BIOLOGICAL SCIENCES AND TECHNOLOGY

Seed Money Fund

Simulation of Biomolecular Complexes: Advanced Computational Sciences at the Molecular Level

B. E. Hingerty,[1] T. Vo-Dinh,[1] S. Broyde,[2] and M. Cosman[3]
[1]*Life Sciences Division*
[2]*Biology Department, New York University*
[3]*Biology & Biotechnology Research Program, Lawrence Livermore National Laboratory*

We have developed advanced theoretical modeling to study the interactions between bioreceptors with target species at the molecular level. (1)The pyridyloxobutylating (POB) agents derived from metabolically activated tobacco-specific nitrosamines can covalently modify guanine bases in DNA at the O^6 position. The adduct formed, O^6-[4-oxo-4-(3-pyridyl)butyl]guanine ([POB]dG), results in mutations that can lead to tumor formation, posing a significant cancer risk to humans exposed to tobacco smoke. A combined NMR-molecular mechanics computational approach was used to determine the solution structure of the [POB]dG adduct within an 11mer duplex sequence d(CCATAT-[POB]G-GCCC)·d(GGGCCATATGG). (2) Six DNA hybridization structures were calculated using the program DUPLEX to compare the effects of triplet base repeats and the effects of a single base substitution on DNA structure and energy. (3) Protamine is a small (~50 amino acids), highly basic (pI ~ 12), arginine-rich (~50%) protein that binds to and condenses DNA in the sperm of most vertebrates. The NMR studies have been completed for one complex, and work on two more peptide/DNA hairpin complexes, as well as the full-length protamine/DNA hairpin complex, is under way. We expect this work to continue while we obtain outside funding.

Introduction

Biological molecules have evolved exquisite signaling and detection mechanisms to communicate with one another. Recently, much effort has been put forth to develop improved bioprobes and to mimic these natural systems in developing robust artificial receptors, or biosensors, as sensitive and highly specific detectors. A first step toward this goal would be the ability to understand the fundamental principles by which these molecules recognize and bind to each other. Thus the aim of this proposal is to develop new computer algorithms and programs to better enable modeling of DNA/protein or protein/protein complexes, which in turn will allow rational design of smart biosensors, with applications toward detectors for diseases, toxins, and other agents.

Technical Approach

Although commercially available software packages are capable of determining biomolecular structures and interactions, they rely heavily on acquiring accurate experimental data. Thus biomolecular interactions that are transient, weak, or otherwise difficult to characterize experimentally but are of critical importance to function are not easily studied. There is, therefore, a need to develop software that can independently provide information about these difficult systems and allow the development of biosensors capable of detecting biologically relevant agents under native conditions. We began by studying a peptide/DNA hairpin complex. Our recently improved sigmoidal distance-dependent dielectric function,[2] which enables our force field to model structures and energetics of carcinogen-DNA adducts to within 1–2 kcal/mol of the experimental NMR benchmarks, was employed at this stage to implicitly model a solvent. Counterions were modeled via reduced partial charges on the non-linked phosphate oxygens according to counterion condensation theory. We had previously developed methods for determining structures of carcinogen-damaged DNA.[1] Our newly derived algorithms were tested with an experimental peptide/DNA complex and against commercially available packages. The experimentally derived NMR distance restraints for the peptide/DNA sample were provided by Dr. Monique Cosman (Biology & Biotechnology Research Program, Lawrence Livermore National Laboratory). Our model building program, DUPLEX,[1] using improved potentials that accurately reproduce B-DNA double helical structure,[2] in concert with commercial software, was used to perform docking experiments whereby the bioreceptor is complexed with the specific analyte under question. A typical example would be a DNA/protein complex, an antibody/antigen complex, or a DNA hybridization study. The torsion space minimization procedure has a much greater radius of convergence than Cartesian minimization

routines, so stable minima are more likely to be found far from the starting position.

Results and Accomplishments

(1) The **pyridyloxobutylating (POB) agents** derived from metabolically activated tobacco-specific nitrosamines can covalently modify guanine bases in DNA at the O^6 position. The adduct formed, O^6-[4-oxo-4-(3-pyridyl)butyl]guanine ([POB]dG), results in mutations that can lead to tumor formation, posing a significant cancer risk to humans exposed to tobacco smoke. A combined NMR–molecular mechanics computational approach was used to determine the solution structure of the [POB]dG adduct within an 11mer duplex sequence d(CCATAT-[POB]G-GCCC)·d(GGGCCATATGG) (Fig. 1). Gaussian 98 was used to determine quantum mechanically optimized geometry and charges for the G-POB adduct, and DUPLEX was used in the combined energy/NMR restrained optimization. In agreement with the NMR results, the POB ligand is located in the major groove, centered between the flanking 5'-side dT·dA and 3'-side dG·dC base pairs, and thus in the plane of the modified [POB]dG·dC base pair, which is displaced slightly into the minor groove. The modified base pair in the structure adopts wobble base pairing {hydrogen bonds between [POB]dG(N_1) and dC(NH_4) amino proton and between [POB]dG(NH_2) amino proton and dC(N_3)}. A hydrogen bond appears to occur between POB carbonyl oxygen and the partner dC's second amino proton. The modified guanine purine base, partner cytosine pyrimidine base, and POB pyridyl ring form a "triplex" via this unusual hydrogen-bonding pattern. The phosphodiester backbone twists at the lesion site, accounting for the unusual phosphorus chemical shift differences relative to those for the control DNA duplex. The helical distortions and wobble base pairing induced by the covalent binding of POB to the O^6-position of dG help explain the significant decrease of 17.6°C in melting temperature of the modified duplex relative to the unmodified control. This work is to be published in the December 2, 2003, issue of *Biochemistry*.[3]

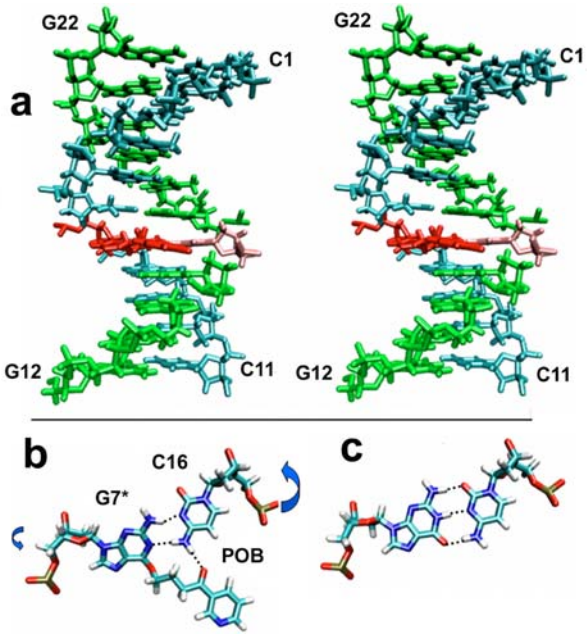

Fig.1. (a) Stereo view looking in the major groove of the lowest energy structure with the modified strand (C1-C11) in blue, complementary strand (G12-G22) in green, [POB]dG7 in red, and dC16 in pink. (b) The bases of the modified guanine, partner cytosine, and POB pyridyl ring form a "triplex" through an unusual hydrogen-bonding pattern (dashed lines). In order to accommodate the larger aromatic surface area while still retaining the normal overall dimensions of the DNA helix, the phosphate backbone at the dC16-dA17 step and, to a lesser degree, the [POB]dG7-G8, twist (arrows), in agreement with the phosphorus chemical shift differences (Fig. 5d). (c) The dG7·dC16 base pair in the control DNA duplex is shown for comparison.

(2) The illustrated six **DNA hybridization structures** were calculated using the program DUPLEX to compare the effects of triplet base repeats (Fig. 2) and the effects of a single base substitution (Fig. 3) on DNA structure and energy. The results indicate that once such slipped structures form, very little additional energy is required to produce either triplet repeat extensions or mutations and deletions.

(3) **Protamine** is a small (~50 amino acids), highly basic (pI ~ 12), arginine-rich (~50%) protein that binds to and condenses DNA in the sperm of most vertebrates. It inactivates the sperm genome by condensing it into toroids until it can be uncondensed and reactivated following fertilization in the egg. Previous studies have shown that the basic arginine residues in protamine bind in the major groove of the DNA helix and interact predominately with the acidic phosphodiester backbone of both strands of DNA in a sequence-independent manner to form a neutral, insoluble complex. In order to gain insight into how protamine interacts with DNA at the molecular level, we

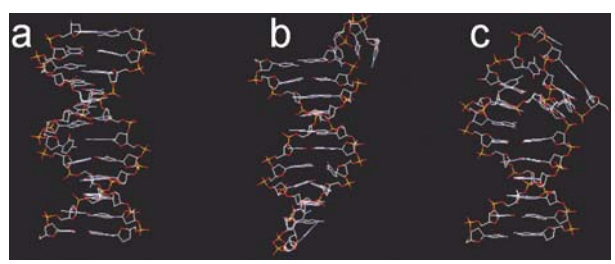

Fig. 2. Triplet repeats are seen in such diseases as fragile X syndrone, mytonic distropy, and Huntington's disease: (a) control duplex, d(CTGCTGCTGCTG)·d(CAGCAGCAGCAG), energy = –495.01; (b) 3′ three-base single strand overhang on each strand, energy = –465.42; and (c) a duplex containing a 3′ three-base single strand overhang on one strand and a central three-base loop out region in the other strand, energy = –456.84.

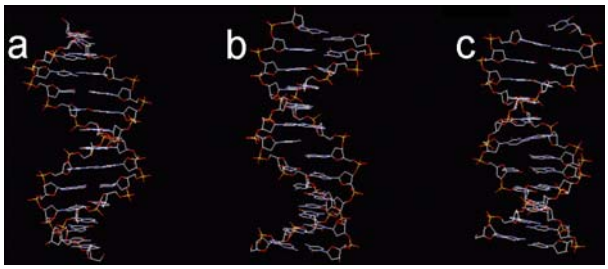

Fig. 3. Single-base substitutions effect DNA structure as shown above: (a) control duplex, d(GGGGGGGGGGGG)·d(CCCCCCCCCCCC), energy = –542.99; (b) a central substitution of a guanine with adenine in one strand, energy = –450.61; and (c) deletion of a single cytosine opposite one guanine, with a 3′-C overhang, energy = –448.09.

have prepared soluble complexes composed of short (14–16 amino acids) protamine segments and DNA hairpins to study by a combined NMR/computation method. The complex remains soluble because the protamine binds and neutralizes the duplex stem of the DNA hairpin, while the hairpin loop retains its negative charges, thereby keeping the complex charged as well.

The NMR studies have been completed for one complex (Fig. 4), and work on two more peptide/DNA hairpin complexes, as well as the full-length protamine/DNA hairpin complex, is under way. The addition of peptide parameters to DUPLEX and refinement of the DNA hairpin structure calculated using DUPLEX is also work in progress. We will also use the package CYANA, which has been optimized for proteins with DUPLEX in these calculations.

Summary and Conclusions

The [POB]dG adduct posed several challenges in structural determination. There are numerous degrees of freedom of rotation due to a longer butyl chain and pyridyl ring, which can interact with the DNA nonspecifically, resulting in a heterogeneous mixture consisting of a number of conformers. Yet, despite these expected challenges, it was possible to determine a single conformation for the [POB]dG adduct within an 11mer duplex sequence. A combined NMR-molecular mechanics computational approach was necessary to provide a detailed structural view of the types of distortions that can be induced into DNA by a bulky and flexible alkylating agent, such as POB, at the O^6 position of guanine (Fig. 1) because of the broad overlapped lines in the NMR spectra due to the motions. We had to work iteratively between the NMR analysis and the computational results in order to obtain the best structures with lowest energies and best fit to the data. This work resulted in one publication.[3]

The DUPLEX program has been given the ability to analyze hybridized double strands of DNA. DUPLEX is capable of analyzing the three-dimensional structures and energetics for such DNA hybridization studies. DUPLEX also has the ability to study hairpin loops such as those found in the stems of tRNA and in the DNA-protamine complex.

From the chemical shift perturbation studies of the G8HP+$R_3FGR_3SR_2VA_2R_2$ and the changes induced upon complex formation in the intensities of the proton, carbon, and phosphorus resonances we observe in the DNA, as well as proton and carbon in the peptide, we can say that the peptide binds to the 5-base pair DNA duplex terminal end, and the peptide is oriented so that its N-terminal end is located at the DNA duplex terminal end. In addition, we can say that the conformation of the peptide in the complex adopts an extended conformation and that a helical conformation, as has been postulated to be the case by others, is not observed. From NMR temperature studies, we have been able to form a hypothesis about why the binding is specific to the terminal end of the DNA duplex based on dynamic factors. We also show evidence that the PHE residue is intercalating between the two terminal DNA base pairs. However, the peptide and DNA are too short, resulting in a very weak binding constant (~200 µM) for the complex, and thus not allowing us to determine enough contacts between the peptide and the DNA to

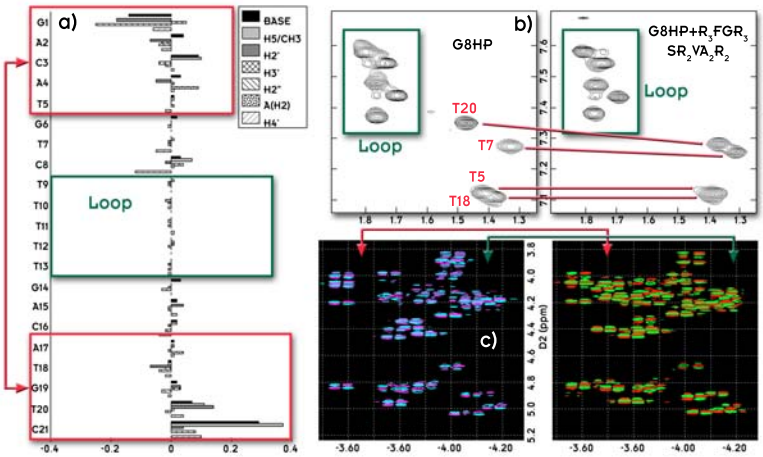

Fig. 4. (a) Summary of DNA chemical shift perturbations (control-complex) for the d(GACATGTCTTTTTGACATGTC) (dG8HP) +$R_3FGR_3SR_2VA_2R_2$ complex. (b) Expanded region of the 300-ms NOESY spectrum of the G8HP+$R_3FGR_3SR_2VA_2R_2$ complex showing the thymine methyl region. Note that the thymines in the hairpin loop do not shift, while T20 and T5 in the duplex stem in the complex shift relative to their positions in the control. (c) ^1H-^{31}P-heteronuclear correlation spectra of the control (left) and complex (right). The arrows indicate changes in the spectra that occur upon binding.

allow us to obtain a high-resolution structure. Thus computational calculations will be critical in helping to determine the structures of these peptide/DNA hairpin complexes. This work is still in progress as of this report.

Once we are successful with a solution of the DNA/protamine complex by combined NMR-computational methods, we feel confident that we will be successful in obtaining NIH or DOE funding, especially with the expansion of the Genomes to Life (GTL) initiative at DOE. Much of this is in relation to the future direction of biological research since the completion of the Human Genome sequence.

References

[1]S. Broyde and B. E. Hingerty, "Effective Computational Strategies for Determining Structures of Carcinogen-Damaged DNA," *J Comput Phys* **151**, 313–332 (1999).

[2]L. Wang, B. E. Hingerty, A. R. Srinivasan, W. K. Olson, and S. Broyde, "Accurate Representation of B-DNA Double Helical Structure with Implicit Solvent and Counterions," *Biophysical . J.* **83**, 382–406 (2002).

[3]L. A. Peterson, C. Vu, B. E. Hingerty, S. Broyde, and M. Cosman, "Solution Structure of an O^6-[4-oxo-4-(3-pyridyl)butyl]guanine Adduct in an 11mer DNA Duplex: Evidence for Formation of a Base 'Triplex'," *Biochemistry* (in press).

Development of a Generic Computational Method for Biological Data Clustering

D. Xu, V. Olman, and Y. Xu
Life Sciences Division

Data clustering is fundamental to mining large-scale biological data. We further developed a general clustering method based on the representation of multidimensional data as a minimum spanning tree. The presentation opens a new door for clustering, which converts a clustering problem to a problem of partitioning a one-dimensional profile (a sequential representation) derived from the Prim algorithm for constructing a minimum spanning tree. Such a method facilitates partitioning and identification of "dense" clusters from a noisy background. We have successfully applied our methods to a number of data mining problems for high-throughput biological data, including cluster identification in gene expression profiles, regulatory binding site identification, two-hybrid protein-protein interaction data, and signal transduction pathway construction.

Introduction

The mass production of biological data is unprecedented, especially with high-throughput protocols at the genome scale. Types of data include (1) sequence data generated by the Human Genome Project and other sequencing efforts, (2) structural data produced by the NIH Structural Genomics Initiative, (3) gene expression data from microarray experiments, (4) mass-spectroscoopy data for proteins, and (5) two-hybrid data.

In addition to the experimental data, huge amounts of computational data are also being produced through annotations and predictions. The rates of data generation far exceed our current capability of analysis and interpretation. Analyzing these data is highly challenging, since they are often noisy and incomplete. New ideas and approaches are urgently needed to establish greatly improved capabilities for biological data analysis. Data clustering is fundamental to mining a large quantity of biological data. Many problems in mining and analysis of high-throughput data are clustering or they can be formulated as clustering problems. The main objective of this project was to develop generic clustering methods to help biological discovery using the high-throughput biological data. The research tasks included two parts: (1) developing algorithms for a general-purpose clustering method that can be applied to various biological data and (2) testing the algorithms on actual biological clustering problems.

Technical Approach

Our methods are based on the concept of minimum spanning tree. A *spanning tree* is a connected graph that contains every vertex but does not contain any cycle. A *minimum spanning tree (MST)* is a spanning tree with the minimum total distance between vertices. An MST can be efficiently constructed by classical Kruskals algorithm. In essence, our MST-based approach tries to capture our intuition about a cluster so that distances between neighbors within a cluster should be smaller than any inter-cluster distances. We have shown that each cluster of the multidimensional data corresponds to one subtree of the MST. This property can generally convert a multi-dimensional clustering problem to a tree-partitioning problem. In addition, MST-based clustering has higher tolerance to noise in the data than other clustering methods. Due to these properties, the MST representation is very suitable for general clustering problems.

Based on our early development in minimum spanning tree–based clustering, we discovered a new approach to identify clusters from noisy background based on a rigorous relationship between data clusters and a sequential representation derived from the Prim algorithm for constructing a MST of a data set.[1] This approach converts a clustering problem to a problem of partitioning a one-dimensional profile (a sequential representation), which facilitates easy identification of "dense" clusters from a noisy background. A statistical model was developed to assess the confidence level of a cluster. We have applied the method to a number of data mining problems for high-throughput biological data, including cluster identification in gene expression profiles and regulatory binding site identification. The results have shown that our generic clustering methods can identify underlying patterns from large-scale biological data effectively and efficiently.

Results and Accomplishments

MST-Based Clustering Methods and Implementation in the EXCAVATOR Package

We implemented the sequential representation–based cluster identification method, as described above. We also explored different features, such as different distance measurement between gene expression profiles, different objective functions, etc. We incorporated all the new developments into the computer package, EXCAVATOR[2] (Fig. 1, http://compbio.ornl.gov/structure/excavator/), for clustering gene-expression profiles. This program has a number of unique features, including capabilities for (i) data-constrained clustering, (ii) identification of genes with similar expression profiles to prespecified seed genes, (iii) cluster identification from a noisy background, and (iv) computational comparison between different clustering results of the same data set. EXCAVATOR can be run from a Unix/Linux/DOS shell, from a Java interface, or from a Web server. The clustering results can be visualized in colored figures and two-dimensional plots. Moreover, EXCAVATOR provides a wide range of options for data formats, distance measurements, objective functions, clustering algorithms, methods to choose number of clusters, etc. The effectiveness of EXCAVATOR has been demonstrated on several experimental data sets. Its performance compares favorably against the popular K-means clustering method in terms of clustering quality and computing time.

Application of the MST-Based Clustering Methods in Regulatory Binding Site Identification

Transcriptional binding sites are short fragments in the upstream regions of genes, to which transcription factors bind to regulate the transcription of genes into mRNA. Computational identification of transcriptional binding sites remains a challenging and unsolved problem though a great amount of effort has been put into the study of this problem. We have recently applied the sequential representation-based cluster identification method for identification of binding sites from a set of upstream regions of genes[3,4] that could possibly be transcriptionally co-regulated and hence might share similar transcriptional binding sites. By utilizing two key features of such binding sites (i.e., sequence similarities among them and their relatively high frequencies compared to other sequence fragments), we have implemented the cluster identification algorithm and a new statistical significance analysis method as a computer package CUBIC for regulatory binding site identifications. Extensive testing on CUBIC has been carried out.

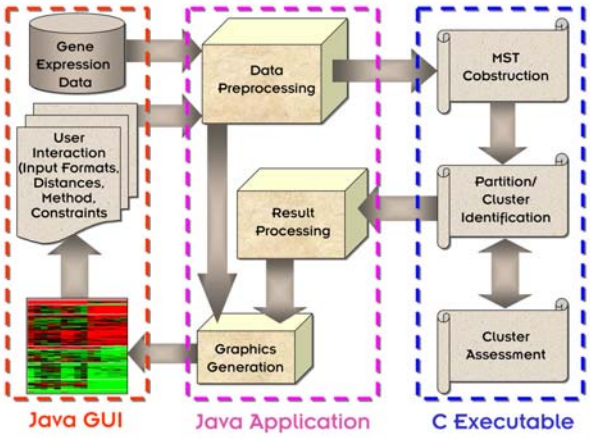

Fig. 1. The flowchart for the design of EXCAVATOR.

Application of the MST-Based Clustering Methods for Pathway Prediction

To further validate our generic clustering methods, we applied our methods in studying a particular biological system, that is, the signal transduction pathway in the amino acid and peptide transport in yeast *S. cerevisiae*. In collaborating with experimentalist Dr. J. Becker at the University of Tennessee, together with Yu Chen, a graduate student from the University of Tennessee doing his thesis under the supervision of Dong Xu, we constructed a global pathway model[5] using various information including data from DIP (http://dip.doe-mbi.ucla.edu), BIND (http://binddb.org), and gene expression data. Regulatory region analysis (clustering of binding site motifs) and gene expression clustering analysis validate the pathway model by showing how the selected transcription factors control the amino acid transporters and how the cross-talks between the amino acid transport pathway and the other related pathways are achieved (Fig. 2). We also extend this work to general signal transduction pathway construction using high-throughput biological data.[6,7] A Web server (http://compbio.ornl.gov/structure/expression/pathway) has been constructed for public use. This work demonstrates the power of our analysis methods for biological discovery from high-throughput biological data.

Summary and Conclusions

In this project, we developed MST-based clustering methods and successfully applied our methods to a number of data mining problems for high-throughput biological data, including cluster identification in gene expression profiles, regulatory binding site identification, and signal transduction pathway construction. The results of this work were incorporated into our software packages, that is, EXCAVATOR for clustering gene expression profiles, CUBIC for regulatory binding site identification, and a pathway prediction Web server. We also applied our methods in constructing the amino acid transport pathway model in yeast.

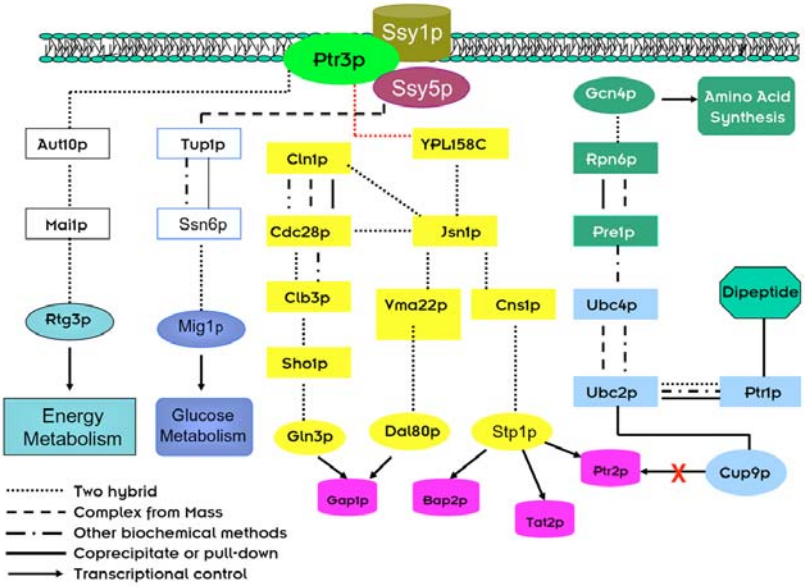

Fig. 2. Our constructed model for the signal transduction pathway of amino acid and peptide transporters in yeast, where the different shades of gray scale indicate different pathways. The different line patterns represent protein-protein interactions as identified by various experimental techniques. The ovals represent transcription factors, and the boxes represent intermediate proteins between the SPS sensor and transcription factors.

The above work resulted in seven publications listed as references. Based on our results, we have submitted an NSF proposal (Ying Xu, Victor Olman, and Dong Xu, entitled "Biological Data Clustering Using Minimum Spanning Trees"), which has been funded. This work also provided some foundation for other DOE-funded research projects, in particular, the U.S. Department of Energy's Genomes to Life program (www.doegenomestolife.org) under project "Carbon Sequestration in *Synechococcus Sp*.: From Molecular Machines to Hierarchical Modeling" (www.genomes-to-life.org).

References

[1]V. Olman, D. Xu, and Y. Xu, "Solving Data Clustering Problem as a String Search Problem," in *Proceedings of Conference on Statistical Data Mining and Knowledge Discovery*, Hamparsum Bozdogan, ed., Chapman & Hall/CRC, in press.

[2]D. Xu, V. Olman, L. Wang, and Y. Xu, "EXCAVATOR: a computer program for gene expression data analysis," *Nucleic Acid Research* **31,** 5582–5589 (2003).

[3]V. Olman, D. Xu, and Y. Xu, "Identifications of Regulatory Binding Sites Using Minimum Spanning Trees," pp. 327–338 in *Proceedings of the 2003 Pacific Symposium on Biocomputing (PSB)*, 2003.

[4]V. Olman, D. Xu, and Y. Xu, "CUBIC: Identifications of Regulatory Binding Sites through Data Clustering," *Journal of Bioinformatics and Computational Biology* **1**, 21–40 (2003).

[5]Y. Chen, Y. Liu, K. M. Goldstein, J. M. Becker, Y. Xu, and D. Xu, "A Computational Study on the Signal Transduction Pathway for Amino Acid and Peptide Transport in Yeast: Bridging the Gap between High-Throughput Data and Traditional Biology," *Applied Genomics and Proteomics* **2**, 43–50 (2003).

[6]Y. Chen and D. Xu, "Computational Analyses of High-Throughput Protein-Protein Interaction Data," *Current Protein and Peptide Science* **4**, 159–181 (2003).

[7]Y. Chen, T. Joshi, Y. Xu, and D. Xu, "Towards Automated Derivation of Biological Pathways Using High-Throughput Biological Data," pp. 18–25 in Proceeding of the 3rd IEEE Symposium on Bioinformatics and Bioengineering (2003), IEEE/CS Press.

Nanostructures for Spatially Resolved Macromolecular Delivery to Viable Whole Cells

T. E. McKnight,[1,2] A. V. Melechko,[3] G. D. Griffin,[4] V. I. Merkulov,[1] M. A. Guillorn,[1,4] M. L. Simpson[1,4]
[1]*Engineering Science and Technology Division*
[2]*Condensed Matter Physics Division*
[3]*University of Tennessee, Center for Environmental Technology*
[4]*Life Sciences Division*

The goal of the project was to evaluate the potential for parallel delivery of DNA to groups of cells using arrays of vertically aligned carbon nanofibers (CNF). CNF arrays are comprised of individual, solid nanofiber elements, with tip diameters of tens of nanometers and lengths of several microns. The needlelike aspect of these fibers proves to be well suited for penetration and delivery of large molecules to the intracellular domain of targeted cells. Specifically, plasmid DNA is spotted and adsorbed onto localized subregions of nanofiber arrays. These arrays are then integrated with cellular matrices using a variety of methods including centrifugation of cells onto fibered substrates and even direct pressing of fibered substrates into monolayers of adherent cells. The subsequent desorption of DNA from the fibers into the intracellular, and potentially nuclear, domains of cells provides a highly efficient material delivery method that is easy to implement with a variety of cell lines and types.

Introduction

Among the many anticipated applications of CNF devices, their potential as cellular, subcellular, and even molecular-scale probes and actuators is perhaps the most intriguing. The size scale of nanofibers (20+ nm radius and micron lengths), their carbon-based nature, and the ability to synthesize them in vertically aligned arrays (Fig. 1) inspire their integration with cellular and tissue matrices to provide a massively parallel electrical, electrochemical, or biochemical interface for interrogation and control of cellular processes. While parallel interfacing has been reported with extracellular devices, nanofiber arrays provide an interface with the potential for massively parallel intracellular integration, where the fibers are directly inserted into live cells and used by the researcher to interrogate or manipulate the intracellular domains of these integrated cells.

In this project, we explored one approach to such a parallel cellular interface—the insertion, retention, and assimilation of nanofiber arrays in live cell matrices. We demonstrated this integration by using the fibers to introduce a new gene to the cellular targets, where gene expression (green fluorescent protein—GFP production) not only denotes successful fiber insertion but also continued viability of the interfaced cell. To further demonstrate this integration, we covalently bound the gene vector to the fibers and observed long-term expression via transcription off tethered plasmid within interfaced cells, thus verifying the continued presence of the nanofiber element within the intracellular domain and the continued biochemical manipulation of the interfaced cell.

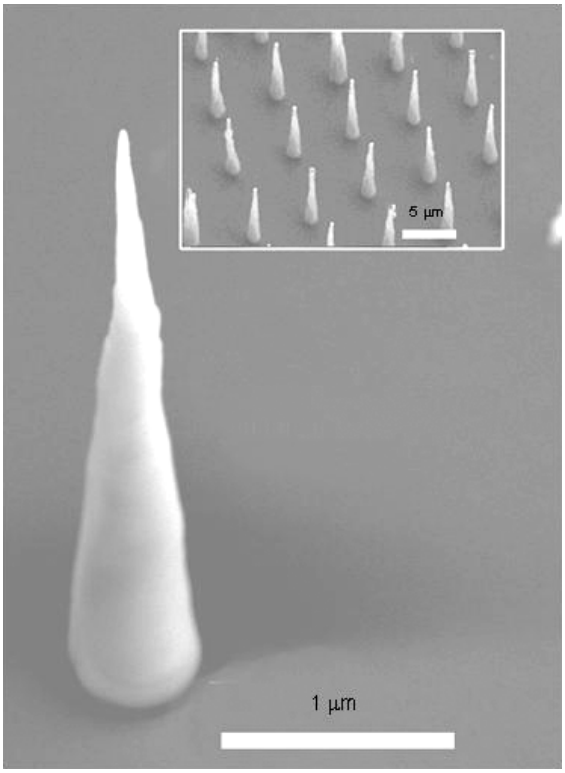

Fig. 1. Carbon nanofiber array photolithographically defined and synthesized in a plasma-enhanced chemical vapor deposition (PECVD) process. Fiber height, radius, location, and surface composition (carbon/nitrogen ratio) may all be defined based upon the initial catalyst thin-film layer and PECVD parameters, such as gas composition, temperature, and growth time. Here, a grid array of 5-µm-spaced, 10-µm-tall fibers with tip diameters of 30 nm was synthesized over a 50- × 50-mm region of a 100-mm silicon wafer.

Technical Approach

An enabling factor for the potential application of CNF as a cellular interface is the ability to synthesize nanofibers in a highly deterministic manner, at specific physical locations.[1] This ability provides the integration of nanoscale, and ultimately *molecular-scale*, elements (nanofiber tips) with mesoscale devices (i.e., silicon substrates), and, in turn, with the macroscale physical world of the researcher. Using a plasma-enhanced chemical vapor deposition (PECVD) process, carbon nanofibers may be synthesized at specific locations upon a substrate, as defined by the lithographic patterning of catalyst metals, such as Ni, Co, and Fe. In catalyst-"tip-based"[2] growth, the size and shape of a nanofiber may be well defined[3] and large arrays of vertically aligned nanofibers may be synthesized using large-area (100-mm wafer) processing techniques. In this effort, we synthesized nanofiber arrays from photolithographically defined nickel catalyst dots on 100-mm, n-type silicon wafers. Following synthesis, fibered substrates were cleaved into smaller pieces to enable convenient handling.

An overview of our methodology for integrating cells with fibers is shown in Fig. 2. In brief, cells were centrifugally pelleted out of suspension onto 3-mm × 3-mm fibered substrates oriented normal to the pelleting force. As such, cells were spun down and impaled upon the vertically aligned fibers grown on the substrate—similar to throwing a ball against a sparsely populated bed of nails [Fig. 2(a)]. Optionally, following the spin, the fibered substrate and the mass of cells upon it were gently pressed against a flat, wetted surface, such that the cells were further integrated with the fiber matrix [Fig. 2(b)]. The fibered substrate was then placed face up in growth media in a culture dish and incubated to allow cell recovery and proliferation [Fig. 2(c)]. The success of this method to provide for macromolecular delivery to intact cells was dependent upon adsorption and retention of plasmid vectors on the fiber surfaces and cell survival to the manipulations.

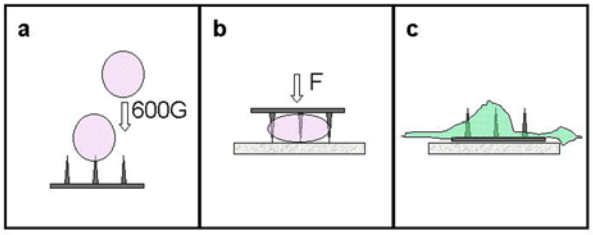

Fig. 2. Schematic of process for integration of adherent cells to a nanofiber array. (a) Cells are spun out of a suspension in phosphate buffered saline down onto a fibered substrate. (b) Following the spin down, the substrate may optionally be pressed against a wetted, flat surface. This further integrates the cells with the fibers, potentially pressing fibers deeper into the cell. (c) Following these manipulations, the fibered substrate and the cells upon it are placed face up in a culture dish, covered with growth media, and incubated for at least 24 h prior to fluorescent imaging.

Adsorption and Retention of Plasmid Vectors

Successful delivery of plasmid vectors to targeted cells using this fiber-mediated method requires that plasmid vectors be retained by fibers during interactions with the cell but remain available for transcriptional activity following the integration. We investigated a variety of coating methods to provide adsorption of DNA onto fibers, retention during cell interaction, and desorption following the integration procedure. We initially experimented with suspensions of plasmid in dilute solutions of agarose (0.1%) but also evaluated a variety of surface coatings on the fiber (sputtered-gold, plasma-enhanced chemical-vapor-deposited SiO_2), thermally reactive coatings (poly-*n*-isopropyl acrylamide, or pNIPAM), and DNA complexing agents (spermidine and $CaCl_2$ precipitation). Plasmids and plasmid cocktails were spotted directly onto fibered substrates with a micropipettor. Plasmid retention on fibers was evaluated qualitatively by staining with propidium iodide and observing the degradation of fluorescence of fibered regions over time when immersed in aqueous buffers. In general, the agarose method provides a very diverse coating methodology, relatively independent of the surface charge effects that often limit DNA/surface interactions. Dried spots of DNA in 0.1% agarose provide fairly uniform coating of fibered surfaces, only slightly diminishing the high-aspect feature of the nanofibers. The agarose also helps to provide a slow release of the DNA from the fiber surface. Unfortunately, the presence of agarose on fibers during interaction and penetration of fibers into cells appears to considerably stress cells—resulting in poor subsequent growth and markedly different morphology. It is unclear if this is a cytotoxic effect of the agarose, or merely an artifact of the ability of the cell to reseal following interaction with an agarose-coated fiber. Surface treatments of fibers can also appreciably affect DNA retention. In general, the native carbon nanofiber is a negatively charged surface. The negatively charged CNF surface will tend to electrostatically repel the anionic phosphate backbone of non-complexed DNA in dilute solutions. Dilute solutions of DNA will coat fibers upon evaporation but will quickly be shed upon rehydration. This is unacceptable for material delivery, as plasmid vectors will be lost to solution upon wetting with the cellular suspension. However, this repulsive effect can be decreased by complexing the DNA with cations (i.e., Ca^{++}), or by modifying the surface charge of the fiber with coatings—such as a sputtered gold or tungsten.

DNA may also be complexed to fiber surfaces using a variety of techniques that have been developed in the microprojectile bombardment community, where plasmid-

coated microprojectiles are physically accelerated onto plates of recipient cells. In these methods, DNA is typically complexed to the particle using a cationic lipofective reagent (i.e., spermidine) and a highly concentrated precipitant ($CaCl_2$). Of all the methods employed, the agarose and spermidine/$CaCl_2$ treatments resulted in the best retention of plasmid DNA on native CNF. However, each method merits further investigation, particularly the controlled release methodologies.

Finally, in addition to post-synthesis modification of CNF, the fibers themselves may be synthesized with higher nitrogen content by simply altering the ratio of carbonaceous gas and the nitrogenated etch gas during PECVD synthesis. Highly nitrogenated fibers may be synthesized that can bear an overall positive charge in plasmid solutions, and thus electrostatically retain plasmid DNA without requiring any additional processing steps.

Another feature of nanofiber tips is a short segment below the catalyst of uncontaminated carbon, typically approximately 100 nm long. Elemental (electron dispersion X-ray) analysis reveals that even with a highly nitrogenated source gas mixture which will produce heavily nitrogen-doped fibers, the extreme tip of the fiber will be virtually nitrogen free. Modification of this region can thus provide another effective handle for DNA retention. One example is the modification of these carbon-rich regions with an oxygen plasma, thus producing oxide functionalities within the carbon tip, including carboxylic acids. With these handles, we have demonstrated that DNA may be covalently bound to the fibers using an EDC-mediated [1-ethyl-3-(3-dimethylaminopropyl) carbodi-imide] condensation reaction between the fiber carboxylic acids and the DNA base amines.

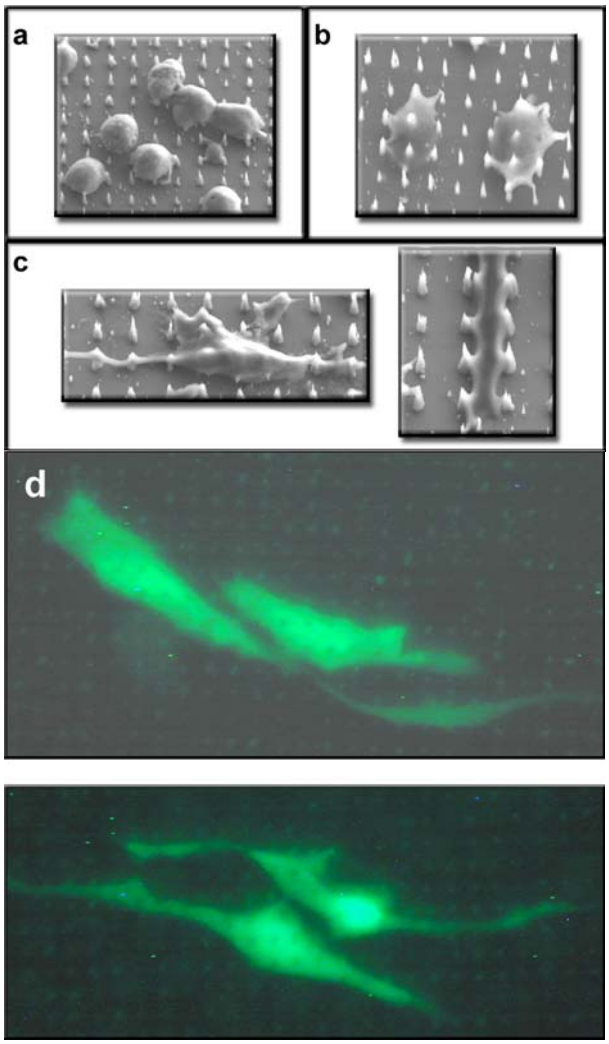

Fig. 3. Scanning electron micrographs of cells following (a) pelleting, (b) press, and (c) culture for 48 h. (d) Production of GFP from the delivered plasmid can be observed as early as 4 h following the integration. Here, GFP+ cells are observed with fluorescent microscopy at 48 h. Large colonies may be observed after several weeks of culture, apparently due to stable genomic insertion of the delivered plasmid.

Plasmid Delivery and Cell Survival

Delivery and expression of GFP coding plasmids was highly dependent on a number of factors including fiber geometry (length and diameter), spacing, and chemistry. In general, untreated carbon fibers were not successful with plasmid delivery. However, nitrogen-containing fibers, exposed catalyst fibers, and plasmid-derivitized (covalent) fibers all provided some amount of GFP+ expression in targeted cells (Fig. 3). The failure of untreated carbon fibers appears to be due to poor retention of DNA on the fiber surface.

Centrifugation of cells at 3000 rpm (150 G) for 1 min onto plasmid-coated and plasmid-derivitized nitrogen-bearing fibers resulted in transient expression of green fluorescent protein at efficiencies of approximately 1% (where efficiency is defined as GFP+ cells vs total number of cells on the fibered substrate). These efficiencies could be improved by approximately a factor of 5 if, following the spin, the cell/fiber matrix is pressed lightly against a flat, wetted surface such as a microscope slide or the bottom of a tissue culture dish. Efficiencies with fibers that had been spotted with DNA after emancipating the nickel catalyst at the tip resulted in even higher transfection efficiencies using this spin and press method, with typically 20–50% of cells on the substrate expressing GFP during subsequent evaluations 24+ hours following the transfection experiment. In each of these scenarios, transfection efficiency is based on GFP+ cells vs total cells on the fibered substrate, and not GFP+ cells vs total number of cells spun down during the experiment.

Covalent attachment of plasmid DNA to fibers and integration with cell matrices also results in plasmid expression within integrated cells. In reaction volumes of 1 µg/mL of plasmid, transfection efficiencies were

approximately 1% using the method described in the paragraph above. These efficiencies could be dramatically improved (10–30% efficiency) by reducing the reaction volume (increasing plasmid concentration) during EDC condensation of DNA to fiber scaffolding.

Expression of GFP following cellular integration with nanofiber arrays has been followed for at least 30 days, indicating long-term survival of interfaced cells.

Results and Accomplishments

1. Demonstration of intracellular integration of arrays of CNF with cellular matrices and long-term biochemical modification of the integrated cells (expression of GFP from delivered plasmid vector).
2. Demonstration of a relatively high efficiency (up to 50%) macromolecular delivery system that is easy to implement with a variety of mammalian cell lines.
3. Development of nanofiber synthesis methods that result in nitrogen-doped fibers that are well suited for retention of anionic macromolecules (DNA) during cellular interactions and subsequent desorption of these molecules within the intracellular domain.
4. Development of a new CNF growth technique capable of producing arrays of single CNFs using only photolithographic definition of catalyst patterns. This eliminates the previous need for electron-beam lithography—thus providing a scalable nanofiber growth method, no longer limited by serial e-beam lithographic processes.
5. Demonstration of covalent linkage strategies for tethering macromolecules (plasmid DNA) to discrete regions of carbon nanofibers.
6. Demonstration of transcription from covalently bound plasmid tethered to nanofiber arrays.

Summary and Conclusions

This project has successfully demonstrated that CNF arrays may be integrated with cellular matrices to provide a massively parallel biochemical interface to live cells. This interface may be utilized as a macromolecular delivery vehicle, as demonstrated via the delivery and expression of plasmid DNA or, by leveraging previous work by our research group,[4] as an electrical/electrochemical interface. In addition to providing an adsorption/desorption–based delivery method, nanofibers may be synthesized or modified to feature rich surface chemistries that may be used for covalent attachment of macromolecules. We have demonstrated that covalently bound plasmid on fibers may be introduced to the intracellular domain—and transcribed while remaining tethered to the fiber scaffold. Transcription from tethered plasmids provides the ability to exploit cells as transcriptional factories, while minimizing the potential for escape of plasmid or genetically modified organisms to the environment. Finally, as physical (i.e., microinjection-based) approaches, these techniques should be applicable over a diverse array of cell types, potentially extending even into the plant and microbial kingdoms.

The following proposals have been generated as a direct result of this seed money effort:

NIH EB-03-011, *Synthetic Nanostructure Gene Delivery Vectors*, April 2003. $250K requested.

NIH PAR-03-045, *Nanoscale Architectures for Intracellular Biochemical Manipulation*, August 18, 2003. $1.8M requested.

NIH PA-03-058, *Nanostructured Gene Delivery Arrays,* October 1, 2003. $250K requested.

Joint USDA/DOE Biomass Program, *Enhanced Biomass Production: Novel Transformation Approaches to Engineering the Cellulose Content of Loblolly Pine,* collaborative proposal between ORNL, the Institute of Paper Science and Technology, North Carolina State University, and Clark Atlanta University, April 2003. $1.5M requested ($350K ORNL).

References

[1]V. I. Merkulov, D. H. Lowndes, Y. Y. Wei, G. Eres, and E. Voelkl, *Appl. Phys.Lett.* **76**, 3555 (2000).

[2]V. I. Merkulov, A. V. Melechko, M. A. Guillorn, M. L. Simpson, and D. H. Lowndes, "Alignment Mechanism of Carbon Nanofibers by Plasma-Enhanced Chemical Vapor Deposition," *Applied Physics Letters* **79**(18), 2970–2972 (October 2001).

[3]V. I. Merkulov, M. A. Guillorn, M. L. Simpson, D. H. Lowndes, and E. Voekl, "Shaping Carbon Nanostructures by Controlling the Synthesis Process," *Applied Physics Letters* **29**(8), 1178–1180 (Aug. 20, 2001).

[4]M. A. Guillorn, T. E. McKnight, A. V. Melechko, D. W. Austin, V. I. Merkulov, M. L. Simpson, and D. H. Lowndes, "Individually Addressable Vertically Aligned Carbon Nanofiber-Based Electrochemical Probes," *J. Appl. Phys.* **91**, 3824 (2002).

Prediction of Sepsis Onset in Trauma Victims Using Advanced Nonlinear Analysis of Time-Serial Physiological Data Including Heart Rate Variability

L. M. Hively,[1,2] V. A. Protopopescu,[3] and N. B. Munro[4]
Collaborators: M. D. Karlstad,[2] W. M. Vassy, M.D.,[2] B. Enderson, M.D.,[2]
B. J. Daley, M.D.,[2] and W. Holland, R. EEG. T[5]
[1]*Computational Sciences and Engineering Division*
[2]*Department of Surgery, UTMCK*
[3]*Computational Science and Mathematics Division*
[4]*Life Sciences Division*
[5]*Department of Neurology, UTMCK*

The primary cause of death in trauma patients is sepsis and septic shock. This situation has not improved significantly over the past 30 years for several reasons. First, reliable predictors for sepsis do not exist to allow early interventions. Second, the patient's condition often deteriorates rapidly after sepsis onset, even with aggressive medical interventions, frequently causing the failure of multiple organ systems and eventually leading to death. Also, recent medical improvements have been offset by the dramatic increase in antibiotic-resistant infections, which are hard to diagnose and harder still to fight. This work is the first step towards sepsis forewarning by the acquisition and analysis of time-serial physiological data. Physiological (ECG) data from endotoxin-exposed rats were acquired by collaborators at the University of Tennessee Medical Center in Knoxville (UTMCK). Oak Ridge National Laboratory (ORNL) analyzed the data via nonlinear dynamical methods, which previously have provided up to an hour's forewarning of epileptic seizures from scalp brain waves. The results show a clear superiority of the phase-space dissimilarity measures in comparison to traditional nonlinear measures, in terms of robustness, timeliness, and discriminating power.

Introduction

Despite considerable advances in treatment, sepsis remains a common source of morbidity and mortality in surgical patients. Early treatment of sepsis is paramount for minimizing complications and preventing death. Unfortunately, timely *identification* of sepsis development prior to clinically detectable onset remains an outstanding challenge. Complex systems analysis of physiological data is a new and powerful computational method to discover interactions and emergent properties of physiological systems. These methods provide a means to study systemic properties of disease and injury in critical care patients for evaluation and differentiation of their complex nonlinear systems. Such complex nonlinear systems are ubiquitous; they include weather patterns, the stock market, biological systems including the central nervous system, and the complex host response to trauma, sepsis, or shock. For example, a change in heart rate (HR) variability, measured as the change in interbeat intervals, is thought to be due to changes in the sympathetic and parasympathetic control of the heart. Standard complex systems analysis methods have shown that reduction in HR variability is associated with increased mortality rate after myocardial infarction,[1] sepsis, and critical illness in pediatric patients.[2,3] Blood pressure variability is also reduced in septic shock.[3] Griffin and Moorman[4] showed reduced baseline HR variability and short-lived decelerations in HR up to 24 h before clinically observable deterioration in a prospective study of infants at risk of sepsis. Heart rate variability has been shown to increase during recovery from sepsis in pediatric patients.[1,2] Advanced nonlinear analysis of electrocardiogram (ECG) data has been shown to forewarn of cardiac fibrillations[5,6] but has not been applied to sepsis detection or prediction. Improved interpretations of physiological data have recently been reenergized by advances in nonlinear and chaotic dynamics[7–10] (and references therein). Here we show that advanced nonlinear analysis of ECG data can be used for pre-symptomatic detection of sepsis.

Technical Approach

UTMCK initially acquired four-channel ECG patient data at a sampling rate of 500 points per second. ORNL's analysis revealed 1- to 2-h gaps during off-line procedures, after which the nonlinear signatures were dramatically different. Consequently, the team agreed that human data were inappropriate for this study, due to the confounding effects of patient treatments.

Subsequently, UTMCK acquired the same data from 23 male Long-Evans rats, which were anesthetized via 2% inhaled isoflurane. The experimental protocol began with a 30- to 60-min acclimation period in a nose-only inhalation chamber, followed by exposure of 17 rats to aerosolized *Salmonella enteriditis* endotoxin for 30 min. Subsequent monitoring of the physiological response continued for 1 to 2 h. The bacterial endotoxin induces an inflammatory response, which should be detectable as changes in the ECG data under our hypothesis. Six other rats were exposed only to deionized water for 30 min, as controls. Table 1 characterizes the 23 datasets. The immune response was verified subsequently by lung lavage and counting of neutrophils in the lavage fluid. Further verification was accomplished by histopathological examination of lung sections after removal at euthanization.

Data were analyzed by the ORNL-developed phase space dissimilarity measures (PSDM) approach.[5,11] The method first removes confounding artifacts (breathing artifacts in ECG data) by a zero-phase, quadratic filter. The artifact-filtered data then are converted to a discrete distribution function (DF) that captures the essential features of the underlying dynamics. Condition change assessment compares test case DFs to a base case (normal state) DF via novel measures of dissimilarity. Forewarning is indicated by several sequential occurrences of the dissimilarity measures above a threshold. While traditional nonlinear measures compare averaged (global) quantities, the enhanced discrimination power of our measures is achieved by focusing on the absolute difference between the two DFs. This last feature enables biomedical event forewarning from (near) real-time, noisy biomedical data of numerous types.

Table 1. Characterization summary

Date	Exposure	ID	T0	T1	T2	T3	SZ(MB)
08/15	Endotoxin	9V	15:23:20	15:53:14	16:23:54	17:24:37	131325
08/20	Endotoxin	9W	13:35:35	14:35:55	15:05:55	17:05:37	227619
08/22	Endotoxin	9X	12:46:34	13:46:31	14:16:29	16:16:47	227818
08/23	Endotoxin	9Y	12:39:38	13:39:33	14:09:38	16:09:41	227637
08/29	Endotoxin	9Z	13:59:28	14:59:35	15:29:32	17:29:27	227565
09/10	Endotoxin	A1	12:09:22	13:09:17	13:39:17	15:39:21	227565
09/12	Endotoxin	23	12:55:38	13:55:31	14:25:31	16:25:34	221531
09/13	Endotoxin	A4	13:26:49	14:26:46	14:56:46	16:56:50	227601
10/08	Endotoxin	A5	12:57:13	13:57:18	14:27:08	16:27:11	227547
10/09	Endotoxin	~1	11:17:26	12:17:20	12:47:39	14:31:13	227565
10/10	Di H_2O	~2	12:31:15	13:31:30	14:01:07	16:01:25	227764
10/15	Di H_2O	A9	15:04:07	15:44:01	16:14:08	18:14:43	206554
10/17	Di H_2O	AA	11:39:41	12:39:36	13:09:36	15:09:40	227565
10/23	Di H_2O	BD	11:34:31	12:34:34	13:04:28	15:04:29	191108
10/24	Di H_2O	AE	12:29:39	13:29:34	13:59:34	15:59:37	227547
10/28	Endotoxin	AF	11:21:31	12:21:28	12:51:25	14:51:29	227547
11/05	Di H_2O	AG	14:29:12	15:29:12	15:59:12	17:59:15	227529
11/11	Endotoxin	AH	13:10:21	14:11:09	14:40:28	16:40:20	227565
11/13	Endotoxin	AI	11:29:35	12:29:30	12:59:30	14:59:37	228071
11/14	Endotoxin	AJ	12:46:12	13:46:06	14:16:07	16:16:10	227547
11/19	Endotoxin	AK	12:33:33	13:33:28	14:03:30	16:03:43	227764
11/20	Endotoxin	AL	12:19:27	13:20:37	13:52:41	15:49:25	227547
11/21	Endotoxin	AM	11:37:52	12:37:47	13:07:47	15:07:50	227457

DATE = 2002 date of this experiment in MM/DD format.
EXPOSURE = either endotoxin or de-ionized water.
ID = dataset identifier.
T0 = starting time of the experiment in HH:MM:SS format.
T1 = beginning of the exposure period.
T2 = end of the exposure period.
T3 = finish time of the experiment.
SZ = size of the dataset in megabytes (MB).

Results and Accomplishments

The 17 endotoxin-exposed animals showed significant post-exposure condition changes in ECG dynamics, as compared with 6 control animals. Figure 1 shows results for one endotoxin exposure case. The raw ECG data in subplot (a) have 14 distinct heartbeats with additional quasi-periodic (nonlinear) features. No indication of condition change is displayed by either the correlation dimension in subplot (b), or by the Kolmogorov entropy in subplot (c). Condition change is shown clearly by both $U(\chi^2)$ in subplot (d) and $U(L)$ in subplot (e), which remain low, then rise abruptly after the exposure onset, remain high for the next 20 minutes, then decrease slowly, presumably as the immune response counteracts the bioagent effects. Table 2 summarizes the analysis results for all 23 animals.

The phase-space dissimilarity measures provide indication in at least one of the four ECG channels for all 17 of the endotoxin-exposure cases. The recovery response is consistent with other physiological observations during the test. Pulmonary inflammation in endotoxin-exposed animals was verified by increased neutrophils in bronchoalveolar lavage (13.2 ± 4.2% versus 3.6 ± 1.8%, $p < 0.02$). The phase-space dissimilarity measures give no indication in at least one of the ECG channels for all six of the non-exposure cases. The earliest detection of condition change occurs within 1 h of the exposure onset in 16 datasets. Dataset #19 (AI) shows condition change after 91 min. The criterion for condition change requires several sequential occurrences, N_{OCC}, above a threshold value, U_{TH}, for a specified number, N_{SIM}, of renormalized PSDM in any one of the four ECG channels. This criterion is a very weak one[11] because no one channel provides adequate indication. Moreover, Table 2 shows instances of correct indication in one channel that is accompanied by contrary indications in other channel(s), including later indication time(s), as well as false positive (FP) and/or false negative (FN) indications. While this initial success is encouraging, much additional work is necessary to provide channel consistency, as well as robustness and reliability that is needed for clinically useful diagnostics.

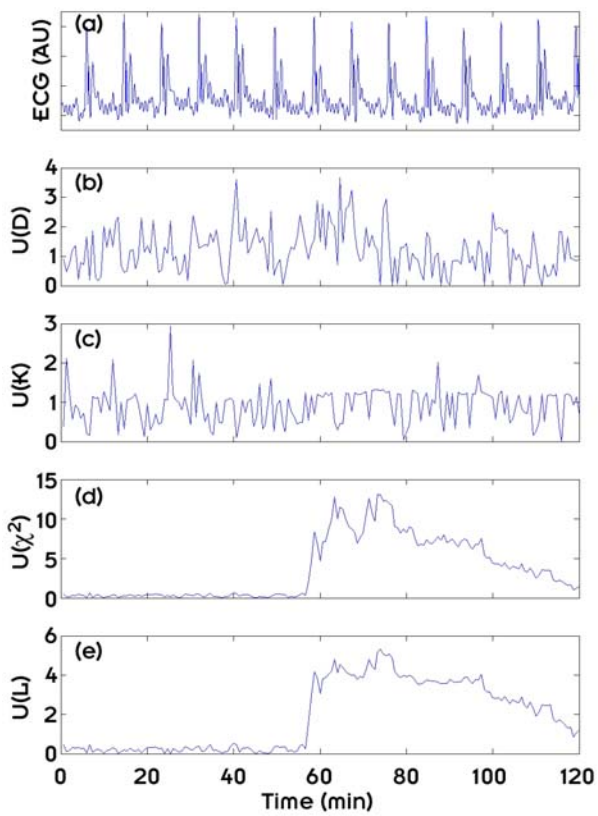

Fig. 1. Results for data set #33209V, showing time-serial plots for (a) 2.4 s of ECG data collected at 500 Hz; (b) correlation dimension, D; (c) Kolmogorov entropy, K*; (d)* U(χ^2)*; and (e)* U(L)*. The phase-space dissimilarity measures in subplots (d) and (e) were computed for d = 2, S = 2, λ = 30, after removal of breathing artifacts. Each cutset has* N *= 20,000 points, corresponding to 40 s.*

Table 2. Summary of results

CS	ID	T3-T2	T3-T0	Indication time			
				CH1	CH2	CH3	CH4
1	9V	1794	5332	270	FN	FN	FN
2	9W	3620	10707	1153	FP	5238	FN
3	9X	3597	10750	FP	FP	488	FP
4	9Y	3595	10750	705	2769	FP	4059
5	9Z	3607	10707	1768	FN	FP	435
6	A1	3595	10750	FP	FN	FP	1952
7	23	3593	10406	1997	FN	FP	FP
8	A4	3597	10750	FP	FP	832	FP
9	A5	3605	10707	FP	2458	FN	2071
10	~1	3594	10750	FP	FP	319	4275
11	~2	0	10707	FP	TN	FP	TN
12	A9	0	10750	TN	TN	FP	TN
13	AA	0	10750	TN	FP	FP	TN
14	BD	0	8729	TN	TN	TN	FP
15	AE	0	10707	FP	FP	TN	FP
16	AF	3597	10750	FP	FP	FP	3240
17	AG	0	10707	FP	FP	TN	TN
18	AH	3648	10707	4350	FN	FN	781
19	AI	3595	10707	FN	5478	FN	FN
20	AJ	3594	10750	FP	FN	FN	663
21	AK	3595	10750	FP	FN	361	FN
22	AL	3670	10750	7080	802	FP	FP
23	AM	3595	10664	490	FP	FP	FN

CS = case number (1 through 23).
ID = identifier from Table 1.
T3-T2 = exposure time (seconds) from Table 1.
T3-T0 = total duration (seconds) of the experiment from Table 1.
CHn = ECG channel number, with the value of n between 1 and 4.
FN = false negative (no indication when exposure occurred).
FP = false positive (indication of change without exposure, or indication of change before exposure).

Summary and Conclusions

Using ORNL's previously developed model-independent analysis to detect condition change (dissimilarity) in nonlinear time series, we examined the ECG data from endotoxin-exposed rats. The phase space indicators of condition change measure the difference between suitably discretized forms of the invariant distribution functions on the attractor for the base case and test case, as χ^2 statistics and L_1 distance. Use of these metrics magnifies the differences between the process dynamics, and avoids the inner cancellation effects due to averaging that occur when calculating the traditional nonlinear measures (TNM). As a result, significant changes in model dynamics are clearly detected by the PSDM, as the parameters vary. On the other hand, these changes are either barely detected or undetected by the TNM, such as the Kolmogorov entropy or the correlation dimension (see subplots b–c in Fig. 1). These preliminary results show that advanced nonlinear analysis of rat ECG data is a sensitive and specific identifier of inhalational endotoxin exposure. This novel tool represents a significant step toward identifying early sepsis in the hospitalized patient. Specifically, the novel nonlinear measures of rat ECG show small values of the PSDM in the normal state, followed by a notable monotonic rise and/or by successive occurrences of the PSDM above the "normality threshold" that indicates forewarning or onset of sepsis, as shown in Fig. 1. The renormalized PSDM are consistently more robust and timely than those provided by the TNM. An important next step is multichannel analysis to strengthen the criteria for correct forewarning.

ORNL stands to benefit substantially from this work. This project has laid the groundwork for further investigation of changes in nonlinear dynamics associated with sepsis onset for the purpose of early identification and treatment of sepsis. This work has fostered a serious collaboration between ORNL and investigators at UTMCK including Dr. Michael Karlstad and several trauma surgeons. A proposal to the National Institute of General Medical Sciences (NIGMS) of the NIH will be prepared. Dr. Scott Somers of NIGMS has indicated to Dr. Karlstad that such a proposal would be received with great interest.

References

[1]D. Toweill et al., "Linear and Nonlinear Analysis of Hemodynamic Signals During Sepsis and Septic Shock," *Crit. Care Med.* **28**(6), 2051–2057 (2000).

[2]M. S. Ellenby et al., "Uncoupling and Recoupling of Autonomic Regulation of the Heart Beat in Pediatric Septic Shock," *Shock* **16** (4), 274–277 (2001).

[3]D. T. Doweill and B. Goldstein, "Linear and Nonlinear Dynamics and the Pathophysiology of Shock," *New Horizons* **6** (2), 155–168 (1998).

[4]M. P. Griffin and J. R. Moorman, "Toward the Early Diagnosis of Neonatal Sepsis and Sepsis-Like Illness Using Novel Heart Rate Analysis," *Pediatrics* **107**(1), 97–104 (2001).

[5]L. M. Hively and V. A. Protopopescu, "Detection of Changing Dynamics in Physiological Time Series," *Proc. ANS Topical Conf. on Nucl. Math. Comput. Sci.* (Gatlinburg, Tennessee) April 6–10, 2003.

[6]J. E. Skinner et al., "Application of Chaos Theory to a Model Biological System: Evidence of Self-Organization in the Intrinsic Cardiac Nervous System," *Integr. Physiolog. Behav. Sci*. **31,** 122–146 (1996).

[7]J.-P. Eckmann and D. Ruelle, "Ergotic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.* **57**, 617–656 (1985).

[8]E. A. Jackson, *Perspectives of Nonlinear Dynamics*, Cambridge University Press, Cambridge, vol. 1989; vol. 2, 1990.

[9]S. H. Strogatz, *Nonlinear Dynamics and Chaos*, Addison-Wesley, Reading, 1994.

[10]H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer Publ., New York, 1996.

[11]L. M. Hively and V. A. Protopopescu, "Channel-Consistent Forewarning of Epileptic Events from Scalp EEG," *IEEE Trans. Biomed. Engr.* **50**(5), 584–593 (May 2003).

Genetic Variation in Cellular Responses to Low-Dose Radiation

E. M. Rinchik,[1,2] K. T. Cain,[1] L. Branstetter,[1] and P. R. Hoyt[1,3]
[1]*Life Sciences Division, Oak Ridge National Laboratory*
[2]*Department of Biochemistry, Cellular, and Molecular Biology*
[3]*Graduate School of Genome Science and Technology, The University of Tennessee*

The molecular responses of cells exposed to biological stimuli or insults (such as radiation, or a chemical or biotoxin) vary with the type of exposure, dose and dose rate, the type and genetic constitution of exposed cell(s), and the tissue/organ site of the responding cell. Our research goal is to use genomic technologies to address how individual genetic constitution affects tissue responses to insults, starting with response to low-dose radiation (LDR). This Seed project began standardizing conditions for validating the whole-genome, molecular LDR response, using microarray technology and quantitative gene-expression techniques, in tissues from genetically diverse inbred mouse strains. These experiments will build a conceptual and technical framework for future applications of mouse genetics to studying variation in gene responses that may underlie differences in human susceptibility to toxic exposures.

Introduction

A primary mission of the DOE's Life Sciences Program historically has been the assessment of hazards from radiation. Cells exposed to radiation undergo a number of complex molecular responses that result in some kind of cellular consequence or action. Although some responses are shared between high- and low-dose exposures, other responses differ. The DOE Low-Dose Radiation Research Program (http://lowdose.org/) is interested in the cellular and molecular consequences of exposure to low-dose (or low-dose-rate), low–linear energy transfer radiation (LDR). It has become increasingly clear over the past 10 years that there are specific cellular and molecular responses to LDR (defined here as ≤10 cGy) and that the tools of molecular biology, as well as newer and evolving tools of genomics, can be exploited to study the highly complex LDR response. In this project, we focused on how an individual's molecular tissue response to LDR may vary because of differences in individual genetic constitution.

Much of the work previously done on the molecular LDR response has been done in vitro, on cell lines in tissue culture; some, but not nearly enough, has been done in whole animals or complex tissues. It is known that exposure of cells to low doses of ionizing radiation induces or represses certain gene products, some with very rapid kinetics (i.e., within hours).[1–3] A number of different responses, feed-in points, and effector mechanisms may characterize LDR responses, and responses in tissues in whole animals may differ from those in cultured cells. Moreover, variation in any of a number of gene products, even minor variation, due to differing genetic constitutions among unique individuals, may have significant effects on molecular responses as well as on overall outcomes.

This project was designed to provide preliminary data for using techniques and resources of mouse genetics and genomics to investigate issues of genetic susceptibility (in this specific case, to low-dose radiation). The project concentrated on (1) building a substantial bank of tissues from genetically different animals exposed to both high- (200-cGy) and low- (10-cGy) doses of acute X rays, a significant investment that could be drawn on for all future LDR-related genomics projects; (2) standardizing global genomic measurements of the LDR response reproducibly on site, at high throughput; and (3) gaining experience with data-analysis issues to identify bioinformatics needs relevant to DOE and other federal agencies interested in low-dose radiation exposure.

Technical Approach

We used DNA microarray technology[4] and other quantitative techniques for measuring gene expression among individuals (represented by different inbred or mutant strains of mice). If we could identify such LDR-response differences, we could subsequently define gene-expression profiles associated with LDR, and, importantly, learn how these profiles can vary from individual to individual. We began the project by using DNA microarrays with 1152 genes ("1K chip") enriched in skin-specific genes and oncogenes to analyze RNA samples (representative of the gene-expression compartment) from irradiated and control tissues. As the project progressed,

we introduced the use of oligonucleotide microarrays representing 22,000 mouse genes, as well as initiated development of quantitative polymerase chain reaction (PCR) (qPCR) technology to verify individual microarray results and to assay specific genes previously known to be involved in cellular stress-response pathways.

Results and Accomplishments

Tissue Bank

One project goal was to create a tissue bank from both high- and low-dose-exposed C57BL/6J and DBA/2J tissues. Almost 2000 individual samples were collected, and all exist in storage either as tissues preserved for subsequent RNA preparation, or as prepared RNA samples. Table 1 provides a general summary of what constitutes this tissue bank as of the end of the project.

Initial Microarray Experiments

At the beginning of the project, we grew, amplified, purified, and verified 1152 clones and controls to be used as the "1K" microarray for this project. This array included approximately 40 skin-specific genes, and at least 100 genes involved in the progression of cancer.

Preliminary results showed that exposing mice to a 10-cGy dose of acute X rays did produce a change in the expression of genes in skin. Results of typical microarray data utilizing the ORNL "1K" chip for 10-cGy exposures in C57BL/6JRn (B6) mice are shown in Fig. 1.

Two general conclusions can be drawn from these studies. First, a transcriptional response due to LDR was detected using our microarrays, as indicated by the subset of genes differentially expressed at >95% confidence levels. The second conclusion is that the LDR response involves unexpected cellular systems in vivo as judged by the lack of "predicted" genes, known to respond in vitro[1–3] (e.g., *Cdkn1a*). We then moved to compare (by microarray) transcriptional responses in dorsal skin among distinct strains B6, C57L/6-p^{un}/p^{un} (abbreviated PUN), and C3H-*a/a* (abbreviated C3H-a). B6 and PUN differ by only one gene, with PUN having diluted amounts of eumelanin pigment in the skin. We were very interested in assaying the effects of pigmentation on the LDR response in skin, as it is known that melanin pigment can protect against higher doses of X rays in cells in culture. (We also bred a large population of B6-Tyr^{c-2J} animals, which are essentially B6 mice with no pigment, but could not analyze these within the project period; however, they are included in the tissue bank in Table 1.) B6 and C3H-a are both black (having black eumelanin but no yellow phaeomelanin in the skin) but have entirely different genetic backgrounds and could differ by thousands of minor variants. Thus, we could compare LDR responses due to variation in genes other than those involved in pigmentation. Preliminary results from these studies are summarized in Table 2.

At 95% confidence levels, each mouse strain displayed a nonidentical, but overlapping, set of LDR-responsive genes. When the strains were compared, C3H-a showed the highest number of genes differentially expressed and also showed the highest number of unique LDR responses (genes responding only in C3H-a). An interesting aspect of these data was the high number of response genes with opposite transcriptional effects. For example, there were 13 common genes found to respond in both B6 and C3H-a mouse strains, using 10-cGy radiation. However, 14 of those genes were upregulated in one strain, while being downregulated in the other. While the mechanism of LDR response is beyond the scope of this study, these differences certainly warrant further study. It should be pointed out that the extensive unique response in C3H-a as compared to the B6 strain could be the result of many genetic differences, or could simply be due to one major difference that reverberates through one or more LDR-response pathways. Such preliminary data will be important in considering future projects.

These data were also examined by looking at those genes with the highest levels of differential expression in order to ascertain their usefulness as strain-specific, robust biomarkers for LDR response, and to provide preliminary data for future funding opportunities. These results are summarized in Table 3.

Table 1. General summary of current bank of tissues from mice of various genotypes exposed to both low (10-cGy) and high (200-cGy) doses of acute X rays

Number of samples in tissue bank	1939 total samples
Strains represented	A/J; C57BL/6J (B6); B6-Tyr^{c-2J}/Tyr^{c-2J}; B6-*Trp53–/Trp53–*; BALB/cJ; BALB-*Tyr+/Tyr+*; DBA/2J; C3H/HeJ, C3H/Rl; C3H/HeJ-*a/a*; B6-p^{un}/p^{un}; C3H-p^{J}/p^{J}; B6C3F1
Doses represented	Zero cGy (sham exposures); 10 cGy; 200 cGy
Tissues represented	Dorsal skin, spleen, testis—all strains Bone marrow, thymus—some strains
Time sampled after exposure	3.5 h—all strains and tissues 1 h—some strains, tissues, and exposures
Tissue preservation for RNA	All strains for all exposures and times

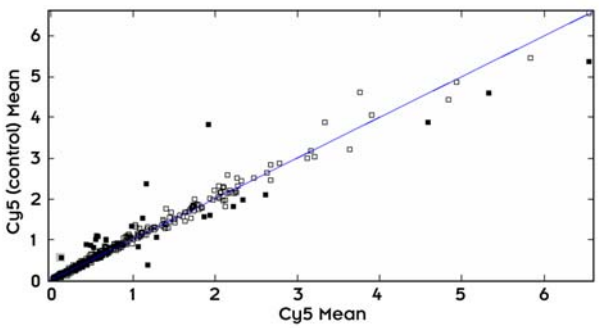

Fig 1. Scatterplot of differential gene expression in C57BL/6JRn mouse skin exposed to 10 cGy of acute X rays. Data were generated using the ORNL "1K" microarray. Control (sham exposed) B6 mouse transcriptional response data are on the y-axis, while the exposed B6-mouse skin data are plotted on the x-axis. The diagonal line represents a 1:1 ratio of gene expression (no response). Genes whose response falls above the diagonal are repressed by the exposure, while induced genes are indicated below the diagonal line. Filled squares are the genes whose response was determined to be significant at the 99% confidence level.

When specific gene systems are analyzed in detail, we noticed that certain genes that might have been thought to have been induced (such as *Cdkn1a*, a gene induced by p53 in a damage/stress response) were not induced in skin according to our microarray results. In retrospect, these data are consistent with other array data reported by other laboratories doing microarray research on the LDR response, reported at the DOE Low Dose Radiation Research Program Workshop IV, held recently in Washington, D.C. Detecting these types of differences in in vivo tissues and perhaps among genetically diverse individuals is what was driving our whole project and indicates potentially enticing data that will need to be confirmed by repetition of microarrays and by other quantitative gene-expression techniques. Because the in vivo response could be dramatically different from the already published in vitro response, we decided, for all mouse microarray applications (not just for LDR), to transition to larger 22,000-gene oligonucleotide arrays. The larger arrays would provide a more complete view of the LDR response, and the use of oligonucleotide probes may also provide additional information about alternate splicing events (if they occur). These arrays would validate the results seen using the ORNL "1K" chip and eliminate possible false data resulting from chimeric cDNA elements that sometimes occur in cDNA libraries.

Summary and Benefits

Using in-house microarrays, a transcriptional response to LDR of 10 cGy was detected and analyzed. In general, when looking at mouse skin, approximately 10% of the genes on the array may be statistically determined to respond to LDR at 3.5 h post exposure. Different strains showed different, but overlapping, responses, with C3H-a showing the greatest number of response genes. The gene-expression experiments, as well as the considerable tissue bank that was amassed during the project period, have provided important preliminary data as well as an important resource for future work. Although some genes predicted to be involved in low-dose response were detected, many genes were not, indicating that the in vivo process has unexpected complexity. The project was successful at determining whether strain-specific LDR response genes might be identified, and a number of presumed genotype-based response differences observed in one early experiment funded by this project were included as preliminary results in a proposal submitted to the DOE's Low-Dose Radiation Research Program.

We continue to believe that an important post-genome niche for ORNL will lie in studying the genetics of susceptibility among individuals to exposures of all types. This focus area would thus leverage ORNL's Mouse Genetics Program into solving problems in human susceptibility, beginning with investigating LDR-response variation in whole-animal tissues. This project clearly helped facilitate this goal. These capabilities for measuring

Table 2. Summary of strain comparisons for differentially expressed genes using 10-cGy radiation exposure

Mouse strain	LDR-responsive genes	Common genes showing opposite transcriptional effects
B6	61	
PUN	82	
C3H-a	179	
Common to all three strains	10	
Common to B6 and PUN	3	2
Common to PUN and C3H-a	33	14
Common to B6and C3H-a	13	8
In B6 ONLY	34	
In PUN ONLY	37	
In C3H-a ONLY	124	

Table 3. Summary of strain comparisons of easily detectable differentially expressed genes
(Genes showing 2X or greater differences)

Strain	Total	Induced	Repressed
B6	5	4	1
PUN	11	4	7
C3H-a	22	4	18

gene expression, as well as the building of a large bank of LDR-exposed tissues from a number of mouse strains, which can be tapped far into the future, have set the stage for determining exposure, sampling, and molecular-screening criteria for future larger projects.

Likewise, we feel that the approach to studying genetic differences among individuals exposed to an agent (either radiation, chemicals, or biotoxins), pioneered in this project, will position ORNL to study genotype-based differences in responses to other types of exposures that will be relevant to funding by the National Institutes of Health [for cancer susceptibility (NCI), susceptibility to chemical exposures (NCI, NIEHS)], and possibly by the Departments of Defense and Homeland Security (for bioterror agents).

References

[1]A. V. Prasad, N. Mohan, B. Chandrasekar and M. L. Meltz, "Induction of Transcription of 'Immediate Early Genes' by Low-Dose Ionizing Radiation," *Radiat. Res.* **143**, 263 (1995).

[2]S. A. Amundson, K. T. Do, and A. J. Fornace Jr., "Induction of Stress Genes by Low Doses of Gamma Rays," *Radiat. Res.* **152**, 225 (1999).

[3]S. A. Amundson, R. A. Lee, C. A. Koch-Paiz, M. L. Bittner, P. Meltzer, J. M. Trent, and A. J. Fornace, Jr., "Differential Responses of Stress Genes to Low-Dose-Rate Gamma Irradiation," *Mol. Cancer Res.* **1**, 445 (2003).

[4]P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee, and J. Quackenbush, "A concise guide to DNA microarray analysis," *Biotechniques* **29**, 548 (2000).

Developing a New Core Competency for ORNL: Macromolecular Neutron Crystallography

G. J. Bunick
Life Sciences Division

The project compared the neutron and the ultra-high resolution synchrotron X-ray structures of D-xylose isomerase (XI) in order to ascertain the relative merits of each technique in locating the critical hydrogen atoms and water molecules in macromolecular structures. The working hypothesis was that neutrons could locate the positions of hydrogen atoms from moderate resolution data that might lead to an understanding of the enzymes mechanism, whereas X rays might not be able to provide the same information. Neutron diffraction data were measured by the Protein Crystallography Station (PCS) at the Los Alamos Neutron Sciences Center (LANSCE). Subsequent solution of the structure validated the hypothesis. XI is the largest protein studied to date by time-of-flight (TOF) neutron diffraction. The second objective of the proposed research was to validate the operational capabilities of the PCS. The project also demonstrated the use of spallation-generated neutrons for macromolecular crystallography and thereby provided scientific justification for the future macromolecular neutron diffractometer (MaNDi) at the Spallation Neutron Source (SNS).

Introduction

The hydrogen bond lies at the heart of biochemistry, with hydrogens comprising roughly half of the atoms in biological materials such as proteins and DNA. Hydrogen ions are involved in the molecular actions of virtually all enzymatic reactions. Within macromolecules, biochemical mechanisms can be elucidated and modeling studies facilitated from knowledge of hydrogen positions, waters of solvation, and ions.

The principal techniques currently used for experimental structure determination of macromolecules are X-ray diffraction and nuclear magnetic resonance (NMR). It is often stated that a combination of these two techniques can provide all the information needed to determine the atomic positions within a macromolecule, including the positions of hydrogen atoms. This is simply not the case.

Neutron diffraction can provide important and unique information for macromolecular structure-function studies. The ability of neutrons to reveal the positions of hydrogens even at moderate resolution (2.0–2.3 Å) is the foundation of the scientific justification for neutron diffraction of biological samples. The protonation and ionization states of atoms, which are necessary to determine the atomic charges and pK_as of proteins, can be determined from a neutron structure. Neutron structures can also provide the orientation of polar groups such as hydroxyls and amides, which can reveal hydrogen-bonding patterns. Water structure is essential to understanding protein structure and function. Neutron density maps can provide detailed information about the water structure in protein active sites and the solvation shells around proteins. From neutron diffraction studies, details of enzyme catalytic mechanisms and protein molecular recognition sites can be determined. Knowing such details are of great value for rational drug design, ligand design, and rational protein design.

Technical Approach

The ultra-high-resolution X-ray structure of XI required that the crystal be flash-cooled below the glass transition temperature of water to reduce radiation damage. Previous X-ray diffraction studies of XI did not utilize flash cooling and synchrotron radiation.[1] A suitable cryoprotectant was found and a flash cooling procedure was developed for XI crystals grown using ammonium sulfate as the precipitant.[2] The X-ray measurements were treated by standard crystallographic procedures to produce an electron density map of the protein. The program Xfit of XtalView[3] was used to fit and interpret the electron density map. The program Refmac5 of the CCP4[4] suite was used to refine the X-ray structure.

Large crystals were required for the neutron diffraction studies because of the low flux that is a characteristic of all neutron sources in comparison to X-ray sources. For example, our laboratory rotating-anode X-ray generator provides an order-of-magnitude-higher flux than any neutron source. Figure 1 shows a large single crystal of XI that was grown in a counter diffusion cell (CDC). This device was developed by the National Aeronautics and Space Administration (NASA) for the

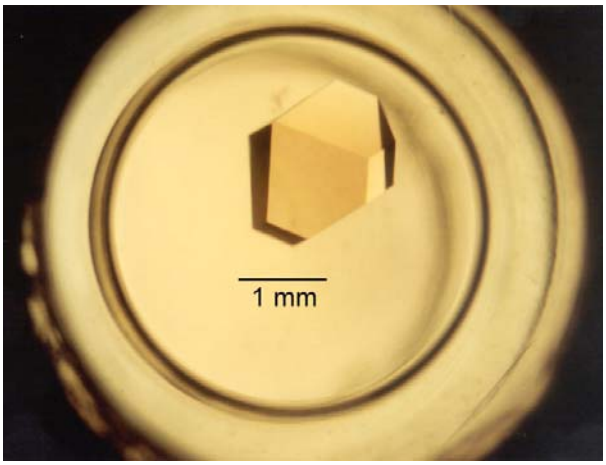

Fig. 1. A single large crystal of XI grown in a counter diffusions cell designed for macromolecular crystallization in microgravity. The crystal was approximately 3 mm × 2 mm × 1.5 mm when it was transferred to a capillary for the neutron diffraction measurements.

growth of macromolecular crystals in microgravity. The concentrations of solutions in the two chambers of the CDC had to be determined experimentally in order to grow large high-quality crystals.

The integration of the measured reflections and the subsequent neutron intensity data scaling, merging, and wavelength normalization required modifications to several different software programs in order to utilize the time-of-flight information encoded in the data. The XI data set was used to debug the programs. The statistics of the data set suggest that other programs will have to be tried to apply absorption corrections to the data. The analysis of the XI data points to the areas where better algorithms for processing TOF neutron data will be needed in the future.

The neutron structure was visualized and fitted using XtalView[3] and refined using the program SHELX.[5] SHELX is the only crystallographic program able to control the refinement process atom by atom and to handle the replacement of hydrogen atoms by deuterium atoms as the refinement proceeds.

Results and Accomplishments

Neutron diffraction data (Fig. 2) were measured on the PCS at LANSCE and the data were integrated, wavelength normalized, and scaled. The quality of the neutron diffraction data were then assessed. The values of R_{merge} for the XI data for most resolution bins are larger than values reported to date in neutron studies of other proteins. The two main differences between this study and previous ones are the size of the system under investigation and the wavelength range of neutrons used. With a unit cell volume of approximately 1×10^6 Å,[3] this study

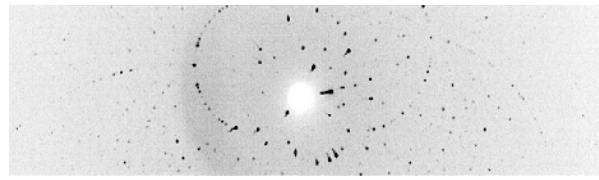

Fig. 2. Neutron diffraction data from one crystal setting for the native XI crystal. The image is an overlay of all time intervals for the setting. Therefore, weaker reflections from longer wavelength diffraction can be lost in the background from the short wavelength time slices. However, since the data from each time slice is analyzed separately, these reflections can be measured against the background within their time slice. This reduces the time needed to measure the signal for each reflection and improves the signal/noise ratio for the weak data.

represents by far the largest system studied to high resolution with neutron diffraction.

Previous neutron data sets on other proteins have been collected at steady-state neutron sources using either monochromatic neutron beams or quasi-Laue neutron beams with a wavelength band of approximately 1 Å, whereas the TOF neutron data set has been collected over a large wavelength range of 0.7 Å to 7 Å. In order to investigate the effects of using a large wavelength band on the data merging statistics, the data were normalized and then merged over small wavelength bands of 1 Å. The results show that over equivalent wavelength ranges the accuracy of the data is at least as good as previous data collected on other neutron instruments. It is postulated that the behavior of the data merging statistics as a function of the wavelength bandwidth may be explained by variations in incoherent scattering from hydrogen, which result in an effective absorption factor that is a function of both wavelength and crystal shape. The scaled intensities of equivalent reflections collected at the same wavelength will differ because of the different associated path lengths through the crystal and the size of this difference will change with wavelength. The unusually high redundancy in the data may help compensate for this effect and result in accurate averaged intensities. However, it is clear that methods should be developed for properly correcting for this effect so that the accuracy of the data can be further improved.

The ultimate test of the quality of the data set is its ability to locate single deuterium atoms. A 2Fo-Fc sigma-A map was calculated using phases from the room temperature structure determined from a 1.55 Å X-ray data set[1] with all of the water molecule atoms and labile hydrogen atoms omitted, and using neutron amplitudes in the resolution range 10 Å–2 Å. There are clear indications of the protonation states of residues around the catalytic site. The direct observation of these states is of crucial importance in determining the details of catalytic mechanism. Figure 3 shows density at the end of a lysine

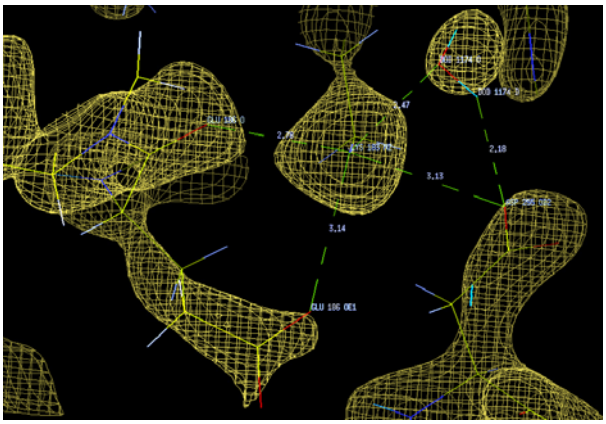

Fig. 3. The terminal nitrogen atom of lysine 183 is triply deuterated in the neutron density map contoured at 2 σ. Hydrogen bonds are made from the nitrogen atom to the main chain oxygen of Glu186, a carboxylate oxygen of Glu186, and a D_2O molecule. The D_2O molecule makes a hydrogen bond to Asp255, which may share one of the deuterons on the terminal nitrogen with Glu186 OE1. Lysine 183 may function as a hydrogen donor during catalysis.

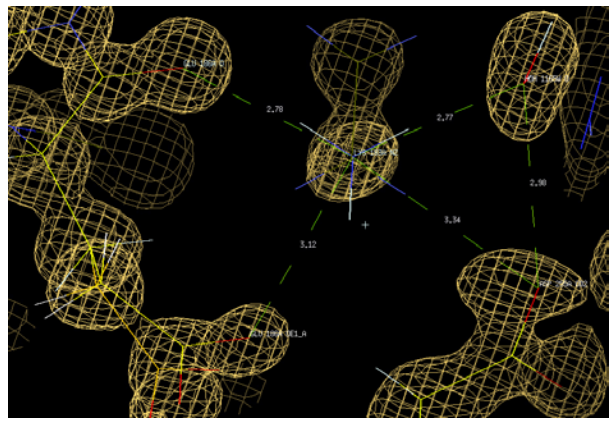

Fig. 4. Electron density for the terminal nitrogen atom of lysine 183 contoured at 2 σ from the 0.94 Å X-ray structure. The protonation state of the terminal nitrogen atom cannot be determined in the electron density map but is clearly determined in the neutron density map of Fig. 3.

residue (Lys183) that defines the orientation of the terminal ND_3 group. This particular orientation allows the individual deuterium atoms in the ND_3 group to hydrogen bond to a main chain oxygen atom from Glu186, a carboxylate oxygen atom from Glu186, a carboxylate oxygen atom from Asp255, and a water molecule. The proximity to the catalytic site suggests that this residue is important. Figure 4 shows the same lysine residue in the 0.94 Å resolution X-ray structure. The protonation state of the lysine cannot be determined from the electron density map.

Summary and Conclusions

The metalloenzyme D-xylose isomerase forms well-ordered crystals that diffract X rays to ultra-high resolution (<1 Å). However, structural analysis with X-ray diffraction data has been unable to differentiate between several postulated mechanisms that describe the catalytic activity of this enzyme. Neutrons with their greater scattering sensitivity to hydrogen atoms could help resolve this by determining the protonation states within the active site of the enzyme. As a comparison of the relative merits of X-ray and neutron structures for locating hydrogen atoms, neutron data to 1.8 Å were measured from the unliganded protein at the LANSCE PCS. Measurement of these neutron diffraction data represents several milestones: this is one of the largest biological molecules (a tetramer, MW ~160,000 daltons (Da), with unit-cell lengths around 100 Å) ever studied at high resolution using neutron diffraction. It is also one of the first proteins to be studied using time-of-flight techniques.[6]

The success of the initial diffraction experiments with D-xylose isomerase demonstrate the power of spallation neutrons for protein crystallography and should provide further impetus for neutron diffraction studies of biologically active and significant proteins.[7] This research provided results that were used in the scientific justification for the macromolecular neutron diffractometer (MaNDi) instrument to be constructed at the SNS. The need for a specialized suite of software to process and solve macromolecular neutron structures was apparent as the project proceeded. It was also determined from presentations of this research that the majority of X-ray crystallographers do not have the basic information to start a neutron study of a protein. A guidebook on the subject will be most useful to help the X-ray crystallographer prepare for a neutron diffraction study.

Additional data are being measured on crystals of XI with bound substrates and inhibitors in order to provide the specific information needed to clarify the catalytic mechanism of this enzyme. Follow-on funding will be sought in the form of grants submitted through the University of Tennessee to NIH and possibly NASA or NSF.

References

[1]H. L. Carrell, H. Hoier, and J. P. Glusker, "Modes of Binding Substrates and their Analogues to the Enzyme D-Xylose Isomerase," *Acta Cryst.* **D50**, 113–123 (1994).

[2]B. L. Hanson, J. M. Harp, K. Kirschbaum, C. A. Schall, K. DeWitt, A. Howard, A. A. Pinkerton, and G. J. Bunick, "Experiments testing the abatement of radiation damage in D-Xylose Isomerase Crystals with cryogenic helium." *J. Synchrotron Radiation* **9**, 375–381 (2002).

[3]D. E. McRee, "XtalView/Xfit—A Versatile Program for Manipulating Atomic Coordinates and Electron Density." *J. Structural Biology* **125**, 156–165 (1999).

[4]"CCP4, Collaborative Computational Project, Number 4." *Acta Cryst.* **D50**, 760–763 (1994).

[5]G. M. Sheldrick and T. R. Schneider, "SHELXL: high resolution refinement," pp. 319–343 in *Methods in Enzymology*, 227, edited by C. W. Carter, Jr. & R. M. Sweet, Academic Press, San Diego, 1997.

[6]LANSCE Highlight, "High-Resolution Neutron Diffraction Data Collected on the Protein Crystallography Station at LANSCE Exceed Researchers' Expectations." (2003) http://lansce.lanl.gov/news/features/Xylose.htm

[7]C. Day, "Search & Discovery: Neutron Diffraction Overcomes Flux Limits to Resolve a Large Protein Structure," *Physics Today*, pp. 17–19, November 2003.

Novel Treatment for Biological Warfare Pathogens

C. E. Easterly,[1] G. D. Griffin,[1] M. Z. Hu,[2] T. L. Nichols,[3] and A. A. Vass[1]
[1]*Life Sciences Division*
[2]*Nuclear Science and Technology Division*
[3]*University of Tennessee Medical Center*

For the first time, engineered antimicrobial nanoparticles have been demonstrated as the basis for an extremely broad-spectrum therapy for life-threatening microbial agents such as might be used in a bioterrorist attack. This work was intended to test whether or not elemental or small-molecule antimicrobials in nanoparticle form can be therapeutic against pathogenic infections. If elemental or very-small-molecule biocidal ions can be transported to bacteria at the site of infection, these ions may provide novel therapies. "Stealth" liposomes appear to provide innate targeting to inflammation sites by taking advantage of natural processes that accompany inflammation-associated infections. The work reported herein entails a series of experiments designed to produce biocidal nanoparticles, make carrier-targeting long-circulating liposomes, and conduct proof-of-principle testing of liposome-encapsulated nanoparticles with bacteria-infected macrophages and bacteria-infected amoebae. Both of these systems are phagocytic, engulfing organisms. Two replicates of these experiments support our hypothesis that liposomes containing antimicrobial nanoparticles do in fact rescue macrophage cells infected with *Legionella.*

This project focuses on the need for therapies against biological warfare pathogens that have been selected or engineered to be pathogenic to humans and immune to modern antibiotics. (Medical use on resistant organisms frequently encountered in modern society is also anticipated.) Our objective was to demonstrate, with a proof-of-principle test, an approach to guide the development of a broad-spectrum therapy against such pathogens. A number of elemental as well as very small molecular ions are known to be toxic to pathogens, but after the discovery of penicillin and the development of subsequent generations of antibiotics, these materials have lacked interest by the medical community. Given the correct form and delivery mechanism, a number of elemental ions have the potential to provide killing mechanisms different from current antibiotics and, in combinations, their diverse modes of attack make future evasive development highly unlikely. Our task is to combine a novel killing methodology together with a carrier and targeting system and demonstrate the recovery or increased survival of pathogen-infected cells.

Technical Approach and Findings

The approach entailed the use of elements or very small molecules in ionic form known to be toxic to pathogens, delivered in a package that was also a targeting agent. This seed project focused on a first demonstration using nanoparticles as the source of killing ions and "stealth" liposomes as the delivery and targeting mechanism. Certain elemental and very-small-molecule ions have long been known to be toxic to pathogens, but the means for providing a constant-concentration internal dosage that is therapeutic but not overtly toxic to the host has not been developed. Nanoparticles offer a means for providing a constant concentration of ions, because the solubility of nanoparticles can be controlled, thus providing a steady source of ions. Thus the first step is to develop a nanoparticle that is stable and biologically active against pathogens. Next, appropriate liposomes must be constructed and loaded with nanoparticles. Finally, a series of biological experiments must be carried out. We have chosen to use *Legionella* as the biological warfare surrogate. The bacterium, *Legionella pneumophila,* is the causative agent of Legionnaires' disease and related respiratory ailments. It is a facultative intracellular pathogen with a similar pathogenicity to many biological warfare (BW) agents. The bacteria are able to infect, multiply within, and kill human macrophages in a fashion similar to BW agents. We have used two phagocytic organisms to demonstrate the potential therapeutic efficacy of the proposed therapeutic approach: J774, a mouse-macrophage cell line, and an amoeba (*Acanthamoebae*), which are noted for their phagocytic activity.

Synthesis of Nanoparticles

Nanoparticles made from elements and/or very small molecules are intended to be encapsulated inside liposomal vesicles. Since the sizes of the "stealth" liposomes will be in the range of 80–100 nm, the nanoparticles must be less than this dimension in order to be encapsulated within the liposome vesicles. An aqueous nanoparticle sol is preferred for compatibility with the liposome formation conditions. Typically, the production of nanoparticles involves the use of steric capping molecules or stabilizers (such as trialkylphosphine/amine, alkanethiols, long-chain unsaturated carboxylates, CS_2, or quaternary ammonium disulfide) to control the size and stabilize the nanoparticles.

Nanoparticles were produced using a process currently under patent application. By virtue of their solubility characteristics in biological media, these nanoparticles were able to maintain a killing concentration of ions for bacteria, but not their host cells (see below). The overall concentration of the killing agent in the sol was approximately 2 ppm, including ions and nanoparticles, as determined using laser-induced breakdown spectroscopy (courtesy of Madhavi Martin, Environmental Sciences Division). However, we have not yet determined the size vs number density for the nanoparticles, nor the concentration of ions in the sol.

Nanoparticle sols were stored in low-potassium scintillation vials because storage in other glass containers resulted in degradation of the nanoparticle size. This phenomenon is thought to be associated with the decay of K-40 in other containers, leading to radiation-induced deterioration of the self-assemblies to particles larger than the liposomes. We have monitored samples, and the prepared sols have remained consistent in size and concentration for at least 2 months.

Liposome Preparation

For intravenously introduced liposomes, it is generally accepted that the prolonged circulation time of polyethylene glycol (PEG)–coated liposomes as compared to conventional non-coated liposomes is a result of the PEG coating. Methods for preparation of PEG-coated liposomes described by Schiffelers et al.[1] were followed. In addition to prolonged circulation times, preferential localization of these liposomes at sites of infection or inflammation has been demonstrated in a variety of experimental models.[1] Schiffelers et al.[1] incorporated two antibiotics into their liposomes.

Liposomes were prepared using procedures described in the literature with slight modifications. Briefly, the following lipids were dissolved in a mixture of chloroform/methanol at molar ratios of 1.85:1:0.15: partially hydrogenated egg phosphatidyl choline, cholesterol, and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[polyethylene glycol 2000]. The solvents were evaporated under nitrogen and the lipids dried and redissolved in 2-methyl-2-propanol, followed by shell-freezing and subsequent lyophilization. Following hydration, the lipid dispersion was sonicated to produce the liposomes. (Sonicator Model 112SP1G, 80KC, 80 Watts, Laboratory Supplies Co., Hicksville, New York).

Liposomes were then separated from unincorporated nanoparticles with a Sephadex G-25 column using distilled, deionized water as the eluant. This column allows larger particles to progress faster through the column than smaller particles. After passage of the liposomes, the next 3 mL were collected. Analysis of this volume indicated that the nanoparticle concentration was 25 times less than the concentration in the nanoparticle sol that was encapsulated by the liposomes. Thus the nonencapsulated material in the liposome preparation is less than 0.1 ppm.

Biotesting of Nanoparticles and Nanoparticle-Containing Liposomes

The first step was to evaluate the concentration of nanoparticle solution required to kill cultures of *Legionella*. Cultures of *Legionella* were grown to a concentration of 10^6/mL and exposed to the nanoparticles. *Legionella* and nanoparticles were combined in a ratio of 1:3 with a final overall concentration of about 2 ppm. After 30 min, all *Legionella* were killed as determined by re-examination of plates incubated 48 h after exposure to determine if any breakthrough growth occurred. Repeating the measurement with 10× diluted nanoparticle sol (about 0.2 ppm) produced about 99% kill after 1 h, but even after 48 h the kill was not 100%. Replication using similar nanoparticle concentrations showed complete killing with about 1 ppm within 1 h and very incomplete killing at 0.1 ppm. Thus, nanocrystalline materials in the low-ppm range provided good killing of *Legionella,* thus allowing for the final stage of this effort.

The next step was the proof-of-principle experiment. This involved the infection of cells with *Legionella* and subsequent exposure of the cells to the liposomes containing the nanocrystals. Two phagocytic cell lines were used: J774 mouse macrophages and *Acanthamoebae*. Prior experiments established the concentration of nanoparticles that was effective against *Legionella* and what concentration of nanoparticles and liposomes did not exhibit toxic effects on the cell lines. The two proof-of-principle experiments used liposomes that were hydrated in pure water containing about 2-ppm nanoparticles. Experiment one used a preparation of liposomes, which were on the order of 800 nm in size. Experiments were set up in duplicate with two flasks of *Acanthamoebae royreba* and line J774 set up as controls (nothing added). Both cultures were plated initially at densities of 10^4/mL in a

total volume of 2 mL. To the eight test culture flasks (four of *Acanthamoebae* and four of J774), 10µL of a 1.2 x 10^8/mL culture of viable *Legionella* (final concentration 1.2 x 10^6/flask) was added 2 days prior to treatment with the liposomes in order to ensure adequate time for phagocytosis. After 2 days, the media was changed in all flasks and 100 µL of the liposome preparation was added to two flasks of each set resulting in two control, two *Legionella*-infected, and two *Legionella*-infected plus liposome-treated flasks for each cell line. These flasks were monitored daily over a period of 2 weeks, and the media was changed as needed. After media changes, liposomes were re-added to ensure adequate treatment. Ultimately there were no survivors in the untreated samples of either organism, but all treated samples recovered.

Experiment two was set up in a similar manner, but used a liposome preparation which had an average 85-nm particle size. Because of slight differences in liposomal preparation, we added 160 µL of the liposome preparation to obtain the same quantity of nanoparticles as the previous experiment. After 8 days, the liposome-treated amoebic cultures were fully recovered, while the untreated test flasks had no obvious survivors. By day 12, the liposome-treated J774 line was nearly fully recovered while the untreated cells had no obvious survivors. The results of the two experiments with different size liposomes tend to indicate that size is important in how the cell lines react to the liposomes, mostly in terms of the initial recovery phase. For both experiments, both J774 and amoebae cultures that were infected with *Legionella* and subsequently treated with the liposomes recovered. Those cultures that did not receive liposomes were completely destroyed (Figs.1 and 2).We estimate that the maximum concentration of nanoparticles outside of the liposomes initially present in the treated flasks was less than 0.005 ppm. This is well below the 0.1 ppm that failed to provide complete killing in pure cultures of *Legionella*. Thus we have confidence that the effects seen with the liposome treatment are due to nanoparticles phagocytized by the macrophages and amoebae in close proximity to the engulfed *Legionella* that provide the recovery of the treated organisms.

Summary and Conclusions

We have demonstrated for the first time that it is possible to use engineered nanoparticles encapsulated within liposomes to act as a "stealth" bacteriocidal agent that can destroy bacteria in growing mammalian cells, and in a non-mammalian phagocytsing cell line, the amoeba. This is a substantial achievement, considering the large number of variables that these experiments entailed.

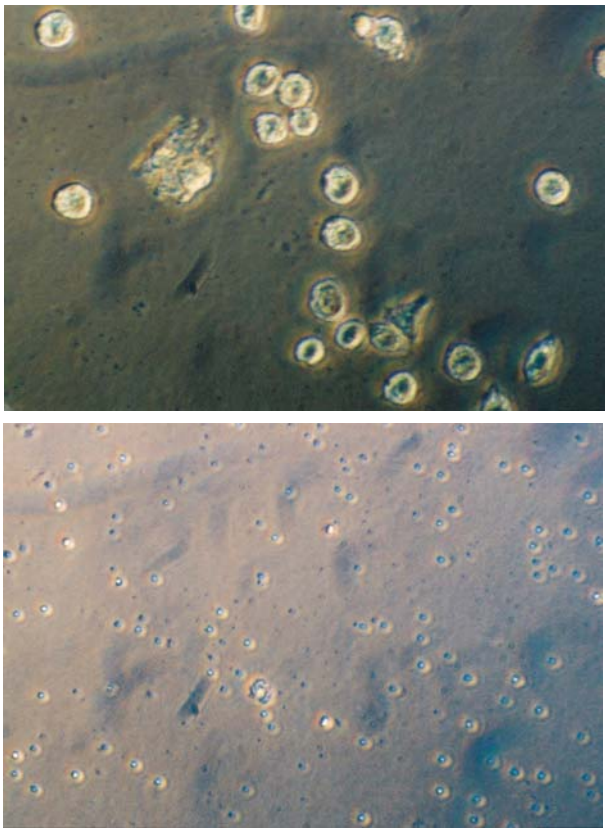

Fig 1. 400× view of J774 macrophage cells infected with Legionella (above) treated with nanoparticle-filled liposomes and (below) untreated.

- The nanoparticles must be stable.
- They must be of the correct size and physico-chemical composition to release ions in the correct active form resulting in an adequate concentration to kill engulfed bacteria but not the host cells.
- The liposomes must effectively incorporate the nanoparticles and simultaneously lend themselves to inclusion into phagocytic cells, either by membrane fusion or by engulfment by phagocytic mechanisms.
- The liposomes must release their bacteriacidal kill package intracellularly.
- The pathogen-toxic ions must traverse the cell interior to kill the bacteria.

These successful proof-of-principle experiments demonstrate the first of many potential antimicrobial materials that could be identified, prepared, and inserted into stealth liposomes.

Experimental results demonstrate an approach to solve one aspect of a critical problem—what could be done

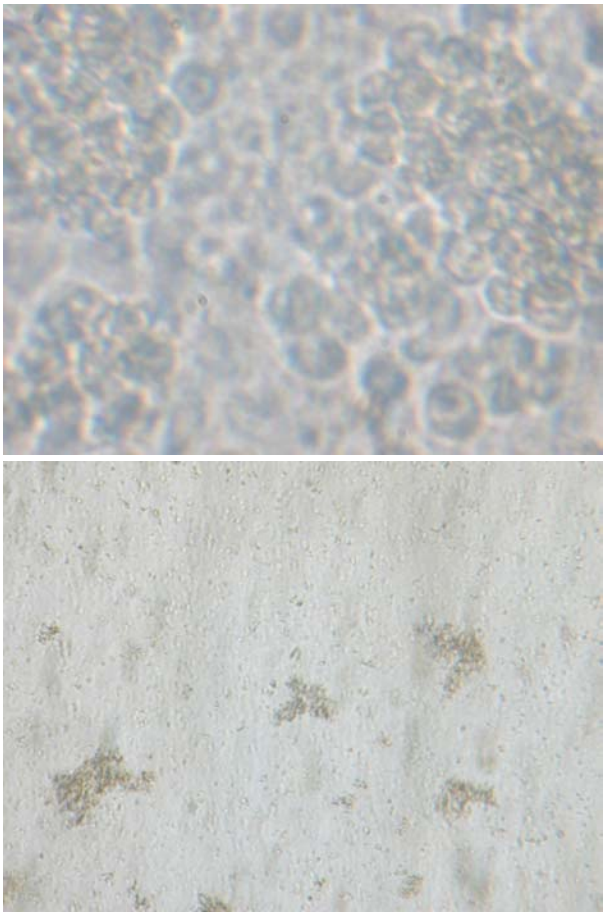

Fig 2. 400× view of amoebae infected with Legionella (above) treated with nanoparticle-filled liposomes, and (below) untreated.

in the event of a terrorist attack with bacteria that have been selected or engineered to be antibiotic resistant. Proof-of-principal therapy for such an organism using in vitro means has been demonstrated. Since the delivery mechanism (stealth, or long-circulating liposomes) has been demonstrated repeatedly in animal models using conventional antibiotics,[1] there is little risk in assuming that they will target inflammation sites carrying different kill materials.

Two of ORNL's key science objectives, nanotechnology and biotechnology, are integral to this project. Observations during the course of work suggest that additional nanoparticles may be engineered using methods similar to those developed during our work. In addition, a substantial number of elements or extremely small molecules may be engineered into nanoparticles for use as additional components in future "kill packages."

A system consisting of delivery, targeting, and a kill package with the kill package capable of eliminating microbial organisms immune to normal therapies should be useful for defense strategies against biological warfare agents as well as organisms found in hospitals that are insensitive to current antibiotic strategies. Given the current fluidity within the federal agencies, the exact target for further developing this concept is not known at this time, but DARPA is a likely candidate, as is the NIH. However, considerably greater quantification of the concept will be required to enable a successful NIH proposal.

References

[1] R. M. Schiffelers, G. Storm, M. T. Ten Kate, L. E. T. Stearne-Cullen, J.G. Den Hollander, H.A. Verbrugh, and I. A. J. M. Bakker-Woudenberg, "*In Vivo* Synergistic Interaction of Liposome-Coencapsulated Gentamicin and Ceftazidime," *J. Pharm. Exp. Therapeut*. **298**(1) 369–375 (2001).

Enhanced Biological Hydrogen Production Using Three-Phase Systems

K. T. Klasson
Life Sciences Division

This work focused on the proof of principle that a three-phase (gas-aqueous-oil) microbial system can yield higher hydrogen production rates than current biohydrogen techniques. The hypothesis to be tested was that improved conversion of CO and water to H_2 and CO_2 via the water-gas shift reaction should be noted in systems where a co-solvent is used. Some results obtained in this study were promising; however, the results were not conclusive, as different results were obtained in different systems.

Introduction

Biological production of hydrogen by microalgae has been known for at least 100 years and other systems have been known for decades. Overall, however, very little progress has been made to make this an applied technology. Current biohydrogen efforts continue to be focused on fundamental research, addressing the shortcomings of biological hydrogen production.

The main approaches to biological hydrogen production are

1. biological water-gas shift reaction,
2. fermentative hydrogen production from organic substrates,
3. photosynthetic hydrogen production from organic substrates, and
4. water-splitting by algae.

Each of these approaches has advantages and disadvantages, and it is clear from the literature that the potential for biological hydrogen has been demonstrated, although the technical feasibility and several key questions remain to be answered.

Of these approaches, the water-gas shift reaction holds the most immediate promise. The water-gas shift reaction ($CO + H_2O \rightarrow H_2 + CO_2$) has been proposed for production of hydrogen from carbon monoxide–containing synthesis gas. Synthesis gas can be produced from a variety of carbon-rich materials such as coal, petroleum products, and renewable biomass. Gasification is an established technology, and the water-gas shift reaction can be carried out by traditional catalysis or by biocatalysts. Limitations for the traditional catalysis are associated with catalyst cost and catalyst poisoning from sulfur present in many low-Btu coals and raw materials. The biological approach does not suffer from this limitation but suffers from rate limitations due to mass transfer, as the reactant is gaseous and reaction is carried out in an aqueous medium. Our proposed work was designed to demonstrate the enhancement of biological hydrogen production in mass transfer–limited systems by the introduction of third phase, silicone oil. The silicone oil would aid in creating new interfacial surface area. This would allow for rapid transfer of CO reactant in the gas phase to bacteria in the aqueous phase.

Technical Approach

Our work used a bacterial strain of *Rhodopseudomonas palustris* P4, kindly provided by Dr. Sunghoon Park, Pusan National University, South Korea. This organism can grow under a variety of conditions (e.g., photosynthetic or in the dark, aerobically or anaerobically, and on a variety of substrates) and can be induced to carry out the shift reaction by switching the growth mode to anaerobic and exposing the cells to CO.[1]

Growth conditions of the culture were established and the conditions for hydrogen production determined through the testing of different media compositions. Once these conditions were determined, cultures were grown and placed in the refrigerator for later use as seed inoculums. The size of these starter cultures varied from 50 to 700 mL.

The first set of experiments was conducted in sealed 160-mL serum bottles and used different quantities of inoculums, providing cell initial concentrations of 0 to 2.0 g dry cell weight/L. The gas phase was 0.14 atm CO in N_2. The conversion of CO was monitored, and the conversion rate and mass transfer properties were determined.

The second set of successful experiments included the addition of two co-solvents, silicone oil or hexadecane, at two different levels. The experiments were conducted in sealed 160-mL serum bottles, and the total liquid volume

and quantities of cells were the same in each experiment. The conditions were chosen so that mass transfer limitations were experienced.

The third set of successful experiments included the addition of three new co-solvents, perfluorocarbons FC-40, FC-72, or FC-77, obtained from 3M. The experiments were conducted in sealed 160-mL serum bottles, and the total liquid volume and quantities of cells were the same in each experiment. The conditions were chosen so that mass transfer limitations were experienced.

Experiments were also conducted in a larger-scale stirred batch fermenter. The total volume of liquid in this system was 1000 mL, of which 30% was either one of two solvents (silicone oil or FC-40) or just additional water (in the case of the control experiment). A continuous stream of gas (47% CO, 10% CO_2, 36% H_2, 2% N_2, and 5% CH_4) was purged through the system, and the conversion of CO was determined from the analysis of the inlet and exit gas streams. The agitation was kept constant until a constant conversion was noted and then increased to investigate if the system was mass transfer limited. (An increased conversion when agitation is increased suggests that mass transfer limiting conditions exist at the lower agitation rates.)

Results and Accomplishments

The conversion of CO in the first set of experiments is shown in Fig. 1, and the calculated mass transfer rate has been plotted in Fig. 2 as a function of the partial pressure in the gas phase of the serum bottles. From the results in Fig. 1, it is noted that CO conversion was not noted in the experiment without cells. From the results in Fig. 2 it can be deduced that the transport of CO from the gas (to the liquid and cells) became proportional to the partial pressure of CO at approximately 0.05 atm CO. It should also be noted in Fig. 2 that the rate/partial pressure relationship was the same for cell concentrations above

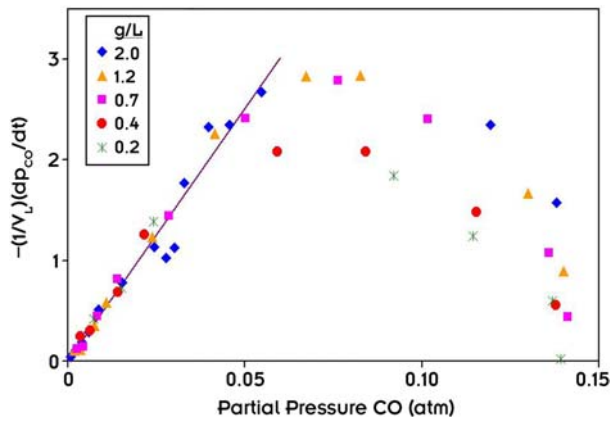

Fig. 2. CO transport rate as a function of CO partial pressure in experiment with different cell concentrations. The rate of consumption was calculated by numerical methods and curve fitting.[3]

0.4 g/L. The slope of the straight line at low partial pressures is proportional to the mass transfer coefficient.[2] Thus, it was concluded that mass transfer conditions in these types of serum bottle experiments should be occurring at cell concentrations of at least 0.7 g/L and at partial CO pressures below 0.05 atm. Experiments in serum bottles were conducted so that these conditions were met by using approximately 0.7 g dry cell weight/L and an initial CO partial pressure of 0.08 atm.

The conversion of CO in the absence and presence of different quantities of either silicone oil or hexadecane is shown in Fig. 3. From the results presented in Fig. 3 it is noted that the presence of either oil did not affect the rate of CO conversion; that is, the curves fall on top of each other for all cases. There was a shift in the CO conversion profile in the experiment with 30 volume % hexadecane, when after one sampling event this serum bottle was not returned to its agitated condition for approximately 30 min. Overall, this experiment proved our hypothesis wrong.

The conversion of CO in the absence and presence of 10 volume % of three perfluorocarbons is shown in Fig. 4.

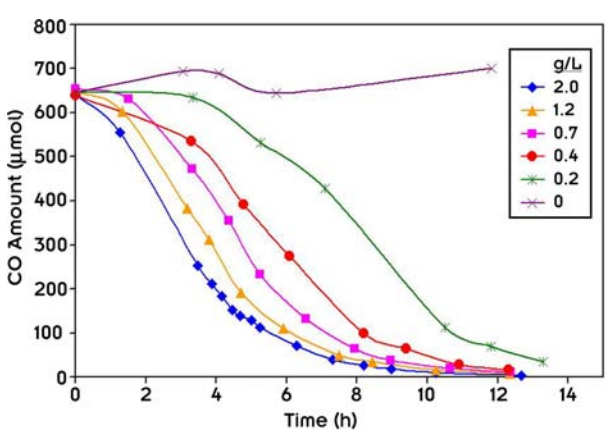

Fig. 1. CO concentration profiles in experiments with different cell concentrations.

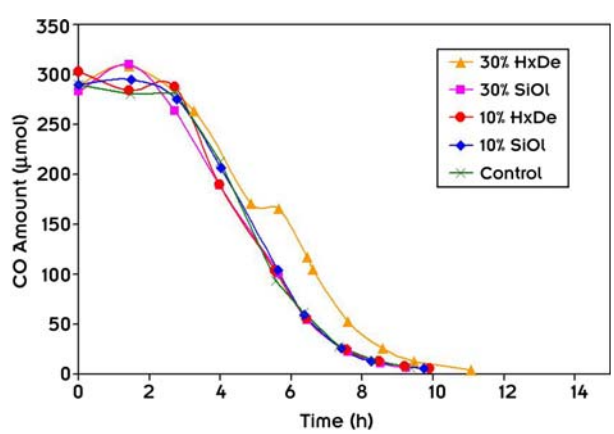

Fig. 3. CO conversion profiles in the absence and presence of either silcone oil (SiOl) or hexadecane (HxDe).

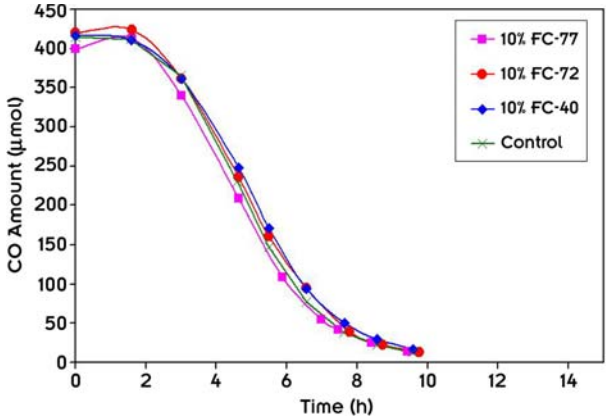

Fig. 4. CO conversion profiles in the absence and presence of different perfluorocarbons.

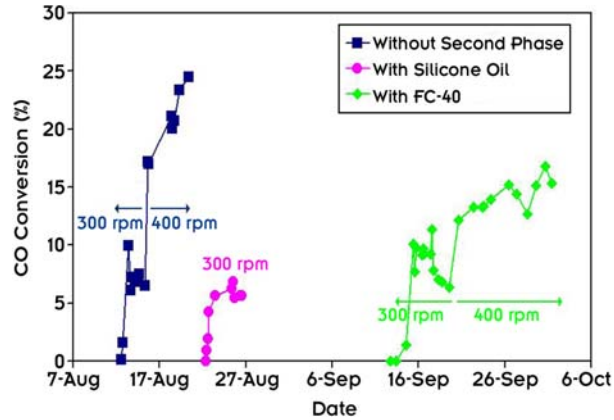

Fig. 5. CO conversion in a stirred batch fermenter without co-solvent (■), with silicone as a co-solvent (●), and with perfluorocarbon FC-40 as a co-solvent (◆).

From the results presented in Fig. 4 it is noted that the presence of either oil did not affect the rate of CO conversion; that is, the curves fall on top of each other for all cases. Overall, this experiment proved our hypothesis wrong. The rationale for investigating perfluorocarbons was that they have a density higher that water, while silicone oil and hexadecane have densities less than that water.

While all the small-scale experiments proved our hypothesis incorrect, it was also noted that the agitation and mixing of the organic and aqueous phases was less than complete. In order to achieve a more complete mixing, three experiments were conducted in a 1-L fermenter with mechanical agitation. The results are shown in Fig. 5. The data to the left correspond to an experiment without any co-solvent present. After several days of fermentation, the CO conversion rate had stabilized at approximately 7%. The agitation was then increased from 300 to 400 rpm and the conversion increased. This suggested that the conversion rate at 300 rpm was limited by mass transfer. The data in the center of Fig. 5 correspond to an experiment where 30% silicone oil was present and the agitation was 300 rpm. Again the CO conversion stabilized around 6 to 7%, indicating that silicone oil did not increase the CO conversion. The data to the right in Fig. 5 correspond to an experiment where 30% of perfluorocarbon FC-40 was present and the agitation was 300 rpm. The CO conversion this time was 9–10% for several days before dropping to around 7%. At this point the agitation was increased to 400 rpm to make sure the culture was not loosing its CO converting abilities. The conversion increased, indicating the culture was still performing its function. The large-scale fermenter data suggest that FC-40 addition had a positive impact on the CO conversion, indicating that the mass transfer was higher in this case and resulted in increased conversion of CO. It is, however, disturbing that the conversion dropped off in a manner which could not be explained.

Summary and Conclusions

Several experiments were performed as part of this investigation. Most of these experiments failed to prove the hypothesis that co-solvents will enhance the mass transfer in biological CO conversion systems. One experiment proved the hypothesis; however, not conclusively. As part of this project, several collaborative contacts have been made which may result in biohydrogen proposal writing activities.

References

[1]T. H. Lee, J.-Y. Park, and S. Park, "Growth of *Rhodopseudomonas palustris* Under Phototrophic and Non-phototrophic Conditions and Its CO-dependent H_2 Production," *Biotechnol. Lett.* **24**, 91–96 (2002).

[2]J. L. Vega, E. C. Clausen, and J. L. Gaddy, "Study of Gaseous Substrate Fermentations—Carbon-Monoxide Conversion to Acetate. 1. Batch Culture," *Biotechnol. Bioeng.* **34**, 774–784 (1989).

[3]K. T. Klasson, "Experimental Data Analysis: An Algorithm for Smoothing of Data and Determining Enzyme and Microbial Kinetic Rates," *Appl. Biochem. Biotechnol.* **63/65**, 339–348 (1997).

Robust Segmentation of Telomeres in Metaphase Fluorescence In Situ Hybridization Images

T. P. Karnowski and J. R. Price
Engineering Science and Technology Division

In this work we explored the use of graphical models to develop a robust segmentation and labeling algorithm for automated, high-throughput telomere quantification in fluorescence in-situ hybridization (FISH) imagery. We first segmented a set of test data and compiled statistics. We developed an independence diagram for the data. Finally, we compiled test results using an algorithm based on the independence diagram.

Introduction

Telomeres, which are the end sections of linear chromosomes, are vital to proper cell maintenance and play a role in cancer and aging. Fluorescence in situ hybridization (FISH) is a procedure used to image telomeres and chromosomes through the use of fluorescent probes that are bound (hybridized) to DNA. Generally, through the use of different fluorescent probes, the telomeres will be imaged by one color channel (e.g., red) and the chromosomes by another (e.g., blue). Scientists analyze these FISH images to examine telomere and/or chromosome characteristics to answer various biological questions. Figure 1 illustrates the basic idea of FISH. Figure 2 shows a composite image where the telomere probes are shown as red and the chromosome as blue.

While some semi-automated software exists to identify and label telomere and chromosome structures, this software suffers from user-interface issues that force scientists to discard significant amounts of experimental data. Furthermore, the image processing algorithms of the existing software are based upon heuristic techniques that do not properly exploit the known structural relationships between the telomeres and chromosomes. In 2002 we implemented a graphical user interface (GUI) for processing telomere/chromosome image pairs. For expediency, this software used only heuristic image processing algorithms. The software has been well received and noted as an improvement by prominent telomere researchers, but the currently employed, simplistic segmentation approach limits broad applicability, hence the need for this project.

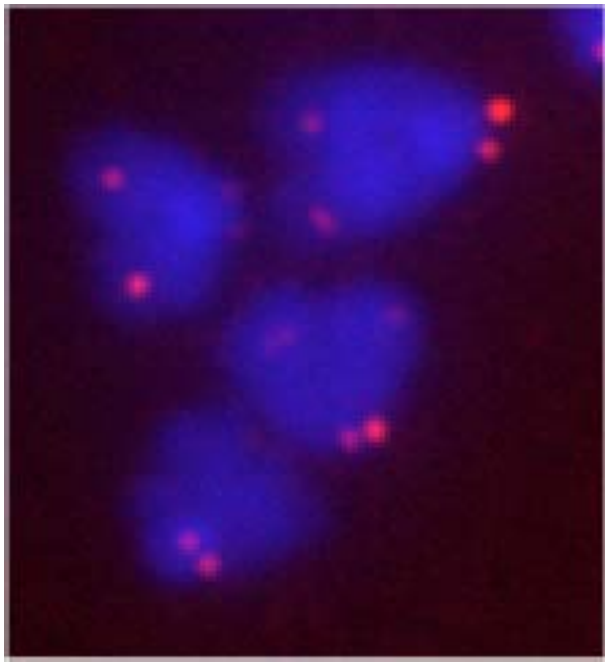

Fig. 2. Chromosomes and telomeres. The chromosomes are blue; the telomeres are red. There are four telomeres per chromosome.

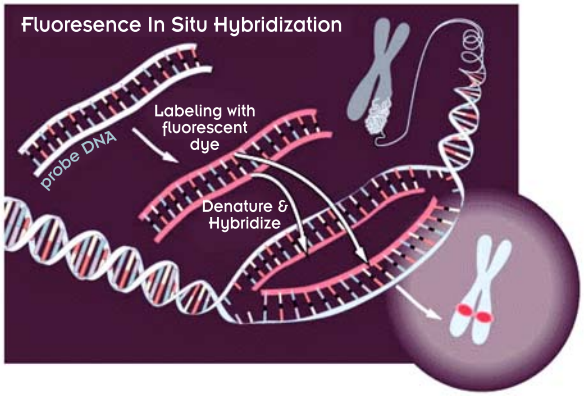

Fig. 1. Fluorescence in situ hybridization concept. (Image from http://www.accessexcellence.org/AB/GG/fish.html.)

Technical Approach

Our technical approach consisted of three main tasks.

Segment Test Data and Compile Statistics

We first segmented existing test data to determine important Bayesian statistics and relationships between pixel levels for both telomere and chromosome pixels of images.

Create Independence Diagram Models and Solution Methods

We next used the statistics compiled as well as our own knowledge of the chromosome and telomere relationships to create a basic independence diagram. Our survey of the literature in this field revealed that independence diagrams have been used in image processing for a variety of purposes, but their application is markedly different from that suggested by the artificial intelligence field.

Compile Test Results

We then applied our new algorithms and statistics to the test images and determined their effectiveness compared with the "ground truth" results of manual and assisted labeling.

Results and Accomplishments

Segment Test Data and Compile Statistics

The existing ORNL FISH image-processing tool was used to segment a set of seven sample telomere and chromosome image pairs. The tool processes the images through a sequence of heuristic algorithms that include image processing operations such as labeling, dilation, and linear filtering. MATLAB programs were written to take the test results and compile statistics, including histograms, to model conditional probabilities. The telomere and chromosome images are made using different staining compounds and illumination sources. As a result, one image causes only telomere material to illuminate and the other image only shows chromosomes. We can combine the two together effectively for display purposes by setting the red plane of an image to the telomere and the blue plane to the chromosome. We compiled histograms of the R,B values to model the probabilities that are required. The histograms are computed on each image and then they are all added together to get an approximation to the probability density function required. We also performed some interpolation between histogram bins, however, to handle cases where a particular R,B pair may exhibit zero probability due to the limited training data. This was done by convolving the histogram with an interpolating low-pass filter.

Create Independence Diagram Models and Solution Methods.

As the project progressed, we realized a very complex independence diagram could be constructed for the telomere—chromosome segmentation problem. The only really observable values are the individual pixel intensities of the telomere and chromosome, but the nature of the structures of the telomeres and chromosomes indicate clique statistics would also be extremely valuable. Unfortunately, the limited scope of this project prevented extensive study of this intriguing possibility, but a greatly simplified independence diagram as shown in Fig. 3 was developed. Even this simple independence diagram is difficult to use, however; it is a multiple connected graph due to the "loop" formed between the telomere and chromosome nodes with the (Red, Blue) value node. We simplified the diagram even further by using a method called "clustering", where we combine the loop to form the diagram in Fig. 4.

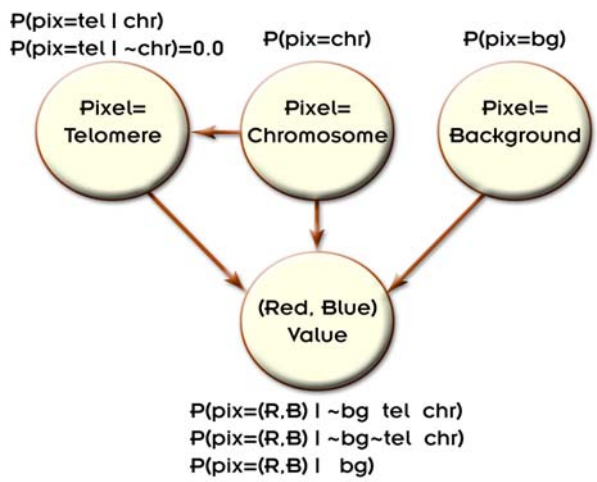

Fig. 3. Original simplified independence diagram with loop.

Design and Conceptualize Algorithms and Hardware for High Throughput

Our segmentation was simple. We merely take each pixel R,B value and determine which is most likely: background, telomere, or chromosome. However, we performed our testing with a hold-one-out procedure. In this process we take one of the seven telomere/chromosome training image pairs, "hold it out," then compile the sample probability functions on the remaining six images. We then label the "held out" image and compare its results to the manual labeled case. Some subsequent image processing is performed to remove noise, but the objective is to keep this to a minimum.

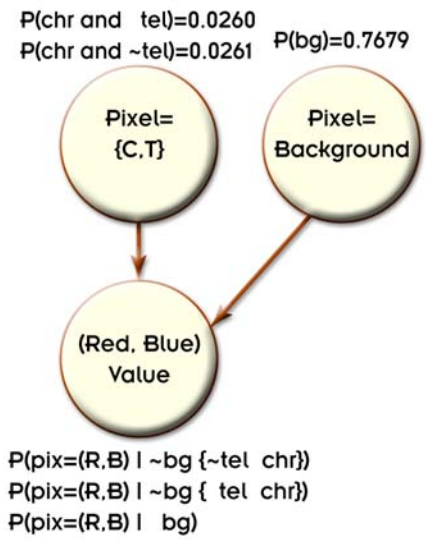

Fig. 4. Revised independence diagram with loop replaced by clustered node.

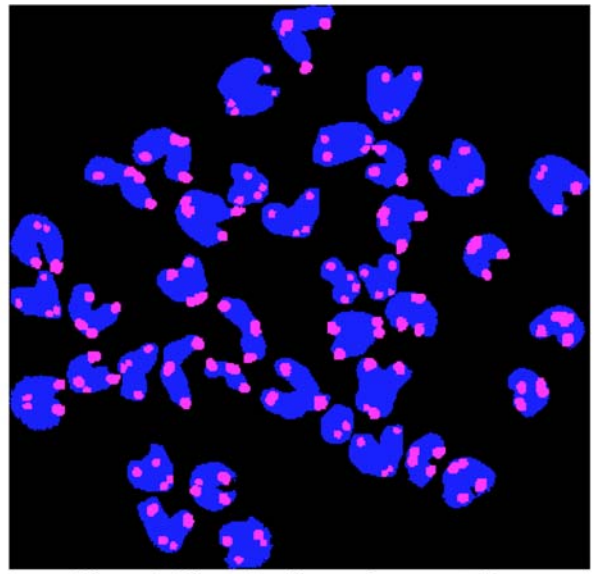

Fig. 5. Example of manual segmentation.

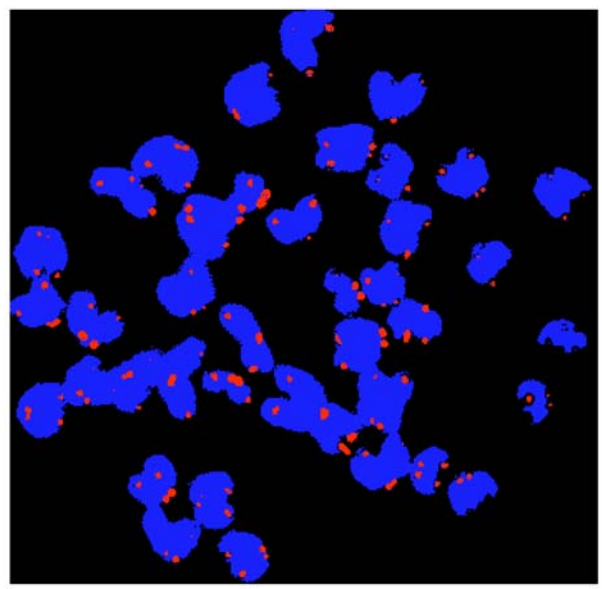

Fig. 6. Example of automatic segmentation. Many entities are joined; some telomeres are not found; and generally the shapes are not as large as the manual segmentation.

Summary and Conclusions

An image pair and example segmentation is shown in Figs. 5 and 6 after simple probabilistic processing. In lieu of more sophisticated processing, we simply removed all blobs from the chromosome and telomere-labeled images that had very small areas (less than six pixels). In Table 1 we show the results of processing. The "Num Chr" and "Num Tel" indicate the number of chromosomes and telomeres that were labeled in the original images. The "Found…" columns indicate the number of these labeled blobs that were intersected by a blob from the segmentation process. We then show the percent of manually labeled chromosomes and telomeres. Finally, we also compare the total area of the chromosomes and telomeres.

Our analysis revealed some advantages and some problems. First, locating chromosomes is rather successful, but finding telomeres is slightly less so, particularly for the first two and the eighth image. The failure to consistently find the telomeres may indicate a sensitivity to the parameters on the FISH process; in other words, images 1, 2, and 8 may have had different image acquisition or chemical processing conditions than the other images, making them somewhat unrepresentative of the true statistics of the image distributions. Secondly, we seem to consistently under-size the chromosomes and telomeres. Finally, our labeling does not split chromosomes or telomeres that are joined together.

These problems may be addressed through heuristic approaches or through more adequate models of the Bayesian relationships between telomeres and chromosomes. The expected area of the chromosomes and telomeres, for example, could help to separate the structures, as can the spatial distribution of the intensity levels within a "joined chromosome" or "joined telomere" region. Clique-based models, where we look at adjacent pixels in bounded areas, should improve these problems.

Our work will further two ultimate goals: (1) to disseminate ORNL image processing expertise, through software distribution and conference publication, to

Table 1. Results of segmentation algorithm

Image	Num Chr	Num Tel	Found Chr	Found Tel	% Chr	% Tel	Area Chr	Area Tel	Found area Chr	Found area Tel	% Chr area	% Tel area
1	39	155	39	100	100	64.51	45995	5199	15227	3520	33	68
2	39	140	39	103	100	73.57	63859	5230	21963	3291	34	63
3	41	161	41	161	100	100	58192	6972	11786	1783	20	26
4	41	164	41	164	100	100	58079	5925	12512	809	21	14
5	40	160	40	160	100	100	52293	6645	11650	1030	22	16
6	40	160	40	157	100	98.12	68486	6911	9434	1370	14	20
8	40	157	40	103	100	65.6051	26671	4991	5239	3736	20	75

scientists interested in analyzing intracellular structures and (2) to generate preliminary data and publications that can be used to support a planned NIH grant application (PA-03-106 Innovations in Biomedical Computational Science and Technology) in 2004. For the NIH grant application , we intend to address automated, high-throughput telomere analysis of interphase cell images. This project dealt only with metaphase cells. Although telomeres in metaphase cells are easier to quantify since they are usually well separated, the preparation and image acquisition is very difficult. Interphase cells, on the other hand, are much easier to prepare and acquire, and therefore provide a more tractable approach for the high-throughput processing that will be required for drug-screening and robust statistical studies. We believe that the planned follow-on effort will be of great value to the pharmaceutical industry for drug-screening applications.