Title

Microorganism Detection by Multiple Non-Specific Oligonucleotide Probes or "Gene Probe Spectroscop"

Authors

Grant, Stanley B
Duong, Michelle H

Publication Date

1994-10-01

Microorganism Detection by Multiple Non-Specific Oligonucleotide Probes or "*Gene Probe Spectroscopy*"

By
Stanley B. Grant and Michelle H. Duong
Department of Civil and Environmental Engineering
University of California, Irvine
Irvine, CA 92717
(714) 856-7320

TECHNICAL COMPLETION REPORT

Project Number UCAL-WRC-W-788

October, 1994

University of California Water Resources Center

The research leading to this report was supported by the University of California, Water Resources Center, as part of Water Resources Center Project UCAL-WRC-W-788.

ABSTRACT

The advent of molecular-based diagnostics has radically improved our ability to analyze the prevalence and activities of microorganisms in environmental systems. In this research, a new molecular tool (gene probe spectroscopy or GPS) was tested which may further enhance the capabilities of existing molecular techniques for carrying-out environmental analysis. With GPS, a microorganism (or a group or related microorganisms) is characterized by its hybridization to a set of non-specific oligonucleotide probes in a dot-blot format. The hybridization intensities for each probe/microorganism combination is determined experimentally and stored in a matrix database ($\mathbf{A}$). Samples containing unknown mixtures of microorganisms are then hybridized with the same set of probes to obtain a set of composite hybridization intensities ($\vec{b}$). By employing linear inverse theory, the concentration ($\vec{c}$) of each microorganism present in the unknown sample may be computed by matrix inversion techniques: $\vec{c} = \mathbf{A}^{-1}\vec{b}$. With an award from the Water Resources Center, we investigated this approach using a model system consisting of four different bacteriophage (T2, M13, lambda, and T7) and six non-specific oligonucleotide probes each of which cross-reacts with two or more phage. In a series of tests, phage concentrations were correctly predicted by the method, illustrating its efficacy for identifying and enumerating microorganisms in fluid samples. Its application to "real" environments will require the use of DNA chip technology in which oligonucleotide probes are arrayed on glass slides. Future work will be directed toward the merging of the techniques developed in this project, namely GPS, with the DNA chip format.

KEY WORDS: Bacteria, Bioindicators, Biomonitoring, Biotechnology, Viruses, Water Quality Monitoring

CONTENTS

TITLE PAGE..... i

ABSTRACT..... ii

CONTENTS..... iii

LIST OF FIGURES & TABLES..... iv

I. PROBLEM AND RESEARCH OBJECTIVES..... 1

A. Introduction & Background..... 1

B. Overview of GPS..... 4

C. Applications of New Technology..... 6

D. Expected Advantages of GPS..... 7

E. Potential Limitations of GPS..... 8

II. RESEARCH METHODOLOGY..... 9

III. RESULTS AND DISCUSSION..... 11

A. Optimization of Mixed Phase Hybridization..... 11

B. Initial GPS Studies..... 16

IV. FUTURE RESEARCH..... 21

V. CONCLUSIONS..... 22

VI. BIBLIOGRAPHY..... 23

LIST OF FIGURES and TABLES

Figure 1. Proposed procedure for detecting microorganisms..... 5

Figure 2. Hybridization intensities as a function of transfer solvent pH..... 13

Figure 3. The cross-reactivity of each probe (A-F)..... 16

Figure 4. Procedure for quantifying hybridization..... 17

Figure 5. Normalized hybridization intensities..... 19

Figure 6. Comparison of actual and estimated phage concentrations..... 20

Table 1. Oligonucleotide probes used in GPS experiments..... 15

I. PROBLEM AND RESEARCH OBJECTIVES

I. A. Introduction and Background

Microorganisms are found in most, if not all, natural environments where they exist as natural inhabitants or as contaminants introduced from waste treatment facilities. The fact that microorganisms are so ubiquitous has important implications for many fields of environmental engineering and science. Water used for irrigation and municipal supply frequently contains pathogenic viruses, bacteria, and parasites originating from septic tanks or other sources of sewage effluent[1]. This contamination has direct and measurable effects on human health. It is estimated that over half of the 62,273 cases of waterborne disease reported in the US. between 1946 and 1977 were caused by the ingestion of sewage contaminated water[2]. Indigenous populations of bacteria also play an important role in carrying-out key chemical transformations in aquatic, soil and groundwater environments. In natural systems polluted with refractory or xenobiotic compounds, these microorganisms may represent an important mechanism for *in situ* removal or mineralization of contaminants[3].

Understanding microbial activities in the environment requires, as a *sine quo non*, the ability to detect the microorganism of interest. Currently available methods for detecting microorganisms either identify a phenotype which is unique to a given microorganism, or they detect a unique genotype directly. Examples of the former approach include culturing for specific bacteria on selective media, detecting viruses by their ability to form plaques on a specific host, the use of antibodies to detect proteins specific to a given microorganism, and the use of biochemical assays to characterize, or fingerprint, environmental isolates. All of these techniques depend on expression of the detected phenotype at the time the assay is being conducted. There is accumulating evidence that such methods seriously underestimate the pool of target microorganisms. For example, fewer than 20% of the approximately one billion cells per liter found in marine pelagic environments can be cultured using plate counts suggesting that culture methods are

inadequate for studying microbial community composition[4,5]. Even if a microorganism's physiological niche is known and can be duplicated experimentally, culture techniques may still be inadequate as a detection tool. Pathogenic bacteria in natural samples can enter a state called "viable but nonculturable" in which viable cells cannot be cultured using standard techniques. However, given the right conditions these bacteria can revert back to their culturable state[6].

Similar problems are encountered in detecting viruses using plaque forming units (PFU) techniques. Successful detection of viruses in fluid samples using PFU methods requires that a suitable host system is selected and that the resulting virus infection is phenotypically expressed as plaques, or visible cytopathic effects (CPE). In many marine systems, the appropriate hosts for the most abundant viruses (probably bacteriophage) are unknown, largely because the hosts themselves cannot be cultured[7,8]. Even in the case where the host is known, virus infection may not necessarily lead to plaque formation. Temperate bacteriophage, for example, can lysogenize a host after infection with the result that either no plaques or turbid plaques are formed[9]. For these and other reasons, PFU assays can seriously underestimate the concentration of targeted virus present in environmental samples. PFU estimates of bacteriophage abundance in aquatic systems are, in some cases, more than four log units below electron microscope estimates of bacteriophage-like particles[10, 11], although there is recent evidence that a sizable fraction of these particles may be inactivated[12]. In the case of mammalian viruses, PFU and CPE measurements are usually a factor of 100 or below the actual virus particle concentration[13]. Antibody based assays are also not effective for environmental samples because they lack sensitivity, and require the production of antibodies for each microorganism detected[14].

The relatively recent advent of molecular technology for detecting microorganisms has greatly improved our ability to study natural microbial populations and to detect low numbers of microorganisms in the environment[15,16]. Although many different molecular methods have evolved, they all have in common the detection of a specific genotype with

probes based on DNA-DNA or DNA-RNA hybridization techniques. The obvious advantage of these techniques over phenotype detection is that the nucleic acid corresponding to a particular gene will be present whether or not the gene is being expressed phenotypically. Perhaps the most widely used molecular method for environmental analysis is mixed phase hybridization in which nucleic acid present in a fluid sample is denatured, immobilized onto a membrane by manifold filtration (using a dot or slot blot), and then the membrane is hybridized to a fragment of DNA (gene probe) that has a nucleotide sequence complementary to a specific region of the target organism's genome. The probe is covalently linked to a detectable reporter group in the form of a radioisotope, fluorescent compound, or dye. After the excess probe is removed by washing, the bound probe is detected by assaying for the reporter group. This technique has been used in a large number of environmental studies to detect the presence of both bacteria and viruses (for a review, see ref. 16). Amplification of the target DNA using polymerase chain reaction (PCR) can also be used to increase the sensitivity of mixed phase hybridization. For example, using a combination of PCR and dot blot methods, Bej et al.[17] report the detection of single *E. coli* cells in water. In addition to the mixed phase hybridization methods described above, fluorescent RNA probes can also be used to label individual cells by *in situ* hybridization techniques[18,19].

While current hybridization techniques are a promising alternative to phenotype-based assays, there are still problems which limit the utility of molecular technology for studying microorganisms in environmental samples. Because gene probes detect specific sequences of nucleic acid, they cannot be used to distinguish between organisms that share the same sequence in their genome. In cases where the probe detects a certain catabolic gene, for example, it may not be important to know whether this gene is found in one microorganism only, or distributed between many organisms. However, in cases where samples are being screened for a specific organism (e.g., a human pathogen), the probe must be complementary to DNA (or RNA) found only in that organism. Finding a DNA

sequence unique to a given organism can be an enormous task, and proving that the resulting probe will not "cross react" with other organisms present in the sample can be virtually impossible. PCR may partially address this last problem, because this technique identifies a microorganism by its hybridization to two specific oligonucleotide primers and the size of the PCR "product" that results from thermal cycling[20]. However, the process of developing primers is difficult and time consuming, requiring that the target microorganism's genome be at least partially sequenced. Sequence data are also required if oligonucleotide probes are used to detect microorganisms by mixed phase hybridization. Techniques which do not require sequence information (like nick translation or random priming) usually result in relatively large probes (>100 bases in length) which can be nonspecific for the target microorganism[16].

I. B. Overview of GPS

In this research we investigated an alternative approach that addresses the dilemmas posed above. The basic idea is illustrated in Figure 1. Traditional hybridization protocols use a single probe at high stringency to detect a specific nucleic acid sequence immobilized on a membrane. In gene probe spectroscopy (GPS) the nucleic acid in a sample is immobilized on several different membranes, and each membrane is subsequently hybridized to a different oligonucleotide probe. Unlike traditional hybridization techniques, these probes may cross react with more than one microorganism, because it is the pattern of hybridization signals from the different probes that identifies an organism in this system. To understand the logic of this approach, first consider what happens if the sample contains only genomic DNA from a single microorganism, *A*. Probes that have a high degree of homology to some portion of the genomic DNA of *A* will bind strongly and give a strong hybridization signal. Probes with less homology to *A*'s genomic DNA will yield weaker hybridization signals. The set of hybridization signals from the different probes, or gene probe spectrum, forms a pattern that uniquely identifies organism *A* for a fixed stringency condition.

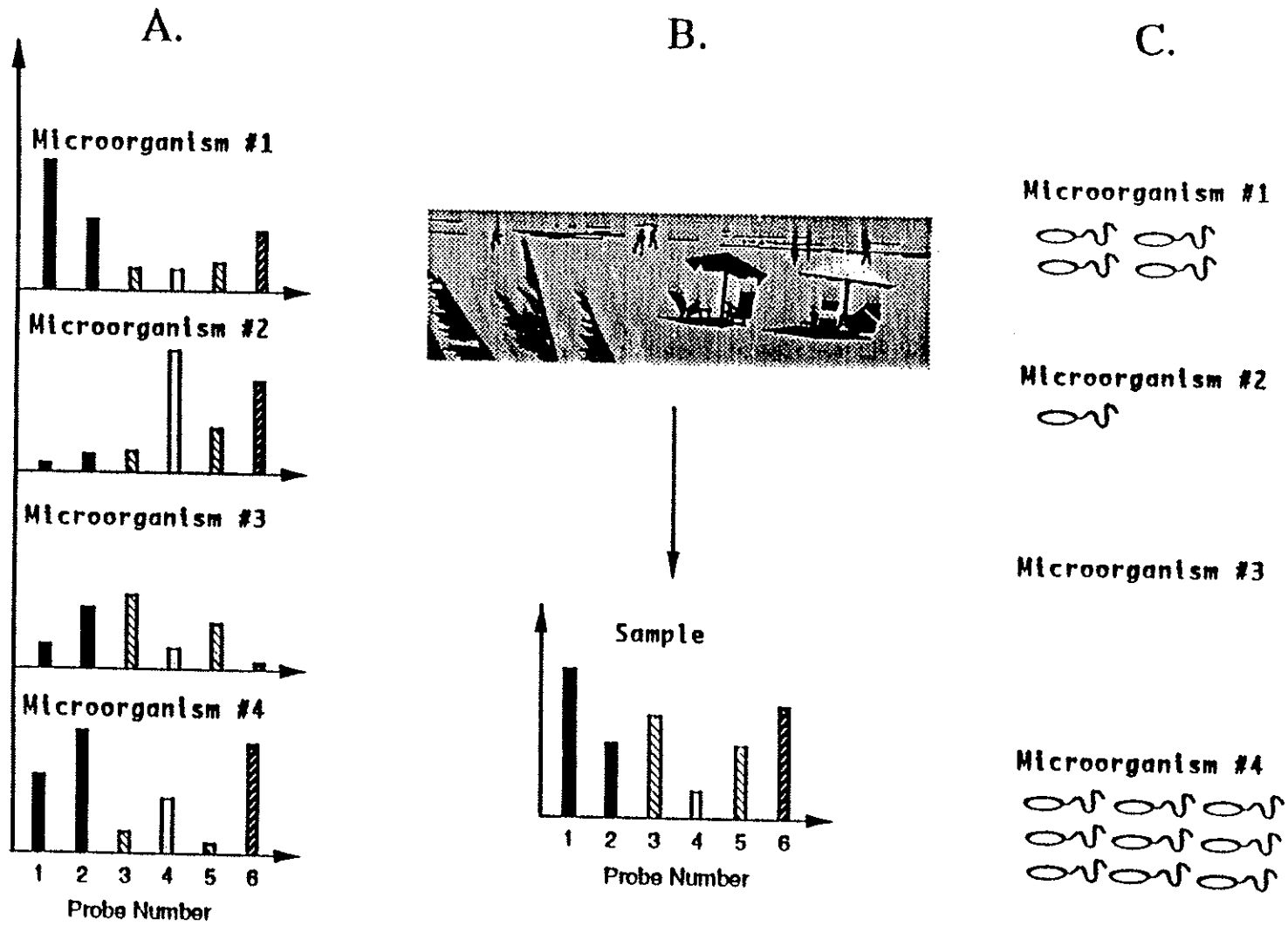

Figure 1. Proposed procedure for detecting microorganisms in environmental samples using gene probe spectroscopy. (A) A database is developed that consists of the gene probe spectrum for each microorganism (or group of microorganisms) potentially present in the environment. (B) The environmental sample is then probed with the same set of probes. (C) This sample spectrum is inverted (using Linear Inverse Theory) to give the concentration of the individual microorganisms present in the sample.

Now consider what happens if the sample contains more than one microorganism. If the hybridization process is carried-out so that the probe is in sequence excess over immobilized DNA, hybridization signals should vary linearly with the amount of nucleic acid immobilized on the membrane[21]. Therefore, the hybridization signal obtained from probing a mixture of microorganisms is simply the sum of hybridization signals contributed by the individual microorganisms, weighted by their concentration in the filtered sample. Put another way, each microorganism contributes to the final hybridization signal *in exact proportion to its concentration in the fluid sample*. Since hybridization signals are linearly correlated with the concentration of nucleic acid in the sample (subject to certain constraints discussed in Section III.A. and Section IV.), linear inverse theory (LIT) can be used to estimate the concentration of individual microorganisms present. To carry-out this inversion, a database is required that contains the gene probe spectra for all the microorganisms that might be present in the sample.

I.C. Applications of New Technology

As conceived, GPS would not replace conventional molecular techniques. Instead, the assay would be used in conjunction with conventional methods to provide additional information about the microorganisms present in a sample. In principle, GPS could be implemented in at least two possible ways, depending on the nature of the probes and the complexity of the microbial community of interest.

(i) For environments characterized by low microbial diversity (e.g., bioreactors or other engineered systems), probes with essentially arbitrary nucleotide sequences could be used. Hybridization experiments would be carried-out with low stringency conditions to allow hybridization between non-complementary strands. The hybridization signals obtained from a set of "arbitrary" probes would constitute the gene probe spectrum for a specific microorganism. Currently, gene probe development is one of the most difficult and time consuming steps in probing. The ability to use arbitrary probes might, in some

cases, significantly reduce the effort required to develop a molecular-based detection system.

(ii) For more complex microbial populations, 16S rDNA probes could be used to monitor classes of microorganisms (such as phylogenetic groups) in a given environment. In this case, each "microorganism" shown in Figure 1 would represent a group of related microorganisms, and the linear inverse step would reveal the relative abundance of these different groups in the sample of interest. The advantage of using GPS for this application is that the 16S rDNA probes would not have to be completely specific for any single group. This might prove particularly important for probes that detect very broad classes of microorganisms. For example, oligonucleotide sequences characteristic of the three primary kingdoms have been proposed; however, sequences characteristic of one kingdom are also found, to a limited extent, in the other two kingdoms[9]. Because this "cross reactivity" is accounted for in the inversion step, GPS might provide an elegant way of analyzing the kingdom-level composition of natural samples, a goal which will be difficult to realize using currently available methodologies. Similar strategies can be envisioned to characterize the relative abundance of physiological groups of microorganisms, for example those using a specific terminal electron acceptor.

I.D. Expected Advantages of GPS

The GPS approach has several inherent advantages over traditional hybridization technologies.

i) As described above, probes used in this system need not possess nucleotide sequences unique to specific microorganisms.

ii) Each microorganism (or group of microorganisms) is characterized by the hybridization signals corresponding to a number of probes, rather than the hybridization signal of a single probe, as is currently done. This means that GPS is less likely to yield false positives for the detection of a specific microorganisms because there is inherently more information to distinguish one organism from another.

iii) Given the gene probe spectrum of a sample and an appropriate database, LIT provides a best estimate of the concentrations of microorganisms present in the sample, based on a chi-square (χ^2) analysis (see section II). From the χ^2 value corresponding to a given inversion, a probability (p) can be calculated that indicates the likelihood that the nucleic acid detected in the sample is adequately represented in the database. If p is close to one, the sample is well-characterized and the concentration estimates obtained by LIT can be trusted; if p is close to zero, then the concentration estimates are poorly constrained and should be disregarded. In most fields of biology, statistical tests are routinely used to determine the significance of a set of results. LIT provides a way to apply the same statistical standards to the detection of microorganisms in environmental systems.

iv) Each sample to be analyzed can be seeded with a known amount of DNA from a control microorganism that is complementary to some or all of the probes. The concentration of this microorganism can be compared with the estimate based on LIT, thereby providing another level of confidence in the concentration estimates obtained by the inverse step.

I.E. Potential Limitations of GPS

GPS (in its current form) also has several limitations when compared to conventional probing.

i) For the LIT to work, there must be *at least* as many probes as there are "detection categories" (i.e., microorganisms or groups of microorganisms). Because each additional probe involves another hybridization experiment, there are practical limits on the number of detection categories that can be analyzed.

ii) The proposed assay requires a database of gene probe spectra for all microorganisms (or groups of microorganisms) that might be detected in a sample. This database can be developed in several different ways, all of which require a certain amount of effort. If there are relatively few detection categories in the sample being tested, DNA from each detection category potentially present in the sample can be hybridized along with

the sample. Alternatively, the gene probe spectrum corresponding to each detection category can be determined by hybridization and referenced to a DNA standard present on the same membrane. These normalized spectra can then be stored in a computer database, much in the same way that nucleic acid and protein sequences are currently stored. The second approach was used in this study.

II. RESEARCH METHODOLOGY

GPS can be used to estimate the concentration of individual microorganisms (or groups of microorganisms) present in a sample as follows. Suppose that m different probes are used to detect n different microorganisms ($m > n$). Let I_{ij} represent the hybridization intensity obtained when nucleic acid from the jth organism is hybridized with the ith probe; C_j the concentration of the jth microorganism in the sample; and B_i the hybridization intensity obtained when nucleic acid from the sample is hybridized with the ith probe. Because hybridization intensity is linearly related to the concentration of DNA in the sample (see discussion on pg. 6), the following linear set of equations can be written,

$$\begin{aligned}
C_1 I_{11} + C_2 I_{12} + C_3 I_{13} + C_4 I_{14} &= B_1 \\
C_1 I_{21} + C_2 I_{22} + C_3 I_{23} + C_4 I_{24} &= B_2 \\
C_1 I_{31} + C_2 I_{32} + C_3 I_{33} + C_4 I_{34} &= B_3 \\
C_1 I_{41} + C_2 I_{42} + C_3 I_{43} + C_4 I_{44} &= B_4 \\
C_1 I_{51} + C_2 I_{52} + C_3 I_{53} + C_4 I_{54} &= B_5 \\
C_1 I_{61} + C_2 I_{62} + C_3 I_{63} + C_4 I_{64} &= B_6
\end{aligned} \qquad (1)$$

This set of linear equations can be cast in matrix notation,

$$\begin{pmatrix} I_{11} & I_{12} & I_{13} & I_{14} \\ I_{21} & I_{22} & I_{23} & I_{24} \\ I_{31} & I_{32} & I_{33} & I_{34} \\ I_{41} & I_{42} & I_{43} & I_{44} \\ I_{51} & I_{52} & I_{53} & I_{54} \\ I_{61} & I_{62} & I_{63} & I_{64} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \end{pmatrix}.$$

If we let

$$\mathbf{A} = \begin{pmatrix} I_{11} & I_{12} & I_{13} & I_{14} \\ I_{21} & I_{22} & I_{23} & I_{24} \\ I_{31} & I_{32} & I_{33} & I_{34} \\ I_{41} & I_{42} & I_{43} & I_{44} \\ I_{51} & I_{52} & I_{53} & I_{54} \\ I_{61} & I_{62} & I_{63} & I_{64} \end{pmatrix}, \quad \vec{c} = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix}, \text{ and } \vec{b} = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \end{pmatrix}$$

then the above set of equations (1) can be simply written as

$$\mathbf{A}\vec{c} = \vec{b}. \tag{2}$$

Physically, $\mathbf{A}$ is a database which characterizes the affinity of each probe for each microorganism (or group of microorganisms), $\vec{c}$ represents the concentrations of microorganisms actually present in the sample, and $\vec{b}$ represents the hybridization intensities obtained when the sample is hybridized to the set of m probes.

The matrix equation (2) is an inconsistent system (i.e., there is no solution vector $\vec{c}$ that exactly satisfies the set of equations) of m equations in n unknowns. The goal of LIT is to find an estimate for the vector $\vec{c}$ which best satisfies (2). There are several different approaches that can be used, the simplest of which is the "linear least squares method"[22]. In this approach, the best set of concentration estimates, or solution vector $\vec{c}^{\,est}$, is found by minimizing the chi-square function (χ^2) which represents the squared difference between the right and left hand sides of (2):

$$\chi^2 = \left|\mathbf{A}\vec{c} - \vec{b}\right|^2 \tag{3}$$

where the vertical lines represent the "norm" of a vector. In addition to estimating the set of concentrations $\vec{c}^{\,est}$, a probability (p) can also be computed which indicates how well the hybridization data fit the linear model given by (2). If p is small, then the estimated concentrations from LIT can be statistically rejected. For a particular inverse, p can be calculated from the minimum χ^2 value and the number of degrees of freedom (υ) defined, in this case, by the number of probes minus the number of microorganisms in the database ($m-n$):

$$p(\chi^2|\upsilon) = \frac{\Gamma(\upsilon/2, \chi^2/2)}{\Gamma(\upsilon/2)}$$

where Γ represents the transcendental gamma function.

III. RESULTS AND DISCUSSION

A two year grant from the U.C. Water Resources Center allowed us to conduct preliminary experiments aimed at investigating the feasibility of GPS. The results of 2 years of research effort are summarized below.

III.A. Optimization of Mixed Phase Hybridization.

Initial studies indicated that hybridization signals obtained from mixed phase hybridization experiments could be quite variable, particularly if the samples being probed had slightly different chemical compositions. Because this variability would make interpretation of GPS data more difficult, we initiated a set of experiments to identify factors affecting hybridization signals obtained by mixed phase hybridization. The results of this preliminary investigation were recently published[23] and are described below.

Mixed phase hybridization has become a routine method for identifying DNA or RNA of a specific nucleotide sequence in virtually all fields of basic and applied biology. One of the most important factors in determining the success of a mixed phase hybridization is the procedure used to immobilize the nucleic acid onto a membrane. Alkaline blotting, in which nucleic acid is transferred to the membrane using a high pH solution, appears to provide optimal results for the capillary transfer of DNA restriction fragments from agarose gels to positive charge modified nylon (PCMN) membranes[24,25,26] to hydrophobic polyvinylidene difluoride (PVDF) membranes[27] and for the immobilization of unfractionated samples using a dot or slot blot filtration manifold[25]. In addition to promoting dissociation of complementary strands, the alkaline solution may also induce covalent fixation of nucleic acid directly to PCMN membranes, eliminating the need for further treatment with *in vacuo* heat or UV irradiation[24,26]. However, Cannon et al.[28] found that UV irradiation of DNA that had been blotted to a PCMN membrane under alkaline conditions made the immobilized DNA more resistant to removal when the membrane was subsequently washed in a Tris buffer. Protocols have

also been reported for alkaline transfer of unpurified nucleic acid samples derived from tissue[29] and viruses[30]. These procedures involve suspending the biological material to be analyzed in alkaline solution (0.1-0.4 M NaOH), heating the sample to release the nucleic acid into solution and filtering the unpurified sample directly onto a membrane.

Many published protocols also recommend adding EDTA (1 to 10 mM) and NaCl (0 to 3 M) to the transfer solution[31,32,33]. Presumably, the EDTA is recommended to inhibit nuclease activity prior to manifold filtration whereas the NaCl is added to promote binding of the DNA to the membrane, although the effect of these additions on the ability of bound nucleic acid to form hybrids with a probe has not been systematically documented. To address these issues we performed a series of experiments to examine the effect of transfer solvent composition on the detection of lambda (λ) DNA by mixed-phase hybridization.

In preliminary experiments (data not shown) we found that the hybridization intensity obtained by mixed phase hybridization was strongly affected by the concentration of both NaOH and EDTA in the transfer solution. We also observed that the highest (optimal) hybridization intensity occurred when the concentrations of NaOH and EDTA in the transfer solvent were in a specific ratio. The following experiments were devised to investigate this phenomenon in more detail, and to determine if transfer solvent pH might be a controlling variable for hybridization intensities. A fixed amount of purified lambda genomic DNA (Boehringer Mannheim) was suspended in 450µl of transfer solvent (final concentration 0.1 µg/ml) containing pre-specified amounts of EDTA, NaOH, and NaCl. These samples were heat treated at 95 ^{o}C for 5 minutes, filtered onto a PCMN membrane (Zetaprobe, Biorad) using a dot blot apparatus (Minifold II, Schleicher and Schuell), and the membrane was UV irradiated for a cumulative exposure of 120,000 microjoules in a UV Stratalinker (Stratagene). The membranes were hybridized to a lambda-specific probe (5'---AAA AAA CAG GGT ACT CAT---3') radiolabeled at the 5' terminus with γ-^{32}P dATP to a specific activity of 1.4 to 5.4 x 10^4 mCi/mmol using bacteriophage T4

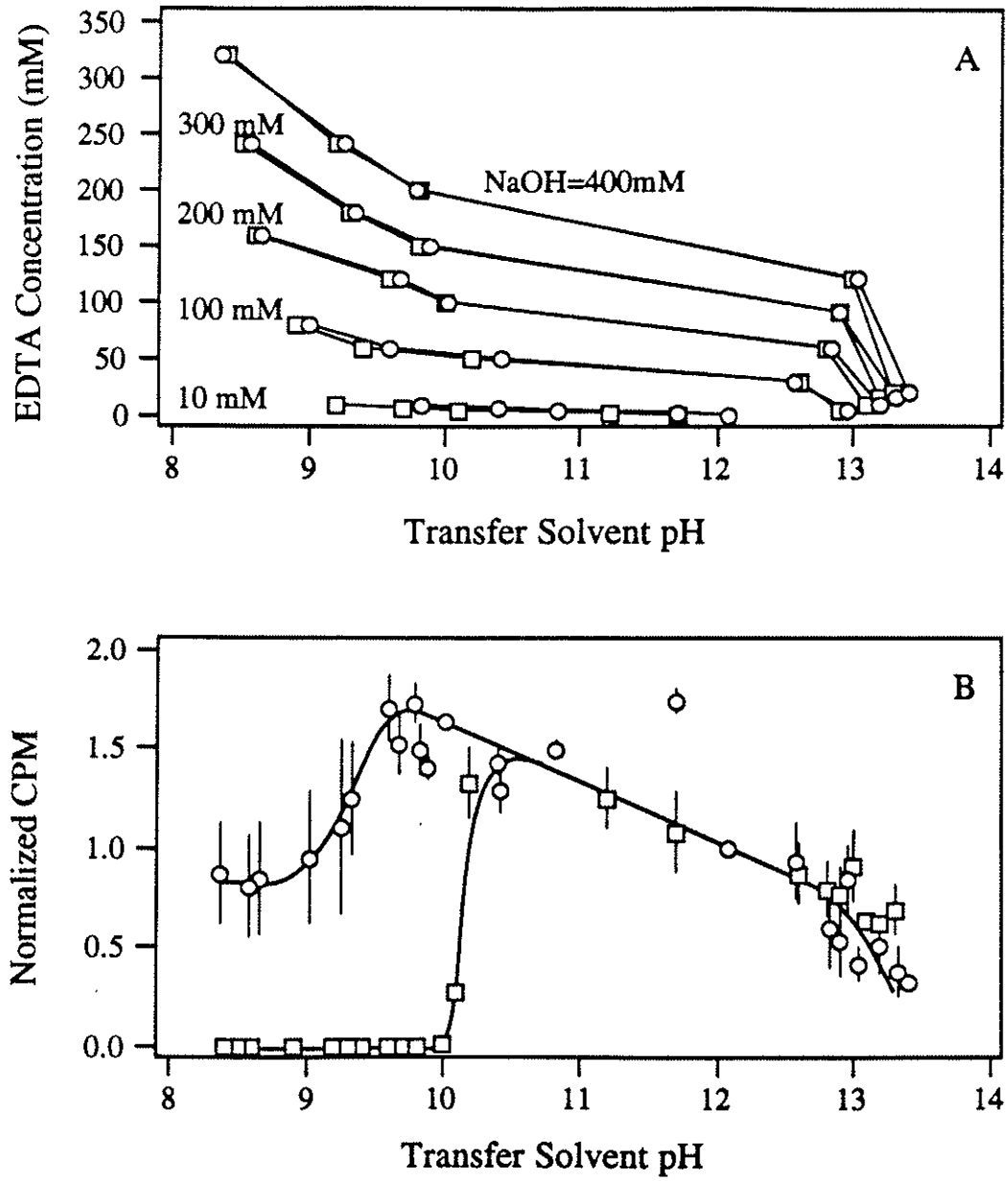

Figure 2. Hybridization intensities as a function of transfer solvent *p*H and ionic composition: (○) no added salt, (□) 50mM NaCl. (A) Effects of EDTA and NaOH concentrations on solution *p*H. Each line in the figure corresponds to the indicated NaOH concentration. (B) Effects of transfer solvent *p*H on hybridization signals from 45 ng lambda DNA per blot. Hybridization signals corresponding to each solution condition were normalized relative to the hybridization intensity obtained when the transfer solvent consisted of 10 mM NaOH and 0.5 mM EDTA (*p*H=12). Points represent average normalized signals obtained from three separate hybridization experiments. Standard deviations are indicated by error bars.

polynucleotide kinase[34]. Hybridization was carried-out at 35 °C overnight in hybridization fluid (5x SSC, 10x Denhardt Solution, 19 mM Na_2HPO_4, 7x SDS) followed by three successive washes in two wash solutions (wash steps #1 & #2: 4x SSC, 1x Denhardt Solution, 29 mM Na_2HPO_4, 6% SDS; wash step #3: 1x SSC, 1% SDS). Radiolabeled DNA hybrids were visualized by direct beta counting (DBC) using a radioimager (Ambis), and counts per minute (CPM) corresponding to each blot were determined. These counts were normalized with respect to a standard immobilized on the same membrane. In all of the experiments the standard consisted of purified λ DNA diluted to a final concentration of 0.1 μg/ml in 10 mM NaOH and 0.5 mM EDTA. Normalized hybridization intensities from three replicate hybridizations were pooled to obtain average values and standard deviations.

Figure 2A shows the effect of EDTA (left axis) and NaOH (lines in figure) on the pH of transfer solvents containing either 50mM NaCl (squares) or no added salt (circles). As expected, transfer solvent pH decreases with increasing EDTA concentration for each fixed concentration of NaOH tested; pH is only slightly affected by the addition of salt. Figure 2B shows the observed relationship between normalized hybridization intensity (left vertical axis) and the pH of transfer solvents containing the different combinations of NaOH, NaCl and EDTA shown in panel A. When plotted in this way, hybridization intensity appears to be a simple function of transfer solvent pH and salt concentration. For experiments conducted with NaCl in the transfer solvent (squares), hybridization signals varied over 4000% depending on the relative concentration of NaOH and EDTA; hybridization signals varied over 500% for the set of experiments in which NaCl was not added to the transfer solvent (circles). For pH values below 10.5, hybridization intensity is reduced to background level in the case of added salt. For pH values above 13, hybridization intensity decreases rapidly with increasing pH. This has important practical implications because blotting protocols typically call for extremely alkaline transfer

Table 1. Oligonucleotide probes used in GPS experiments

Probe Name	Sequence (5' to 3' direction)	Melting Temp, T_m (°C)[a]	Hybridization Temp (°C)
A	CGG CTG GCT TTG TGG	66.9	38
B	GGC TGG TGG TAC TCG	66.9	35
C	ACC AGA CCC GCC GCC	72.2	38
D	AGC GGG GTT TTG CT	59.9	45
E	GGC TGG ATT GGC GAA CGT GC	76.1	43
F	GCA AAA TGA CCA GTA TCA CC	67.9	40

[a] calculated for complementary duplexes using $T_m = 81.5 + 16.6\log[Na] + 0.41(\%GC) - 600/L$, where L equals probe length in nucleotides

solutions (pH>13) containing very high concentrations of NaOH (400mM) and relatively low concentrations of EDTA (10mM)[25,29,31,32]. Our results suggest that this combination of NaOH and EDTA may be far from optimal. Our data also indicate that salt in the transfer solution exerts a strong negative effect on hybridization intensity. The presence of 50mM NaCl in the transfer solvent was sufficient to reduce hybridization intensities to background levels in the experiments with transfer solvent pH between 8 and 10. However, the negative effects of salt on hybridization intensity can be minimized by using transfer solvents with pH>10.5.

In addition to providing a framework within which to optimize mixed phase hybridization experiments, these results also demonstrate that hybridization signals obtained from samples containing subtle differences in their chemical make-up should be interpreted with caution. Previous studies have shown that mixed phase hybridizations can be used to obtain quantitative estimates of the concentration of specific nucleotide sequences present on the membrane provided that the probe is in sequence excess over bound nucleic acid during hybridization[35,36,37,38]. These researchers point out that linear relationships between nucleic acid concentration and hybridization intensity (as measured by dot density on autoradiograms, liquid scintillation counting, or direct beta counting) can

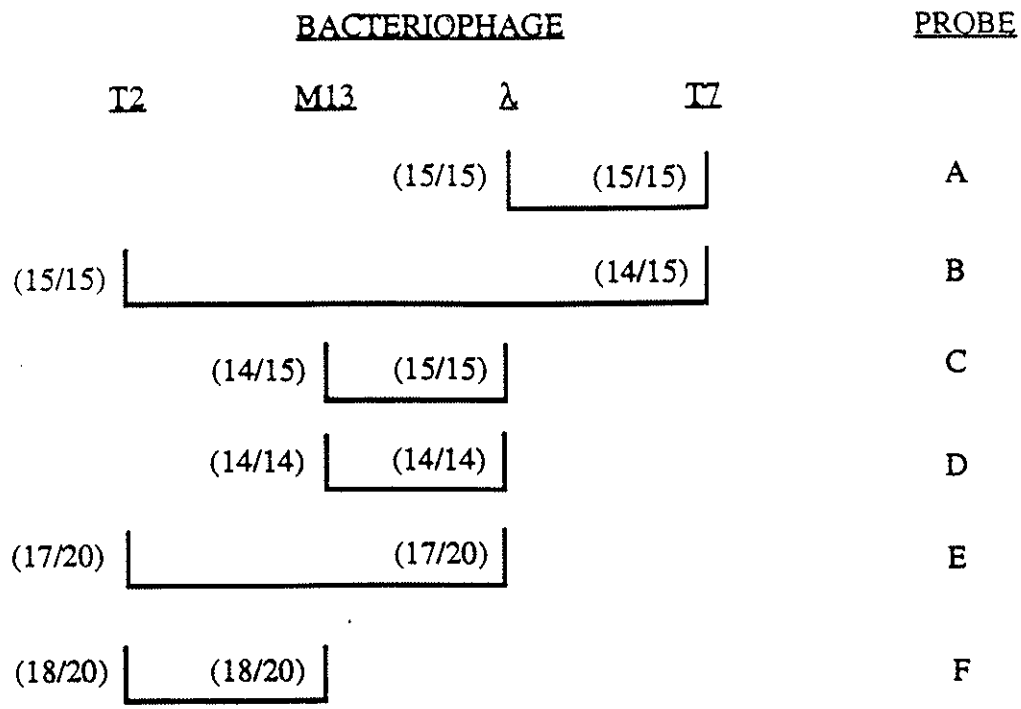

Figure 3. The cross-reactivity of each probe (A-F) with the four bacteriophage investigated in this study: T2, M13, λ and T7. Ratios in parentheses represent the fraction of base pair matches.

be obtained, suggesting that concentration estimates for samples containing unknown quantities of a specific nucleic acid sequence can be found by referencing hybridization signals to nucleic acid standards present on the same membrane. However, our results suggest that such a comparison is valid only if the transfer buffer is chemically identical for all of the samples, including the nucleic acid standards.

III.B. Initial GPS Studies.

With lessons learned from the above study, we set-out to test the GPS method. The model experimental system consisted of four bacteriophage (T2, T7, M13, and λ) and six non-specific oligonucleotide probes, each of which hybridized to two of the phage. The six probes (Table 1) were selected by performing pair-wise homology searches between the different bacteriophage to identify nucleotide sequences that appeared in at least two out of the four phage (Figure 3). Our initial efforts focused on experimentally characterizing the database matrix **A**. In this example, there are six probes and four microorganisms, so **A** has a dimension of 6 rows by 4 columns.

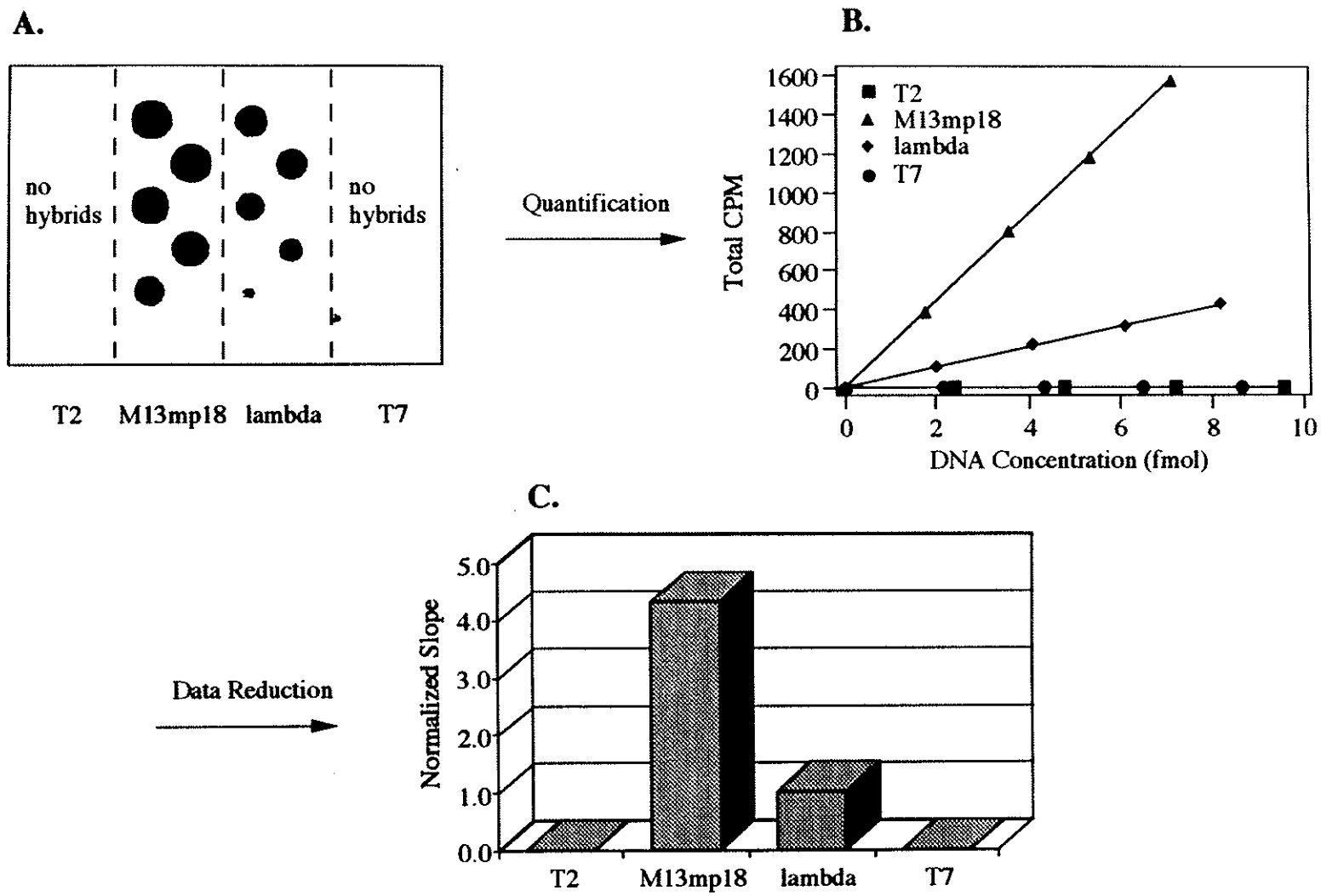

Figure 4. Procedure used for quantifying the hybridization of probe D to bacteriophage T2, M13, lambda, and T7. (A) Image of dot blot; (B) Correlation between counts and immobilized DNA concentration; (C) Normalized hybridization intensities for each of the bacteriophage tested. See text for details.

The process of characterizing a single row of **A** is shown for probe D in Figure 4. Probe D was radiolabeled and hybridized to a membrane containing separate dilution series for T2, M13, λ, and T7 genomic DNA. The radioactivity remaining on the membrane after several wash steps was imaged using a radioimager (Ambis), and is shown visually in Figure 4A. As expected from the cross-reactivity chart in Figure 3, probe D binds strongly to M13, less strongly to λ, and does not form hybrids with the either T2 or T7 (see Table 1 for stringency conditions). Using the radioimager, counts per minute (cpm) corresponding to each blot were determined by direct beta counting, and the resulting counts are plotted against immobilized DNA concentration in Figure 4B. These data show a strong linear correlation between cpm and the amount of DNA immobilized on the membrane. The slopes of the lines in Panel B are a measure of the affinity of probe D for each of the bacteriophage: the higher the slope (in units of cpm/fmol), the stronger the probe's affinity for the specific bacteriophage. The slopes shown in Figure 4B were normalized by the slope corresponding to bacteriophage λ, and the resulting normalized slopes (or hybridization intensities) are plotted in Figure 4C. This last panel represents our first gene probe spectrum, and shows the relative extent to which probe D binds to each phage.

The process outlined in Figure 4 was repeated twice for all six probes, and the resulting normalized slopes are depicted in Figure 5. Each grid of colored squares in this figure represents a "visualization" of the hybridization intensities in the **A** matrix. For example, probe B hybridizes strongly to phage T2, as indicated by the dark red color for that position in the grid. The two different grids represent replicate sets of hybridization experiments, or different "realizations" of **A**. By comparing the patterns of colors in the two different grids, it is apparent that the normalized slope values obtained for each probe/bacteriophage combination are highly reproducible.

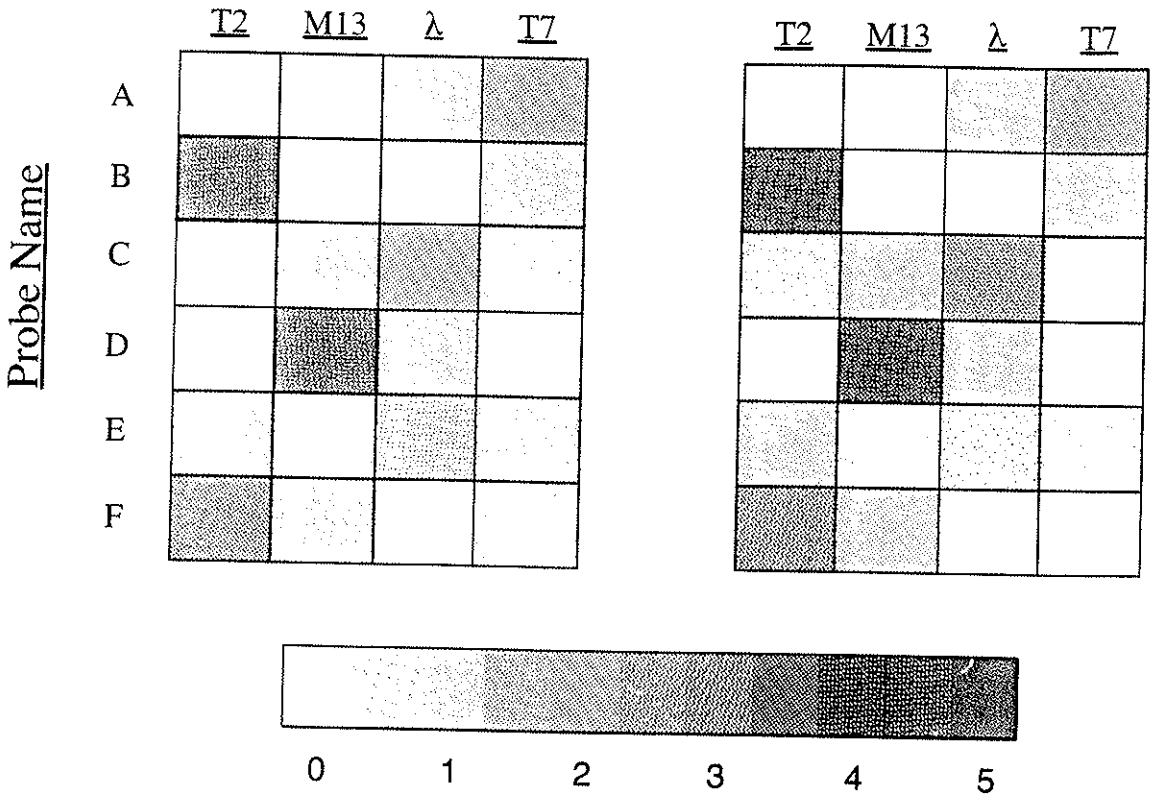

Figure 5. Normalized hybridization intensities for each probe/bacteriophage combination. For each probe a different bacteriophage was used as a normalizing standard (shown in parentheses): Probe A (λ), Probe B (T7), Probe C (M13), Probe D (λ), Probe E (T2), Probe F (M13). The two different grids represent two replicate sets of hybridizations.

In the next set of experiments, we wanted to test the ability of GPS to estimate the concentration of phage seeded into a fluid sample. Fluid samples were prepared with pre-specified concentrations of the four bacteriophage. These samples were blotted in triplicate onto a set of six membranes and hybridized to probes A-F. The resulting counts were normalized using a standard present on each membrane, and the normalized hybridization intensities (in units of (cpm of sample)/(cpm/fmol of standard)) were used to build-up the $\vec{b}$ vector. The database matrix $\mathbf{A}$ and the vector $\vec{b}$ served as input into the linear inverse scheme outlined in Section II, from which estimates of the phage concentrations were obtained, $\vec{c}^{\,est}$. A comparison of actual and estimated phage concentrations for three

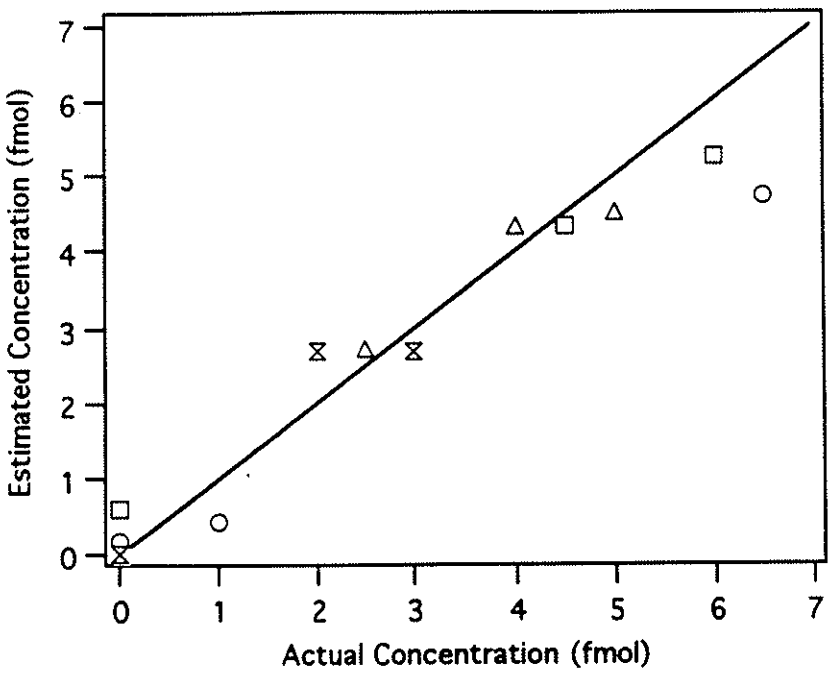

Figure 6. Comparison of actual and estimated bacteriophage concentrations: (□) T2, (Δ) M13mp18, (⧖) lambda, and (○) T7.

different samples is shown in Figure 6. Although there is some scatter in these data, overall the concentrations estimated from LIT correspond well with actual phage concentrations over the concentration range investigated. This is particularly remarkable given that these results were obtained from a single "realization" of the $\vec{b}$ vector (i.e., each sample was hybridized only once with each probe). It is reasonable to assume that the correspondence between estimated and actual phage concentrations would improve as hybridization intensities in the $\vec{b}$ vector were refined through additional hybridizations. These experiments clearly demonstrate that GPS can identify and enumerate microorganisms in the relatively idealized model system employed here.

IV. FUTURE RESEARCH

Additional research is being carried-out in my laboratory to refine the GPS approach, and make it more amenable to field applications. Some specific goals are listed below.

(i) Test the idea of using suites of 16S rDNA probes to estimate the concentration of groups of phylogenetically related microorganisms by GPS.

(ii) Explore the possibility of extending GPS to systems with low microbial concentrations, by using universal or "broad range" PCR primers for amplifying specific regions of the 16S rDNA[39] prior to hybridization.

(iii) Investigate the possibility of using arbitrary oligonucleotide probes to monitor microbial concentrations in systems with low microbial diversity.

(iv) Investigate the possibility of implementing GPS using alternative hybridization formats. There are essentially two problems with the current GPS format which limits its applicability to environmental systems. First, in conventional dot blot systems the surface of the membrane can become saturated with DNA during the filtration step leading to a break-down in the linearity between DNA concentration in the sample and counts on the membrane after hybridization. A non-linear hybridization response would violate one of the fundamental assumptions in the inverse step outlined in Section II, and lead to erroneous concentration estimates. Second, an unrealistic number of different hybridization steps are required for systems where the number of detection categories is large, as might be expected for most natural environments. Both of these problems may be addressed by using a reverse probe format in which oligonucleotides are arrayed on a surface (typically a glass slide) and hybridization is carried out with the target DNA in the solution phase. Reverse probing would reduce the number of hybridizations required to experimentally characterize the database matrix $\mathbf{A}$ and the sample vector $\vec{b}$. Because all of the probes are immobilized on a single substrate, separate hybridization experiments are not needed for each separate probe. Techniques are currently being developed for producing

large arrays of oligomers attached to solid supports, primarily in the context of DNA sequencing by hybridization, or SBH. Arrays containing as many as 65,000 oligomers have been reported using lithographic procedures[40], suggesting that GPS might eventually be extended to include spectra involving 10,000 or more probes. We are currently developing such oligomer arrays, and testing the feasibility of using GPS with this new format.

(v) Investigate the use of silicon-based electronic biosensors to detect DNA/DNA hybrids in the reverse probe format. This work is being carried out in collaboration with several faculty in electrical engineering (one at UCI and another at Harvey Mudd College). The idea is to develop a biosensor that detects the presence of DNA/DNA hybrids by changes in the local electrical potential surrounding the immobilized probes. This detection scheme completely eliminates the need for any DNA labeling step, and thereby significantly simplifies the hybridization process. It would also permit the detection of hybrids during the hybridization step (real-time hybridization), and the response of hybrids to different temperatures. The last point is important as each immobilized probe in an array has a different optimal temperature for discriminating between complementary and non-complementary strands.

V. CONCLUSIONS

In the past two years we have made significant progress in developing and testing GPS. While the technique appears to work for relatively idealized model systems, considerably more work is needed to make the technique "field ready". Practical application of GPS will probably require using reverse probing to reduce the number of separate hybridizations required, and biosensors to detect hybrids. Ultimately, the combination of these technologies will yield a new generation of diagnostic tools with broad applications in environmental analysis.

VI. BIBLIOGRAPHY

[1]Keswick, B.H. and C.P. Gerba (1980) *Environ. Sci. & Tech.* **14**:1290-1300.

[2]Keswick, B.H. (1984) in "Groundwater Pollution Microbiology", eds. Bitton, G. and C.P. Gerba, John Wiley and Sons: New York.

[3]Madsen, E.L. (1991) *Environ. Sci. & Tech.* **25**:1663-1673.

[4]Ward, D.M., R. Weller, and M.M. Bateson (1990) *Nature* **345**:63-65.

[5]Giovannoni, S.J., T.B. Britschgi, C.L. Moyer and K. Field (1990) *Nature* **345**:60-62.

[6]Colwell, R.R., P.R. Brayton, D.J. Grimes, D.B. Roszak, S.A. Huz, and L.M. Palmer (1985) *Bio/Tech.* **3**:817-820.

[7]Fenical, W.H. (1992) *ASM News* **58**:186.

[8]Fuhrman, J.A., personal communication

[9]Brock, T.D., M.T. Madigan, J.M. Marinko, and J. Parker (1994) "Biology of Microorganisms, 7th Ed." Prentice Hall: New Jersey.

[10]Proctor, L.M. and J.A. Fuhrman (1990) *Nature* **343**:60-62.

[11]Bergh, O., K.Y. Borsheim, G. Bratbak, and M. Heldal (1989) *Nature* **340**:467-468.

[12]Suttle, C.A. and F. Chen (1992) *Appl. Environ. Microbiol.* **58**:3721-3725.

[13]Sharp, D.G. (1965) in "Transmission of Viruses by the Water Route", Berg, G., ed., John Wiley & Sons: New York.

[14]Margolin, A.B., M.J. Hewlett, and C.P. Gerba (1991) *Wat. Sci. Tech.* **24**:277-280.

[15]Dockendorff, T.C., A. Breen, O.A. Ogunseitan, J. Packard, and G.S. Sayler (1992) in "Microbial Ecology: Principles, Methods, and Applications", Levin, M.A., R.J. Seidler, and M. Rogul, eds. McGraw Hill:New York.

[16]Sayler, G.S. and A.C. Layton (1990) *Annu. Rev. Microbiol.* **44**:625-648.

[17]Bej, A.K., R.J. Steffan, J. DiCesare, L. Haff, R.S. Atlas (1990) *Appl. Environ. Microbiol.* **56**:307-314.

[18]Amann, R., N. Springer, W. Ludwig, H.D. Gortz, and K.H. Schleifer (1991) *Nature* **351**:161-164.

[19]Amann, R.I., L. Krumholz, and D.A. Stahl (1990) *J. Bacteriol.* **172**:762-770.

[20]Sambrook, J., E.F. Fritsch, and T. Maniatis (1989) "Molecular cloning: A Laboratory Manual, 2nd ed." Cold Spring Harbor Press: Cold Spring Harbor, N.Y.

[21]Anderson, M.L.M. and B.D. Young (1985) in "Nucleic Acid Hybridisation: a Practical Approach", Hames, B.D. and S.J. Higgins, Eds. pp. 73-111, IRL Press:Oxford.

[22]Press, W., S. Teukolsky, W. Vetterling, and B. Flannery (1992) "Numerical Recipes in FORTRAN, 2nd Ed." Cambridge Univ. Press: NY., p. 665.

[23]Duong, M. H., S.B. Grant, and M.E. Lidstrom. (1994) *Analytical Biochemistry* **220**:431-433.

[24]Reed, K.C., and Mann, D.A. (1985) *Nucl. Acid Res.* **13**, 7207-7220.

[25]Chomczynski P., and Qasba, P.K. (1984) *Biochem. Biophys. Res. Comm.* **122**, 340-344.

[26]Khandjian, E.W. (1987) *BioTech.* **5**, 165-167.

[27]Hichks, D.A., and Vecoli, C. (1987) *BioTech.* **5**, 206-210.

[28]Cannon, G. Heinhorst, S., and Weissbach, A. (1985) *Anal. Biochem.* **149**, 229-237.

[29]Reed, K.C., and Matthaei, K.I. (1990) *Nucl. Acid Res.* **18**, 3093.

[30] Richardson, K.J., Margolin, A.B., and Gerba, C.P. (1988) *J. Virol. Meth.* **22**, 13-21.

[31] Meinkoth, J., and Wahl, G. (1984) *Anal. Biochem.* **138**, 267-284.

[32] Kuroda, S., Yasuaki, I., Miyazaki, T., and Fujisawa, Y. (1988) *Biochem. and Biophys. Res. Comm.* **152**, 9-14.

[33] Pepin, R.A., Lucas, D.J., Lang R.B., Lee, N., Liao M. J., and Testa D. (1990) *BioTech.* **8**, 628-632.

[34] Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning, Cold Spring Harbor Press, New York.

[35] Steffens, D.L., and Gross, R.W. (1989) *BioTech.* **7**, 680-686.

[36] Laskey, R.A., and Mills, A.D. (1977) *FEBS* Lett. **82**, 314-316.

[37] Sognier, M.A., Neft, R.E., Roe, A.L., Eberle, R.L., and Belli, J.A. (1991) *BioTech.* **11**, 520-525.

[38] Anderson, M.L.M. and Young, B.D. (1985) *in* Nucleic Acid Hybridisation: a Practical Approach (Hames, B.D. and Higgins, S.J., Eds), pp. 73-111, IRL Press, Oxford, England.

[39] Relman, D.A. (1993) "Universal Bacterial 16S rDNA Amplification and Sequencing" in "Diagnostic Molecular Microbiology", Persing, D.H.; Smith, T.F. Tenover, F.C. and White, T.J., Eds., American Society for Microbiology: Washington DC.

[40] Smith, L.M. (1993) *Science* **262**:530.