

Report to NSF on
Banbury Workshop on
**Designer Molecules for Biosensor
Applications.**

Bud Mishra

Courant Institute, NYU & Watson School, Cold Spring Harbor Laboratory

Jack Schwartz

Courant Institute, NYU



NYU/Courant Bioinformatics Group

Supported by a grant from CISE, NSF (Mita Desai). Principal Investigator: Bud Mishra, Professor of Computer Science & Mathematics, Courant Institute, NYU; Professor, Watson School of Biological Sciences, Cold Spring Harbor Laboratory and Co-director, Center of Comparative Functional Genomics, NYU.

TABLE OF CONTENT

Introduction:..... 3
Background and Motivation: 3
Technology Needed: 4
The biosensor:..... 4
More about the technological perspective: 5
Biosensor systems: technical alternatives:..... 6
 Brief Summary:..... 11
Summary of Discussions: 13
Recommendations:..... 30
Conclusions and Future Avenues: 34
Appendix:..... 36
 Agenda:..... 36
 Participants:..... 39

Introduction:

This report summarizes the discussions and recommendations from a three-day workshop, organized at Cold Spring Harbor Laboratories' Banbury Center aimed at exploring the feasibility of the state-of-the-art and potential near-, medium-and longer-term technologies that can be used to quickly detect pathogenic microorganisms and toxins. The stated goals were to germinate potential multi-disciplinary collaborations, accelerate maturation of existing technology, encourage industrial participation, and structure policy issues for multiple funding agencies to participate cooperatively.

Towards this end, we first identified appropriate scientific and non-scientific participants, who could judge the pros and cons of various technologies, gauge the risks of successful execution of new research plans and identify a structure for integrating existing and potential technologies. With considerable help from Dr. Jan Witkowski and his office at Cold Spring Harbor Laboratory, we organized the workshop at the Banbury Center of Cold Spring Harbor Laboratory during a three-day period 12-14 of August 2002.

Background and Motivation:

In a review of technologies for the detection of biological terror organisms and toxins, (Iqbal, Mayo, Bruno, Bronk et al, *Biosensors and Bioelectronics*, **15** (2000), 549-578), Iqbal and Mayo note that "The ease with which small countries and terrorists can now obtain biological warfare agents has escalated the need to provide the war fighter and civilian alike with miniature, easy to use, disposable instruments for detection and identification of potentially hazardous biological agents." Radiological and chemical attacks appear similarly feasible, for much the same reason.

The potential biological weapons should be categorized into two groups: *organisms* (replicative) and *toxins* (non-replicative). The sensors to be employed for each category may differ, as do the recovery and containment processes, after an attack of a particular kind has been launched. Also different forms of attack differ substantially in the speed with which their effects would become visible, and, within each of these attack categories, effects will differ by types of biological weapons.

Chemical toxins such as Sarin will have visible effects within an hour and often less. Some biological toxins, e.g. botulinum toxin, will act less rapidly and only manifest days after the exposure. The effects of radiological weapons will develop much more slowly, but once detected, its presence and geographic spread will be relatively easy to determine. The most insidious weapons are infectious biological organisms such as anthrax spores or the smallpox virus, whose effects develop over a period of days so that their presence may not even be suspected until days after an attack has occurred, by which time the agent may have dispersed widely. In many cases, a difference of a few days or even hours warning in knowing that the attack has occurred and the nature of the attack may have a large effect on the ability to manage its consequences, and on the ability of the hospital system to respond adequately.

Technology Needed:

The arguments presented above show the critical importance of identifying chemo-bio-warfare organisms and toxins that have been used as early as possible, and certainly once a surge in patients presenting for medical treatment has begun to make it clear that a bioterrorist attack may have occurred. A few days gain in identification speed might well ultimately reduce the size of the affected population in half. The technical task of detecting organisms and toxins in minute quantities, e.g. from likely attack sites or in swabs taken from suspected attack victims, pose interesting inter-disciplinary problems. Note that the environmental sensors or diagnostic devices for monitoring the environment vs. detecting a disease in a population should be made distinct as they pose different challenges, are effective in different ways and require different technologies. However, these technologies in totality must address the following questions: *Can attacks in progress be detected swiftly enough to give useful warning? Once an attack has occurred, how quickly can the attack agent (toxins or organisms) be identified? What is the cost of testing per single patient or maintaining daily surveillance per single site? What rate of false positives and false negatives will be found acceptable by the first response and public health officials in determining a course of action?*

Though sensitive, rapid, and cheap sensors constitute the technical core of the full warning system needed, attention must be paid to how the outputs of such sensors must ultimately be transmitted to the first response and public health officials, including police, and how the medical system can best begin to prepare its response. *Hence the sensors must be integrated into a comprehensive system.*

In a civilian setting, sustainability of surveillance and avoidance of false alarms is perhaps more important than in a military setting. This civilian application requirement puts a premium on low cost and detection accuracy. Our project will concentrate on systems appropriate for the civilian (urban and rural) setting in which homeland defense authorities may need to continue daily surveillance for the foreseeable future.

The biosensor:

One could consider the design of biosensors with two guardedly optimistic hypotheses. The first is that simple aerosol detectors, not much more expensive than standard smoke detectors, can be developed to provide a fast, cheap, timely, moderately reliable indication that an attack with an aerosolized pathogen has occurred. The second is that recent progress in biological detection techniques makes it possible for a deployable pathogen identification technology based on rtPCR (real-time polymerase chain reaction) to be demonstrated within one year of the start of a sufficiently energetic development of the necessary component technologies and systems elements. The participants in this workshop examined these and other possibilities, and concluded, for many reasons, that a pilot study based on rtPCR possesses many merits.

It is highly desirable that the research on biosensors must bring us closer to producing laboratory prototypes of detector systems of the two types described above (aerosol sensors and broad-spectrum pathogenic organism sensors) suitable for full-scale development and subsequent deployment.

Nonetheless, the Banbury workshop kept its particular focus—in terms of a three year

goal—on the underlying molecular biological technologies and to only a much smaller degree, on the electronic, computational and information technologies, as the most critical development need is for molecules that would be needed for these detectors. Consequently, a significant portion of the discussion at this workshop centered around the biomolecular and chemistry-related science and research, where NSF has and is likely to continue to play a strong role in advancing the research and development.

More about the technological perspective:

Requirements. The biosensor systems to be designed must take various criteria into considerations: coverage, sensitivity, robustness, speed, and cost.

(i) *Coverage:* What attack organisms or toxins is a system capable of detecting? Can it discriminate between these organisms and toxins? Is such discrimination immediate, or are separate analyses required for each organism or toxin discriminated?

The CDC bioterrorism website (<http://www.bt.cdc.gov>) lists 18 biological organisms or toxins in three categories of priority that possess qualities making them plausible weapons. The agents of greatest concern are those “that pose a risk to national security because they can be easily disseminated or transmitted person-to-person; cause high mortality, with potential for major public health impact; might cause public panic and social disruption; and require special action for public health preparedness.” Six first priority agents (six organisms and one biological toxin) are listed: *smallpox* (a virus), *anthrax* (a bacterial spore), *botulinum* (a toxin), *plague*, *tularemia*, and the hemorrhagic fevers (including Ebola and Marburg fevers, caused by several distinct families of viruses). Attack organisms and toxins listed in the CDC’s ‘second priority’ group include Undulant Fever (*Brucellosis*), *Clostridium* (a food-poisoning bacterium), *Glanders* (occurring principally as an animal disease), *Melioidosis* (a water and soil-borne bacterium), parrot fever (*Psittacosis*), Q fever, *Ricin* (a toxin produced from castor beans), *Staphylococcal enterotoxin B*, *Typhus*, *Viral encephalitis*, *Cholera*, *Cryptosporidium* (a protozoan), *Nipah virus*, and *hanta virus*.

(ii) *Sensitivity:* How few bacteria, viruses, or individual fungi, or how few nanograms of a toxin, is a sensor able to detect? Can it be used to detect organisms and toxins (a) collected from the environment: air, water or earth; (b) collected from swabs or blood samples taken from a newly infected patient; (c) collected from swabs or blood samples taken from a large number of patients? How reliably will a sensor perform, given the background of dust, earth, blood, saliva, and other organic materials likely to be present in samples collected for examination?

(iii) *Robustness:* How complex and error-prone are the steps involved in obtaining a reading from a given sensor? Can persons with limited training perform these in the field, or do they require special laboratory skills, supplies, and equipment? If a surveillance system involves distributed collection of samples and shipment to centralized laboratories for analysis, how much training and equipment is required for proper sample collection? Does shipment involve any special technical requirements?

(iv) *Speed:* How rapidly can a given sensor yield its readout: Within minutes, hours, one day, several days? This will determine whether it can be used for attack warning, outbreak anticipation, or only attack identification once an outbreak has become evident.

(v) *Cost*: What is the cost of a test conducted with a given sensor, per single test and per attack organism or toxin? What is the capital cost of the sensor?

Biosensor systems: technical alternatives:

Many promising approaches to biosensor design, some based on technologies first described quite recently, are now available. Biosensor designs can be arranged in a matrix of several dimensions:

(1) *Affinity systems vs. single-molecule systems.*

Then, for the affinity systems,

(A) *Kind of target molecule to be detected*

(B) *Kind of probe molecule to attach to it*

(C) *Means for observing molecular attachment*

<i>Things on Sensors</i> →	Nucleotides (N)	Proteins (P)	Other Molecules (O)
<i>Things in Environment</i> ↓			
Nucleotides (N)	E.g., Microarrays	—	Possible future combinatorial approaches
Proteins (P)	E.g., Protein Aptamers	E.g., Antibodies and Protein Aptamers	Possible future combinatorial approaches
Other Molecules (O)	E.g., Nucleic Acid Aptamers	E.g. Antibodies and Protein Aptamers	Possible future combinatorial approaches

For (A) and (B) three basic possibilities exist, namely one can employ or target (i) DNA/RNA molecules (nucleotides), (ii) Proteins, or (iii) other molecules. This leads to the following cases, for instance:

(NN) Nucleotide in Environment; Nucleotide on Sensors:

Microarray-based sensors

(PN) Protein in Environment; Nucleotide on Sensor:

Aptamer systems

(PP) Protein in Environment; Protein on Sensors:

Antibody and peptide aptamers- systems.

The Antibody Approach. The oldest and most established of these approaches to pathogen detection is the (PP) or antibody (immunological) approach. Antibody systems are used in a wide variety of standard laboratory tests for infectious agents and disease conditions, sometimes in simple portable kits on which a swab or drop of blood can be spread and some colored blot or pattern detected if a suspect molecule is present.

Techniques such as phage display for constructing antibodies that bind very strongly and selectively to target molecules have become standard. Thus, antibody techniques, which afford no way of amplifying the target molecule population, need new research to overcome several technological obstacles to “amplifying signals” while working over a wide range of proteins. The “tadpole” system described subsequently in this report suggests one such approach. However, they may have promise for use with environmental sensors. If inexpensive sensors, capable of separating and concentrating droplets, in a size range of tens and hundreds of microns, can be developed, say onto a removable glass surface, it may be possible for attack agents to be captured from air in a form concentrated enough for simple and very fast antibody based tests to show their presence. Another approach using peptide aptamers with similar capabilities is discussed in a later subsection.

The Pure Genomic Approach. The detection technique at the opposite pole to the (PP) antibody technique is the (NN) approach. This approach works by extracting the nucleic acid from the sample (if it is RNA converting it to DNA), amplifying this DNA using the PCR or polymerase chain reaction technique (developed by Kary Mullis in 1983) and then binding the amplified target DNA to a set of complementary DNA probe strands. As the target DNA can have some kind of reporter element attached, the subset of probes bound to target can be detected by a scanning device and associated with a possible set of target pathogens. By proper design of the probe sets, it is possible to detect a particular target pathogen unambiguously. Many authors have noted the potential advantages of this approach. One arguable advantage of a genomic approach is that the geometry of the interaction is relatively well understood; so that once the pathogens have been sequenced, design of the probes becomes a combinatorial problem amenable to calculation. Another approach along this line could be based on rtPCR, where for example minute amounts of pathogen DNA can be detected by PCR amplification in the presence of molecular beacons capable of binding to the PCR product.

The Aptamer Approach. Nucleic acid aptamers are single-stranded DNA molecules designed to possess a very high affinity to proteins or smaller organic molecules. Recently, there has been much progress in developing peptide aptamers or protein aptamers that can be used in a similar manner and for a similar purpose. Aptamers may also have organic catalytic (enzymatic) activity, in which case they are called ‘aptazymes.’ The (PN) (Aptamer) approach requires that the aptamer probes to be used must be identified by starting with an artificially randomized nucleotide population and winnowing it down by wet-lab selection of the randomized nucleotides possessing useful affinities. The initial nucleotide randomization can be performed chemically.

If, in the limit, one were trying to track multiple thousands of proteins, then an aptamer microarray design would seem to be in order, where the population of aptamers to be arrayed can be selected accordingly.

The Final Detection Stage. Depending on the sensitivity of the final detection method used, the fact that DNA can be greatly amplified using PCR may give decisive advantage either to the genomic (NN) or aptamer (PN) approach. In principle, PCR allows a single target molecule of DNA to be detected. Many techniques are available for the detection of limited numbers of organic molecules. These include the use of fluorescent attachments, fluorescence resonance energy transfer (FRET), molecular beacons, time

resolved fluorescence, affinity techniques possibly involving subsequent PCR steps, atomic attachments detectable by X-ray, surface binding modifying optical properties of surfaces, use of up-converting phosphors, electrochemical detection by use of surface waves, mass spectrometry, detection of spectral changes when a DNA oligonucleotide with resonating attachments at both ends binds to a target DNA strand, etc. (See for example “*Disposable microdevices for DNA analysis and cell sorting*” by Chouet al, available at <http://thebigone.caltech.edu/quake/publications/pnasjan99.pdf>).

More About Genomic Approaches and Their Optimization. Genomic oligonucleotide microarrays (GOM) can be optimized computationally once all the biowarfare organisms of interest have been sequenced. GOM can be manufactured in bulk in a small glass slide format. It may be possible to make GOM’s available for earlier deployment than will prove possible for aptamer systems, which involve newer technologies. Genomic detectors based on microarrays, which exploit PCR technology to achieve high sensitivity, can yield readings in fewer than several hours and are essentially available now. The completed system will be relatively inexpensive (in the price range of US \$ 100 – \$1,000). As mentioned earlier, rtPCR-based method, without the additional microarray technology, may still prove highly competitive to the array-based methods, and may be deployable in a shorter period of time.

Finally, the software needed to design the probes required, the hardware needed to produce customizable chips, the hardware needed to read them, and the databases and algorithms needed to store and interpret the data that they produce, etc. already exist in demonstrable form. (For example, versions of all these elements could be supplied collectively by Wigler’s Laboratory at Cold Spring Harbor Laboratory, NimbleGen, Inc. Wisconsin, Garner’s Laboratory in U. Texas, Sequenom, Inc., California and NYU bioinformatics group, among other possible sources.)

In order to assess the feasibility of a fully deployable genomically based bioattack organism detection system, it is instructive to examine some of the key technical parameters of such a system. (i) coverage: the system might be designed to target 100-1000 attack pathogens simultaneously, each with a 5Mb genome size on average. A genome microarray/chip with this capability could be designed to contain about 100,000-many 70-mer probes (single stranded DNA probes synthesized on a glass surface using a maskless customizable process, currently available from NimbleGen and Affymetrix). Thus such an array can contain about 100 probes per organism of interest. Each probe needs to be chosen in such a manner that makes it unique within the virtual “concatenated” genome of size 5×10^9 bp (basepairs) obtained by stringing all the target pathogens end-to-end. Such a microarray is roughly comparable to the kind of “copy-number array” for human genomes that have been used in cancer studies in the Wigler lab. Based on the experience of this laboratory and computation studies conducted in the NYU bioinformatics group, we can expect that such an array will detect a biowarfare organism with acceptably high specificity.

To verify this, part of the research effort could be targeted towards: (i) Faster algorithms for precise choice of the operational protocols and experimental parameters; faster initial probe design; (ii) Algorithms for data analysis; (ii) Mathematical techniques for estimating *false-positive* and *false-negative* rates to be expected; (iii) A methodology for periodic probe redesign and system improvement; (iv) Experimental verification that the

optimized probe slides have the performance expected; (v) Development of the software and networking infrastructure needed to deploy the system widely.

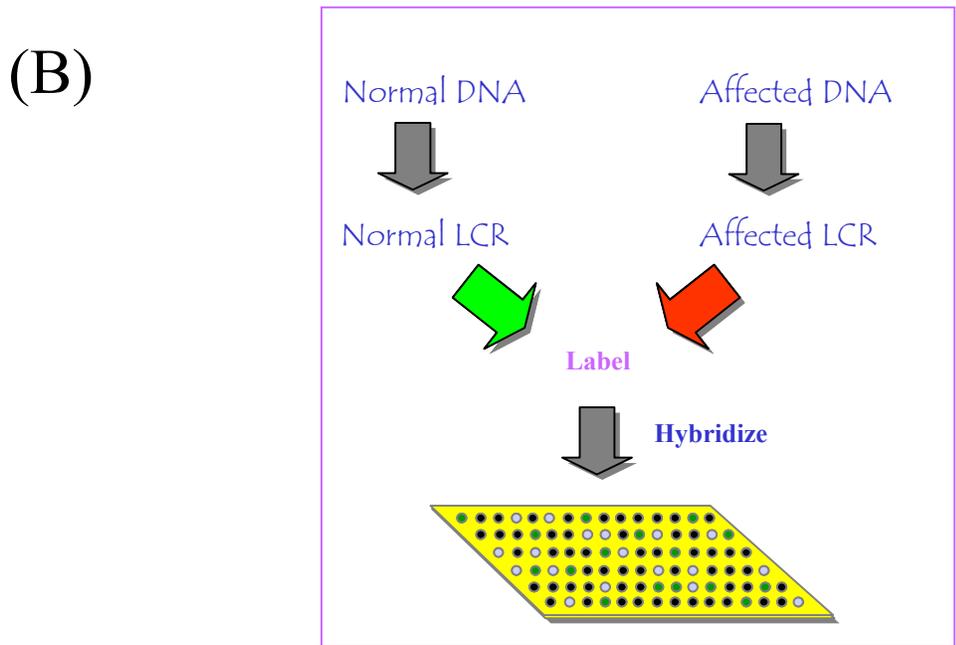
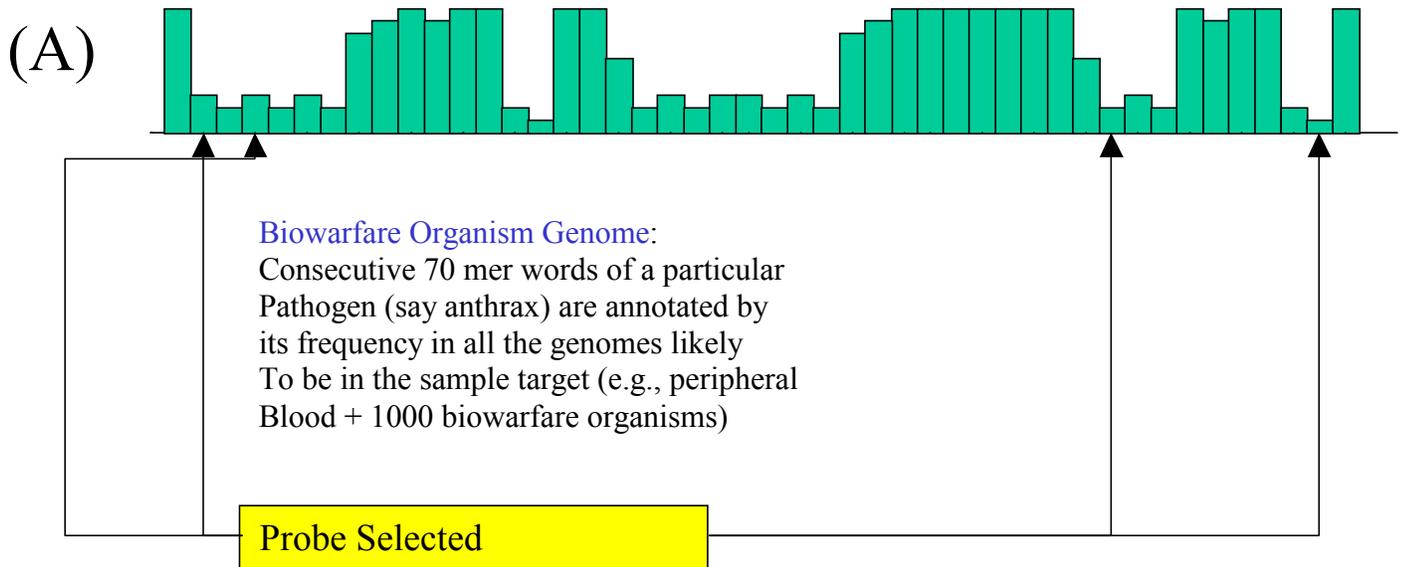


Figure 1. (A) Certain “words” (70 mers) contained in a low-complexity representation (LCR) of the genomes of biowarfare organisms and with good hybridization characteristics (e.g., unique words from the genome, similar base-pair content, etc.) are chosen and the corresponding synthesized oligonucleotides are spotted on the array. (B) Representations are samplings of affected DNA (possibly containing one or more biowarfare organisms) and normal DNA (from a healthy person known not to have been exposed to same agents). The DNAs are given a new format and reduced complexity. This complexity reduction step is achieved by cleaving the genome with a restriction enzyme (e.g., Bgl II) and amplifying small restriction fragments by PCR. These representations, obtained from a normal human genome and a potentially affected human genome, are colored green and red respectively and are hybridized to a spotted microarray. A high ratio of red-to-green at a spot indicates the presence of a particular pathogen

More About Aptamer-Based Approaches and Their Optimization. Nucleic acid aptamers are single-stranded DNA or RNA molecules often selected by repeated screening for their affinity to proteins or smaller organic molecules. Thus, in the (PN) (Aptamer) approach, one selects the aptamer probes by starting with an artificially randomized nucleotide population and winnowing it down using wet-lab selection of the randomized nucleotides selected for high affinities. However, the best solution may frequently not yield aptamers with the affinity needed in a sensor operating with many aptamers in an array format. In order to circumvent this problem, one may consider combining many aptamers to produce polymers that combine numerous weak interactions to produce high avidities.

An example of a significant advance in aptamer technology comes from a group of researchers at Uppsala University, (see Fredriksson et al “Protein detection using proximity-dependent DNA ligation assays,” *Nature Biotechnology*, **20**:473-477, May 2002) and the accompanying news article “Bringing picomolar protein detection into proximity,” (*Nature Biotechnology*, **20**:448-449, May 2002) which gives a clarifying exposition of Fredriksson’s method and makes some very interesting broader comments. This new ‘Aptamer-pair’ technique is to develop, for each protein P of interest, a pair P_a , P_b of aptamers with binding sites near one end and known sequences for use as primers at the other. Once P has bound P_a and P_b , an enzymatic binding step can be used to bridge between their proximate ends by a short inserted oligonucleotide X , yielding a sequence P_a-X-P_b which can be selectively amplified by PCR since both its ends are known. In effect this transduces the presence of a protein P into the presence of an oligonucleotide P_a-X-P_b , thereby partially attaching a nucleotide signature to protein. This technique clearly lets one transduce a mix of proteins P_1, \dots, P_n into a corresponding amplified mix of oligonucleotides $P_{an}-X-P_{bn}$, provided that none of the P_j s has substantial affinity for any other aptamer than the P_a and P_b intended for it. Hence the availability of the following more general technique will allow one to look for one of several pathogen-related proteins against a background. (a) If one only wants to find whether any of a list of pathogens (say some one of 256 pathogens) is present, proceed as follows: Mix the suspect substance with a soup containing 256 left-ends P_{an} and 256 right-ends P_{bn} , bridge, and amplify by PCR; test for elevated oligonucleotide concentration in any convenient way. This should in principle require less than 4 hours. (b) If one wants to identify *which one* of these 256 proteins is present, the process changes slightly: Mix the suspect substance with 8 separate soups each containing 128 left-ends P_{an} and 128 right-ends P_{bn} , bridge, and amplify the eight results in parallel by PCR. This will in principle give an ‘address’ of 8 bits identifying the one P in 256 that is present. Similarly if 2 target substances might be present simultaneously 16 such ‘experiments in parallel’ should serve to resolve the situation, and if 3 are present then 24. Another possibility is to ensure that all the oligos P_a-X-P_b that one needs to identify are of significantly different molecular weight or hydrophilicity, and to separate them by gel electrophoresis, visualizing the result in the usual optical or electrophoresis manner.

If, in the limit, one is trying to track multiple thousands of pathogens, the techniques mentioned earlier can be used with a nucleotide arrays as readout. Furthermore, the identification of pairs of aptamers with the linking oligonucleotides would also require new research: in particular, the extensions of the approach by which aptamers are selected by artificial evolution. Subsequently, we will briefly describe the work of Andrew Ellington and Ronald Breaker in this context, and also examine how the work of Nadrian Seeman on self-assembly may be relevant in this context.

Other Novel Approaches. In addition to the approaches discussed earlier, one can explore how several approaches may be combined in order to improve either the avidity or signal-to-noise ratio. For instance, the approach based on “low-complexity-representation” subsamples the material and thus, eliminates vast amount of material that only add marginally to the signal while amplifying the noise in the detection system. In another example, a “tadpole” may be created by binding an oligonucleotide sequence to an antibody such that when the antibody attaches to a protein, these complexes can be extracted and their signatures can be easily detected by amplifying the corresponding oligonucleotide by cycles of PCR reactions. It is expected that few of such approaches will be selected for a biosensor system to be ultimately designed and deployed.

Brief Summary:

In summary, the preceding discussion has primarily focused on the issues involving the biology and chemistry needed to evaluate the available technology that can be quickly deployed and those suggesting possible novel approaches for the development of advanced technology. Clearly, there are many additional issues that are only alluded to and left open for more in-depth discussions in a follow-up workshop. For instance, there are many open questions involving an even broader inter-disciplinary approach that will bring in researchers from other fields: device physics, electronics, photonics, database design, networking, statistical algorithm and experiment design. This future discussion will focus on signal transduction by electronics or photonics means, data fusion from multiple distributed sensors with diverse operating characteristics, data archival and distribution, real-time analysis of data with robust statistical algorithms and notification of the resulting answer in a comprehensible manner. Eventually, a broad picture will emerge and will delineate how individual components of a large scale monitoring system will ultimately be integrated.

The following is a short list of only those technological goals that this workshop focused upon: a) Development of novel technologies based on genomic arrays and arrays of heteropolymers (e.g. RNA aptamers and other nucleotide-based polymers and Single chain antibodies, protein aptamers), b) Development of a novel particle recognition system based upon geometrical recognition and self-assembly, c) Development of single molecule sensors based upon electrical conductivity properties, fluorochromes & FRET (Fluorescence Resonance Energy Transfer), or time resolved fluorescence.

Additionally, the discussion above raises many questions related to the transducer-architecture and the related bioinformatics/information technology. These questions were only addressed rather briefly.

Similarly, the issues surrounding the design of the transducer raise several questions: It is now possible to design a field effect capacitor or transistors (FETs), whose state could be modulated by the binding event (described later). In a particular design, an array of aptamers may be attached to an array of FETs, and organized in such a manner that if a corresponding aptamer is not bound, it “interferes” with the FET and turns it “on.” The pattern of FET’s in “on” states could be directly decoded to infer the most likely viral particle or protein. Other similar examples may center on the ENFET and ESFET technologies. The data analysis and transmission aspects of the problem raised the following issues: How can we devise the best statistical algorithms that can compare quickly and accurately the pattern of bindings to a small database of possible patterns in order to detect which biowarfare organism is involved, especially if this is a complex mix.

In a distributed implementation of the sensors, there are many technological questions involving data transmission by wireless, archiving the historical data, securely accessing the data over a network, etc. and need to be addressed in-depth.

Summary of Discussions:

1. Desai:

Dr. Mita Desai, a Program Officer from National Science Foundation, Arlington, Virginia, in her address emphasized the need for creating new programs to stimulate basic science research focusing on all aspects of biosensor technology. The new research can address such questions as: *How can existing and novel technologies be further developed to detect pathogens as early as possible? What technologies are immediately available for deployment with modest basic science research and what novel technologies require further significant basic science research and development? Can we design biological materials specifically for certain pathogens of interest and place them on electronic transducers for automatic detection? How can new and reliable technology, combining such biological materials and electronic transducers, be developed quickly such that they can be used for environmental detection at “time $t = 0$ ” (i.e., at the time pathogenic materials may be released and before even the first infection is detected)? Finally, how can computer and information science play a role in the design, deployment and data-interpretation for a distributed system of such sensors? How can NSF address the dichotomy posed by the different basic science needs of these approaches? What are the most promising anti-body based, aptamer based and genome based technologies currently available? How can they best be improved to gain better accuracy, reduce the time of detection, etc.?* Finally, the workshop should also consider various issues involved in creating a plan for NSF to address these problems.

2. Schwartz:

Prof. Jacob T. Schwartz, a Professor of Mathematics and Computer Science at the Courant Institute of Mathematical Sciences, NYU addressed the issues of bioterrorism and the need for a cross-disciplinary approach. In particular, he asked: Is the current technology adequate? What are the major problems with the current technology? Can the existing technology of detectors be easily deployed? Do we have the needed confidence in the current technology? What are the problems with current technology? “Do we know what to do?” There are various criteria for evaluating these sensors: *the ability to detect small amounts of pathogens in air, coverage* (i.e., number of pathogens that these sensors are designed to detect), *sensitivity* (the ability to detect miniscule amount of pathogens in a background of large amount of other biological and non-biological materials), *robustness* (in terms of sensors’ false positive and false negative errors), *speed*, and *cost*.

Prof. Schwartz announced the plans to hold a follow-up second conference funded by Sloan Foundations in near future in Washington, DC. The second conference is likely to be of larger scale than the current meeting and will involve more government policy makers in addition to the scientists. The goal of the second conference will be to identify a major national program that can be quickly launched to address the homeland security problem.

3. Wigler:

Prof. Michael Wigler of Cold Spring Harbor Laboratory in Long Island focused on the issue of detection in a more general manner, not necessarily in the specific narrow context of bioterrorism. As did the previous speakers, he pointed out that detection methods must satisfy many desiderata: *they must be fast, they must be specific and yet flexible*; and finally, *they must be reliable*. The detection methods can be categorized in terms of the nature of the target: (a) *nucleic acids*: DNA or RNA, (b) *protein* and finally, (c) *elements* (supra-structural elements, or organisms)—for instance, viruses, bacteria and other structures. The samples collected may be from (1) air (which may have to be dissolved in fluid), (2) surfaces or (3) body fluid, each case having interesting implications for the detection methods and analysis of the signals. The signals may be of various kinds, but usually, either *electronic* or *photonic*.

His presentation was divided in to three parts: (i) General discussions, (ii) Ideas from his laboratory developed in the context of cancer research, (iii) Several new and interesting (but somewhat speculative) ideas. In particular, Prof. Wigler suggested that a very fruitful novel research direction must be able to exploit the interesting geometric structure common to biological entities, i.e., the fact that they have more “structure.” Examples of such geometry may be based on the distribution of surface proteins on the outer membrane of a bacteria or the regular geometry of a virus. Thus the sensor itself could encode a complementary surface geometry or regular arrays of protein aptamers, which will make the binding of bacterial and viral structures to the sensors more avid than other non-biological structures.

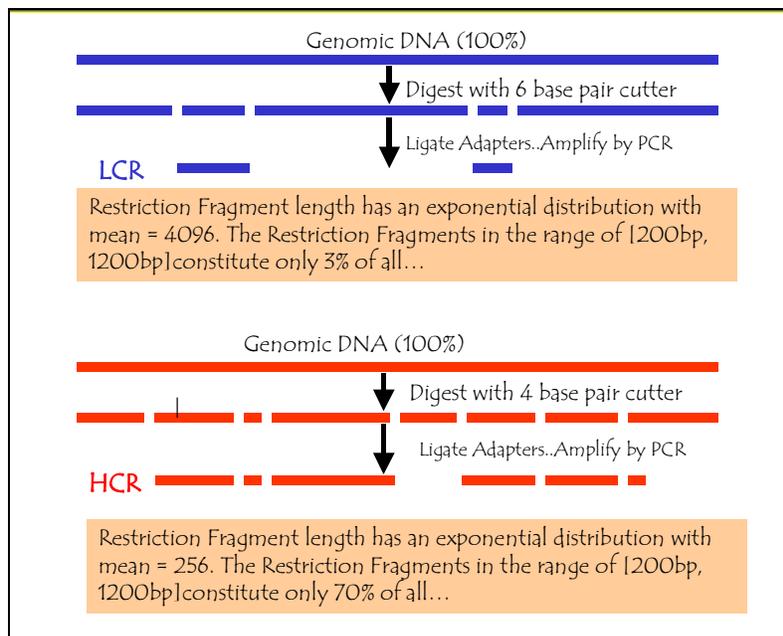


Figure 2. Prof. Wigler's lab has devised a new approach to reduce the complexity of the genomic target by deterministically subsampling it to eliminate genomic regions that should not hybridize to the probes on the array. The approach is based on two steps: In step 1, genomic DNA is cleaved with a restriction enzyme generating restriction fragments of various lengths. The cutting frequency of the restriction enzyme determines the statistical distribution of the fragment lengths and thus ultimately determines the final complexity of the representation. In step 2, only fragments of lengths ranging between 200 to 1200 bps are amplified by PCR reaction, thus eliminating longer fragments. The probes on the array are selected from

the short fragments amplified by PCR. Such a low complexity representation improves the signal-to-noise ratio of the array.

Next, Prof. Wigler described an approach currently used in his laboratory for detecting gene copy number fluctuations in cancer genomes, and how it could be adapted to biosensing. See Figure 2. The method in use employs a DNA based detection system using microarrays, where, by comparing the fold changes for a randomly chosen almost unique set of probes across the “normal” genome, one could detect lesions in DNA, e.g., amplifications and deletions in the genome. In effect, the method implicitly allows a biologist to scan the entire genome with high resolution and with very good signal-to-noise ratio, achieved by reducing the complexity by sampling. The complexity reduction is achieved by selecting small restriction fragments (by PCR that selectively amplifies 200-1200 bp sequences) after the genome has been cut by a 6-cutter restriction enzyme. The method can be easily modified to scan a mixture of genomes containing human, pathogens and other genomes and should be able to detect pathogen genomes even if they are present in extremely low copy number. Prof. Wigler expressed his belief that this method is as fast as anything based on direct PCR or sequencing.

In the final part of his talk Prof. Wigler focused on the challenging problem of “detection of elements (supra-elements).” For instance, if the detector is to detect a whole virus particle with a protein coat, then the detector can achieve much higher accuracy by taking advantage of the regularity of the outer surface and the manner in which these proteins may have been self assembled. Thus the problem of detecting elements must take judicious advantage of the structural and geometric properties of the particle: namely, (i) shape, (ii) scale and (iii) periodicity of the features on the outer surface. In this manner, the detector may achieve high avidity, even when individual components of the detector are limited by their weak affinity. Such a “co-affinity mapping” approach may be based on techniques developed for (a) encoded combinatorial chemistry and (b) self-assembly to create scaffolding with periodicity. Among many research problems such an approach suggests the most important one would be the problem of shape optimization through mathematical modeling and computational optimization or techniques based on “in vitro artificial evolution.”

The last approach generated many interesting suggestions and questions: Dr. Eric David pointed out that aptamers could lead to one of the best mechanisms for element detection—an approach that will be discussed at length in the meeting. Dr. Ellington questioned if elements are the best target to aim for, as there may be very little to gain by detecting the shape. Many participants pointed out that in several situations there may be no nucleic acid involved in the target; PCR based approaches may be more susceptible to background noise, although these problems may be circumvented by hybridization which has a somewhat better performance. Thus future research must distinguish between how to address replicative elements (e.g., organisms) vs. non-replicative elements (e.g., toxins). In cases dealing with organisms, element-based approaches suggest a solution. Dr. Cantor suggested exploring techniques based on “polymer molds” created by Alnis Biosciences, founded by Dr. David Soane, using printed surfaces with patterns. Dr. Seeman suggested research to create the needed scaffolding by self-assembly.

4. Dahlberg:

Prof. James E. Dahlberg of University of Wisconsin, Madison, discussed the four different aspects of the biosensing problem: *Why should we look? What should we look for? Where should we look? How should we look?*

Why should we look? Essentially, there are five broad categories of reasons why the need for biosensor has become so crucial.

- **Bio-terror agents to humans**
In these cases, we may assume the agents of attack are known. They could be through the following: Anthrax spores, smallpox virus, etc.
- **Bio-terror agents to agriculture**
Similar to the ones above. Some examples: Foot and mouth disease, plant viruses, etc.
- **Environmental contamination**
They are motivated by the ability to conduct epidemiological studies or by public health concerns.
- **Monitoring food and water supplies for safety**
- **Military**

What should we look for? Again the proposed biosensors may operate by targeting a particular biological component or some combinations of them. These targets can be categorized as follows, based on whether we are dealing with easily amplifiable nucleic acid or other molecules such as proteins that cannot be easily amplified directly.

- **Proteins:** These can be detected through antigens and by other means exploiting our knowledge of the protein structures (e.g., DNA/RNA based aptamers).
- **Nucleic acids:**
 - *Genomes:* The detectors exploit our understanding of the sequenced genomes of the organisms that may pose a danger.
 - *SNPs:* The single nucleotide polymorphisms (if detectable by the detectors) give us a better understanding of the strain variations and may have many more applications in the context of epidemiology and forensics
 - *Genes:* The detector may target specific genes of the pathogens. Some examples are
 - Pathogenic plasmids: toxins (E.coli 0157/H7)
 - Pathogenic plasmids: Trojan horses
 - Related, non-toxic agents
 - *mRNAs:* The transcriptional states of a cell (e.g. host cells affected by a pathogen) can be monitored through gene expressions and may be used for early and fast detection of biowarfare organisms. However, this approach is ideal for diagnostic devices, and not relevant for environmental sensors.

In the context of mRNA monitoring, it may be useful to focus on a highly targeted group of people such as pediatric nurses or flight hostesses (so called, “canary corp”). Extending this idea, one can imagine future biosensors that can be based on bio-engineered cell lines (“canary cells”) that can react to biowarfare organisms or toxins much more readily. Such engineered cell lines could be modified to have a weakened

immune system or could be nerve cells whose shape of the axon potentials will indicate a possible attack.

Prof. R. Breaker of Yale University suggested that to this list one may also add other bio-materials such as lipids and polysaccharides, even though they are much more challenging in comparison with the problem of detecting proteins and DNA and even though, like proteins, they cannot be easily amplified.

Where should we look? Depending on which particular biowarfare agent is being targeted by the biosensor, the sample collection may be from one of the following locations.

- *Bio-terror related targets:* The samples may be collected from air, water, surface, soil and mail. The sample collection, distribution, etc. could become the responsibilities of FBI, CIA and Homeland defense agencies.
- *Public health related targets:* The samples may be collected from air, water and food. The sample collection process could be delegated to the usual first responders in these situations: e.g., hospitals (emergency rooms), airports, state laboratories and CDC (Center for Disease Control).
- *Military targets:* In these cases, the sample collection is done by various military organizations in locations under their command: e.g., battlefield, field hospital, field laboratories, etc.

How should we look? Note that the sample collection process involves many subcomponents, each one with its special requirements: *Collection*, *Preservation*, *Preparation*, and *Analysis*.

- *Collection:* This process can be carried out using (i) *Air and water filters*, (ii) *Swabs and washes*.
- *Preservation:* In addition to the usual concerns for the material to be well preserved for further analysis, there should be clear records and procedures for “chain of custody.”
- *Preparation:* The preparation process should pay special attentions to the contaminants that may inhibit or mislead.
- *Analysis:* There is a need for well-defined protocols and algorithms for analysis that minimize various kinds of false positive and negative errors.

In conclusion, Prof. Dahlberg, as the previous speakers, enumerated what he considered to be the desirable properties of the sample assay process should be: They should be: *sensitive, specific, reliable, rugged* and *subject to low contamination*. The detection method itself should be: *rapid, easy to learn, adaptable to new targets and agents, scaleable* and *inexpensive*. Furthermore, the method deployed must have *redundancies*, built upon different technologies to avoid being subject to common inhibitors, etc. The detector design must anticipate many different contaminants that could interfere. Thus there is both a need to decide what targets to work on as well as a need to decide what tests to develop. There should be a significant effort expended to interest the private sector to the problem. These inevitably raise many questions: How big is the market? What competition is there? What support is available for research and development?

What is the liability of a mistaken analysis? How quickly does an assay have to be developed? Where can one obtain and test authentic samples?

Finally, Prof. Dahlberg suggested that the report of the “Workshop on Microbial Forensics” organized by American Academy of Microbiology in June 2002 would be of interest to the participants of this workshop.

5. Panel 2: Protein Based Techniques:

This panel, headed by the discussion leader, Dr. Roger Brent of The Molecular Science Institute, Inc., Berkeley, California, discussed various approaches for detecting a protein target using DNA/RNA based aptamers or antibody and peptide aptamer based “tadpole technology,” developed recently at the Molecular Science Institute. In addition to the discussion leader Dr. Brent, the panel consisted of Prof. Ronald R. Breaker of Yale University and Prof. Andrew Ellington of University of Texas, Austin, two experts on Aptamer technology, Dr. Ian Burbulis of The Molecular Science Institute, one of the inventors of the “tadpole technology” and Prof. Meera Sitharam of University of Florida, a theoretical computer scientist, who has developed computer algorithms for satisfying a set of geometric constraints.

Prof. Brent introduced this approach to biosensing as one that works by “adding structure to the problem.” A major advantage of this approach is that it provides a powerful approach to analyze molecules in the environment in a very direct way. These techniques are rather novel and require substantial amount of research before being deployable in a practical implementation. Nonetheless, the aptamer and “tadpole” based approaches are likely to extend to be rather general and flexible for applications involving manipulation of molecules. Thus, one may convincingly argue that nucleic acid and protein aptamers are truly the “Designer molecules for Biosensors” of the future.

Profs. Breaker and Ellington described how aptamer-ribozyme switches are able to detect small molecules because a conformational change to the switch can be detected through a cleavage event of small RNA fragment. Thus one may think of aptamer as a “switch” that targets specific molecules. One can create more complicated switches (e.g., an “AND” or “OR” gates) out of the basic system described earlier, since through a suitable change to the ribozyme one can create chemical machinery necessary to recognize a molecule by binding to two different molecules (or either of the two) before the ribozyme can cleave. These switches have been shown to work quite well in organic solvents, but remain untried for the operation in air.

The aptamer switch needed for a particular biosensing application is difficult to design *ab initio* based on a physical modeling, as the computational problem is essentially infeasible. Based on a theory that since the early RNA’s must have acted as switches interacting with molecules involved in metabolism, one can simply look for the appropriate structures through an “*in vitro* evolution and selection” method. Such attempts in the recent past have resulted in the design of many new aptamers and hence, one can expect that these methods may suffice to create the many different aptamer components that can be assembled to build an aptamer array.

As mentioned earlier, in principle one can combine the aptamers to create simple combinatorial logic gates and hence, it is possible to improve their collective-specificity not just through chemical/structural means, but also through the combinatorial schemes.

Furthermore, as Prof. Ellington described, one can combine molecular beacon structures with aptamer switches in a fairly simple manner such that a conformational change in aptamer has the effect that the fluorophore, which is normally adjacent to a quencher molecule, is separated in the bound state, thus enabling the fluorophore to light up. Such a simple scheme allows the biosensing to occur in a test-tube without the need for a more complex array design.

Next, Dr. Burbulis described a “tadpole” scheme for biosensing. In this technology, an antibody or peptide aptamer is bound to an oligonucleotide sequence (forming the “tadpole” structure), such that different antipodes antibodies or peptide aptamers have recognizably distinct oligonucleotides attached to them. Once the target protein attaches to the antibody of a particular kind, those complexes can be separated and they can be recognized with high fidelity by PCR amplification and the oligonucleotides are recognized by chemical or optical means.

Finally, Dr. Sitharam described a software system that can be used effectively to determine configurations satisfying a set of “geometric constraints.” As the algorithm uses tools from symbolic computation, it can be guaranteed that the software gives topologically consistent answers. A real-time demonstration of the system showed how the system explores the possible configurations and how it interactively treats commands from a user. It also illustrated the potential use of the system to navigate molecular conformation spaces and macromolecular assembly pathways.

6. Panel 1: Genome Based Techniques

This panel, headed by the discussion leader, Dr. Fred R. Kramer of The International Center for Public Health in Newark, New Jersey, discussed various approaches for detecting a collection of genome targets using oligonucleotide based probes or beacons. The panel consisted of Dr. Charles R. Cantor of Sequenom, an expert on PCR and mass-spectrometry based approaches to detect base-composition specific signature (or, variations) of a genome, Prof. Lloyd Smith of University of Wisconsin in Madison, an expert on invader-based assay, Prof. Jingyue Ju of Columbia Genome Center in New York, an inventor of a novel massive sequencing technology and Dr. Fred R. Kramer of The International Center for Public Health in New Jersey, an inventor of the molecular beacon technology.

Dr. Kramer presented a framework for describing assays that work well in practice. The assay based on the molecular beacon technology is one of three ways to detect rtPCR reactions, and one of the two best ways as it enjoys many advantages, enumerated below:

- *Sensitivity*

The signals are highly dependent on the target molecules (genomic DNA or PCR product). Since, in this method, targets are amplified and so are the signals, the signal-to-noise ratios are dramatically improved. The other advantage comes from

the fact that this method allows for direct detection (through instrumentation).

- *Functionality*

This system's improved functionalities are due to: (i) Speed, (ii) Simplicity, and (iii) Reduced Cost,

- *Specificity*

This system's high specificity is due to its ability to select and isolate target and is further improved through better probe design (through a detailed modeling of the physical properties).

- *Organization*

One achieves improved systems performance through: (i) Multiplexing (mono, mini, mega), (ii) Identification processes, (ii) Quantification processes and (iv) Internal controls.

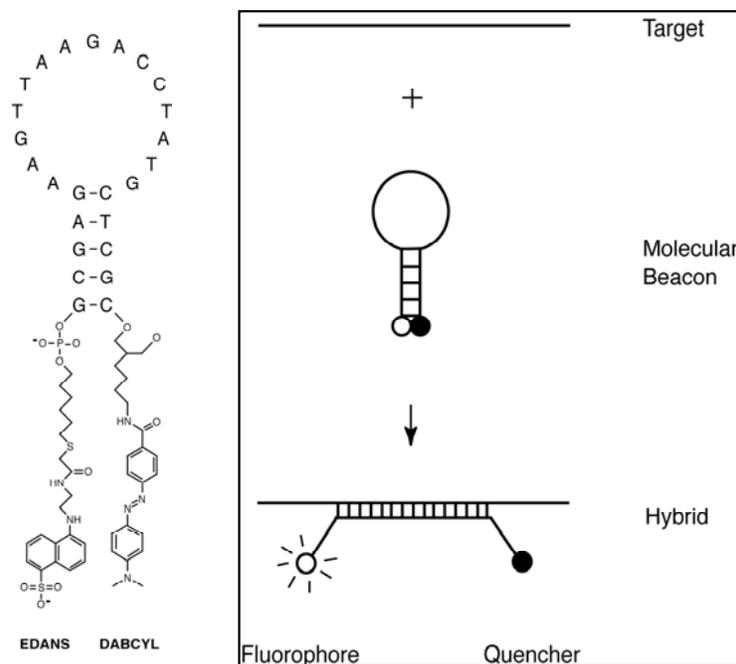


Figure 3. Molecular beacon technology in operation.

The basic idea of the system is presented in the diagram above, illustrating how molecular beacons are used to quantify PCR products in real time. The probe, in this particular implementation, has two complementary “stem” oligonucleotides at both ends of an internal “loop” sequence of interest. In the absence of any other related molecules, these probes have a “stem-and-loop” or “hairpin” structure. At the two ends of the “stem” oligonucleotides are the “fluorophore” and “quencher” molecules. If the probe does not

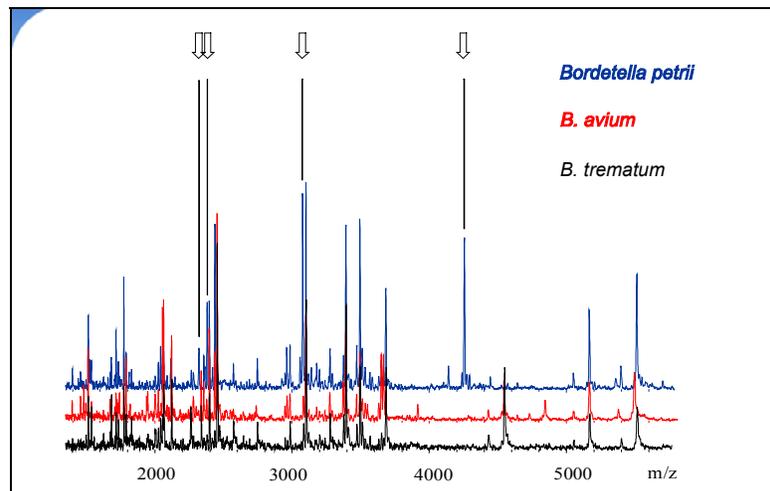
hybridize to any molecule and takes the hairpin shape, then fluorophore does not fluoresce because of the adjacent quencher molecule. On the other hand, in the presence of the appropriate target molecules, the probe hybridizes to the target, forming a rigid double-stranded complex, and also separating the quencher from the fluorophore, thereby putting the probe in a state that allows it to fluoresce when excited at the appropriate frequency. A competing approach for detecting rtPCR product is using the “Taqman” approach: it is well developed, works well, and is the default standard polymerase and should be explored further.

Next, Dr. Charles Cantor of Sequenom, Inc. of San Diego discussed how the PCR and mass spectrometry based methods, now used to detect allelic differences among people, could be deployed almost immediately for monitoring distribution of microbial organisms in the environment. The advantage of the methods he discussed are manifold, but most prominent among these are the advantages due to high degree of automation and the proven accuracy of the system in detecting and distinguishing various pathogens of interest. In terms of the improvement to the technology, he emphasized the following: high-speed flow cell for PCR, very high-speed proprietary PCR chip, use of base specific cleavage to re-sequence pieces of DNA, and finally, making DNA into RNA to make the pieces fly better.



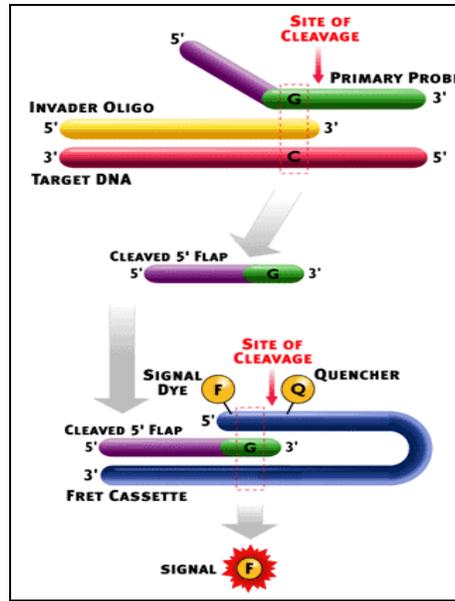
This basic system can be used to detect target DNA molecules as follows: a) one could select a set of very specific PCR probes that can be used to extract a small number of “signature” PCR products from each pathogen via PCR amplification. The signature PCR products are specific to the pathogen, in that it only occurs in the pathogen of interest and not in other pathogens or the human genome (e.g., absent in the background material). b) Next, the PCR products are identified by resequencing. First, the PCR product is cleaved in a base-specific manner (for instance, cleaving at every site where a “C” occurs), thus breaking the PCR product into many small fragments (each containing only the base pairs

other than the one used for cleaving). The PCR product is characterized by the masses of these fragments by means of Sequenom's mass-spectrometry instruments.

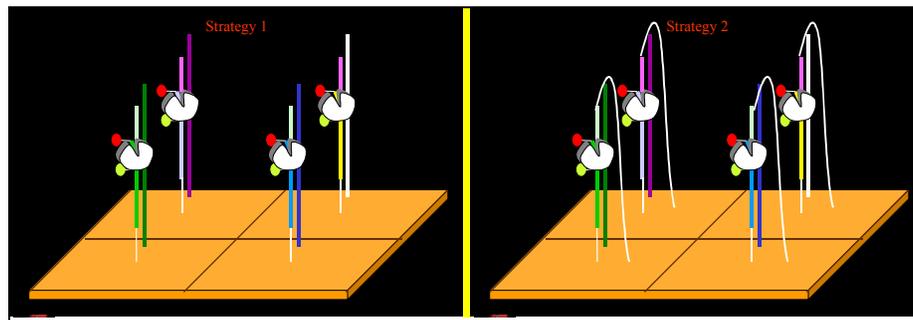


Third speaker of this panel, Prof. Jingyue Ju a Professor at Columbia University's Genome Center in New York City described new technologies with potential applications to biosensing. He described his recent experiments in which two different fluorophores were put onto tags with different distances between the fluorophores in order to obtain different FRET signatures and thus, distinguish tags. For example, by using only three colors, it was possible to get the effect of eight distinguishable tags, and it was shown to be able to tell pieces of DNA apart from one another by color combinations. These ideas could allow for detection of thousands of pathogens using FRET technology with only handful of colors. Prof. Ju in his earlier experiments had already demonstrated how to use these different tags on PCR primers to perform SNP detection—the different tags confer slight differences in mobility as well on gels, an effect that could be fruitfully exploited in the design of a sensor. Prof. Ju also explained a novel single molecule sequencing method he is currently developing.

The final speaker of the panel, Prof. Lloyd Smith of University of Wisconsin, Madison and a founder of the Third Wave Technology, described the invader assay technology that they have developed for SNP (single nucleotide polymorphism) detection—a technology that can in principle be suitably modified to create a biosensor. Among many advantages this system enjoys, the primary one is that the required amount of DNA need not be as high as it is in other approaches.



The approach described by Prof. Smith is based on the “invasive cleavage reaction.” An invader oligonucleotide and a probe oligonucleotide attach to the target DNA in a manner determined by the base constitution of the target. In particular, in the single base position, where the invader and probe oligonucleotides overlap, if the bases in the probe and target are complementary, then the third oligonucleotide (“flap oligonucleotide”) belonging to the probe gets cleaved off by FLAP (an enzyme). This flap oligonucleotide is detected by a combination of the “molecular beacon” technology and the invader assay technology; the flap oligonucleotide with just a single base from probe on it, invades a hairpin structure DNA with a fluorophore and a quencher and when the hairpin gets cleaved by FLAP (again in the second reaction), the signal is detected. This is also called “invader squared” and reflects how the combination of these two events improves the specificity of the system. The technology described above can be transferred to a “microarray-like” system, where many reactions can be carried out simultaneously.



7. Panel 3: Miscellaneous

This panel, headed by the discussion leader, Prof. Nadrian C. Seeman of New York University, New York, discussed various other approaches and issues related to

biosensing. The panel also consisted of Dr. Harold R. Garner of University of Texas South Western Medical Center, Dallas, an expert on “mask-less gene-chips” (DOC), Dr. Eric Michael David of The Rockefeller University of New York, a physician with considerable practical experience with medical emergency situations, Prof. Bud Mishra of Courant Institute of Mathematical Sciences in New York, a computer scientist with expertise in genome analysis algorithms, Prof. Animesh Ray of Keck Graduate Center and University of Rochester, a biologist with expertise in DNA-based computer technology and finally, Dr. Scott Manalis of MIT Media Laboratory, an expert on the design of electronics (e.g., a novel fabricated device, Field Effect Capacitor) to interface directly with biological materials.

Dr. Nadrian Seeman of the Chemistry Department of NYU described the research work of his laboratory, centering around a nanoscale self assembly approach with DNA molecules. For instance, potentially, this approach can be extended to devise self-assembly of a mold for protein detection by ligation or for finding patterns on the surface of a particle.



Figure 4. A cube structure created with branched DNA.

For instance, by using immobile branch junctions, designed with doubled stranded DNA with sequence composition minimizing sequence symmetry, one can create a small structure whose branch point acts as a node of a graph and whose branches act as arms with ability to connect to specific other arms by self-assembly to produce edges of this graph. Such a graph structure (either finite or potentially infinite) will have physical structures based on the edge geometry and can create rigid polytopes of certain shapes with connecting structures (DNA nanorobots), space lattices with certain texture or even a circuit or an embodiment of a computation (computationally Turing equivalent). See figure 4.

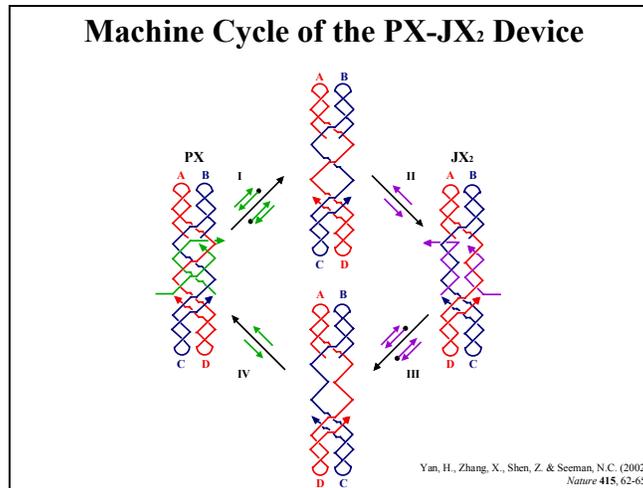


Figure 5. A switch with two different shapes determined by a hybridizing target.

Furthermore, Prof. Seeman also demonstrated a novel approach in which the individual geometry can be locally and reversibly modified by hybridization with a small oligonucleotide. See figure 5. He also described how this new technology could be used in creating programmable geometry for detecting particles.

Dr. Harold R. “Skip” Garner of University of Texas South Western Medical Center, Dallas described a mask-less programmable microarray system where the photosensitive oligonucleotide synthesis process is controlled by a programmable Texas Instrument MEMS based micro-mirror-array chip. Similar systems are also available from two other competing groups, one of the other being a commercial venture called Nimblegen from Madison, Wisconsin.

Dr. Garner also described a surface plasmon detector system that accomplishes biosensing by observing optical properties of proteins, is competitive to mass-spectrometry-based systems, and is small enough to be easily deployable.

The next speaker, Dr. Eric Michael David of The Rockefeller University of New York, described various problems that the biosensors face because of the issues involving sample preparation. He described various experiments conducted to study septic shock (Sepsis) using high-throughput microarray analysis and the statistical problems unearthed through ANOVA test of the genomic expression data. The key problem seemed to be due to the manner in which sample preparation was conducted—picking up materials from different parts of the body has huge effect on the non-stationarity of the statistical data. The various components of the problem can be categorized and isolated as follows: 1) *Sampling*, 2) *Sample Preparation*, 3) *Sensing*, 4) *Analysis*, and 5) *Result*. He also described some results from Anthrax data analysis, where the data were collected by emission and absorption spectra obtained from photo-spectrometer data. He also described the statistical data analysis algorithms used. He raised several ethics-related questions about sampling a population without their consent.

The next speaker, Dr. Scott Manalis of MIT’s Media lab discussed various MOSFET (Metal-Oxide-Semiconductor Field Effect Transistor) like devices with direct applications to biosensing as transducers. These devices are created out of silicon surfaces with a silicon-dioxide insulating layer, and they act like field effect devices with

the major difference being that its capacitive properties play a more major role than its switching properties. In one other implementation, the FET device is designed as a cantilever structure made of silicon dioxide. In each case, the oligonucleotides, linked to the silicon-dioxide surface, change the electrical properties of the surface depending on whether the surface oligonucleotides are bound to an ambient target DNA by hybridization or not. Dr. Manalis showed that these systems are capable of detecting 10^4 DNA's per $2 \times 400 \mu^2$ area. He also demonstrated other schemes that make use of similar other mechanisms: for instance, one based on measuring displacements of cantilever-like structure by a diffraction grating.

The last speaker of the day, Prof. Bud Mishra of Courant Institute (NYU) and Cold Spring Harbor Laboratory, spoke about the design of the biosensor devices, the experimental conditions and their implications to the design of the statistical algorithms. In particular, he described how the interaction between the experimental parameters and the algorithmic problems has a serious implication in terms of the underlying computational complexity. He drew an interesting example from his earlier work on restriction mapping using single molecules, where various unavoidable and serious errors produced by the chemical experiments (e.g., partial digestion, sizing errors, false sites, etc.) may in some cases make it impossible for the algorithm to draw any meaningful conclusions. He also described various “computational phase transitions” that are hallmarks of these experiments and how this effect can be exploited to design experiments with accompanying simple and efficient algorithms.

8. Brent:

The last speaker of the meeting, Dr. Roger Brent, provided a personal perspective and summarized all the essential topics that came up during the meeting. In his presentation, he broadened the concept of “sensor” to cover all possible “means of detecting things.” Thus, now, one has many different ways to slice up the “sensor” space; they can be categorized as follows: (A) Where the sample originates: (i) Sensors that sense things inside the human body or, more frequently, in samples drawn from the human body. (ii) Sensors that sense things from the environment. Now, (B) for the environmental sensors, the sensors can be characterized based on where the sample is acted on: (i) Sensors that detect some property of molecules still in the environment (remote sensors) (ii) Sensors that act on molecules taken from the environment. Of course, the final characterization could now be based on (C) What you detect: (i) Sensors that detect the presence of nucleic acid sequences (e.g. specific sequences, all sequences, etc.), (ii) Sensors that detect other molecules.

Sensors that sense things inside the human body: (or, more frequently, in samples drawn from the human body). These are frequently used as diagnostic devices, a typical example being a cDNA microarray, in which an array of DNA sequences, drawn from thousands of human genes, can detect and quantify absolute expression of those genes in terms of mRNA. Specific patterns of gene expression in white cells (human PBMCs) can be quite indicative, for example of infection with different diseases. Furthermore, these devices can also be used as prognostic device (e.g., cancer classification—more specifically, classification of leukemia into ALL vs. AML). Dr. Brent pointed out that we already have these devices for nucleic acids, and will soon have for proteins. Hence these devices can be and should be deployed soon for diagnosis after a bioterror attack.

Sensors that sense things in the environment: As discussed earlier, they fall into two classes: (a) sensors that act on molecules taken from the environment, vs. (b) sensors that detect some property of molecules still present in the environment (remote sensors). An example of the first category is a filter disk in a ventilation system of a public building, from which DNA can be extracted and analyzed by PCR using specific primers. An example of a remote environmental sensor would be a “hyperspectral” imaging system placed in reconnaissance satellites, or the LIDAR detection of particles in aerosol clouds. Nonetheless, Dr. Brent cautioned against high false hopes, as “the idea that any currently understood physical phenomenon can be parleyed into ‘standoff’ techniques that can detect the low numbers of entities needed in biology is purest fantasy.”

Remainder of the talk focused on different types of sensors for monitoring molecules in samples taken from the environment—much of the technology remains relevant to medical/diagnostic sensors as well.

Sensors that detect the presence of specific nucleic acid sequences:

- **For DNA:** These could be PCR with defined primers to detect specific DNA sequences. Accurate quantification, while preserving existing “dynamic range,” is quite difficult for straight PCR. But one may do considerably better in quantifying the data, if attention is also given to the number of amplification cycles required before a product is first observed. Real Time PCR, rtPCR, to produce fluorescent PCR products is another common approach. A reasonably good quantification can be achieved by measuring number of cycles required before fluorescence product signal exceeds background. Another approach could be based on high-throughput microarray technology, which requires algorithms for deconvolving mixtures of PCR products hybridized to DNA array.
- **For RNA: Similar to the ones discussed earlier.** RT (*reverse transcriptase*) PCR can be used to convert RNA into DNA and perhaps, only amplify specific resulting DNA sequences. Once the DNA is available, one may proceed as before.

“Sensors” that detect the presence of any sequence: These sensors are most definitely not devices (for the moment) but rather methods.

- **For DNA:** The DNA is first broken into PCR-sized small fragments; they can then be ligated on “universal” primer linkers; they are further amplified by PCR and the resulting DNA is sequenced. This method has many advantages: it identifies suspected sequences (and hitherto unsuspected ones); provides information on relative abundance; individual small laboratories or large industrial-scale biotech laboratories can participate in this process at different scales.
- **For RNA:** Similar to the techniques used above, except that RNA must be converted into DNA by reverse transcription.

However, in order to improve the specificity of these techniques, several pre-processing or screening steps will be necessary: for both DNA and RNA, one must filter out (either physically or computationally) subsequences that play no role (i.e., increase the complexity without adding information), before the analysis either by sequencing or hybridization. In the future, it is possible that these systems will find applications not just

for bio-defense, but also for large-scale “bio-monitoring” of the genes in the environment.

Sensors that detect the presence of non-nucleic acid molecules in samples: Here, it makes more sense to focus on two aspects of these sensors: (a) recognition molecules, and (b) schemes to generate signal resulting from the binding event.

a) Recognition molecules: Examples are: Polyclonal antisera, Monoclonal antibodies, aptamers, nucleic acid, protein or other special purpose molecules (e.g., Straptavidin, fancy combinatorial successors), etc.

b) Detection schemes:

- **Pre-WWII technology:** Ouchterlony, “Rocket”, antibody-coated latex beads with dye on the latex, etc. Currently deployed detection technology is still based on these methods!
- **Post-WWII technology:** RIAs (Radio-immune assays), ELISAs (Enzyme Linked ImmunoSorbent Assays), etc. These technologies may not be appropriate for environmental detection or early-warning systems as they are only able to detect millions of molecules, but generally fail for smaller numbers (say, in thousands).
- **1990s technology:** Evanescent Wave, Surface Acoustic Wave, Cantilevers, Field Effect, etc. These systems seem to depend on accumulation of a fair amount of matter on a surface (perhaps, with possible exception of FET), and have same order of sensitivity as RIAs and ELISAs.
- **2008 technology:** Engineered protein based approach or perhaps, systems based on ion channels.
- **2020 technology:** Engineered organisms containing engineered channels, etc.

Requirements:

- **Requirements for the molecules that do the recognizing:** (i) High affinity, (ii) High specificity and (iii) Ability to work with amplification schemes. Thus, we see that the stretches of DNA now used as PCR primers provide a superb platform for detection. Antibodies are, by contrast, not of comparable affinity, although they are pretty specific. While nucleic acid and especially protein aptamers appear highly promising, they may need much more work to get high enough affinity, although they may already be specific enough.
- **Requirements for the scheme that detects the binding event:** (i) High signal to noise, (ii) ability to get a signal out of <100 binding events, etc. At present, PCR and other related methods perform well, but in the near future, engineered ion channels may become strong challengers.
- **Requirements for the “back end” that generates the signal after the binding event:** (i) Automated analysis of amplified nucleic acid sequences by hybridization and/or by sequencing, (ii) Automated analysis of the object

generating the signal from binding proteins and other molecules. Eventually, the back end needs to communicate with other devices and may have to be networked for distributed analysis of the data.

In conclusion, Dr. Brent pointed out that the most promising approach for the time being is presented by the methods based on nucleic acids. The methods for proteins and other molecules are likely to be deployable very soon.

Recommendations:

It appeared obvious to this group that the first immediate step for creating biosensors must start with monitoring bio-information systematically, and archiving this data for further detailed statistical analysis. The most challenging task could be understanding and exploiting complex interaction of all possible sources of noise in the biosensors and the statistical algorithms that must achieve sensory fusion of the data from multiple sources, achieve fast reaction to an attack and yet, minimize both false positive and negative errors.

Towards this goal, the panelists made several concrete recommendations that are likely to result in availability of preliminary data, rapid progress, and timely deployment of a usable system. It was agreed that there already exists a technology (based on PCR and analysis of the PCR-products) that, while not ideal in the long run, is immediately deployable and is likely to produce a better understanding of the problem. Nonetheless, the reliance on this technology should be considered temporary. No single “panacea” technology is likely to emerge in the near future. Instead, one may aim for a spectrum of technologies: some of them may be based on large-scale factory-type operations, as opposed to miniaturized distributed remotely accessible and reusable sensors. Thus,

Technological Recommendations

The next set of recommendations centered on key areas of the current technology that pose serious bottleneck to rapid progress:

1. ***Development of more avid method by increasing sensitivity while preserving a great difference in affinity between specific and nonspecific targets:*** Since the biosensor of the future is most likely to perform its recognition by binding an oligonucleotide, aptamer, antibody etc., to target DNAs, proteins or particles, and then separating the bound state from the unbound state, a significant improvement of biosensor’s performance depends on how tight the binding is. The improvement in avidity could be achieved by selecting better binding molecules, by exploiting geometry and regularity of the particles, through combinatorial approaches, by intelligent sub-sampling of the targets, and finally, by optimal selection of ambient operating conditions (e.g., temperature, salt-concentration, etc.)
2. ***Concentration Improvement:*** Another problem affecting the fidelity of the biosensor relate to the sample preparation and improvement of the target concentration. In particular, if one imagines dealing with samples that may contain only handful (say 4-10) of pathogen molecules mixed in with a much larger samples containing other harmless biological materials (e.g., genomic DNA, proteins, etc. from the host cells or other symbioants), the signal corresponding to the biowarfare molecules and detected by the biosensor may be too feeble in comparison to the noise contributed by the “background” molecules. Various biochemical approaches can be used to improve the concentration of the

biowarfare molecules relative to the background, by selectively eliminating significant amount of the background material, in particular the parts of the background that contribute to the noise and/or false positive detection. Alternatively, the same effect can be achieved by selectively amplifying the biowarfare molecules, while leaving the background material unamplified or only weakly amplified. In the context of DNA-based approach, an example would be the complexity-reduction accomplished in Wigler's laboratory by selecting only small fraction of the targets (select small restriction fragments after the entire genome has been completely digested by a restriction enzyme).

3. **Signal Processing:** Finally, in a full-scale system it is likely that there will be multiple distributed set of sensors, each exposed to its own error processes and idiosyncrasies, perhaps operating under related but varying protocols and ambient conditions and with thresholds optimized for different rates of false positive and false negative errors (e.g., imagine that same sensors may be employed differently in civilian applications than in military ones). A critical component then becomes the algorithm in control of processing the raw signals, translating them into common formats, normalizing the signals, sensory integration and finally, interpreting the data into a format that can be understood by an unsophisticated operator.

Recommendations for Funding Organizations.

At this point, it is difficult to anticipate all possible issues. But we can be better prepared by creating an appropriate infrastructure. Thus the funding organizations might focus on creating several task forces, each one in charge of a *focus of effort*. These foci of efforts can be broadly categorized into two groups:

1. Research [**R**]
2. Projects [**P**]

and can be enumerated as follows:

- *Research based on various targets:*
 1. DNA [**R & P**]: Here the targets are DNA molecules obtained from the people, collected in emergency rooms or airports. The recognition can be based on hybridizing oligonucleotide probes on an array or through the selection of PCR primers.
 2. Proteins/Other [**R**]: Similar to the previous approach, the targets are protein molecules obtained from the people, collected in emergency rooms or airports. The targets can also be surface proteins being detected directly from air samples. The recognition molecules can be aptamers or antibodies.
 3. Particles [**R**]: Here, the targets are treated as intact viral or bacteria particles in air sample detected directly by a reusable sensor (e.g., an aptamer array or "molecular velcro.")
 4. Confounding [**R**]: This is probably the most difficult situation, where not

much can be assumed about the potential target (the target may have been cleverly constructed by an adversary capable of creating a false negative or a misleading positive.) In order to address these kinds of attacks, the best solution is to consider many different and complementary technologies and to combine their signals in a statistically robust manner. Once their interactions are understood better and systematized, one could proceed to automating the combined technology in one sensor.

- *Automation of existing laboratory methods [P]*: This step is critically needed as most of the current lab protocols are manual and hence, tedious, and fail to produce results in a timely manner.
- *Data Analysis [P]*: There is also a critical need for computer software designed to handle large-scale design problems, perform statistical analysis of the data and manage large amount of real-time data.

Field Pilot Study and Data generation: The workshop participants also recommended that a field pilot study be initiated *as quickly as possible* with the goal of creating protocols for handling

- Collection of Data on Microbial Populations
- Dissemination, Archival and Analysis of the Data

The idea for this pilot project came from Harold “Skip” Garner, (University of Texas South Western Medical Center, Dallas) and Charles Cantor, Sequenom, Inc, San Diego.

It is widely recognized that repeated testing of the biosphere for potentially very dangerous pathogens that might be introduced by hostile action can be a valuable way of detecting bio-terrorist attack quickly, and thus of limiting the consequences of such an attack.

Error-Quantitation

This pilot study directly addresses the following question: *Do there exist systems with the required capabilities presently capable of use at large enough scale to begin serious error-rate quantitation?* We believe the answer to this question is yes. For instance, Sequenom Inc, of San Diego has developed a fully automated PCR system capable of handling 200,000 PCR reactions/day. After automated PCR this system automatically submits the generated PCR product to a mass-spectrometric analysis (MALDI-TOF), producing a digital output whose very detailed nature should allow very small false positive rates to be attained.

It is anticipated that using such a system it is possible to collect vast amount of data once samples have been gathered from many different sources in a “fully automated, end-to-end, samples-in to digital-data-out service.” In a proposed scheme, using the sequenced genomes of biowarfare pathogens, several PCR primers are designed in a manner such that a particular primer pair selects only a DNA fragment, which is specific to the particular pathogen and absent in the general background genomic material (e.g., not occurring in the human genome or the genomes of the other pathogens). These

“signature” DNA fragments are selected and amplified rapidly using Sequenom’s PCR technology and the signatures are recognized accurately using mass-spectrometric analysis.

Based on this technology, a group of researchers from New York University has proposed to organize a large-scale study, extending over a 1-year initial period, which will test daily samples from the following sources,

- Air samples collected from a few filters located in high-traffic New York public spaces, buildings, and transportation hubs.
- Oral swabs collected in dentist’s offices.
- Oral swabs collected in hospital emergency rooms.
- Blood samples collected from patients going in for routine hospital blood work.
- Artificial samples, generated to contain known (but harmless) pathogenic DNA. These will be designed to quantitate the system’s false negative rate.

General Recommendations to Funding Agencies: The funding agencies such as NSF should aim at meeting the need for stable, long-term support focusing on the basic science research for biosensor in a manner that more mission oriented agencies (e.g, DARPA or Homeland Defense Agencies) are unlikely to address. These NSF programs should be much more broadly based and emphasize the development of a multi-disciplinary science that combines ideas from chemistry, biology, device physics, and information technology. Such an approach will yield results that will have far reaching implications. For instance, a technology for detecting minute amount of proteins by aptamers will result in a high-throughput protein array technology that can be used to monitor the translational states (proteome) of a cell, which when combined with microarray-based transcriptional states (transcriptome) will strengthen current methods in functional genomics.

More concretely, following areas should receive special attentions:

- *Focused effort on integration [R]:* NSF may initiate new programs that emphasize novel integration schemes. For instance, one such project may address how to integrate aptamer-based sensors with electrical devices that immediately report the bound state of the aptamer.
- *Real time continuous monitoring [R]:* Another NSF program may address the information technology related issues geared at devising real-time statistical data analysis that continually improve both false positive and negative error rates.
- *Direct embedded systems [R]:* A third NSF program may address the device and communication infrastructure needed to connect a large number of these sensors such that any individual sensor is remotely accessible.
- *Literature database for bio-sensing [P]:* A program to create a sharable database for the emerging biosensing community. Such a database would not only make all the relevant literature readily accessible, but also aim at disseminating information about various announcements, novel devices, materials and industrial technologies as soon as they become available. The proposed system could be

modeled after the existing pub med database, which is widely used by biologists.

- *New program for biosensors [R]*: There should be additional NSF programs with the goal of bringing together biologists and non-biologists developing new technology.

Conclusions and Future Avenues:

The threat of biological attack is now quite real and might, for example, develop further consequent to current military conflicts. After any such attack the pressure to respond would be overwhelming. Even now, we need to move with all prudent speed to strengthen our defenses. Early detection and pre-symptom detection of biowarfare organisms can greatly alter the outcome of an attack in terms of morbidity and lethality. Biosphere surveillance systems are an essential part of this. One requires technical means capable of assaying samples which are sufficiently reliable and sensitive to reduce false positive and false negative rates to acceptable levels, and which are capable of examining samples for an adequately wide range of pathogens. We need means for routine sampling of the many spaces, which people traverse in large numbers. A crucial policy question is whether any of the very many technical options currently reported in the literature and reviewed in our conference is ready for large-scale development. Informed opinion seems, in agreement with the general consensus of our conferees, to hold that the answer to this question is 'Yes.' Several recently developed biological assay technologies suggest themselves as attractive candidates for this application. Real-time PCR using molecular beacons is one such. Another is selective PCR amplification of target pathogens, followed by mass spectrometry as a means for detecting the amplified DNA product.

If this view is accepted, the next step, to be taken as soon as possible, should be to organize a limited number of major National Demonstrations, each centered on one of the assay technologies identified as most promising, and aiming to establish their adequacy in field settings. Real world rather than laboratory use of such systems must deal adequately and reliably with a spectrum of problems that only such demonstrations can assess adequately. A crucial problem is to determine protocols, which steer appropriately between the twin dangers of false-positive and false-negative errors. Environmental samples often contain contaminants that can interfere with basic PCR-based amplification and cause PCR to fail or mislead. High background levels of cross-contamination can foil immunoassays even more readily. This problem grows even worse when complex samples are preserved for any length of time. Aside from all these issues, organizing large-scale sample collection, preparation, and large-scale daily assaying poses many practical problems. But it should be possible to overcome all these problems, and we need to begin building the necessary laboratory capabilities and systems before an attack occurs.

Acknowledgments:

The authors wish to acknowledge help, advice and editorial remarks from Mita Desai, Roger Brent, Jan Witkowski, Mike Wigler and other participants. In particular, Roger Brent worked with the authors through many versions of this document with a keen eye on the scientific accuracy of the material and its readability. We also thank Jane Mishra, Joe West, Toto Paxia and Raoul Daruwala for many constructive remarks.

Appendix:

Agenda:

Designer Molecules for Biosensor Applications

The Banbury Center, Cold Spring Harbor Laboratory, 12-14 August 2002.

Organized by: Bud Mishra, Courant Institute of Mathematical Sciences, New York University, New York and Cold Spring Harbor Laboratory, Long Island.

Program

Monday, 12 August

Afternoon	Arrival at Robertson House, Banbury Center, for Registration and Room assignment
6:00 pm	Reception at Robertson House
7:30 pm	Dinner at Robertson House

Tuesday, 13 August

7:30-8:30 am	Breakfast at Robertson House
8:50-9:00 am	Jan A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory, New York, <i>Welcome</i>
9:00-10:30 am	Session 1: Chair: Bud Mishra, Courant Institute of Mathematical Sciences, New York University, New York
9:00-9:10 am	Mita Desai, Program Officer, National Science Foundation, Arlington, Virginia
9:20-9:30 am	Jacob T. Schwartz, Courant Institute of Mathematical Sciences, New York University, New York
9:30-10:00 am	Michael Wigler, Cold Spring Harbor Laboratory, New York

10:00-10:30 am	James Dahlberg, University of Wisconsin, Madison
10:30-11:00 am	Break
11:00-12:30 pm	<p>Session 2: Discussion Panels Panel 1: Genome Based Techniques Chair and Discussion Leader: Fred R. Kramer, The International Center for Public Health, Newark, New Jersey</p> <p>Charles R. Cantor, Sequenom, Inc., San Diego, California James Dahlberg, University of Wisconsin, Madison Jingyue Ju, Columbia Genome Center, New York, New York Fred R. Kramer, The International Center for Public Health, Newark, New Jersey Michael Wigler, Cold Spring Harbor Laboratory, New York</p>
12:45-2:00 pm	Luncheon at Robertson House
2:00-3:30 pm	<p>Session 3: Discussion Panels (cont'd) Panel 2: Protein Based Techniques Chair and Discussion Leader: Roger Brent, The Molecular Science Institute, Inc., Berkeley, California</p> <p>Roger Brent, The Molecular Science Institute, Inc., Berkeley, California Ronald R. Breaker, Yale University, New Haven, Connecticut Ian Burbulis, The Molecular Science Institute, Inc., Berkeley, California Andrew Ellington, University of Texas, Austin Meera Sitharam, University of Florida, Gainesville</p>
3:30-4:00 pm	Break
4:00-5:30 pm	<p>Session 4: Discussion Panels (cont'd) Panel 3: Miscellaneous Chair and Session Leader: Nadrian C. Seeman, New York University, New York</p> <p>Nadrian C. Seeman, New York University, New York Harold R. Garner, University of Texas South Western Medical Center, Dallas Eric Michael David, The Rockefeller University, New York, New York</p>

Bud Mishra, Courant Institute of Mathematical Sciences,
New York University, New York
Animesh Ray, University of Rochester, New York

6:00 pm Reception at Robertson House

7:00 pm Dinner at Robertson House

Wednesday, 14 August

7:30-8:45 am Breakfast at Robertson House

9:00-10:45 am **Session 5: Conclusions**
Chair: Bud Mishra, Courant Institute of Mathematical
Sciences, New York University, New York

9:00-9:30 am Roger Brent, The Molecular Sciences Institute, Inc.,
Berkeley, California. *Personal Perspective.*

9:30-10:45 am **Session 6: Panel Summaries and Review**
Discussion Leaders and participants

10:45-11:15 am Break

11:15-12:15 pm **Session 7: Key Points for Report and Recommendations**
Mita Desai, National Science Foundation, Arlington,
Virginia
Steve Katona, Department of Health and Human Services,
Washington, DC
Jacob T. Schwartz, Courant Institute of Mathematical
Sciences, New York University, New York

12:30 pm Luncheon at Robertson House

Afternoon Departure

Participants:

- **Ronald R. Breaker**, Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT.
- **Roger Brent**, The Molecular Sciences Institute, Inc., Berkeley, CA.
- **Ian Burbulis**, The Molecular Sciences Institute, Inc., Berkeley, CA.
- **Charles R. Cantor**, Sequenom Inc., San Diego, CA.
- **James Dahlberg**, Biomolecular Chemistry, University of Wisconsin, Madison, WI.
- **Eric Michael David**, The Rogosin Institute, The Rockefeller University, New York, NY.
- **Mita Desai**, National Science Foundation, Arlington, VA.
- **Andrew Ellington**, Department of Chemistry, University of Texas, Austin, TX.
- **Harold “Skip” R. Garner**, Department of Biochemistry, U Texas Southwestern Med Center, Dallas, TX.
- **Jingyue Ju**, Department of Chemical Engineering, Columbia Genome Center, New York, NY.
- **Steve Katona**, Department of Health and Human Services, Office of Public Health Preparedness, Washington, DC.
- **Fred R. Kramer**, The Public Health Research Institute, The International Center for Public Health, Newark, NJ.
- **Scott Manalis**, MIT Media Laboratory, Cambridge, MA.
- **Bhubaneswar “Bud” Mishra**, Courant Institute of Mathematical Sciences, NYU, New York, NY & Cold Spring Harbor Lab, Long Island, NY.
- **Animesh Ray**, Keck Graduate Institute, Claremont, CA.
- **Jacob T. Schwartz**, Courant Institute of Mathematical Sciences, NYU, New York, NY.
- **Nadrian “Ned” Seeman**, Department of Chemistry, New York University, New York, NY.
- **Meera Sitharam**, Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL.
- **Lloyd M. Smith**, Department of Chemistry, University of Wisconsin, Madison, WI.
- **Michael Wigler**, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

