

AD-A126 761

MASS SPECTROMETRIC RAPID DIAGNOSIS OF INFECTIOUS
DISEASES(U) STATE UNIV OF NEW YORK AT BUFFALO SCHOOL OF
MEDICINE M ANBAR ET AL. JAN 79 DAMD17-78-C-8035

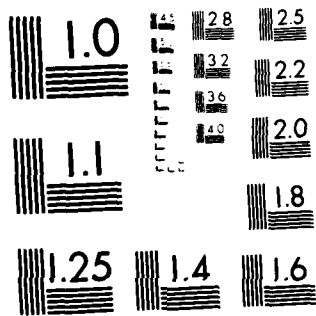
1/1

UNCLASSIFIED

F/G 6/5

NL

END
DATE
FILMED
5 - 85
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

ANNUAL PROGRESS REPORT



Back
6C'

ADA 126761

Michael Anbar
James McReynolds
Michael Scanlon
William Rieth

January 1979

MASS SPECTROMETRIC RAPID DIAGNOSIS OF INFECTIOUS DISEASES

Supported by

U. S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Maryland

Contract DAMD17-78-C-8035

School of Medicine
State University of New York at Buffalo
Buffalo, New York

Approved for public release; distribution unlimited.

The findings of this report are not to be construed
as an official Department of the Army Position unless
so designated by other authorized documents.

UNCLASSIFIED

Department of Biophysical Sciences
School of Medicine
State University of New York at Buffalo
Buffalo, New York 14214

83 04 12 142

DTIC
ELECTE
APR 13 1983
S D D

DTIC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM																					
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-112676	3. RECIPIENT'S CATALOG NUMBER																					
4. TITLE (and Subtitle) MASS SPECTROMETRIC RAPID DIAGNOSIS OF INFECTIOUS DISEASES		5. TYPE OF REPORT & PERIOD COVERED Annual Progress Report 6/1/78 - 1/31/79																					
7. AUTHOR(s) Michael Anbar Michael Scanlon James McReynolds William Rieth		6. PERFORMING ORG. REPORT NUMBER																					
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Medicine State University of New York At Buffalo Buffalo, New York 14214		8. CONTRACT OR GRANT NUMBER(s) DAMD17-78-C-8035																					
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Medical Research & Development Command Fort Detrick, Maryland		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	12. REPORT DATE 1/31/79																				
14. MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office)		13. NO. OF PAGES 18																					
16. DISTRIBUTION STATEMENT (of this report) Approved for public release, distribution unlimited		15. SECURITY CLASS. (of this report)																					
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE																					
19. SUPPLEMENTARY NOTES		<table border="1"> <tr> <th colspan="2">Accession For</th> </tr> <tr> <td>NTIS GRA&I</td> <td><input checked="" type="checkbox"/></td> </tr> <tr> <td>DTIC TAB</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Unannounced</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Justification</td> <td></td> </tr> <tr> <td colspan="2">By _____</td> </tr> <tr> <td colspan="2">Distribution/ _____</td> </tr> <tr> <td colspan="2">Availability Codes</td> </tr> <tr> <td>Dist</td> <td>Avail and/or Special</td> </tr> <tr> <td>A</td> <td></td> </tr> </table>		Accession For		NTIS GRA&I	<input checked="" type="checkbox"/>	DTIC TAB	<input type="checkbox"/>	Unannounced	<input type="checkbox"/>	Justification		By _____		Distribution/ _____		Availability Codes		Dist	Avail and/or Special	A	
Accession For																							
NTIS GRA&I	<input checked="" type="checkbox"/>																						
DTIC TAB	<input type="checkbox"/>																						
Unannounced	<input type="checkbox"/>																						
Justification																							
By _____																							
Distribution/ _____																							
Availability Codes																							
Dist	Avail and/or Special																						
A																							
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Clinical diagnosis, infectious diseases, metabolite profiles, urine, field ionization mass spectrometry, multicomponent analysis, molecular weight profiles.																							
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Using new chemical procedures, mass spectrometric instrumentation and appropriate computerized data analysis, the diagnosis of infectious hepatitis, through the molecular weight profile of neutral metabolites urine, was demonstrated,																							

I. INTRODUCTION

This is the first annual progress report on a research program entitled, "MASS SPECTROMETRIC RAPID DIAGNOSIS OF INFECTIOUS DISEASES", under Contract No. DAMD177808035, sponsored by the Department of the Army, U.S. Army Medical Research and Development Command, Fort Detrick, Maryland. This first progress report covers just the period from June 1, 1978 - January 31, 1979 and is, therefore, rather limited in scope.

The capability of making a rapid and reliable diagnosis of infectious diseases at an early stage and at low cost would be of especially great value to the military where large numbers of soldiers are often stationed in confined areas and their continuing health is crucial to carrying out their objectives. Early and reliable diagnosis of an infectious disease could prevent the spread of disease to large groups of soldiers and civilians on the post.

Multicomponent analysis may be used to identify in the host's reaction a characteristic metabolic pattern associated with general infection, with bacterial or viral infection, or with specific infections. The multiscan mass spectrometric method offers three types of uses in the diagnosis of infectious diseases. First, multicomponent analysis by mass spectrometry may be used as such a diagnostic tool. Second, the characteristic components identified by the pattern recognition approach, can be chemically characterized by the FI-CID technique leading to an understanding of the biochemical nature of the host's reaction. Third, the quantitative determination of a small number of identified metabolites by non-mass spectrometric analytical techniques (e.g., glc, hplc or specific fluorometric determinants) may prove advantageous for routine diagnosis from the standpoint of cost per analysis.

During the first 7 months of this second phase, we have achieved a number of critical objectives. All of the mass spectrometric systems to be

used in this project have been put into routine operation and a new dedicated computer system, acquired by SUNY/AB, has been interfaced with each of the mass spectrometers and is now being used with the multiscanning mass spectrometer, thus significantly augmenting our data handling capabilities. This computer system has also been interfaced with the SUNY central computer, which carries out our diagnostic data analysis. A new field ionization source has been developed which is significantly sturdier and also less expensive than our previous source. The sample preparation techniques have been retested and reevaluated by the new personnel and, with slight modifications, found adequate for routine analysis. A series of diagnostic tests of the analytical procedures has been carried out to ascertain the reproducibility and to determine the reliability of the methodology. Finally, examination of clinical samples has been started and will continue in the following months. If we obtain worthwhile clinical findings within the coming few weeks, we shall submit them as a supplement to this proposal. We have delayed the start of clinical analysis to allow us to interface the new computer system, which is going to be used routinely on this project. To analyze clinical samples without the computer would have required a major effort to translate and debug CYBER programs to reduce accumulated data from a marginally adequate 1024 channel analyser system. Although in our original proposal we planned to start using the dedicated computer in the second year of the project, we are now about 9 months ahead of schedule in this respect.

II. SUMMARY OF ACHIEVEMENTS

During the first phase of this program we have accomplished the following tasks: (documented in the Final Report of 1977)

1. Selection and development of mass spectrometric instrumentation capable of rapid and reliable multicomponent analysis.

2. Development of rapid and reproducible sample extractions and concentration techniques which enable routine urine multicomponent analysis.
3. Development (in a parallel effort) of instrumentation which allows the chemical characterization of components of interest - those that comprise a characteristic metabolic pathological pattern.
4. Development of appropriate statistical data-handling techniques which facilitate the extraction of diagnostic information from metabolic profiles.
5. Preliminary demonstration of differentiation between bacteria through the metabolic profile of their homogenates.
6. Demonstration of diagnosis of infectious hepatitis through multicomponent analysis of the acidic as well as the neutral metabolites in urine.
7. Demonstration of diagnosis of urinary and of pulmonary infections through multicomponent analysis of neutral metabolites in urine.
8. Demonstration of a general diagnostic pattern associated with infections.
9. Demonstration of differential diagnosis of the patients suffering from the 3 types of infections (hepatic, pulmonary and urinary).
10. Demonstration of the ability of the methodology to differentiate in the same urine between two superimposed pathological patterns.

During the first 7 months of the current phase of the program we have accomplished the following:

1. Recruited and trained adequate project personnel.
2. Got the equipment transferred from California fully operational.
(This involved complete reassembly of two of the instruments plus extensive laboratory renovation to provide essential utilities).

3. Developed a simpler, more rugged and much less expensive field ionization source, adequate for routine multicomponent analysis.

4. Translated the computer programs developed at SRI into languages supported by the CYBER SUNY computer system.

5. Adapted and rechecked the sample preparation procedures for urine samples.

6. Developed an adequate sample preparation procedure for plasma samples.

7. Interfaced the FI multiscanning mass spectrometric system with the INCOS ROVA computer system.

8. Interfaced the INCOS computer with the CYBER system.

9. Tested the new systems including the sample preparation procedures for reproducibility and potential sources of variances.

10. Collected urine samples from patients (children and adults) with a variety of infectious diseases which will be examined in the coming months.

In the following sections we shall describe in some detail a number of these accomplishments.

A. Ion Source Improvements

A new type of field ionization source has been developed that has a number of advantages for analysis of physiological samples. The activated foil of our previously described source has been replaced with a brush of $\sim 20-50$ graphite fibers of 8 micron diameter (Union Carbide "Thornel") mounted on a stainless steel foil with silver conducting paint. A narrow sample feed path past the evenly cut ends of the multiple fibers is defined by a second foil on the opposite side of the source vedge as in the activated foil source. (See 1977 Report). A counterelectrode consisting of a slit 0.25 mm wide or an 80 line per inch nickel grid is placed 250 microns above the ends of the fibers. This is a wider spacing than previously used with the multipoint or activated

foil sources and an ionizing potential of 5 kV is required to obtain optimum ionization efficiency. The larger spacing is necessary to prevent shorting due to broken fibers lodging between the brush and counterelectrode.

An entire source can be assembled in one-half hour and no activation is required for operation. The sensitivity of the source is in the range of 10^{-12} coulombs/ μ g which is approximately 10 - 20 times less than the maximum obtained for activated foil or multipoint sources. Since some of the latter sources lost their efficiency during continued operation, requiring repeated activations, the new stable source is superior in spite of its relatively lower sensitivity. Analyses of over 100 samples of plasma and urine extracts, including some samples of unextracted dried plasma and urine have been performed on a single source with no deterioration in sensitivity.

B. Computerized Data Acquisition System

Within the last month we have adapted our spectrometer for use with a Finnigan Model 2400 data acquisition system. An analog scan signal is used directly to drive a hall probe controlled magnet power supply. A scan of 15 sec up and 3 sec down between 1 and 450 amu with a 1 sec hold time at the upper and lower ends, has resulted in good long-term mass assignment stability.

Statistical analysis can be applied over the extended mass range, making use of any additional diagnostic peaks not found in the 50 - 350 amu range previously used. In order to allow reliable mass assignment for field ionization spectra we use a seven compound calibration mixture, volatile at room temperature, covering a mass range from 73 to 298 amu. These FI calibrations, taken at the end of every profile analysis, are used to assign masses to the most recent multicomponent sample, and to monitor any instrument drift between samples.

To allow for more efficient recalibration and change of scan parameters, we plan to install a combined EI/FI source so that recalibration over the

full range using PFK (perflurokerosene) can be performed when desired.

Since the time-temperature profile of each peak is preserved by the data system, we will develop procedures to test the diagnostic value of multiple components at single nominal masses resolved in time by their volatility differences. Using the above scan parameters, a single pure component is now being assayed in less than 10 scans, and 80 scans are used to obtain a full multicomponent molecular weight profile; therefore, several components contributing to the same nominal mass peak may be resolved, each carrying potential diagnostic information.

C. Implementation of Data Analysis Procedures

The objective in processing the experimental data is to determine the degree of validity of the null hypothesis that any observed differences in mass spectra of samples obtained from the pathological and control groups are due to chance. Although it can never be proven that this hypothesis is incorrect, it is possible to determine the odds that this hypothesis is wrong and that, consequently, the mass spectral differences are due to real metabolic differences.

The validity determination requires that (1) raw spectral scans be combined; (2) their masses and areas converted to nominal and normalized values respectively; (3) these data be suitably reformatted for statistical processing; and (4) the Wilcoxon P-values and WNI (weighted non-correlation index) values be obtained. (See the 1977 Report). Below is a brief description of the progress made over the last 6 months in satisfying these four requirements.

Requirements (1) through (3) can all be met by use of our newly-acquired Finnigan-INCOS model 2400 Mass Spectrometer Data System. Although ordered early in the current contract year, delivery and pre-acceptance servicing and adjustment by Finnigan staff were completed by the beginning of January. Ex-

perience in using the system to edit, combine and convert the individual scans has been developing over the past weeks, and will so continue as project personnel learn to employ more of its subtle and important capabilities.

Our dedicated system also is being used to reformat our normalized, FORTRAN-readable data into the sequential form required by our PASCAL-based Wilcoxon statistical program recently developed to substitute for the SRI program. No utility program for this "laundering" of FORTRAN-to-PASCAL format is available on the Finnigan 2400, so the NOVA minicomputer resident in the data system was programmed to accomplish this. In addition, other programming, circuit changes, RS-232 data interfacing and modem installation were completed during this period.

Considerable programming effort has been expended in converting the ALGOL Wilcoxon program (brought from Stanford Research Institute) to a PASCAL-based program suitable for use on the University's CYC CYBER 170 computer. The program has been completed and initial tests indicate successful implementation, based on correct P-values obtained from test sample input data.

Plans for the next period include the refinement of programming to allow redundancy-check data transmission from the dedicated system to the CYBER computer using more extensive count and check-sum information. Also to be developed is a FORTRAN version of the weighted non-correlation index algorithm that will be suitable to carry out an optimization study of weighting functions for the WNI. It is expected that some effort also will be devoted to interfacing the data system with a high-speed multiplex data link to the CYBER, when this feature is made available in the next few months by University Computer Services Department. On the other hand, if the INCOS system will prove to be capable of carrying out the whole pattern recognition procedure, the need for the CYBER use may become superfluous.

D. Preparations of Urine Samples

The method used for separating the organic components from human urine is similar to the one used in our 1977 report. The method makes use of a chromosorb P column to retain the highly polar inorganic portions of the urine while allowing the organic components to pass through the column with the eluting dichloromethane. (See 1977 Report). NaCl saturated urine (1.5 ml) is loaded onto the chromosorb column with nitrogen pressure. A 5 ml volume of dichloromethane (DCM) is then eluted through the column, also under nitrogen pressure. The DCM is collected in a separate vessel and most of the solvent is removed under a stream of N₂. The organic residue is dissolved in about 100 μ l of DCM and is placed on a much smaller chromosorb capillary column for introduction into the mass spectrometer. The entire process, aside from column preparation, takes 25 minutes. This procedure offers greater speed than earlier techniques and greater reproducibility than simple extractions.

Since the last report we have made slight changes in the design of our chromosorb column to assure reproducibility. A smaller volume of glass wool is now used at the beginning of the column to retain the chromosorb. The column (135.0 mm x 8.0 mm O.D. x 6.2 mm I.D.) is packed tightly with a larger volume of chromosorb (2.4 grams). One gram of anhydrous Na₂SO₄ is now packed on top of the chromosorb and the end of the column is tamped with another small glass wool plug. The columns are then carefully washed with several volumes of methanol and DCM and baked at 200°C for 24 hours.

An appropriate amount of NaCl saturated urine (1.5 ml) is then loaded onto the column. This volume is sufficient to just wet the entire chromosorb. The fully wetted column reduces variation caused by reabsorption of materials from the eluting DCM back onto the dry chromosorb. Such reabsorption, may have been responsible for pattern fluctuations sensitive to in flow

during elution. We have found that a fully wetted column produces good reproducible yields with less flow dependent fluctuation in the metabolites isolated. The Na_2SO_4 added to the far end of the column is used to hold back any water that might otherwise accompany the DCM leaving the column. The smaller volume of glass wool at the beginning of the column assures that most of the urine is applied to the chromosorb before elution.

The pH of the urine can be adjusted before absorption onto the chromosorb. At present, we are proceeding with the isolation products of neutral urines which give both good yields and information-rich "fingerprints".

E. Preparation of Plasma Samples

We have tried several procedures for preparing human plasma samples for mass spectrometric analysis.

The simplest and most direct method of plasma preparation consisted of absorbing 50 μl of plasma onto a small capillary column filled with glass beads. The column provides a large surface for absorbing precise amounts of plasma. After absorption, the plasma was dried with a gentle stream of nitrogen, and the sample was introduced to the mass spectrometer via a solid probe specially adapted to accept these 12 m x 1 mm O.D. x .8 mm I.D. columns. This straightforward method may prove suitable for certain applications.

This technique is not, however, without its limitations, for it is indiscriminate in its presentation of plasma metabolites. We sought to rectify this situation and gain a greater degree of control over the metabolites represented in the sample, by using a variety of precipitation, extraction and column techniques. We were particularly concerned with the contribution of cholesterol which contributed a group of large metastable peaks in the 379-385 amu region and which might obscure some small peaks in the same mass region. After confirming the identity of the compound causing the inter-

ference, by the use of CID, we have developed a precipitation procedure to remove the plasma cholesterol.

Cholesterol was removed by precipitation with digitonin. A 1% digitonin (Baker) in absolute alcohol solution was prepared. Two mls of the 1% solution were added to one ml of plasma in a centrifuge tube and mixed on a vortex mixer. The suspension was centrifuged for 10 minutes at 3500 rpm. The cholesterol-free supernate was then poured off. The ethanol of the digitonin solution precipitated some of the plasma protein along with the cholesterol. The removal of the remaining protein was effected by the addition of 8 mls of absolute alcohol. The supernatant and the additional ethanol were mixed and centrifuged as above. This now cholesterol- and protein-free supernatant was removed and concentrated under a stream of nitrogen. The concentrate was then loaded on glass beads or chromosorb and introduced into the mass spectrometric solid probe for analysis.

We have found this procedure effective for two reasons: (1) the cholesterol peaks are selectively removed from the spectrum; (2) the protein-free plasma makes possible a wide variety of further manipulation not previously possible without the interference of protein.

A wide variety of solvents and conditions are now being used in column and extraction methodologies. We have proceeded with further purification of the plasma by chromosorb P and XAD resin columns as well as with extraction with organic solvents. We have acquired considerable expertise with these methods and are well prepared to isolate particular classes of metabolites from plasma. Preliminary results indicate that (1) precipitation of cholesterol by digitonin as described effectively removes the cholesterol without the addition of artifacts; (2) protein precipitation does not markedly alter the mass spectral profile of the low molecular weight components in the sample.

F. Evaluation of Reproducibility of Analytical Procedure

At this time we have analyzed replicate samples from a single individual to determine reproducibility of both the mass spectrometric analysis and the total experimental procedure. Initial tests of the Wilcoxon Test and spectrum normalization programs have been run on the CYBER using previously analyzed test data. The Wilcoxon Test and the weighted non-correlation index (WIN) programs will be used to analyze our current data as soon as we accumulate a sufficient number of control and pathological spectra. A preliminary analysis of the recently analyzed clinical samples has been performed using the library search programs of the mass spectrometer data system.

In order to test the pattern variance due to the mass spectrometric procedure, an 18 ml urine sample from a single individual was prepared in a scaled-up version of our extraction procedure. The 100 ml dichloromethane extract was concentrated to approximately 4 mls. Twelve sample capillaries were loaded by applying approximately 50 μ ls of the concentrated extract to each capillary. This is equivalent to about 15 - 25% of the material extracted from a normal 1 to 1.5 ml urine sample. The total counts for each of these samples was approximately 10^5 and the smallest peak in any sample was greater than 100 counts.

The areas of selected individual peaks, expressed as a percentage of the total ion count, varied by 15 - 20%, with no apparent systematic error related to the mass or intensity. Diagnostic programs are being developed to compute the average variation for the entire spectrum and to accurately test for systematic errors. In addition, this analysis will be performed on the data after removing the largest peaks from the total area used for normalization. (See the 1977 Final Report). This algorithm is currently

imbedded in the program designed to process raw multichannel analyzer data and is being rewritten for use with the reduced data format produced by the Finnigan data acquisition system. This correction is expected to reduce the average coefficient of variation by 6 - 7%, and make it possible to compare the performance of the present procedure and instrumentation with that of the older system described in the 1977 report.

The total spectral pattern from the replicate samples was also analyzed using the library search programs of the data acquisition system. To do this a sub library was created using these replicate spectra. Library entries are generated in the data system by a reduction algorithm that saves up to 50 peaks from the original spectrum using the relative peak intensities to select the most significant peaks in a window that is moved across the entire spectrum. The reduction occurs in two steps, first selecting 40 peaks in a window ± 50 amu wide and next selecting the six largest peaks in a sliding ± 7 amu window. As stated earlier, the library uses the square root of the mass times the intensity in all of its operations.

Using the sub library, 15 pairs of replicate spectra were compared to obtain the FIT parameter ($0 \leq \text{FIT} \leq 1000$) which is proportional to the cosine of the angle between the 50 dimensional vectors represented by each spectrum.

The replicate samples were compared to a second set of five samples of the same urine extracted and analyzed individually. In this set of samples the total intensities were above 500,000 counts per sample. Examining individual peaks, within the set of five, no increase in average variation was detected as compared to the replicates. To obtain a more quantitative measure of the variation between the two sample sets, we again used the library comparison, using 10 pairs of the individual spectra compared to each other. For the first set of 15 comparisons the average value of FIT

was 909 ± 29 and the second set of 10 comparisons gave an average FIT of 949 ± 29 .

Although this provides only a semi-quantitative comparison, the results indicate that the individual samples are more similar to each other than the replicate analysis of a pooled sample. We believe this is primarily due to the larger count rates obtained in the second sample set.

Using these vector dot product measures of similarity, we also examined the match between pairs of the pooled samples and matches between the pooled samples and the individually extracted samples. For the pooled samples the average FIT was 909 ± 29 , yet 12 pooled-individual comparisons gave a FIT of 832 ± 14 : a difference significant beyond the beyond the 99% confidence level. This indicates a detectable difference in the patterns due to the large scale extraction procedure used for the pooled samples.

G. Preliminary Tests of Clinical Samples

Preliminary tests were also carried out using a limited set of samples obtained from Children's Hospital of Buffalo. This initial set of samples included 7 individuals with pneumonia, 1 with bronchitis and 6 individuals hospitalised with traumatic injury or for tonsilectomy. This latter group, used as control samples, may not be completely free of infections in the case of the tonsilectomy patients.

Each sample was analyzed in duplicate by the procedures described above. For each sample the 80 scans were converted to nominal masses and summed to obtain the composite integrated spectrum. This reduced spectrum is equivalent to the output of our earlier data reduction programs designed to handle the raw counts and channel numbers acquired by the multichannel analyzer. (See the 1977 Final Report). An approximate average spectrum was obtained for each class of samples by summing the individual composite

spectra together. Each individual spectrum was included in the sum as many times as necessary to obtain approximately equal total intensity for each sample in the class average sum spectrum. (When our CYEER normalization procedure becomes operative, we will sum the normalized spectra and the above described procedure will be superfluous). In this manner we computed 4 separate class averages taking only one of the duplicate spectra for each individual, thus generating two control averages (designated \overline{CA} and \overline{CB}), and two pathological averages (\overline{PA} and \overline{PB}). We also computed a combined control average using all the control analyses and a similar pathological average deleting the bronchitis sample and one additional pneumonia sample which gave an ambiguous diagnosis in one of our preliminary pattern analysis tests. In order to evaluate these patterns we used the library search programs described previously. For each possible pair of average spectra CA, PA; CA, PB, etc., we obtained the FIT parameter for the other other pair of averages relative to these to reference vectors. In this test a control average should give a higher fit value when compared to a control then when compared to a pathological average. This test is equivalent to a projection of the 50 dimensional unknown vector on to 2 reference vectors. The two reference vectors are not necessarily "orthogonal" in the original 50 dimensional space but may be conveniently used to display their relationship to the unknown vector in 2 dimensions. Fig. 1 shows the result of this type of analysis for each of the possible comparisons made. In this representation, the 45° line represents the expected decision surface indicating an equal similarity to both pathological and control spectra. For the replicate average spectra all test points fell on the expected side of this decision surface although the graph shows that there is a high similarity to both reference spectra for either sample type. This is to be expected since the bulk of the metabolic profile does not

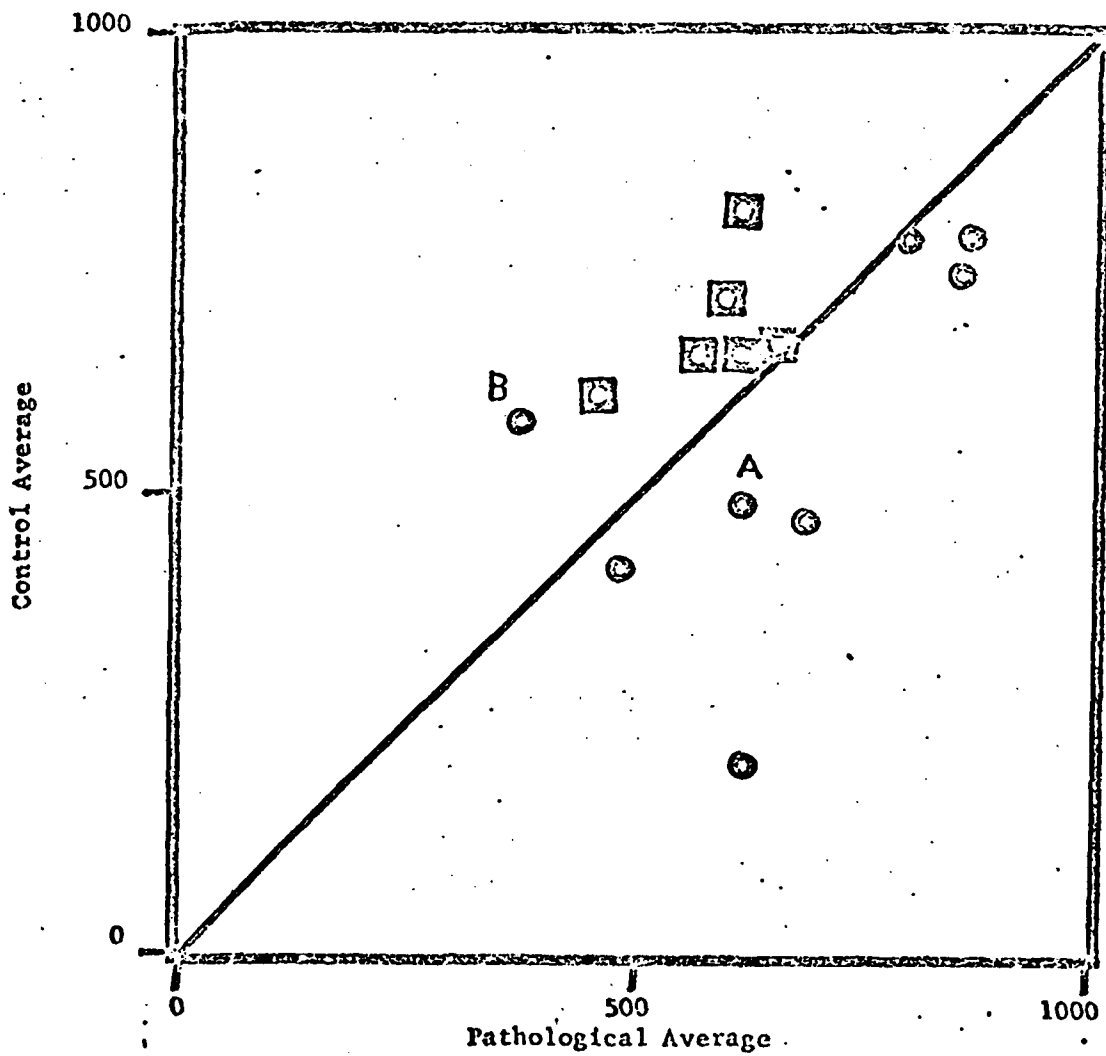


Figure 2. Comparison of Individual Average Spectra to Class Average Spectra

change but is only modified by the disease. The Wilcoxon Test that will be used for larger classes of normal and pathological samples will select the peaks with the higher significant differences and therefore amplify the difference between the pathological samples and controls.

A similar comparison was made by comparing the average of each set of duplicate individual samples to the total class averages. These results, shown on Fig. 2, while highly preliminary, show a significant separation between the two classes. In the traditional supervised learning approach to pattern recognition the expected decision surface would be altered to achieve the best separation.

This graph reveals some other features of the data set that deserve further consideration. First, the point indicated as B on the figure, representing the individual with bronchitis, appears to be more similar to the control group than to the pathological group. When this sample was included in the pneumonia average some of the points representing the control samples moved below the 45° decision surface. One additional sample (A on Fig. 2) analyzed later than the first group of samples, also fell on the decision line when the bronchitis sample was included in the pneumonia average. In Figure 2 these two samples were excluded from the pathological average. The second feature worth noting is that the pneumonia samples appear to cluster in two distinct groups. This may be due to differences in medication, the nature of the infection, (viral or bacterial) or the stage of the illness. At the present time we cannot be certain that this difference is real although use of a number of different individual samples as reference vectors preserved the grouping of these two sets of samples into two distinct clusters.

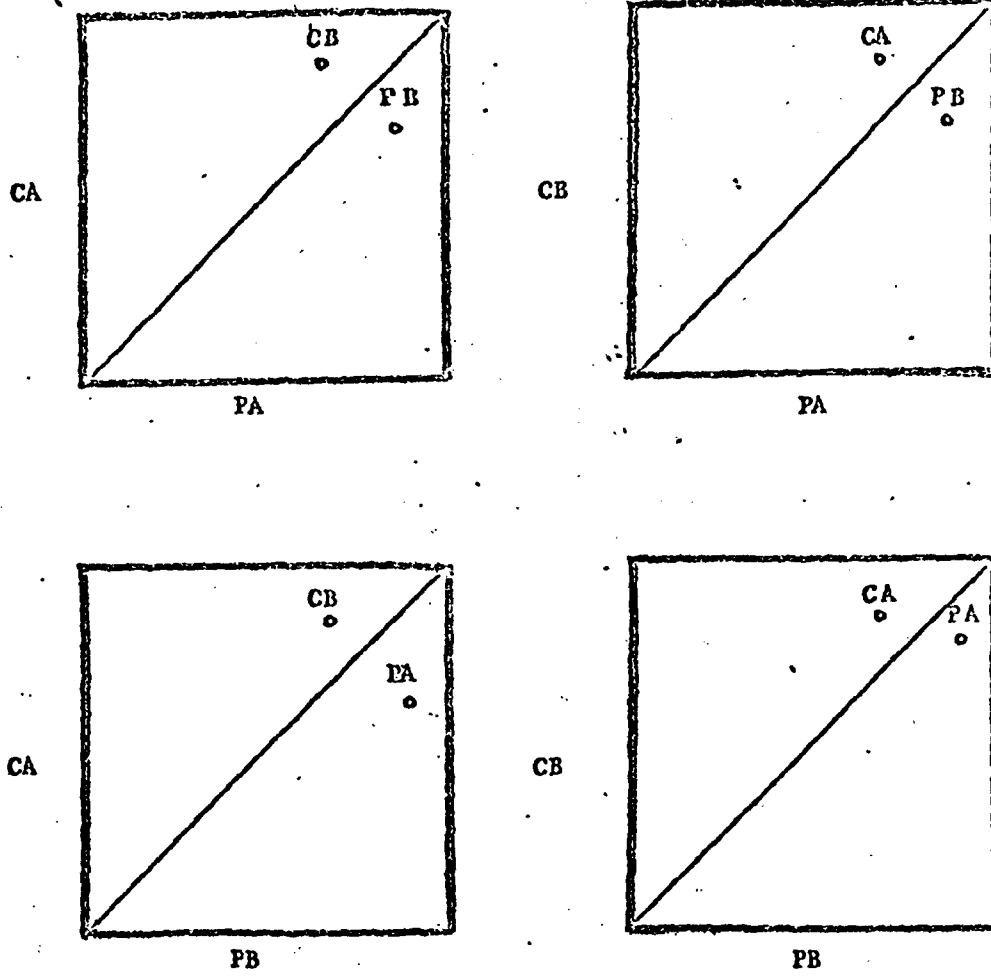


Figure 1. Comparison of Duplicate Class Average Spectra

The successful separation between the two very small sets of pathological and normal samples using all measured peaks (rather than those selected by a Wilcoxon Test) is extremely encouraging. It indicates real differences between the two sets which were not obscured by the majority of metabolites unaffected by the disease.

**ATE
_MED**