

Technical Report No. IRL 1073

**MOLECULAR BIOLOGY APPLICATIONS OF
MASS SPECTROMETRY**

Final Report Covering Period July 1, 1965 to December 31, 1967

For

Air Force Office of Scientific Research

Contract AF 49(638)-1599



**Instrumentation Research Laboratory, Department of Genetics
Stanford University School of Medicine
Palo Alto, California**

MOLECULAR BIOLOGY APPLICATIONS OF MASS SPECTROMETRY

Final Report Covering the Period July 1, 1965 to December 31, 1967

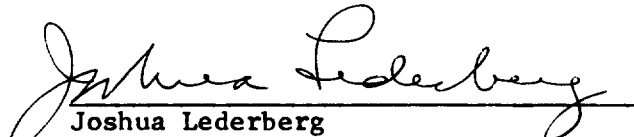
Air Force Office of Scientific Research


Contract AF 49(638)-1599

Instrumentation Research Laboratory, Department of Genetics

Stanford University School of Medicine

Palo Alto, California


Joshua Lederberg
Principal Investigator


Elliott C. Levinthal
Program Director

INTRODUCTION

The research objectives and purposes of our Air Force Program entitled "Molecular Biology Applications of Mass Spectrometry" have been closely related to those of our NASA program entitled "Cytochemical Studies of Planetary Microorganisms - Explorations in Exobiology". This is not too surprising since the same reasons that make mass spectrometry a powerful tool for biological explorations carried out remotely on a planet forty million miles distant from the earth make it a most sensitive and selective method for analyzing organic molecules important in problems of molecular biology germane to modern medicine. It is a potentially powerful method for investigating the unknown structure of unknown molecules in nerve tissue that form the engram which is part of the process of memory as well as Martian surface material. Cytochemistry via mass spectrometry is still a distant and challenging goal. However, significant progress has been made during this period. In addition to building a base for further advances our efforts have yielded results of present value.

The problem and the program can be subdivided into separate areas of concern. First, there is the question of volatilizing the molecules of interest. This can be approached by means of chemical modification of the class of molecules under investigation. We have successfully applied this concept to problems of resolution and identification of optical isomers of amino acids using the combination of a gas chromatograph and mass spectrometer. A second but much more difficult method, which has

the advantage that it is more directly applicable to the goal of cytochemistry, could conceivably utilize electron, heavy particle or photon beam energy. We have investigated the use of both heavy particles and laser photon beams. In the case of the former, we had useful but discouraging results. In the latter case our present results show some real promise.

A second subdivision of this effort addresses itself to acquiring basic data on the mass spectra of a large number of monomers of biological interest. A report covering the work on amino acids has already been published. Work on nucleotides and related products is in progress. Some of these results are referred to and discussed in this report.

Computer control of mass spectrometers describes the third categorization of the program. Full advantage of a mass spectrometer as a biological tool can only be achieved when the instrument is under computer control. The typical processes of calibration and optimization of operating parameters are sufficiently complex that they require automation if it is desired to analyze a large number of spectra in a short period of time. We have published a report which describes a very effective system for computer operation of both a time-of-flight and quadrupole mass spectrometer. This system has been and is continuing to be used for biological research purposes. The system is being elaborated to provide a more sophisticated level of control. In addition, work is underway to achieve some degree of control of high resolution mass spectrometers that use magnetic filters.

Fourthly, there is the question of data retrieval for subsequent computer analysis. The ultimate goal of a two dimensional micro-description of the distribution of molecules in a tissue by means of their mass spectra presents formidable problems of data handling. The bandwidth requirements are at least an order of magnitude greater than color video. High bandwidth data retrieval and buffer storage are required. We have not directly confronted this problem. General advances in the technology of high speed solid state switching devices and information storage methods lend some hope for the future. We have, however, made some modest steps. We have implemented an interface system for a direct data link from a high resolution mass spectrometer to an IBM 360/50 computer.

The fifth and last subdivision of the program really represents most clearly the ultimate goal for which the previously described efforts provide the technological tools. Ultimately the spectra acquired must be analyzed. This requires computer manipulation of chemical hypotheses. This poses a problem in both artificial intelligence and organic chemistry. A great deal of progress has been made in this direction. Most of this research is supported by the Advanced Research Projects Agency of the Office of the Secretary of Defense, Contract No. SD-183, and carried out in collaboration with Professor E. Feigenbaum of the Department of Computer Science.

This final report is divided into five sections appropriate to each of the areas previously described. In most cases it will rely heavily on references to papers previously published in professional journals

and technical reports which in some cases have been submitted to other agencies. For the sake of completeness, some items are included in the bibliographies which predated the Air Force contract.

I. Volatilization of Molecules of Biological Interest

a. Chemical Methods

Molecules of biological interest are characterized by asymmetries at one or more of the carbon atoms incorporated in the molecule. From the viewpoint of the exobiologist this statement is the basis of the well-known significance of optical activity as a clue for the recognition of life. The preparation of volatile diastereoisomers, their separation by gas chromatograph and their further identification by a mass spectrometer provides a method important to both terrestrial and extraterrestrial biology. The general concept is elucidated in the following papers enumerated in the bibliography which is part of this section (references 1.1, 1.2). Initially the technique was applied to the high sensitivity scanning of amino acids for optical activity (references 1.3 to 1.8). Other papers demonstrate the general applicability of the method (references 1.9 to 1.14).

b. Physical Methods of Volatilization

The goal of this aspect of the investigation was to develop a method which ultimately will lead to mass spectrometry image scanning and thus allow the possibility of microidentification and the physical analysis of the cellular and subcellular constituents of biological structures.

The following material is presented in more detail than other sections of the report representing comparable efforts since it does not yet appear in the appropriate professional journals. It has however been described in the semi-annual status reports submitted

under our National Aeronautics and Space Administration grant No. NsG 81. The initial efforts were based on the developments of Professor R. Castaing and George Slodzian of the University of Paris who had developed an imaging, secondary-ion-emission mass spectrometer which produced magnified images of the surface distribution of atomic species with one micron resolution. The intention of our efforts was to ascertain if sufficient numbers of intact molecular ions characteristic of organic particle materials could be evolved to enable the determination of the molecular distribution and structures of biological interest. For these purposes Dr. George Slodzian joined our staff for a period of approximately a year and a replica of the prototype developed by Castaing and Slodzian consisting of an inert gas ionizer, primary ion gun, target support assembly and secondary ion electrostatic lens structure was acquired. The mass filtering was accomplished by use of a 60° signal focusing mass spectrometer that was designed and constructed in our laboratory.

Positive primary ions of argon and hydrogen were employed to bombard polyethylene, nylon, polyvinylpyrrolidone methylcellulose, graphite, OFHC copper, gold, and aluminum. Both positive and negative secondary ions were examined. The incident primary ion energy was approximately 6 keV and 14 keV during the examination of positive and negative secondaries, respectively.

The molecules evolved from the bombarded organic samples were found generally to constitute rearranged configurations not representative of the original sample structure. It seems probable that the

incident primaries produce sufficient local dissipation of energy to devastate the molecular structure. Consequently, the "hottest" of the evolved secondaries come off as rearranged molecular ions and the majority of the material recondenses on the surface, subsequently to be re-evolved, retaining little information regarding the original structure.

As often happens there has been an unexpected finding in this work. At one state in the data analysis it was decided to plot only those mass peaks in any given run that corresponded to singly ionized integral numbers of carbon atoms clumped together. It was thought that this information might provide a useful clue as to the nature of the processes taking place at the surface. When such plots were made for negative secondaries, a distinct pattern revealed itself for all observed combinations of primary ions and organic targets. Namely, there was a preference for the negative secondaries to contain an even number of carbon atoms. This effect has been seen to exist out to 12 carbon atoms, whereupon the effect faded out into the "noise". The effect has been found to persist with the addition of one or two protons to the complex. When graphite was examined it was found that the curve corresponding to one proton affixed to the complex was a replica of the unprotonated curve but displaced downward, on a semilog plot. The addition of two protons, however, so enhanced the effect that it could be observed out to 14 carbons and over 6 decades of intensity whereupon instrumental sensitivity limited further observation. A typical odd-even enhancement in intensity was 10-fold.

A complementary but less pronounced enhancement of odd carbon clumps over evens has been found for positive secondaries.

Arguments based on observed odd-even carbon effects have even recently been offered in the literature in support of the biogenic origin of oil; the rationale being that such preferential effects mediate in favor of an ordered production mechanism. That odd-even carbon effect can be generated by abiogenic mechanisms would appear to weaken this particular argument. Odd-even carbon effects have been reported in the past. They appear, however, not to be widely known. Since the only explanation we can find for these effects calls for the carbon complexes to be in the form of linear chains, we are forced to conclude that most of the evolved ionized carbon complexes are in such form.

Subsequently our efforts were directed to the investigation of the suitability of laser induced vaporization as a mechanism for enabling spacial resolution in mass spectro analysis of organic solids. The additional experiments were conducted with an Optics Technology Model 100 pulse ruby laser coupled to a Bendix time-of-flight (TOF) mass spectrometer (MS).

Our measurements have indicated a lasing threshold of 190 joules input with a laser output of 3.5×10^{-2} joules at the maximum available input of 440 joules. The wavelength of the laser radiation is 6943 Å. Beam divergence at threshold appears to be in approximate accord with the specified figure. We have some evidence that the full angle beam

divergence at maximum output is approximately 5×10^{-3} rad; we have not however, performed accurate determinations of the directional dependence of the beam brightness of the radiation issuing from the laser.

Figure 1.1 illustrates schematically the optical configuration utilized for these mass spectral studies. A portion of the radiation issuing horizontally from the pulsed ruby laser 1 is reflected downward by the beam splitter 2 toward the simple biconvex converging lens 3 of focal length 100 mm, transmitted through the planar glass window 4 to focus at the target 5 supported by probe 6 inside the source chamber in the vacuum environment of the MS.

Alignment and aiming are accomplished by arranging that the quasi-CW radiation from a 1 milliwatt Optics Technology Model 170 He-Ne gas laser 7, reflected off mirror 8, partially transmitted through splitter 2, and reflected off the ruby and mutually parallel mirrors in the laser head 1 returns to the point of origin at the gas laser 7. Since this alignment ensures normal reflection of CW radiation off the laser optics, it follows that the pulsed ruby radiation will be incident at the same point on the target as that illuminated by the portion of the CW radiation following the path 1-2-3-4-5. The point of impact of the laser radiation is adjusted by horizontal translation of the converging lens 3.

Each 100 μ sec the MS fires a 0.25 μ sec duration burst of ionizing electrons transversely across the source chamber immediately above that target. That portion of the material that is vaporized by the laser and

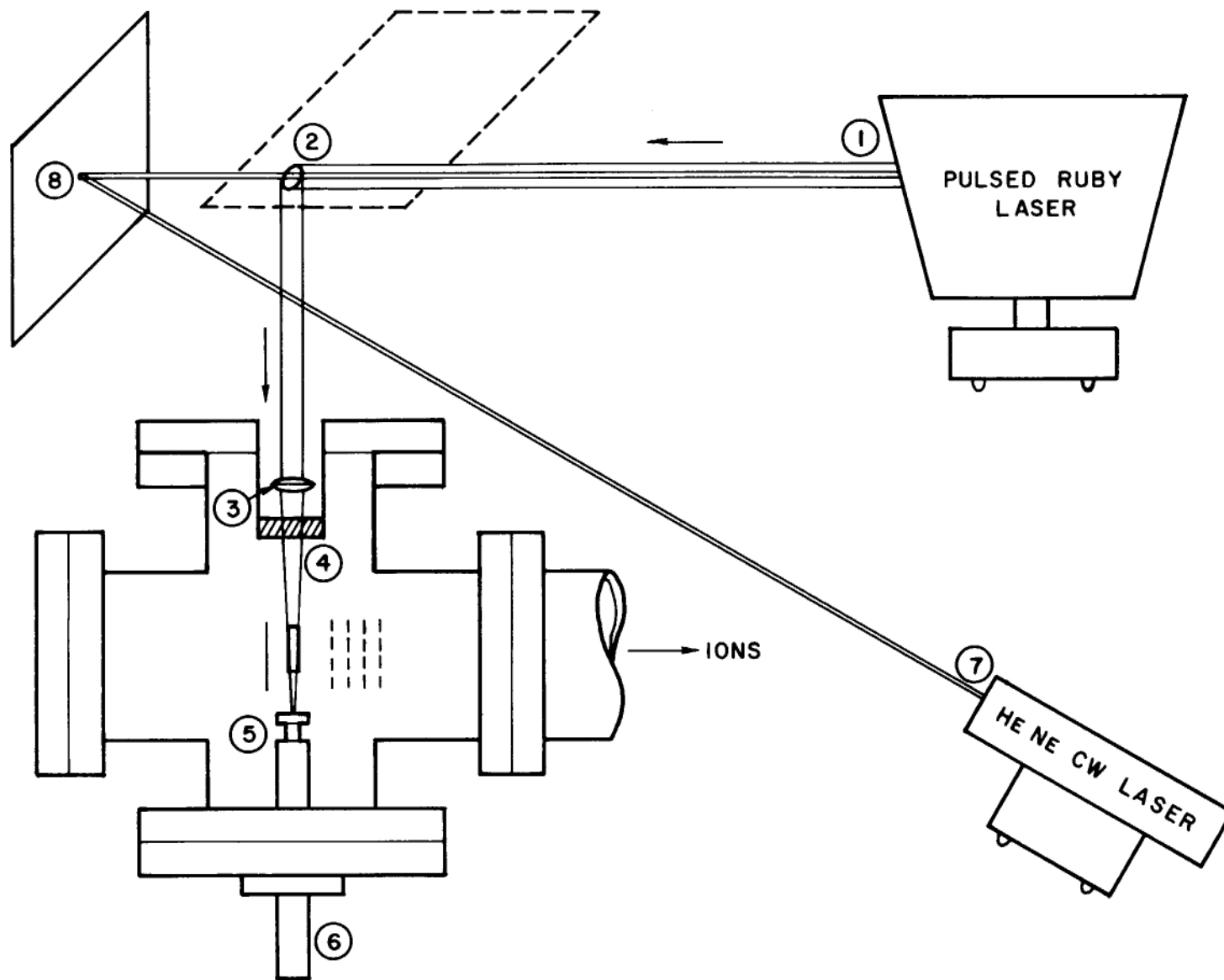


FIGURE 1.1

Schematic of Laser Optical Configuration

Used for Preliminary Studies

is positively ionized by each electron burst is impelled, by application of a 2700 volt kick, into a linear field-free drift space at the terminal end of which is a high speed ion detector.

The voltage kick V imparts to each ion a velocity v given approximately by

$$v \cong \sqrt{2neV/m} , \quad (1.a)$$

where n is the number of electrons removed from the molecule, e is the magnitude of the electronic charge, and m is the mass of the ion. The elapsed time between application of the kick and detection of arrival of an ion is given approximately by

$$t \cong \ell/v , \quad (1.b)$$

where t is the transit time down the drift tube of length ℓ .

Substitution of (1.a) into (1.b) yields

$$t \cong \ell \sqrt{m/(2neV)} . \quad (1.c)$$

Conversion of the mass-to-charge dependent velocity dispersion into ion time-of-arrival dispersion at the detector enables masses to be identified from the cathode ray oscilloscope (CRO) display of ion current versus time. The capture of maximum spectral information over one or a succession of MS repetition cycles (several of the 100 μ -sec intervals) has involved photographing the CRO display of output. Utilization of a procedure based upon the approximate quadratic relationship between m and t expressed in (1.c) has enabled us to associate mass numbers with recorded peaks out to the neighborhood of 400 atomic mass units (amu).

Figure 1.2 is a schematic representation of the initial version of the pulsing circuitry utilized external to the MS to enable the recording of either one or a series of successive spectra. The system allows for the establishment of a controllable delay between firing of the laser and the initiation of the CRO display. The number of successive spectra displayed can be freely varied. Successive spectra are vertically displaced from one another on the CRO. The joint resolution of the CRO and the 10,000 ASA Polaroid film is insufficient to enable unambiguous mass identification over the range 0-400 on a single trace. Recording of this mass range has generally required 5 to 10 laser shots, the mass range window for each shot being successively displaced to higher masses.

The mass spectral analysis of the vapor produced by laser irradiation of a powdered crystalline target of N-dinitrophenyl (DNP)-L-isoleucine is presented in Table 1.1. There is also presented, for comparative purposes in this table, the spectrum obtained by conventional crucible warming of the same sample to 40°C. The structural and graphic formulae for this sample are presented in Figure 1.3.

The photographic recording process and subsequent identification and evaluation of mass peaks is at present a time consuming process that does not readily lend itself to rapid scanning of a heterogeneous target. This limitation is for the moment, however, academic. We found that this initial experimental laser configuration appeared to deliver insufficient focused pulsed areal energy density to vaporize

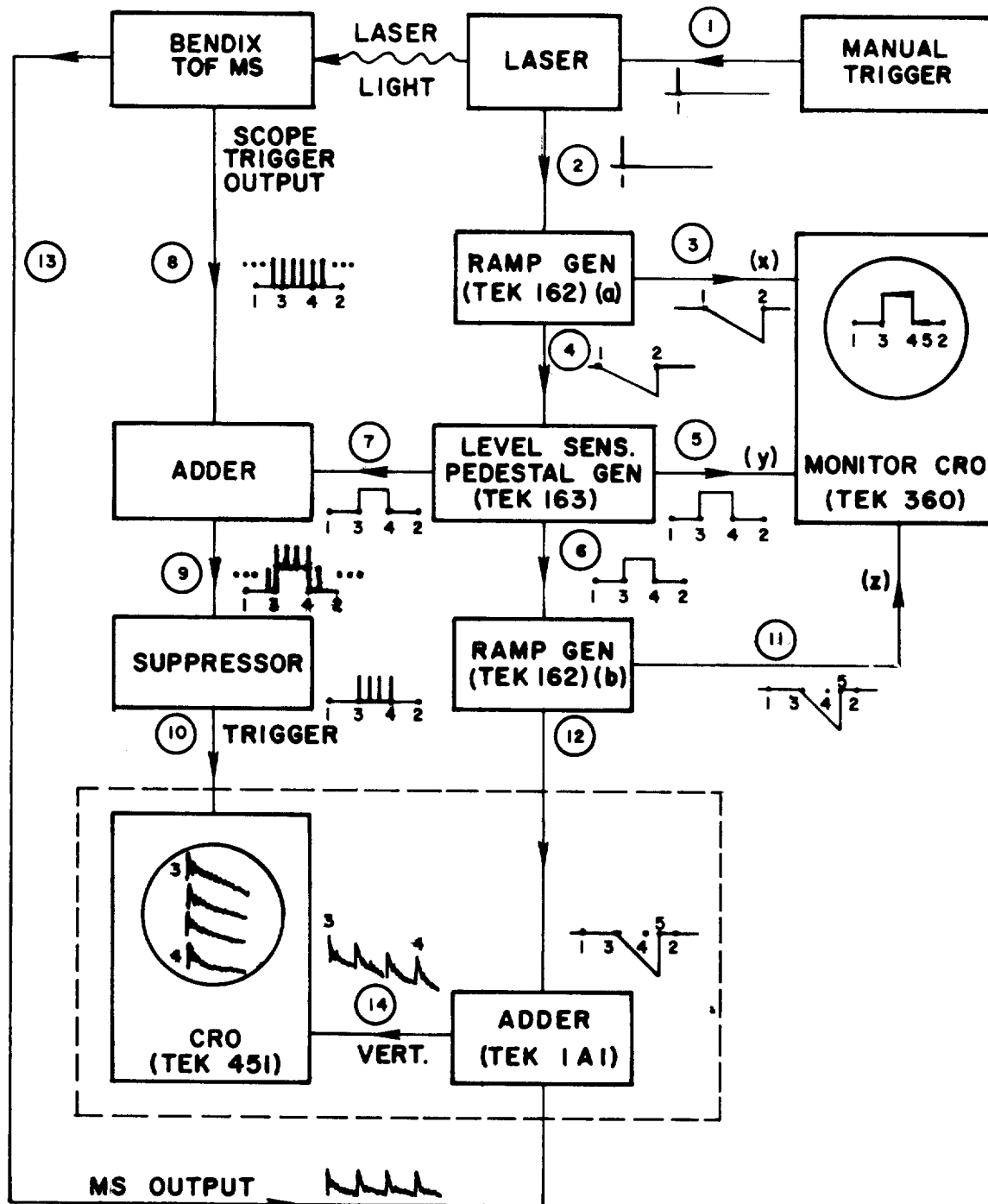


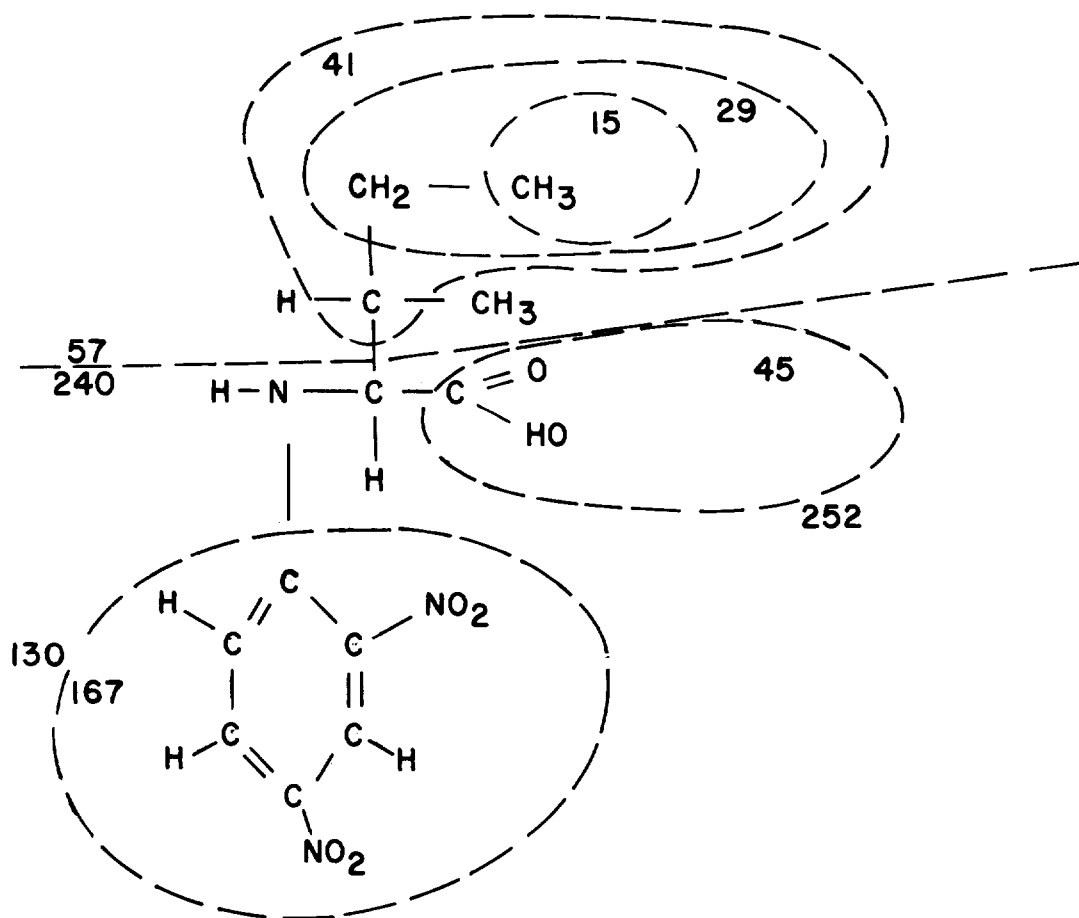
FIGURE 1.2

Block Diagram of Pulse Circuitry Utilized to Enable Recording of Either One or a Series of Successive Mass Spectra

Table 1.1. Relative Intensities of Laser Spectrum and Crucible (40 Deg. C) Spectrum of dinitrophenyl-L-Isoleucine.

Mass Laser Cruc.	Mass Laser Cruc.	Mass Laser Cruc.	Mass Laser Cruc.	Mass Laser Cruc.	Mass Laser Cruc.
1	0	1	51	14	1
2	0	1	52	25	1
3	0	0	53	23	1
4	0	0	54	22	0
5	0	0	55	40	2
6	0	0	56	8	0
7	0	0	57	100	3
8	0	0	58	7	0
9	0	0	59	11	0
10	0	0	60	0	0
11	0	0	61	3	0
12	0	0	62	10	0
13	0	0	63	38	1
14	3	5	64	28	0
15	4	0	65	14	0
16	0	5	66	8	0
17	3	20	67	14	0
18	14	100	68	8	0
19	0	0	69	45	1
20	0	0	70	7	0
21	0	0	71	0	0
22	0	0	72	0	0
23	0	0	73	5	0
24	0	0	74	12	0
25	0	0	75	48	2
26	0	0	76	19	1
27	46	2	77	20	1
28	67	100	78	20	1
29	100	7	79	0	0
30	62	1	80	0	0
31	12	1	81	0	0
32	28	37	82	11	0
33	0	0	83	0	0
34	0	0	84	7	0
35	0	0	85	0	0
36	0	0	86	0	0
37	0	0	87	5	0
38	0	0	88	0	0
39	55	2	89	0	0
40	22	2	90	11	0
41	100	7	91	0	0
42	18	1	92	0	0
43	24	2	93	0	0
44	0	1	94	0	0
45	37	1	95	0	0
46	0	0	96	0	0
47	0	0	97	0	0
48	0	0	98	0	0
49	0	0	99	10	3
50	22	0	100	0	3
			101	10	4
			102	14	1
			103	27	1
			104	17	0
			105	7	0
			106	5	0
			107	8	0
			108	0	0
			109	0	0
			110	0	0
			111	0	0
			112	0	0
			113	0	0
			114	8	0
			115	0	0
			116	0	0
			117	12	0
			118	3	0
			119	12	0
			120	2	0
			121	0	0
			122	5	0
			123	0	0
			124	0	0
			125	0	0
			126	0	0
			127	0	0
			128	0	0
			129	0	0
			130	7	0
			131	7	0
			132	3	0
			133	0	1
			134	14	1
			135	0	0
			136	0	0
			137	0	0
			138	0	0
			139	0	0
			140	0	0
			141	0	0
			142	0	0
			143	0	0
			144	0	0
			145	0	0
			146	0	0
			147	7	0
			148	7	0
			149	0	0
			150	0	0
			151	0	0
			152	0	0
			153	0	0
			154	0	0
			155	0	0
			156	0	0
			157	0	0
			158	0	0
			159	0	0
			160	0	0
			161	0	0
			162	0	0
			163	0	0
			164	10	0
			165	0	0
			166	11	1
			167	0	0
			168	0	0
			169	0	0
			170	0	0
			171	0	0
			172	0	0
			173	0	0
			174	0	0
			175	0	0
			176	0	0
			177	0	0
			178	0	0
			179	0	0
			180	0	0
			181	0	0
			182	0	0
			183	0	0
			184	0	1
			185	0	0
			186	0	0
			187	0	0
			188	0	0
			189	0	0
			190	0	0
			191	0	0
			192	0	0
			193	0	0
			194	0	0
			195	0	0
			196	22	1
			197	0	0
			198	0	5
			199	0	8
			200	0	12
			201	0	16
			202	20	3
			203	0	0
			204	0	0
			205	0	0
			206	24	0
			207	0	0
			208	0	0
			209	0	0
			210	0	0
			211	0	0
			212	0	0
			213	0	0
			214	0	0
			215	0	0
			216	0	0
			217	0	0
			218	0	0
			219	0	0
			220	0	0
			221	0	0
			222	0	0
			223	0	0
			224	13	0
			225	0	0
			226	0	0
			227	0	0
			228	0	0
			229	0	0
			230	0	0
			231	0	0
			232	0	0
			233	0	0
			234	0	0
			235	0	0
			236	0	0
			237	0	0
			238	0	0
			239	0	0
			240	8	1
			241	0	0
			242	0	0
			243	0	0
			244	0	0
			245	0	0
			246	0	0
			247	0	0
			248	0	0
			249	0	0
			250	0	0
			251	0	0
			252	72	5
			253	0	0
			254	0	0
			255	0	0
			256	0	0
			257	0	0
			258	0	0
			259	0	0
			260	0	0
			261	0	0
			262	0	0
			263	0	0
			264	0	0
			265	0	0
			266	0	0
			267	0	0
			268	0	0
			269	0	0
			270	0	0
			271	0	0
			272	0	0
			273	0	0
			274	0	0
			275	0	0
			276	0	0
			277	0	0
			278	0	0
			279	0	0
			280	0	0
			281	0	0
			282	0	0
			283	0	0
			284	0	0
			285	0	0
			286	0	0
			287	0	0
			288	0	0
			289	0	0
			290	0	0
			291	0	0
			292	0	0
			293	0	0
			294	0	0
			295	0	0
			296	0	0
			297	0	1
			298	0	0
			299	0	0
			300	0	0

EMPIRICAL FORMULA OF (DNP)-L-ISOLEUCINE: $C_{12}H_{15}O_6N_3$



GRAPHIC FORMULA OF (DNP)-L-ISOLEUCINE

FIGURE 1.3

Empirical and Graphic Formulae for (DNP)-L-Isoleucine.
 The dashed lines suggest non-rearrangement subgroupings of
 atoms that might give some of observed mass numbers.

many potentially interesting materials. We then undertook a re-instrumentation designed to rectify this deficiency.

Our reinstrumentation was intended to accomplish three purposes:

- (1) Reduction of the focused spot size;
- (2) More precise aiming of the laser;
- (3) An increase of the areal energy density delivered at the focus of the laser beam.

The new system, illustrated in Figure 1.4, has been constructed and is now in use. The design optimizes delivered energy density, with full energy delivery, by maximizing the solid angle of condensed radiation. Practical constraints necessitate a sufficiently long distance from the condensing lens to the working point that a lens aperture larger than the laser beam is required. The laser beam is expanded to fill the condensing lens by placing a diverging lens immediately in front of the pulsed ruby laser.

The system delivers approximately a 1 millisecc - 10 millijoule pulse of ruby radiation to an approximately 70 micron diameter spot. The present optical configuration involves the placement of a planar glass vacuum port within the cone of rays converging to the target. A significant contribution to the observed spot size derives from the spherical aberration contributed by the window.

We have found that the optical pulse is of insufficient magnitude to vaporize bulk reflective samples. Such samples are handled by depositing them in such a manner on a black oxidized copper support

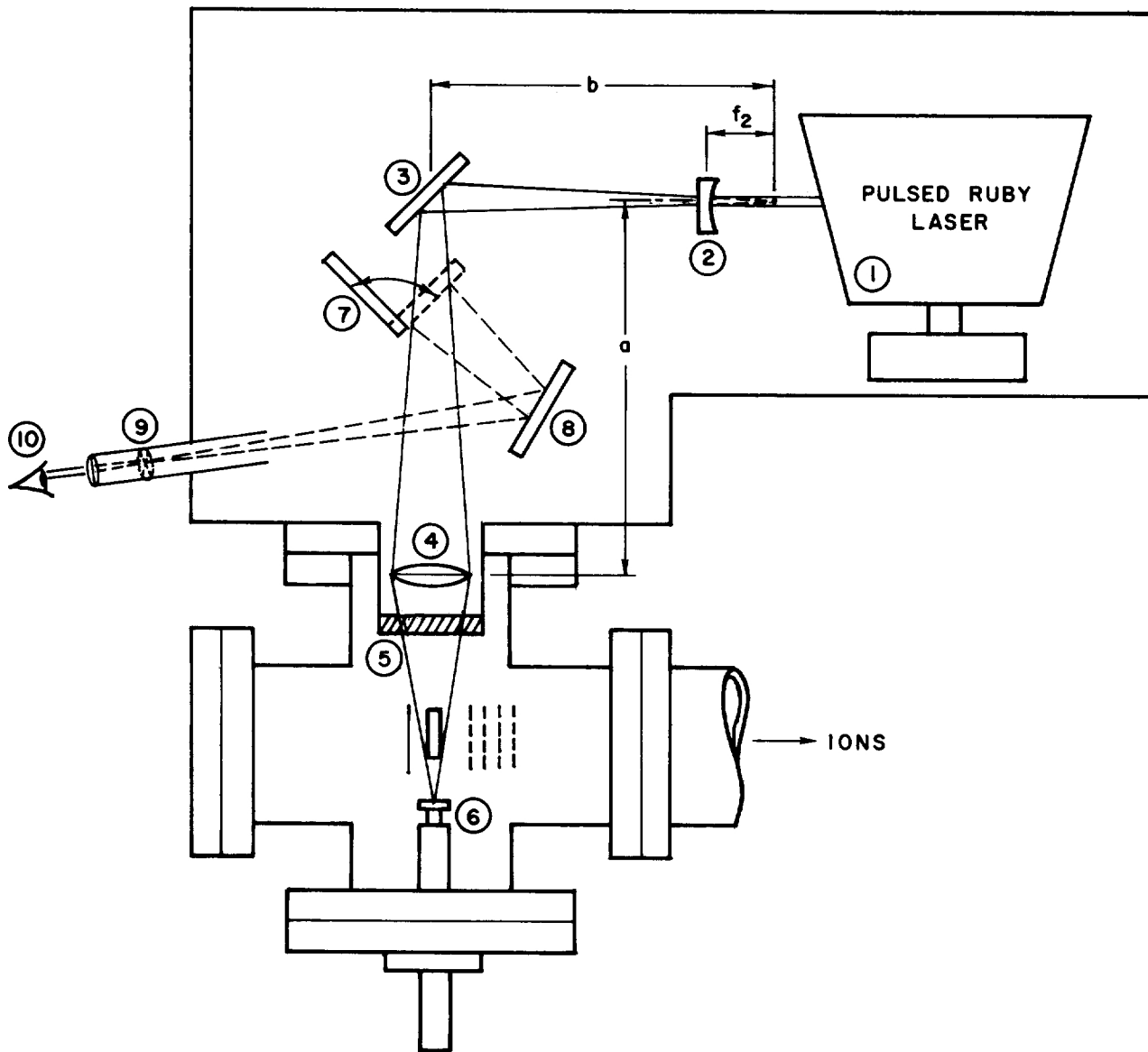


FIGURE 1.4
Schematic of Laser Optical System

that thermal vaporization is accomplished by thermal transfer from the copper to the sample.

Placement of the sample at the point of concentration of the laser radiation is achieved with mirror 7, in Figure 1.4, rotated to its phantom position. The sample target point is placed so that it is imaged at the cross hair 9 of the eyepiece. The optical configuration is such that the portion of the sample imaged on the crosshair will be at the focal point of the condensed laser radiation when the system is fired.

Figure 1.5 illustrates the probe that has been constructed to provide the three degrees of freedom required to properly position the sample. The probe consists of a hollow tube containing a centrally located stem penetrating through a 0.010 inch thick stainless steel diaphragm brazed to the stem and the tube. The sample is placed on the upper end of the removable stem tip. The shoulder on the probe base structure seats on the vacuum lock of the mass spectrometer source. The sample is raised or lowered relative to the shoulder by rotating a concentric wheel engaged to the threaded base of the tube through an intermediate annular key. The sample is displaced perpendicular to the tube axis by two mutually orthogonal micrometers that pivot the stem about the flexible diaphragm. The micrometer driven motion is reduced 10:1 at the sample. The diaphragm tolerates 2×10^{-2} radians of angular displacement from equilibrium before acquiring a set.

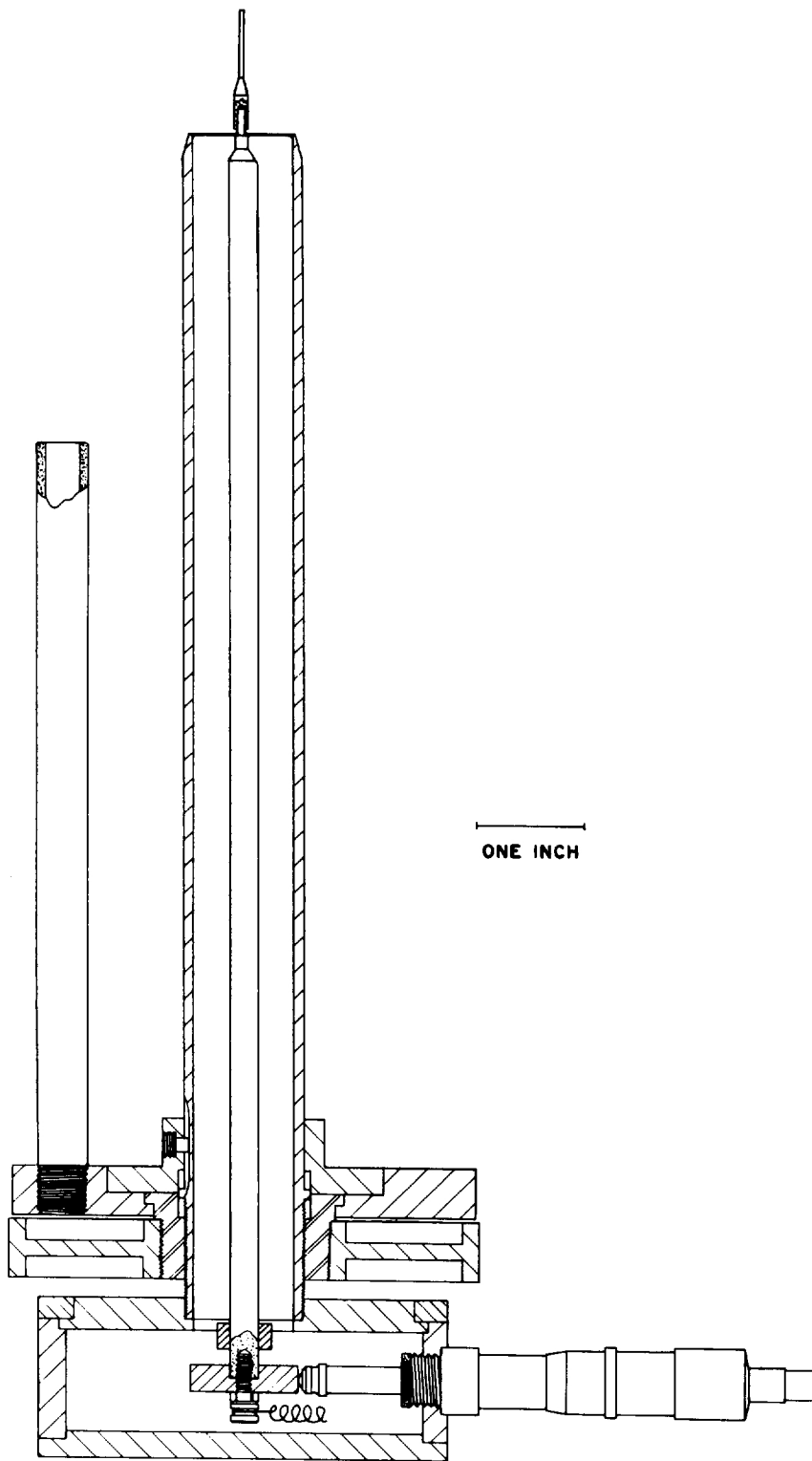


FIGURE 1.5

Mass spectrometer sample probe
for laser induced vaporization

Figure 1.6 is a schematic illustration of a control unit now being constructed to expedite the recording of output from the laser-mass spectrometer system. When the start button is pressed with the unit in "normal" function, a "background" number of reference gas traces is recorded across the top and then again across the bottom of the scope face, the laser is then fired, and after the designated number of mass spectrometer "cycles delayed" (100 microsec/cycle), the "sample" number of successive cycles of mass spectral analysis of the laser induced output is recorded in an equally spaced display between the upper and lower background traces.

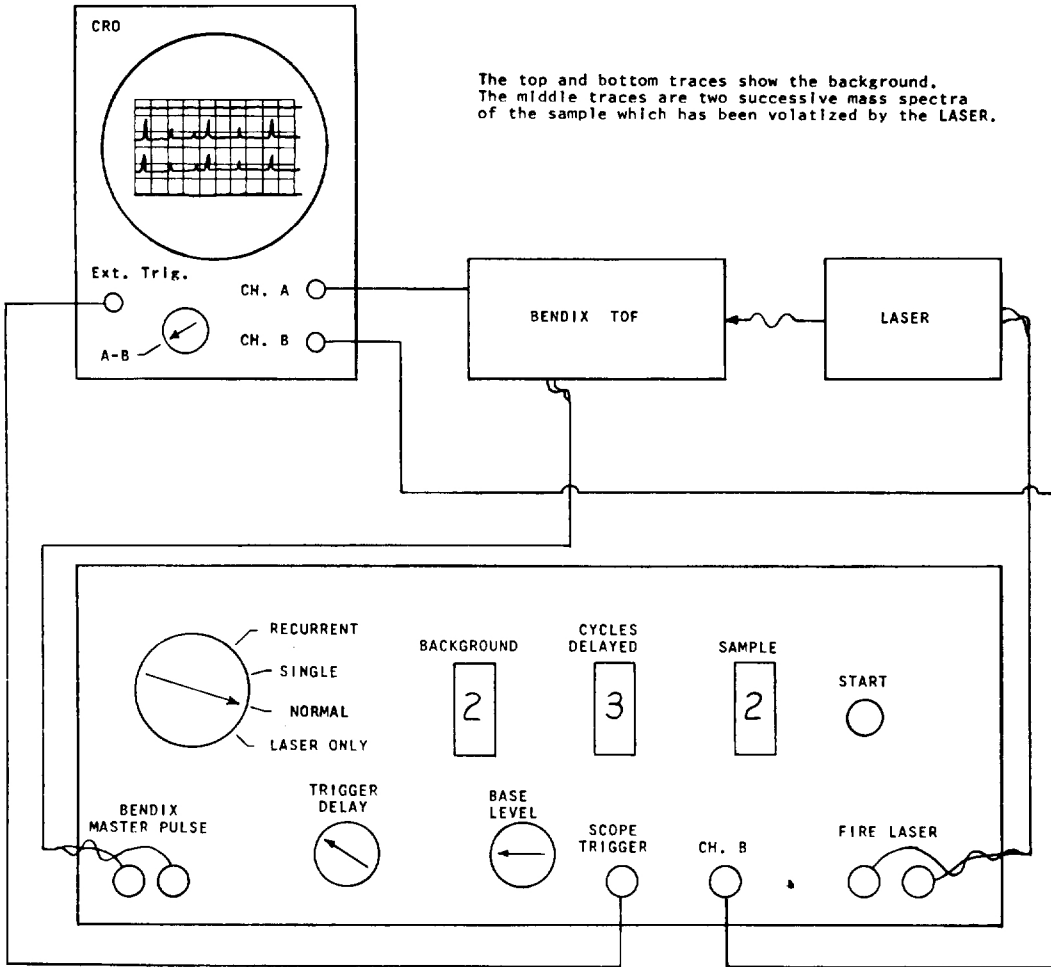


FIGURE 1.6

Control unit to expedite scope display of laser induced mass spectra

REFERENCES

- 1.1. B. Halpern, J. W. Westley, E. C. Levinthal and J. Lederberg, "The Pasteur Probe: An Assay for Molecular Asymmetry." Life Science and Space Research, North-Holland, Amsterdam (1967).
- 1.2. J. W. Westley, "Detection of Optical Activity as a Sign of Life." presented to the Amer. Astro. Soc., (1966). IRL-1047.
- 1.3. B. Halpern and J. W. Westley, "High Sensitivity Optical Resolution of Poly-functional Amino Acids by Gas Liquid Chromatography." Tetrahedron Letters 2283-6 (1966). IRL 1043.
- 1.4. B. Halpern, P. J. Anderson, J. W. Westley and J. Lederberg, "Demonstration of the Stereospecific Action of Microorganisms in Soil by G.L.C." Analytical Biochemistry (1966). IRL 1050.
- 1.5. B. Halpern and J. W. Westley, "High Sensitivity Optical Resolution of D,L Amino Acids by Gas Chromatography." Chem. Comm. 12: 246 (1965).
- 1.6. B. Halpern and J. W. Westley, "High Sensitivity Optical Resolution of D,L Amino Acids by Gas Chromatography." Biochem. Biophys. Res. Comm. 19(3) 361 (1965).
- 1.7. B. Halpern and J. W. Westley, "Resolution of Neutral D,L Amino Acids via Their L-Menthyl Ester Derivatives." Chem. Comm. 18:247 (1965).
- 1.8. B. Halpern, J. W. Westley, I. vonWredenhagen and J. Lederberg, "Optical Resolution of D,L Amino Acids by Gas Liquid Chromatography and Mass Spectrometry." Biochem. Biophys. Res. Comm. 20:710 (1965).
- 1.9. B. Halpern, J. Ricks, and J. W. Westley, "The Stereospecificity of α -Chymotrypsin-catalysed Reactions." IRL 1042.
- 1.10. B. Halpern, J. Ricks, and J. W. Westley, "The α -Chymotrypsin-catalysed Hydrolysis and Alcoholysis of Specific Ester Substrates in the Presence of Added Nucleophiles." IRL 1045.
- 1.11. B. Halpern and J. W. Westley, "High Sensitivity Optical Resolution of Amines by Gas Chromatography." Chem. Comm. 2:34 (1966). IRL 1041.

- 1.12. B. Halpern, J. Ricks, J. W. Westley, "The Stereospecificity of α -Chymotrypsin-Catalyzed Hydrolysis and Alcoholysis of Specific Ester Substrates." Aust. J. Chem., 20:389 (1967). IRL 1049.
- 1.13. 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. 39:228 (1967).
- 1.14. B. Halpern and J. W. Westley, "Chemical Resolution of Secondary (\pm) Alcohols." Aust. J. Chem. 19(8):1533-1534 (1966). IRL 1044.

II. Spectral Analysis of Monomers

In a preliminary study to provide insight into operational problems of computer control a comprehensive set of spectra of amino acids, as observed with solid samples in the Bendix time-of-flight instrument, was collected. This work is described in a technical report prepared by Miss Nancy Martin^{2.1}. This work has been supported jointly by a National Aeronautics and Space Administration Grant NsG 81-60; National Institutes of Neurological Diseases and Blindness, Grant No. NB 04270; and Air Force Grant No. AF-AFOSA-886-65. This Air Force Grant was the predecessor to the present Air Force Contract for which this final report is being prepared.

One of the most important problems of data acquisition, the calibration of mass numbers, had only begun to be handled by the computer system at this stage of development, and assignments given in this report must be regarded as tentative. Furthermore, no attempt was made to assess the ultimate sensitivity of the assay. Nevertheless, these data showed the potentialities of the technique, especially when the distinctive temperature characteristics of each amino acid are considered.

We commenced the mass analysis of selected organic samples with the completed laser optical system described in the previous section. Our first results demonstrated that we were readily able to detect the announced apparent molecular peaks for the four nitrogenous bases - adenine, guanine, thymine, and cytosine. We also demonstrated that the base peaks are conspicuous for the nucleosides deoxyadenosine,

deoxyguanosine, and deoxycytidine. We also found peaks at the base masses in the two nucleotides we have so far examined, deoxyguanosine monophosphate and thymidine monophosphate.

Table 2.1 is a listing of the spectra that have been obtained and are available in the computer data file for display and comparative analysis. Section IV of this report includes a discussion of this aspect of the work.

REFERENCES

- 2.1. N. Martin, "An Investigation of the Mass Spectra of Twenty-Two Free Amino Acids." IRL-1035, September 21, 1965.

Table 2.1

B-Subtilus - DNA
Salmon Sperm DNA - Crucible
Adenylic Acid - Bendix Crucible
Deoxycytidine - Bendix Crucible
3M-FC43 - Bendix Crucible
Deoxyguanosine - Bendix Crucible
Deoxyadeninosine - Bendix Crucible
Deoxyadenosine Monophosphate - Crucible
Thymidine Phosphate - Crucible
Thymine - Crucible
Cytosine - Crucible
Guanine - Crucible
Adenine - Crucible
Deoxycytidine - LAS
Deoxyguanosine - LAS
Deoxyadenosine - LAS
Deoxyadenosine Monophosphate - LAS
Deoxyguanosine Monophosphate - LAS
Thymine - LAS
Thymidine Phosphate - LAS
Cytosine - LAS
Guanine - LAS
Adenine - LAS
Salmon Sperm DNA - LAS
P₂O₅ - LAS
Dextrose LAS
Clean Copper Probe

III. Computer Control of Mass Spectrometers

The question of computer control of mass spectrometers is central to our ultimate objective. The computers used in this part of the program are a LINC and an IBM 360/50. The Instrumentation Research Laboratory was assigned a LINC computer in 1963 as a participant in the LINC evaluation program under the sponsorship of the NIH, NASA, and U.S. Air Force. This was administered under an NIH grant FRO0151-01. The IBM computer is part of the Advanced Computer for Medical Research (ACME) program supported by the NIH, Division of Research and Facilities Resources under Grant FROO 311-01. We have applied these computational and computer control facilities to a Bendix time-of-flight mass spectrometer, an Electronic Associates, Inc., quadrupole mass spectrometer, and an MS-9 high resolution mass spectrometer.

References 3.1 and 3.2 describe the development of some of the software which has been used in this research.

The design and construction of hardware and software for an ACME 360-LINC communication gave us the capability of processing mass spectrometer data gathered by the LINC on the 360/50. This overcame the handicap posed by the limited computational ability of the LINC. This limitation results from the LINC having a 12 bit integer arithmetic system. Most data manipulation steps, therefore, must be done with software routines for double-precision, floating-point arithmetic. The result was that arithmetic operations took place at millisecond rather than microsecond speeds.

With this facility we used the 360 to transfer the data collected from the Bendix TOF mass spectrometer into a more usable form. A computational process which required 20 minutes to complete on the LINC was found to require only two minutes when done under the ACME time-sharing system on the 360. In this case the data was sent to the 360, operated on in allocated "time slices", and returned to the LINC for display on the LINC oscilloscope. About half of the two minutes required was the result of the tape reading and writing operations on the LINC.

A computer interface and computer operating system for the Electronic Associates Inc. (EAI) QUAD 300 mass spectrometer was completed in June 1967. In this system the computer exercised control and acquired spectra from the QUAD 300 quadrupole mass spectrometer. It allowed the analysis of gas chromatograph effluents using a mass spectrometer as a detector. The scientific results of these efforts are discussed in the first section of this report and the associated references.

The details of this computer controlled gas chromatograph-mass spectrometer system are separately reported (3.3). The instrumentation consisted of a quadrupole mass spectrometer, a special small rack of electronics, which were termed the "interface", a LINC computer, and a computer software system developed in our laboratory. The following is the abstract of that report.

"A mass spectrometer-computer system has been devised to utilize the decision-making capabilities of the modern digital

computer. The system described assists the researcher user by allowing a computer to query the researcher for operating parameters. The computer translates these into detailed control functions that operate the instrument. The data acquired from the mass spectrometer is made available to the researcher in an on-line system. The system employs a small digital computer and an integer resolution quadrupole mass spectrometer. A reference gas is valved into the mass spectrometer by computer control to permit automatic calibration. Spectra processing of GLC effluent was demonstrated; the means and results are given."

The design of the QUAD 300 computer interface made provision for the use of the electronic hardware and the computer programs with the Bendix time-of-flight mass spectrometer as well as the QUAD 300. In addition to its use with the QUAD 300 this interface equipment was installed with the Bendix TOF.

A voltage boosting circuit was installed in the Bendix analog scan unit to control the Bendix from the computer interface. This and other rework on the Bendix was done in a manner to allow either conventional or computer operation.

The Bendix time-of-flight could be operated from and by the computer in a manner virtually identical to that accomplished on the QUAD 300. The system allows a reasonably fast data acquisition,

nominally 4 to 8 seconds per spectra, and an on-line plot or presentation of the spectral measurements. The data acquisition time is compatible with obtaining mass spectra during the period of sample peaks in a gas chromatograph effluent. However, as yet no gas chromatograph has been connected to the TOF mass spectrometer.

The completed system of computer control, data acquisition and data presentation is very economical of the mass spectrometer operator's time. It does allow more reliable data acquisition in a fraction, perhaps as little as one tenth, of researcher and technician attention to the mass spectrometer, chart interpretation, and manual data reduction.

Some progress has been made on a project for direct computer acquisition of high resolution mass spectra from a MS-9 mass spectrometer and for eventual computer control of the system. This included the installation of a remote data terminal (270-Y) of the ACME 360/50 system adjacent to the MS-9 mass spectrometer in the Chemistry Department.

The ACME IBM 360/50 is a time-sharing computer located in the Medical Center and has a short cable connection to the Instrumentation Research Laboratory. The MS-9 is physically located in another building 1500 feet (cable length) away. The 270X-Y is a special IBM remote data access system designed to operate at these distances. The intent of the MS-9 development is to take the electrical output of MS-9 detector and, via suitable conversions and transmissions, sample it in

real time with the ACME computer.

The required 270X-Y system has just recently become usable from the ACME terminals. It must still be tested at the rate of data transmission required for the MS-9 operation. The ACME software system must make provision for allocation of the large data storage space required. Both of these requirements are within the specifications of the respective systems. The first, the technical ability of the hardware to achieve the speed, is not expected to cause any problems. The software system necessary to accommodate this data within the framework of time-shared computer use is not as straightforward. This software system is still under development.

REFERENCES

- 3.1. R. K. Moore, "An Operating System for the LINC Computer". Technical Report No. IRL-1038, 1965.
- 3.2. R. B. Tucker, T. Coburn, W. Reynolds, and J. Bridges, "Software for the LINC Computer". Technical Report No. IRL-1055, 1967.
- 3.3. W. E. Reynolds, J. C. Bridges, T. B. Coburn and R. B. Tucker, "A Computer Operated Mass Spectrometer System". Technical Report No. IRL-1062, 1967.
- 3.4. ACME Progress Report to NIH, October 1, 1966 - April 1, 1967, NIH, Division of Research Facilities and Resources under Grant FRO0311-01.

IV. Data Retrieval and Display

Much of what we have done concerning the problem of data retrieval is covered in the previous sections of this report and in the references cited in those sections. The references cited in this section are for background purposes only and predate the period of this contract.

We have not directly confronted the problems posed by the requirement to handle a very high bandwidth data base that would be generated by the ultimate mass spectral image scanning applications we envisage. We have, however, designed, constructed and used systems which allow us to explore with some efficiency small sections of this stream of data. Although it is without the elegance that we ultimately seek and requires a great deal more effort than might be desirable, this does permit an evaluation of the suitability of other elements of the system such as the method of volatilization and the mass spectrometer itself for ultimate biological applications.

Mass spectra from the Bendix TOF are sent through a logarithmic amplifier then digitized and stored on magnetic tape by the LINC computer. A transformation is applied to the time axis of the spectra to place the mass positions at equal intervals on this time axis. Because of certain drifts in the TOF, the mass values of this altered spectrum cannot be accurately determined simply by their position on the new scale. A method was developed for determining their position which was quite time consuming with the limited computational abilities of the LINC and will have to wait for the implementation of the ACME system for practical usage.

A method has been developed for identifying integer mass peaks by utilizing the display scope of the LINC computer to display a portion of a spectrum and a raster of the peak positions. This method displays a portion of the spectrum (after the rough linearization) on the screen with a generated raster indicating the approximate separation of the mass positions (see Figure 4.1). The raster can be stretched and translated using the potentiometers on the LINC console to fit the peak positions of the displayed spectrum. This approach combines the user's ability to distinguish meaningful peaks with the computer's capacity to calculate the raster, store the digitized spectrum, and store the information about the mass peaks once the user has adjusted the raster to fit the particular part of the spectrum being displayed.

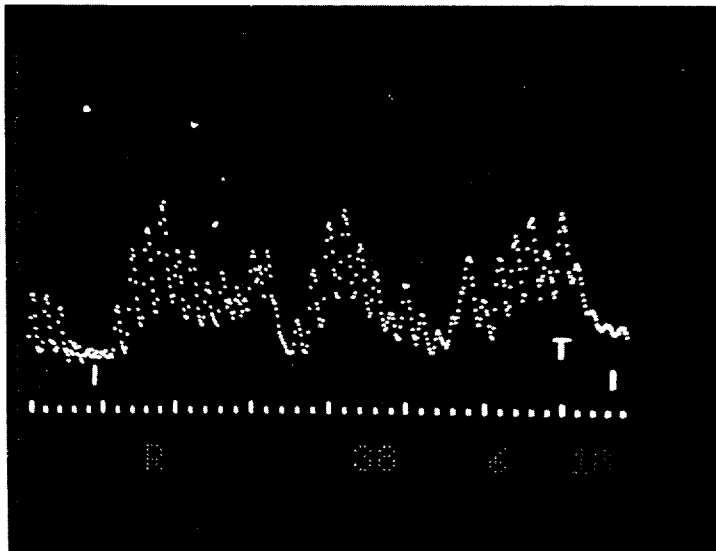


FIGURE 4.1

The program is operated under the locally written LOSS monitor system (see reference R. K. Moore, "An Operating System for the LINC Computer", Technical Report No. IRL-1038) and uses its conventions for handling the data tape (the spectrum). Upon loading the program, the first portion of the spectrum (actually the transformed spectrum as described above) is read into the memory and displayed on the oscilloscope (8 cm by 8 cm display area). About 40 mass positions are within view of the user. Initially the user positions an illuminated "pointer" over a known mass peak and enters the associated mass value using potentiometers on the console. Lifting a console switch identifies this position in the spectrum with the particular mass positions equal to zero modulo five being enhanced. Thus a raster is developed for the set of peaks based on the particular position and mass number entered through the potentiometers. The distance between the elements of the raster may now be expanded until the elements and the mass peaks coincide. This expansion takes place about the reference mass previously mentioned so as not to disturb its position. The user can, however, translate the entire spectrum to improve the matching between the raster and the observed mass peaks by adjusting another potentiometer.

With the matching accomplished the researcher can now either have the amplitudes at the mass positions typed out (still in logarithm form) or, by setting a switch recorded on magnetic tape for further processing (see below). The mass positions considered in this operation are those between a set of vertical bars on the screen which act as parentheses

around the masses under consideration. The user has complete flexibility in determining the portion of the displayed peaks to type or record or may investigate an area more than once to observe changes due to altering the raster. In general, the latter will yield little change since the program searches half a peak on either side of the raster position to find the maximum in the signal.

Having investigated the displayed portion of the spectrum the user moves the reference mass up to a position in the rightmost quarter of the screen by moving the pointer to a mass position on the raster and lifting a toggle switch. Lifting another switch will move the data on the rightmost quarter of the screen to the extreme left and read in the next portion of the spectrum. The raster can again be adjusted and another set of mass values typed or recorded.

In actual practice it has been found that only a slight amount of raster adjustment is needed (less than a peak width) in any one portion of the spectrum and thus the user can move along quite rapidly. This latter condition prompts one to think about automatic adjustment of the scale using an algorithm which examines the spectrum to find the spacing for each successive portion. While this would work well in the lower part of the spectrum (say up to mass 75), the large voids in the higher mass regions make these methods undependable. Once familiar with the system, a user can cover the mass positions up to 300 in three or four minutes if the output is being stored on tape rather than typed.

Once the data is on magnetic tape in the form of mass number and amplitude number pairs, it is readily usable in a number of ways. It can be put on standard seven channel digital tape or possibly sent directly to another computer as intended under the ACME concept. Presently the data is used as input to another LINC program which takes the antilog of the mass amplitudes and produces a bar graph (see Figure 4.2) in which the largest peak is considered to have value 100 and all other peaks are scaled accordingly.

We are commencing to use the Stanford Medical Center IBM-360 ACME time-shared computer system for the acquisition, storage, retrieval, manipulation, visual display, and comparison of mass spectral data.

Mass spectra obtained by laser induced vaporization of solid samples are supplemented, when possible, with the spectra that result from heating the samples in a quartz crucible. The photographic records of the laser induced spectra are visually interpreted and manually transcribed into the computer. The data obtained by crucible heating is generated slowly enough that it may be readily tape recorded. In the latter case a LINC computer is currently used to convert the tape recorded information into a table of mass number vs. intensity, and this information is then transmitted to ACME. We are about to start operating, in the crucible mode, with a computer generated mass scan.

We have been interested in experimenting with the visual comparison of spectra. We have just completed a program which enables this comparison to be performed by observation of a television display driven by ACME. Figure 4.3 shows the display format that appears when the program is first called.

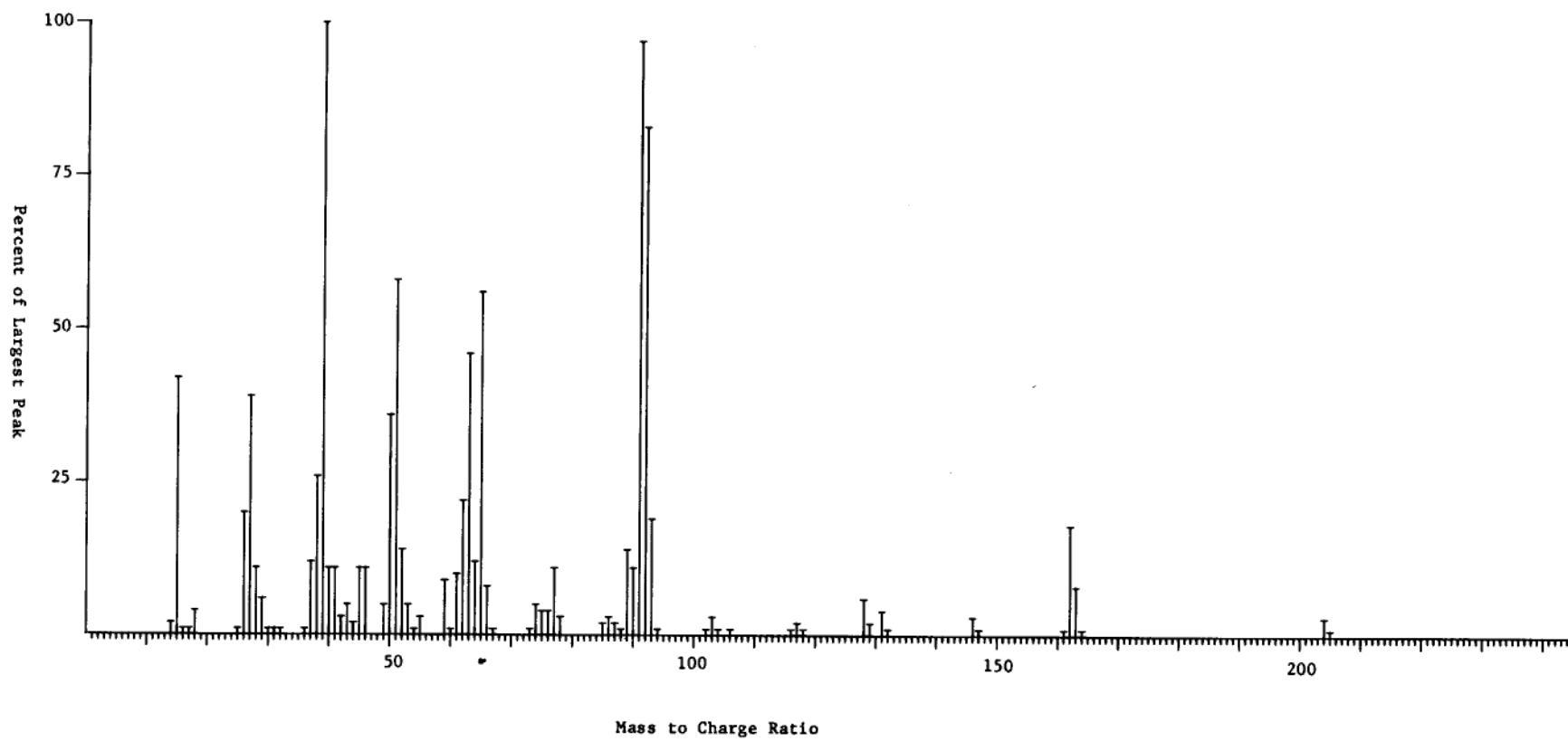


FIGURE 4.2
Mass Spectrum of Phenyl Alanine Methyl Ester HCL

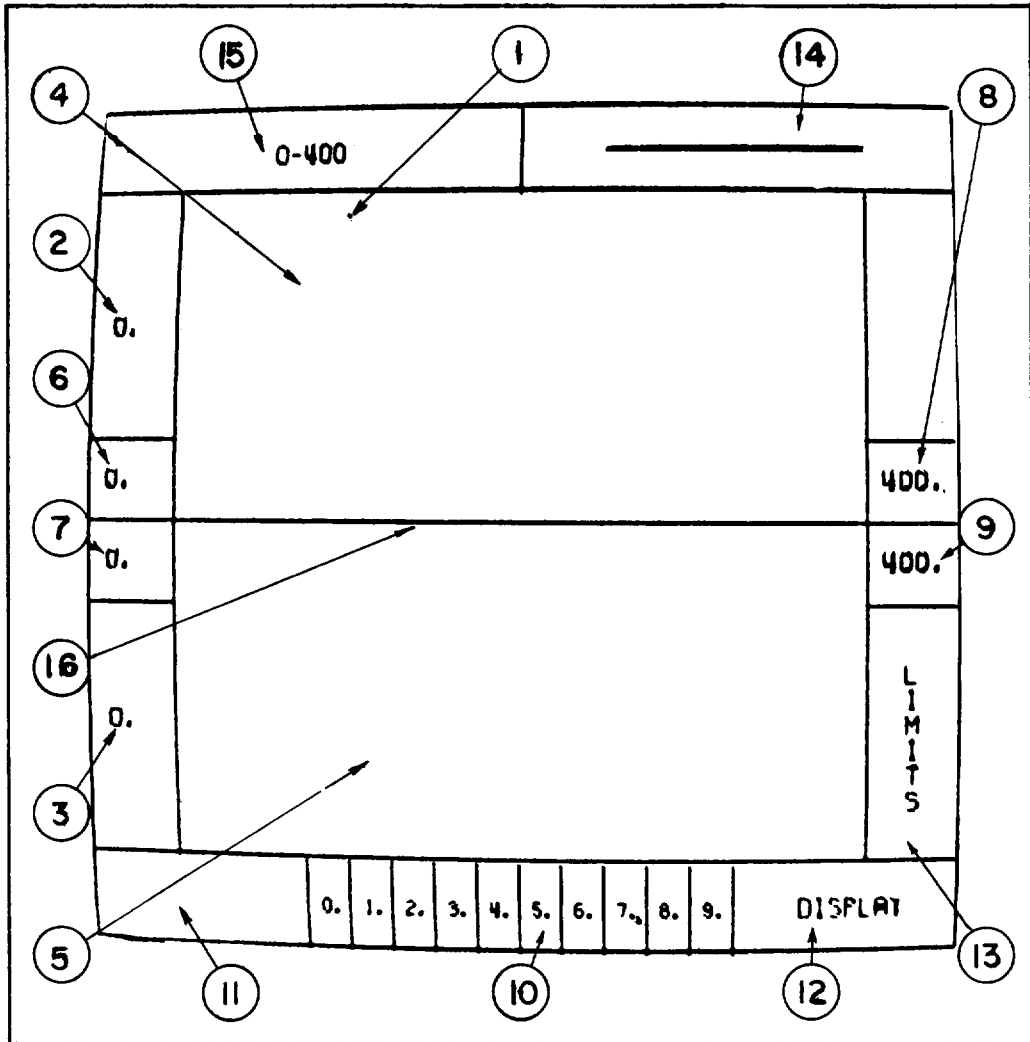


FIGURE 4.3

The image consists of a square pattern divided up into a set of rectangular zones. The straight line 16 is the common base line for two spectra that may be simultaneously displayed. The mass peaks for the spectrum displayed in region 4 will extend upward from the base line and for the spectrum displayed in zone 5 they will extend downward. For each of the two line spectra the mass numbers increase to the right. The numbers in zones 6 and 8 denote the lower and upper limits, respectively, of the mass range for the upper spectrum. The numbers in zones 7 and 9 serve a similar function for the lower spectrum. The numbers in zones 2 and 3 identify the record numbers of the upper and lower spectra, respectively.

All control in execution is accomplished by the use of a deck mounted gearshift-like control with a sensing button on the end of it. For historical reasons, this unit and the blinking spot 1 that it positions on the display are referred to as "mouse". The horizontal and vertical coordinates of the spot 1 on the screen are sensed by potentiometers coupled to the manual control. Reading of these coordinates occurs when the depressed sensing button is released.

When it is desired to select or change either a spectrum record number or the mass limits, mouse is moved into the area consisting of the set of zones 10, numbered 0 to 9. A multiple digit number may be collected from region 10 by successive selections and mouse then used to deposit this collected number in zone 11 for checking. The flashing of bar 14 advises the user that the computer is waiting for mouse. This information is useful when the computer is in heavy use. The selected number may be either corrected or transferred to any one of zones 2, 3,

6, 7, 8, or 9 by directing mouse to the desired zone and releasing the button. After selection is made, mouse may be used to sense zone 12, labeled "DISPLAY", thereby signaling the computer to display the selected spectra over the chosen mass ranges. The photograph reproduced in Fig.4.4 demonstrates the comparison of the base guanine obtained by use of the laser, record 31 displayed up, and the guanine spectrum obtained by use of the crucible, record 35 shown down. Sensing mouse above the base line and at a particular mass peak causes the "v"-like symbol 17 to appear at the mass peak and the value of the mass peak, 151 in this case, to be displayed at location 18. Mass 151 is the guanine molecular peak. Sensing below the base line causes the inverted "v" denoted 19 to appear at the mass peak and the associated mass number 202 to appear at location 20. Mass 202 is one of the mercury isotopes arising from the use of mercury as a pump fluid.

The photograph reproduced in Fig. 4.5 helps to illustrate some of the options available to the user. Record 37, salmon sperm DNA volatilized by the laser system, is displayed over the mass range 0-400 in the upper spectrum. The lower spectrum is a portion of the same spectrum showing in further detail the masses ranging from 100 to 150. The base peaks for cytosine (mass 111), thymine (mass 126) and adenine (mass 135) are present. Guanine evidently got lost in the noise on this shot.

Figure 4.6 illustrates the manner in which the display shown in Fig. 4.5 was obtained. Spectrum 37 in the mass range 0-400 was first

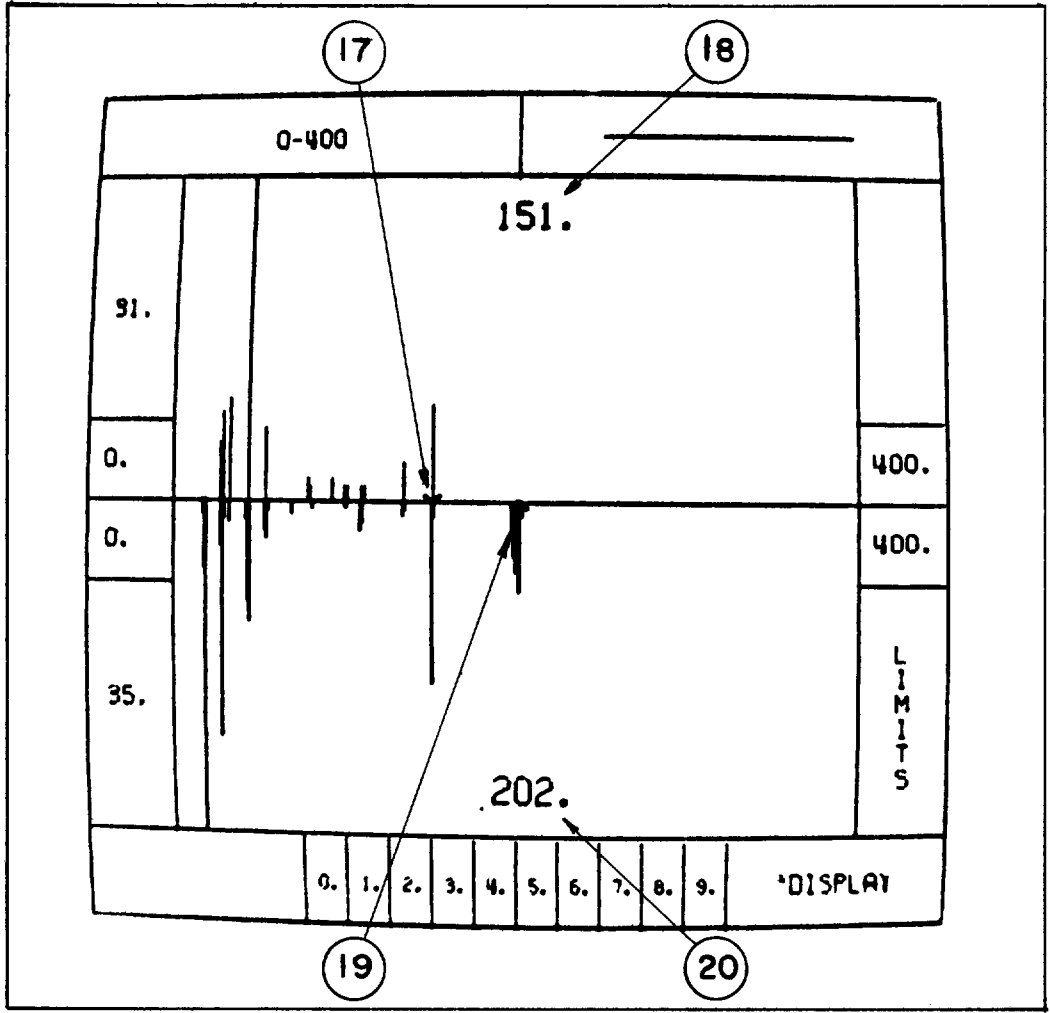


FIGURE 4.4

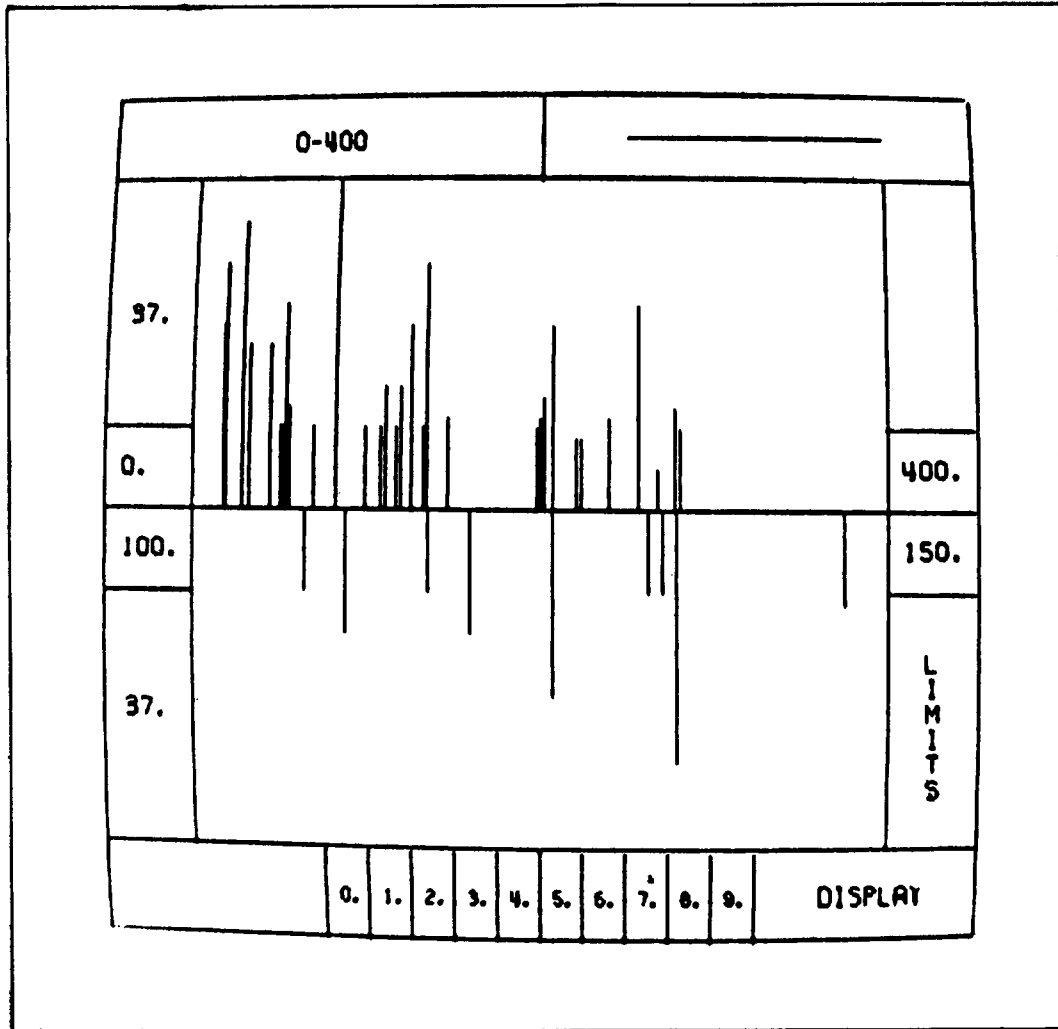


FIGURE 4.5

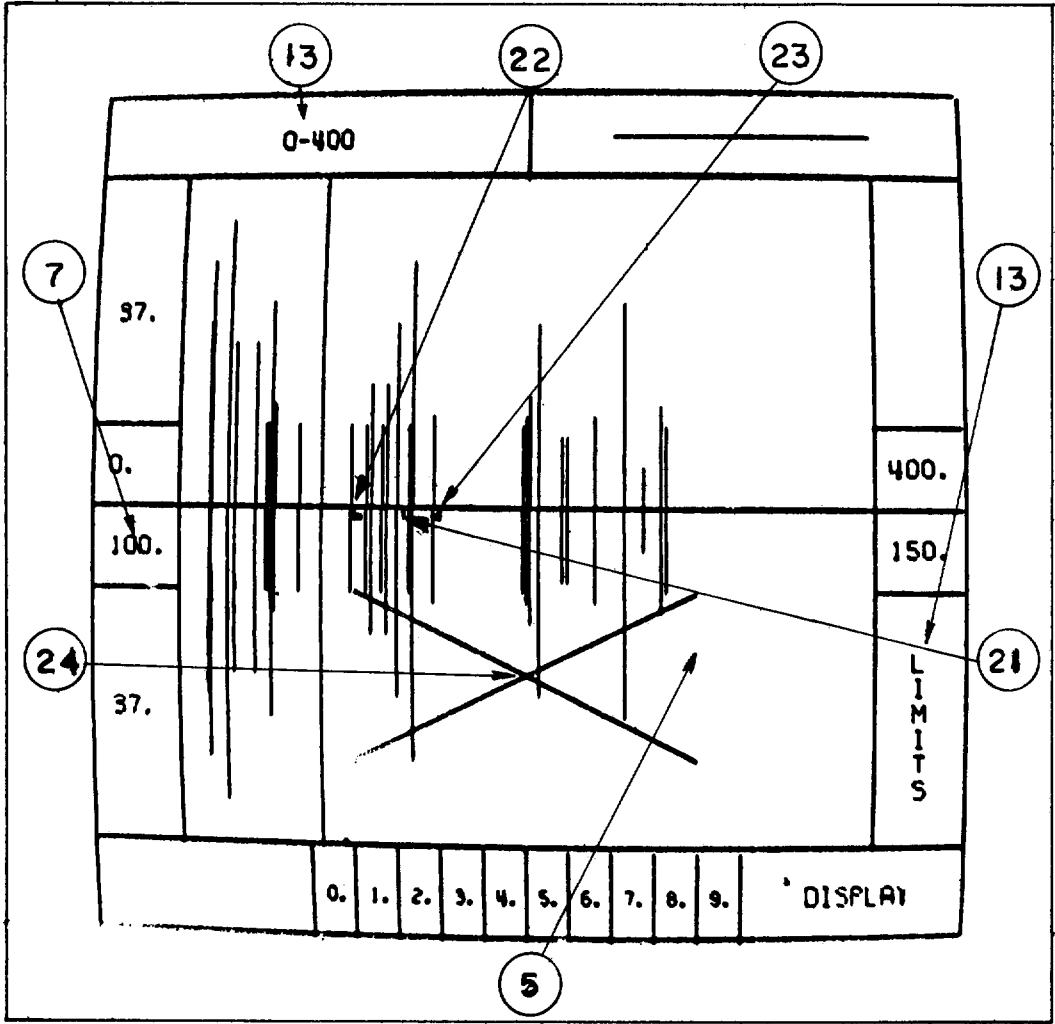


FIGURE 4.6

displayed both up and down. Mouse was then sensed in region 13, labeled "LIMITS". Following the sensing of LIMITS, mouse was sensed in zone 5 at that mass position intended to be an upper limit for the lower bound on the desired expanded mass range. An enhanced bar such as the one shown at 21 then appeared at the position chosen by mouse. Sensing mouse in the lower limit zone 7 then placed the newly selected lower limit in this zone. Because of the discreteness of available point sites on the digitally driven television screen the lower limit is automatically rounded to the nearest multiple of 25 that does not exceed the mass identified by the enhanced bar. Mass 100 resulted. An "L"-like symbol, denoted 22, is created to show the mass location of the new lower bound on the future spectrum. A flashing "X" appears at 24 to tell the user that the displayed spectrum is no longer valid. Sensing mouse successively in the LIMITS zone, in the mass zone, and in the high limit zone provides the new upper limit 150.

Sensing DISPLAY then results in the appearance of display shown in Fig. 4.5. If mouse is subsequently sensed in zone 13, labeled "0-400" the upper and lower limits of either of the spectra may be restored to the range 0-400 by a subsequent sensing of mouse either above or below the central base line of the display.

REFERENCES

- 4.1. D. G. Luenberg, "Resolution of Mass Spectrometer Data." IRL-1021, November 1964.
- 4.2. W. Reynolds, "The Use of a Logarithmic Amplifier in Data Processing of Analog Signals." IRL-1017, April 1, 1965.
- 4.3. J. F. Gibbons and H. S. Horn, "A Circuit with Logarithmic Transfer Response over Nine Decades." IEE Transactions on Circuit Theory, CT-11 (September 1964).

V. Computer Manipulation of Chemical Hypotheses

While the high resolution mass spectrometer is perhaps the most capable single instrument for organic structural analysis, the sheer volume of its signal output poses formidable problems of data reduction and data analysis. These problems would be multiplied by the number of samples that would need to be processed by the micro-scanning mass spectrometer which is our ultimate goal. At one level, the problem is the identification of mass numbers with compositional formulas. However, no mass spectral signal is free of noise and great effort must then be spent to obtain an accurate determination of mass to ultimate resolution. Much of this effort is wasted when it does not answer a concrete question, i.e., which of a set of possible compositions is indicated by a given measurement. Even for all compositions, the corresponding mass numbers are not continuously distributed; they are rather the discrete set of numbers calculated from linear integral sums of nuclidic masses, and represented in the tables.

The tabulations and calculation programs (references 5.1, 2,3, 4, 5, 6, 7, 8) are the first step in a control program for the mass spectrometer. As soon as the peak is identified within a given mass neighborhood, the competing possibilities should be computed, then weighted in accordance with any other available information. This allows the experimental problem to be restated as a choice among competing possibilities, and the signal information need be accumulated only long enough to lead to a meaningful choice among them.

In solving a structure, the chemist hypothesizes a series of trial structures, then matches them with the data (in this case a mass spectrum, but this can be generalized to any data set) and accepts or rejects his trial solutions, usually part by part, in a structure. Much of this tedious effort could be emulated or at least assisted by the computer in a program we call "mechanized induction".

For this purpose, a language has been devised for representing chemical structures in easily computable form; "Dendral '64" (reference 5.2 and 5.5). The development of this language required the filling of a surprising gap; the systematic application of simple topological principles to the field of chemical graphs - that is, a symbolic representation of organic molecules. Existing notations were found to be quite defective as the chemist already knows too well from his difficulties with nomenclature (other organizations like Chemical Abstract Service also recognize the problem and are working on it, but tend to compromise topological rigor for the benefit of established traditions in notation). At any rate, with the help of some theorems on canonical forms of trees, and on Hamilton circuits of planar maps (for acyclic and cyclic structures respectively), a complete system has been worked out. This gives an algorithm by which the computer can generate an exact list of all isomers in a given composition.

By itself this is a futile approach to any but the simplest problems, since the number of possible isomers quickly exceeds the range of a fast computer. Heuristic and symbiotic methods are therefore

called for whereby the computer emulates or cooperates in the use of human problem solving techniques in searching wisely selected parts of the space of possible solutions. Professor Edward Feigenbaum and Dr. Richard Watson of the Computer Science Department participated in a cooperative effort to program efficient displays of structural ideas for conversational interaction with the computer. This was a step to evaluate the chemists problem solving heuristics incorporating them in the machine program. The effort was limited at present by the existing computer facilities (a PDP-1 machine) with inadequate displays.

This work described above is covered in detail in the references indicated and the additional items in the bibliography below. A complete description of the current status is given in an article by J. Lederberg and E. A. Feigenbaum^{5,8}. The following is the abstract of that paper.

"A computer program for formulating hypotheses in the area of organic chemistry is described from two standpoints: artificial intelligence and organic chemistry. The Dendral Algorithm for uniquely representing and ordering chemical structures defines the hypothesis-space; but heuristic search through the space is necessary because of its size. Both the algorithm and the heuristics are described explicitly but without reference to the LISP code in which these mechanisms are programmed. Within the program some use has been made of man-machine interaction, pattern recognition, learning, and tree-pruning heuristics as well as chemical heuristics which allow the program to focus

its attention on a subproblem and to rank the hypotheses in order of plausibility. The current performance of the program is illustrated with selected examples of actual computer output showing both its algorithmic and heuristic aspects. In addition some of the more important planned modifications are discussed."

REFERENCES

- 5.1. J. Lederberg, "Systematics of Organic Molecules, Graph Topology, Hamilton Circuits." IRL 1040.
- 5.2. J. Lederberg, "DENDRAL-64. A System for Computer Construction, Enumeration and Notation of Organic Molecules as Tree Structures and Cyclic Graphs. Part 1." NASA CR-57029. STAR No. N65-13158, IRL-1036, 1964.
- 5.3. J. Lederberg, "Tables and an Algorithm for Calculating Functional Groups of Organic Molecules in High Resolution in Mass Spectrometry." NASA - STAR No. N64-21426, IRL-1019, 1964.
- 5.4. J. Lederberg and M. Wightman, "A Subalgol Program for Calculation of Molecular Compositional Formulas from Mass Spectral Data." NASA - STAR No. N66, 11689, 1964.
- 5.5. J. Lederberg, "Topological Mapping of Organic Molecules." Proc. Nat. Acad. Sci. U.S. 53: 134-139 (1965),
- 5.6. J. Lederberg, "Calculation of Molecular Formulas in High-Resolution Mass Spectrometry." An appendix to Structure Elucidation of Natural Products by Mass Spectrometry (Volume 11: Steroids, terpenoids, sugars, and miscellaneous classes), by H. Budzikiewicz, C. Djerassi and D. H. Williams. Holden-Day, Inc., San Francisco, 1964.
- 5.7. J. Lederberg and M. Wightman, "Calculation of Upper Limit of Hydrogens of an Organic Formula for Analysis of Mass Spectra." Anal. Chem. Soc. 36: 2365 (1964).
- 5.8. J. Lederberg and E. A. Feigenbaum, "Mechanization of Inductive Inference in Organic Chemistry", Symposium on Cognition, Carnegie Institute, John Wiley & Sons, in press (1967).