifuge. Evaluation of the densitometry system will be presented in the next progress report.

---

[20] D. S. Hogness and J. R. Simmons, J. Mol. Biol., 9, p. 411 (1964).

C. PERSONNEL AND ORGANIZATION

There have been additions to our organization and their curriculum vitaes are listed below. Mr. L. Hundley, who was part of our Instrumentation Research Laboratory, is no longer with Stanford.

HENRY R. HULETT

PERSONAL DATA: Born Nespelem, Washington, May 5, 1920. Married, 2 children. American citizen.

EDUCATION:

1941	B.S., Oregon State College, Corvallis, Oregon Major: Chemistry
1942	M.S., Oregon State College, Corvallis, Oregon Major: Chemistry
1943-44	University of Chicago, Chicago, Illinois Graduate work in Meteorology
1944	Harvard, M.I.T., Cambridge, Massachusetts Graduate work in Electronics
1964	Ph.D., Stanford University, Palo Alto, California Major: Chemistry

APPOINTMENTS:

1946-51	Research Engineer at Hughes Aircraft, Culver City, California.
1951-52	Electronics Section Head, Pacific Mercury Research Center, Santa Barbara, California.
1952-57	Electronics Section Head, Detroit Controls Research Division, Redwood City, California.
1957-59	Guidance & Control Subsystem Manager, Staff Scientist, Lockheed Missiles & Space Company, Palo Alto, California.
1959-63	Director of Research, Advanced Technology Laboratories, Division of American-Standard, Mountain View, California.
1963-64	Research Associate, Stanford University, Department of Chemistry, Palo Alto, California.

HENRY R. HULETT

1964-66	Associate Professor, College of Notre Dame, Belmont, California.
1965-66	Group Director, Vidya Division, Itek Corporation, Palo Alto, California.

PUBLICATIONS:

"Effect of Particle Length Distribution on Infectivity of Tobacco Mosaic Virus," H. R. Hulett and Hubert L. Loring, Virology, 25, 418, 1965.

"Spectrophotometric Determination of Iron," J. P. Mehlig and H. R. Hulett, Ind. & Engr. Chem., Anal. Ed., 14, 869, 1942.

Numerous classified reports, five patents.

JAMES C. BRIDGES

PERSONAL DATA: Born South Amboy, New Jersey, April 13, 1941. Single, American citizen.

EDUCATION:

1963	B.S.E.E., Rensselaer Polytechnic Institute, Troy, New York.
1964	Stanford University. Twenty seven graduate credits under a Research Assistantship.

APPOINTMENTS:

1961-63	Co-op Student. Full-time work semesters at IBM interspersed with undergraduate education.
1963-64	Research Assistant. Electron Devices Laboratory, Stanford University.
1964	Peace Corp Training.
1965-66	Development Engineer, Nuclear Instruments, Hewlett-Packard Company, Palo Alto, California.
1966	Research Engineer, Genetics Department, Instrumentation Research Laboratory, Stanford University.

DANUTE E. NITECKI

PERSONAL DATA: Born in Lithuania, April 22, 1927. U. S. citizen.

EDUCATION:

1956	M.S., University of Chicago, Department of Chemistry, Chicago, Illinois.
1961	Ph.D., University of Chicago, Department of Chemistry, Chicago, Illinois.
1961-63	NIH Postdoctoral Fellow with M. J. S. Dewar, University of Chicago, Department of Chemistry, Chicago, Illinois.

APPOINTMENTS:

1963-66	Assistant Research Biochemist, University of California, Department of Microbiology, San Francisco Medical Center.

PUBLICATIONS:

"Reactivity of Cyclic Peptides I. Transannular Histidine-O-Acetyltyrosine Interaction," K. D. Kopple and D. E. Nitecki, J. Am. Chem. Soc., 83, 4103 (1961).

"Reactivity of Cyclic Peptides II. Cyclo-L-Tyrosyl-L-Histidyl and Cyclo-L-Tyrosyl-Triglycyl-L-Histidyl Glycyl," K. D. Kopple and D. E. Nitecki, J. Am. Chem. Soc., 84, 4457 (1962).

"Immunochemical Studies on the Poly-γ-D-Glutamyl Capsule of Bacillus anthracis. I. Characterization of the Polypeptide and the Specificity of its reaction with Rabbit Antisera," J. W. Goodman and D. E. Nitecki, Biochemistry, 5, 657 (1966).

"Immunochemical Studies on the Poly-γ-D-Glutamyl Capsule of Bacillus anthracis. II. The Synthesis of Eight Dipeptides and Four Tripeptides of Glutamic Acid," D. E. Nitecki and J. W. Goodman, Biochemistry, 5, 665 (1966).

"Qualitative and Quantitative Determination of Mixtures of Amino Acids Using 2,4,6-Trinitrobenzene Sulfonic Acid," D. E. Nitecki, I. M. Stoltenberg, and J. W. Goodman, Anal. Biochem., 1967 (in press).

D. PAPERS AND REPORTS

April 1, 1966 to October 1, 1966.

Reports

1. J. Lederberg, "Planetary Exploration and Biological Research." Impact of Space Exploration on Society, Vol. 8, Science and Technology Series, 1966, Amer. Astro. Soc.

2. J. W. Westley, "Detection of Optical Activity as a Sign of Life," presented to the Amer. Astro. Soc., (1966). IRL-1047.

Publications

1. B. L. Karger, R. L. Stern, W. Keane, B. Halpern, and J. W. Westley, "GLC Separation of Diastereoisomeric Amides of Racemic Cyclic Amines," J. Anal. Chem., 39, p. 228 (1967).

2. B. Halpern, J. Ricks, J. W. Westley, "The Stereospecificity of α-Chymotrypsin-Catalyzed Hydrolysis and Alcoholysis of Specific Ester Substrates," Aust. J. Chem., 20, p. 389 (1967). IRL-1049

3. B. Halpern, L. Chew, and J. W. Westley, "Investigation of Racemization during Peptide Bond Formation by GLC of Diastereoisomeric t-BOC-Amino Acid Amides," (submitted to J. Anal. Chem.). IRL-1053.