LA-UR- 09-02323

Approved for public release;
distribution is unlimited.

Title:	LANL Capabilities Towards Bioenergy and Biofuels Programs
Author(s):	Jose Olivares Cliff Unkefer Andrew Bradbury Geoff Waldo Min Park
Intended for:	Multiple Meetings

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

LANL Capabilities

Towards Bioenergy and Biofuels Programs

Algal Production and Harvesting

LANL invented technology for increasing growth and productivity of photosysnthetic organisms, including algae and higher plants. The technology has been extensively tested at the greenhouse and field scale for crop plants. Initial bioreactor testing of its efficacy on algal growth has shown promising results. It increases algal growth rates even under optimum nutrient supply and careful pH control with CO_2 continuously available. The technology uses a small organic molecule, applied to the plant surfaces or added to the algal growth medium.

CO_2 concentration is necessary to optimize algal production in either ponds or reactors. LANL has successfully designed, built and demonstrated an effective, efficient technology using DOE funding. Such a system would be very valuable for capitalizing on local inexpensive sources of CO_2 for algal production operations. Furthermore, our protein engineering team has a concept to produce highly stable carbonic anhydyrase (CA) enzyme, which could be very useful to assure maximum utilization of the CO_2 supply. Stable CA could be used either immobilized on solid supports or engineered into the algal strain.

The current technologies for harvesting the algae and obtaining the lipids do not meet the needs for rapid, low cost separations for high volumes of material. LANL has obtained proof of concept for the high volume flowing stream concentration of algae, algal lysis and separation of the lipid, protein and water fractions, using acoustic platforms.

Organocatalysis and Synthesis Discovery

This capability is targeted toward developing biosynthetics, chiral syntheses, high throughput protein expression and purification, organic chemistry, recognition ligands, and stable isotopes geared toward Bioenergy applications. Areas of expertise include stable isotope chemistry, biomaterials, polymers, biopolymers, organocatalysis, advanced characterization methods, and chemistry of model compounds. The ultimate realization of the ability to design and synthesize materials that mimic or are inspired by natural systems will lead to entirely new applications in the bioenergy areas. In addition, there are new developments in this capability that involve development of catalytic methods for the production of carbon chains from the most abundant carbohydrate on the planet, glucose. These carbon chains will be useful in the production of high density fuels which defined characteristics. In addition, these methods/capabilities will be used to generate feedstocks for industrial processes.

Home to the LANL Component of the DOE Joint Genome Institute

LANL is the second largest partner institution of the Department of Energy's Joint Genome Institute (DOE-JGI), and specializes in high throughput genome finishing and analysis in support of DOE missions in energy, bioremediation and carbon sequestration. This group is comprised of molecular biology labs and computational staff who together

focus on the high-throughput DNA sequencing of whole microbial genomes, computational finishing and bioinformatics. The applications team focuses on the use of new sequencing technologies to address questions in environmental science. In addition to supporting the DOE mission, this group supports the Nation's national security mission by sequencing critical pathogens and near neighbors in support of relevent application areas.

Protein Structure and Enzyme Engineering

As one of the founding members of the NIH funded Structural Genomics Center and Integrated Center for Structure and Function Innovation, Los Alamos has proven track record in protein engineering, structure determination, and software development. It also has strong programs in ligand generation and applications to identify protein structures, protein localization, protein-protein interactions, and protein solubility. These tools will be essential elements in engineering proteins for biofuel production and for understanding the biology of biofuel production.

LANL also has a strong capability to rapidly characterize and evolve enzyme catalysts with improved catalytic properties and increased thermo stability. This is essential to the realization of bioprocessing of biomass into renewable energy. These capabilities include high-throughput cloning, expression, protein purification, structure determination, characterization of enzyme function and high throughput fluorescent screening. In addition, we have the demonstrated capability to evolve enzymes using directed evolution and structure-based rational methods.

Neutron Scattering at Lujan Center

Neutron scattering is an invaluable probe of bio-related materials for fundamental studies of structure and dynamics. In a user program access format, neutron scattering is particularly powerful for hydrogenous materials wherein deuterium substitution can be effected. The Lujan Center also has user labs for chemical and biological syntheses and analyses as well as a range of x-ray scattering instruments. These tools are being applied to the understanding of cellulase activity in the breakdown of cellulose.

Multi-platform Microscopy

The microscopy capabilities within the Integrated Spectroscopy Laboratory within Chemistry division houses a variety of microscopic techniques which can be used to study the structure of bioenergy samples and well as how these structures are altered either during or after treatment, where treatment can involve fungal, chemical, or enzymatic exposure. Example platforms include structural analysis (Optical and Scanning electron microscopes), elemental analysis (XRF and SEM), and chemical profiles (Raman, IR, fluorescence). Each of these techniques is necessary because no single microscope platform is adequate to enable a complete breakdown of the structural and chemical components present in a given sample. The unique aspect of this work is the determination of methods which enable a single sample to be measured across multiple microscopic platforms and that the resulting data can be fused to form a unified view of the system.

Stable-Isotope Capability-based Metabolomics and Proteomics

Due to our Stable Isotope Resource, Los Alamos is the one of the world leaders in stable isotope design, synthesis, and production. This capability has become the foundation of our strong capability in metabolomics; pathway regulation, flux analysis, carbon fate mapping and new language development. By combining our capability in various advanced mass spectrometry and NMR technologies, our metabolomics capability currently focuses on plant metabolism with the focus of cellulosic biomass conversion and algal biofuels. Our mass spectrometry based capability is broadly applied to protein analysis from a standpoint of measuring expression levels, post-translational modifications and structural parameters. Proteomics is a combination of multiple technical approaches and is an important member of any OMICS strategy. This capability is critical to Bioenergy research from the standpoint of analyzing catalytic enzymes and metabolic pathways in micro- or higher order organisms involved in biomass production, conversion and/or storage.

Theoretical Biology & Biophysics

LANL's theoretical groups focus on the modeling of biological systems and the analysis and informatics of molecular and cellular biological data. We house one of the few research groups in the world devoted to mathematical modeling and computational analysis of problems in cellular and molecular biology. Research efforts include understanding dynamics and treatment of viral diseases such as HIV, influenza, and hepatitis; immune system modeling; receptor-ligand interactions and cell signaling; computational aspects of the human genome initiative; pattern recognition in DNA sequences; characterization and prediction of macromolecular structure; protein function and dynamics; and protein folding. LANL has created and is also responsible for the maintenance of the HIV and Influenza Sequence Databases as well as the HIV Immunology Database and the HIV Resistance Database. The HIV/HCV database group at LANL has an excellent worldwide reputation. The group does cutting-edge computational biology research, develops tools, and maintains several databases and an extensive and widely used website. The funding for its sister HCV database project was discontinued and transferred to the Viral Bioinformatics Resource Center (VBRC), and NIH-funded web, database and tool development, one of seven Bioinformatics Resource Centers. Our experience with project participants in creating, maintaining and providing access to innovative and user-friendly tools and databases in the realm of viral genetics and immunology will generalize fairly easily to bioinformatics needs for bioenergy applications.

Energy and Infrastructure Analysis Group

Our mission is to perform basic and applied research focused on increasing the security of the nation's energy infrastructure. We specialize in the development and use of macromodels and microsimulations. These provide detailed, quantified information on the physical, operational, and economic behavior of energy networks, infrastructures, and industries. Our expertise covers the generation, transmission, and distribution of electric power, natural gas, oil, and coal, as well as other, non-energy systems critical to energy security. We ensure the scientific and technological validity of our models by working closely with physicists, engineers, mathematicians, statisticians, computer scientists, and

economists and by employing interdependency, optimization, and risk-assessment frameworks.

Bio-Chemical Reaction Engineering

Reaction Engineering capabilities enable designs to maximize photon flux and control the selectivity during chemical conversion processes through management of thermal profiles. Operational strategies can be elucidated using these techniques to enhance conversion of reactant to product (e.g. by periodic forcing of the reactor), to develop optimal strategies for feed (nutrient) injection, and to make best use of recycle streams. LANL capabilities range from the classical techniques (models for kinetics of complex reaction networks, residence time distribution analysis) to high resolution, spatially resolved interrogation of specific reactor geometries (computation approaches to multi-phase flow with simultaneous chemical reaction and heat transport).

Environmental and Resource Management/Assessment

The environmental management system aims to provide environmental protection strategies, technologies and processes to be adopted for management and assessment of environmental resources utilized and monitoring of effects of processes on the local environment. This resource can support the environmental and other impacts that a bioenergy program can induce.

Distributed Sensor Networks

Los Alamos National Laboratory (LANL) has been developing techniques for distributed sensor networks with collective computation (DSN-CC). DSN-CC represents an innovative approach to in situ sensing. DSN-CCs are networks of sensors capable of intercommunicating and executing complex computations in the field without central control (Figure 2). This method of computation and communication has demonstrated clear advantages over traditional approaches where sensors simply collect and transmit data out of the network.

Unlike a traditional array, there is no central (and vulnerable) control point to which all data must be passed. In the DSN-CC, information and conclusions exist throughout the network. This avoids several problems encountered by the traditional approach. Classically, raw data is transmitted for long distances, requiring large amounts of power and bandwidth. Central processing incurs delays and leaves the system vulnerable to single-point failures. In contrast, DSN-CCs operate only with short-range transmission and in situ data processing. This approach saves communication bandwidth, provides redundancy, eliminates single-point failures, and delivers conclusions rapidly to users.

DSN-CC nodes communicate with neighboring nodes in order to exchange measurements and cooperatively solve a sensing problem or to cue a different sensor. A concise conclusion or cue request is rapidly propagated across the network. Users of the network can obtain this information by listening to any portion of the network. This concept of in situ computation has seen little exploration, but allows for capabilities not possible with traditional sensor arrays due to reliability issues, prohibitive cost, scaling of power, computing, and bandwidth usage. Small, low power sensors with such functionality are now emerging from companies such as Crossbow Technology.

Efforts at LANL are focusing on developing DSN-CC systems in both simulation as well as in hardware using commercial off-the-shelf platforms. To this end, LANL has developed an open source simulation engine for base-lining application-specific networks. Upgrades to the simulator are ongoing as hardware platforms evolve and emerge. Additionally, by focusing on particular applications LANL's DSN-CC team has been able to move from simulations to hardware systems. The ongoing goal of this project continues to focus on the demonstration of in situ collective computation abilities of networks using inexpensive, readily available off-the-shelf technology and hardware.

Protein engineering capabilities at LANL

Protein Engineering and enzyme thermostabilization

LANL has developed a number of different methods to stabilize proteins. One method involves using polypeptide fusions (internal or terminal) that interfere with the folding of a protein, and then evolving the protein to accumulate compensatory mutations so that it is able to overcome the folding interference. Finally these so-called folding interference domains are removed. LANL has two publications showing that proteins engineered this way exhibit improved folding properties that can include resistance to chemical and thermal denaturation, faster folding, and resistance to mutations that otherwise destroy folding: i.e. the evolved proteins are resistant to the effects of increased mutational pressure. These methods are easily applied to proteins with an easily screened or selectable phenotype. We have also developed methods that allow selection for improved thermostability and folding, that are independent of activity assays. These methods rely on the use of "folding reporters": proteins with specific easily screened functions, such as fluorescence. Fused protein domains that misfold result in misfolding of the folding reporter and consequent lack of function, while correct folding results in folding reporter activity. This high-throughput function-independent approach can increase throughput during selection for improved folding, when activity assays are costly or low-throughput.

LANL has also developed a suite of split GFP tags. Here, a small 15 amino acid fragment of GFP is attached to the protein of interest, then detected by the remaining GFP fragment. The small fragment of the GFP has been engineered to not interfere with fusion protein folding. This system can be used to accurately measure soluble expression and protein concentration in living cells, or in-vitro assays. Since the small 15 amino acid tag is non-aggregating, this tagging system can be readily used to quantify the precise amount of soluble enzyme in activity assay, allowing normalization of measured enzyme activity, for example, to yield specific activity. Enzyme variants that retain more specific activity after exposure to heat denaturation and subsequent cooling can be identified, providing an alternative approach to engineering improved protein stability. These methods can be used in conjunction with increased selective pressure for correct folding, using the folding interference domain approach above.

Finally, LANL has extensive automation facilities, including two high throughput picking robots, and an integrated system comprising a highly-flexible incubator farm, fluorescence plate reader, automated centrifuge, and liquid handling system (Beckman). This system can be used to perform evolution, quantify protein stability and folding yield using activity assays and the split GFP and folding reporter platforms, as well as using fluorescent tagging to quantify protein in in-vitro protein-specific activity assays.

Phage display and affinity reagent generation

LANL has extensive experience in the selection of antibodies from phage display libraries, and has selected antibodies with unique specificities, including one able to recognize the sulfotyrosine post-translational modification independently of sequence context and another able to differentiate between two proteins differing by a single amino

acid. We are in the process of developing novel affinity reagents based on fluorescent proteins that will provide intrinsic, direct and real time read out of affinity reagent binding. LANL has experience in the use of both phage and yeast display, for the purposes of both selection from naïve libraries, as well as for affinity improvement of selected clones.

Assay development and mechanistic enzymology

A crucial component to any protein engineering efforts is the ability to devise and implement functional assays to ascertain the kinetic competency and substrate tolerance of native or engineered enzymes. In turn, information obtained from analysis of the chemical mechanisms that underlie enzyme catalysis may be used to direct rational or semi-rational engineering approaches. To this end, LANL researchers have an extensive expertise in the design of enzyme activity assays/product characterization for a number of enzyme systems, particularly those that have their origins in microbial secondary metabolism. This includes, but is not limited to, enzymes from metabolic pathways such as isoprenoid biosynthesis, polyketide synthases, chorismate pathway enzymes, and cellulases. Furthermore, we are capable of chemically synthesizing a variety of small molecules to directly test for enzyme activity, or to probe catalytic mechanisms. LANL researchers have conducted detailed mechanistic analyses and inhibitor binding studies of methylerythritol phosphate pathway enzymes, which required the synthesis of isoprenoid precursors (MEP, DXP, DXP analogues, CDP-ME, HDMAPP, and additional longer chain branched, unsaturated hydrocarbons). This work has led to multiple publications in highly respected peer-reviewed journals (Journal of Organic Chemistry, Biochemistry). In addition, our synthetic schemes were designed for facile incorporation of isotopes, which is important for downstream product identification and characterization.

Enzyme activity assays and bioengineering

LANL's abilities in mechanistic enzymology and assay design form the foundation which enables us to exploit and the basis to engineer the intrinsic catalytic properties of a targeted enzyme for use in a specific process. Our team is devoted to developing 'green' technologies for conversion of non-feedstock material to simple sugars for subsequent fermentation to bioethanol. Quite simply, we are attempting to harness Nature's strategy for degrading recalcitrant lignin by enzymatic means in order to access stored energy in the form of cellulose (polymeric glucose) and hemicellulose (polymeric xylose). Enzyme classes being examined include; lignin peroxidases, manganese peroxidases, copper oxidases, glycosidases, xylanases, and other related sugar processing enzymes. We are using directed evolution to engineer glycosidases with novel catalytic properties. We plan to engineer microbes for production of commodities, pseudo-commodities, and alternative biofuels, focusing on enzymes involved in isoprenoid biosynthesis and aromatic amino acid degradation. In order to minimize environmental impact, reduce production costs, and maximize natural non-depleting resources, biocatalysts tailored to perform a specific function are being aggressively pursued.

High throughput analysis of the metabolome in Algae

Overall approach - Metabolome analysis has the advantage that most organisms have fewer metabolites than genes or gene products. However, the analysis of all of the metabolites in an organism remains a difficult analytical challenge that requires a multiple techniques. At Los Alamos metabolome analysis starts with controlled chemostat growth and an automated robotic sampling so that identical replicate samples can be obtained. For studying algae, ion chromatography is used to determine inter- and extracellular concentration of ions like ammonium, nitrate, and bicarbonate. Non-volatile metabolites are analyzed using liquid chromatography-nano electrospray-mass spectrometry using an LTQ-FT mass spectrometer. Separate optimized chromatography systems (columns and solvents) are used to examine positively and negatively charged metabolites. The Advion Nanomante ion sources splits the sample and collects fractions that can be reanalyzed with signal averaging and allow MS/MS experiments for structure identification. Finally, volatile compounds are analyzed using two-dimensional gas chromatography separation and TOF mass spectrometric detection using a Leco Pegasus IV GCGC TOF. The GC/MS analysis provides both qualitative and quantitative analysis of lipids and isoprene hydrocarbon, which is essential to optimize algal production. Discussed below our metabolomics capability uses stable isotope labeling to provide internal standards for quantitation, metabolic flux analysis and for metabolic pathway discovery by MS and NMR isotopomer analysis. Our capabilities are described in greater detail below.

Isotope Labeling - A major advantage of carrying out this work at Los Alamos National Laboratory is the availability of ^{13}C-, ^{2}H-, and/or ^{15}N-labeled compounds. This project will leverage the National Stable Isotope Resource at Los Alamos (C.J. Unkefer, Director). The Resource has exceptional expertise in the synthesis and analysis of biomolecules labeled with stable isotopes. Because we can synthesize virtually any metabolite with multiple specific ^{2}H, ^{13}C and/or ^{15}N labels, we have the unique ability to trace the fate of carbon-carbon, carbon-hydrogen or carbon-nitrogen bonds through their transformations on a scale appropriate to metabolomic characterization. We have considerable experience analyzing metabolic pathways using stable isotope labeling and NMR spectroscopy, or GC/LC-Mass spectrometry.

Chemostat Growth - Essential to comparative metabolomics, metabolic flux analysis and metabolic regulation studies are steady-state culture conditions. We outfitted four chemostats based on existing New Brunswick Bioflow III fermentors. Briefly, the chemostats have 1 liter, 2 liter and 5 liter vessels. By varying both the vessel size and pumping rate, we can achieve a wide range of dilution rates. We have altered the chemostat to use mass flow controllers to tightly regulate the rate of the addition of three or four gasses including O_2, N_2, CO_2 and NH_3. This allows the atmosphere in the chemostat to be controlled for aerobic or anaerobic growth of heterotrophic or autotrophic growth. The addition of the gasses is regulated and monitored by computer controlled mass flow controllers, which allows us to vary the CO_2 concentration over a wide range. In addition, the chemostat monitors and adjusts temperature, pH, [O_2], stirring rate, and the rate of nutrient addition. The rate of nutrient addition is monitored by the loss of weight in the reservoir as measured using a digital balance. For growth of algae, lights have been added and light intensity and day length can be controlled and measured. All of these parameters are set, monitored and logged using a LabVIEW-based program developed in our laboratory. A screen shot of the chemostat control page is shown in **Figure 1**. As depicted in the upper right corner of **Figure 1**, we have interfaced a LEAP technologies robot to automatically sample the chemostat and store 1 ml samples at 4°C for analysis. We have

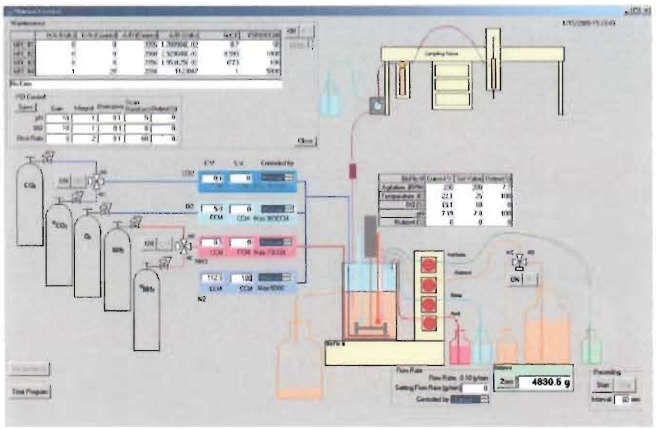

Figure 1) A screen shot of our chemostat control software. The chemostat uses mass flow controllers to regulate the flow of oxygen, ammonia and carbon dioxide for the growth of *N. europaea*.

used these samples to determine the optical density of the culture and monitor nitrate or ammonia in the culture medium. All of the chemostat data are logged and then displayed and analyzed using Excel. Other features of the chemostat control software are: the operator is notified via a digital pager when parameters are out of specification and the chemostat parameters are monitored and set remotely.

Automated chemostat sampling and metabolic quenching – Like steady-state growth, reproducible chemostat sampling and metabolic quenching are essential to comparative metabolomics, metabolic flux analysis and metabolic regulation studies. We developed and have now automated a method for removing samples from the chemostat by separating the cells from the medium and quenching the metabolism.

After rapid removal of a sample from the chemostat, our method collects the cells on a 2-micron filter; the growth medium is collected as filtrate. The sample's metabolism is quenched and the metabolites are released from the cells by treating the cells with acetonitrile/water solutions on the filter. After a short incubation time, the metabolite solution is separated from precipitated protein and cell debris by filtration. The filtrate is dried and prepared for analysis by mass spectrometry. We have automated this process using a TECON robot and a custom plate designed to hold the filter. To minimize metabolic activity during the quenching protocol, the tubing use to move the sample and the filter holder are cooled to 4^0C (**Fig. 2**).

Figure 2) Our customized sampling and metabolic quenching robot. Using a syringe pump, the robot removes culture solutions from the chemostat and recovers the cells by flitration through a 0.2 μ glass filter (white disks). Then the robot moves the filter to the second row and aqueous acetonitrile solutions are used to release metabolites from the cells. The cell debris are retained on the filter; the intracellular metabolites are collected separately in the second tube.

Thermo Finnegan LC/ESI/LTQ FT Mass spectrometer - In addition to its exquisite sensitivity, mass spectrometry has the advantage in metabolite analysis of creating a direct interface to high-pressure liquid or gas chromatography for separation of metabolites. Electrospray ionization used to interface liquid chromatographs and mass spectrometers yields a single molecular ion from each compound. In the positive mode, compounds are detected as their protonated (m+1) species (or sodium (m+23) or potassium adducts (m+39)); in the negative mode, compounds are detected as their deprotonated species (m-1). Because only a single ion for each compound can be observed, it was difficult to assign structure based on nominal mass observed in our LTQ. In the Kegg database, at each nominal mass, a large number of potential metabolites are possible. For example, in the Kegg database of known metabolites, pyruvate, 3-oxopropanoate, butanoic acid, acetoin, 3-methylbutanol all have a nominal mass of 88 and these metabolites therefore cannot be individually resolved in a nominal mass spectrometer. Significantly, high-resolution FT mass spectrometers capable of ppm mass resolution allow prediction of molecular formulae from mass. This reduces the list of possible metabolites to structural isomers, which greatly reduces redundancy of compounds as listed in Kegg. Our laboratory has a dedicated Thermo Finnegan LC/ESI/LTQ FT Mass spectrometer (**Fig. 3**). The second stage of this tandem mass spectrometer is an ICR mass analyzer capable of ppm mass resolution. This ppm mass accuracy makes the difficult job of assigning structures to metabolite profiles possible. We have recently obtained an Advion Nanomate ion source for the LTQ FT mass spectrometer. This Nan spray ion source is designed to minimize ion suppression in the source so that peak intensity is more representative of concentration. In addition the Nanomate splits the sample and collects fraction that can further analyzed off line. Discussed below, commercial MS analysis software is not useful for analyzing high-throughput FT MS data. We therefore teamed with mathematicians at LANL to develop new algorithms to extract information from FT MS data.

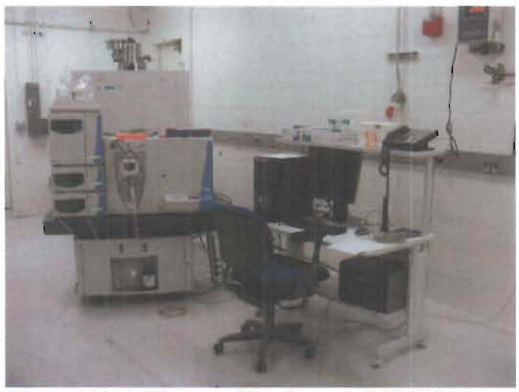

Figure 3) A ThermoFinnigan LC/ESI/FT Mass spectrometer. The ppm mass accuracy allows metabolites to be identified by their masses or their calculated formula weights, thus greatly aiding in the identification of previously unknown metabolites.

Leco Pegasus 4D Mass Spectrometer – Our laboratory also has a dedicated Leco Pegasus 4D GC/GC/TOFF mass spectrometer that we use to obtain high-throughput metabolite profiles. This instrument is capable of comprehensive two-dimensional gas chromatograph separations and has been used for metabolite analysis in several laboratories. The output of the gas chromatograph is coupled through an electron impact source to a rapid scanning time-of-flight mass analyzer. The mass analyzer is capable of obtaining five hundred full-mass-range

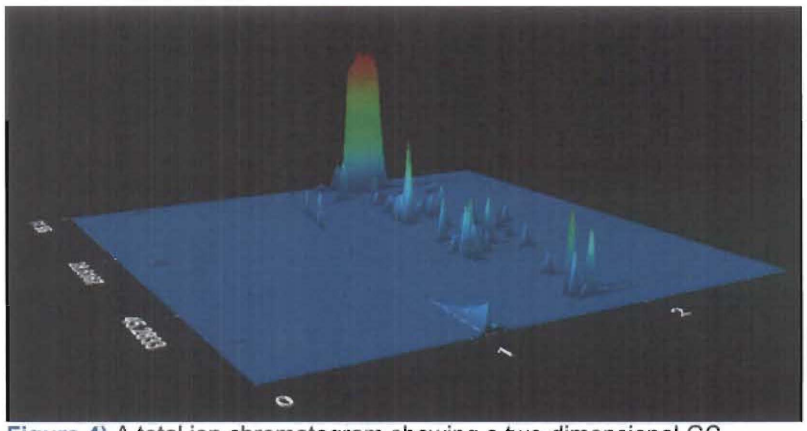

Figure 4) A total ion chromatogram showing a two-dimensional GC separation of our standard mixture of forty matabolites. Compounds that are not resolved by the primary GC column (y axis) are resolved in the more polar column in the second dimension (x axis).

spectra per second. The powerful two-dimensional chromatographic separations coupled with the rapid scanning mass spectrometer give the Pegasus 4D a unique capability to resolve complex mixtures such as all of the metabolites in a cell. This Instrument also has a Gerstel robot on the front end, which is capable of automated sample derivatization and auto injection. To minimize sample carryover, the Gerstel robot also changes the injector liner on the GC. While our Pegasus 4D was installed only recently, the power of two-dimensional GC separations can be seen in a plot of our standard metabolite mixture (**Fig. 4**).

Analysis of on LC/ESI/FT Mass Data - *Metabolomic* science requires new approaches to data analysis. Traditional analysis tools for mass spectrometry data are optimized for guided visualization and analysis of known or suspected metabolic constituents; however, the number and complexity of the spectra generated in the course of high throughput, high accuracy, high sensitivity metabolomic studies, as well as the need for algorithms to support discovery science, make manual analysis prohibitive. Attempts to automate traditional peak-picking approaches frequently fail, as the use of hardwired parameters results in large numbers of false positives, while setting customized parameters for each application makes these algorithms semi-automated at best. Building on LANL's work (GENIE project) in applying machine learning techniques to biomedical imagery analysis and remote sensing imagery of the earth, we have developed a new tool for automated mass spectrometry data analysis that is designed to be adaptive (using machine learning techniques) and robust (high accuracy and reliability). Our ABSTRACT (Automated Biological System Tracking and Classification Tool) algorithm exploits high throughput, high accuracy, high sensitivity data, based on detection and tracking of individual mass spectrometry peaks rather than regular binning of the data. Our technique exploits contextual information across multiple process blanks and data sets, including time courses, and identifies natural and labeled isotopomers. A comparison of our ABSTRACT algorithm to two popular algorithms, one proprietary (ACD9's CODA algorithm) and one open-source (MZmine), using LC-MS data for a reference sample containing 20 amino acids, is presented in Table 1.

Table 1: Comparison of LANL's Algorithm, ABSTRACT, to the MZmine and ACD9 algorithms

	LANL's ABSTRACT	MZmine	ACD9
Number of Analytes in Reference Sample	20	20	20
Number of Detectable Masses[1]	15	15	15
True Peaks Detected	14[2]	12[3]	13[4]
False Negatives (True Peaks Missed)	1	3	2
False Positives (False Peaks Reported)	2	31	> 300

[1] Glycine, alanine and serine are below limits of detection. Leucine and isoleucine are isomers and have identical m/z values. Glutamine and lysine masses are not distinguishable at 0.5 amu.
[2] Cysteine is not detected.
[3] Cysteine, arginine and tyrosine are not detected.
[4] Cysteine and threonine are not detected.

The power of isotopic labeling - Stable isotope labeling and NMR/mass spectral analysis of metabolites can be combined for metabolomic analysis. We can now use uniform stable isotope labeling to compare highly complex metabolite profiles of controls directly with experimental cultures. Metabolites extracted from uniform ^{13}C-labeled control cultures are combined with extracts from unlabeled experimental cultures. Labeled and natural isotopic abundance (unlabeled) metabolites chromatograph together but the mass spectrometer resolves their differences in mass as a function of the number of carbon atoms in the metabolite. The ratio of U-^{13}C-labeled and natural abundance isotopomers (U-^{13}C/N.A.) is

constant for metabolites produced by constitutively expressed enzymes. Metabolites with significantly altered U-^{13}C/N.A. isotopomer ratios result from enzymes that are either repressed or stimulated relative to controls. Employing the isotopomer ratio helps us to reduce certain systematic errors and greatly simplifies the overall issue of calibration of the mass spectrometer for quantification. The advantages of isotope labeling are illustrated in the **Figure 5**.

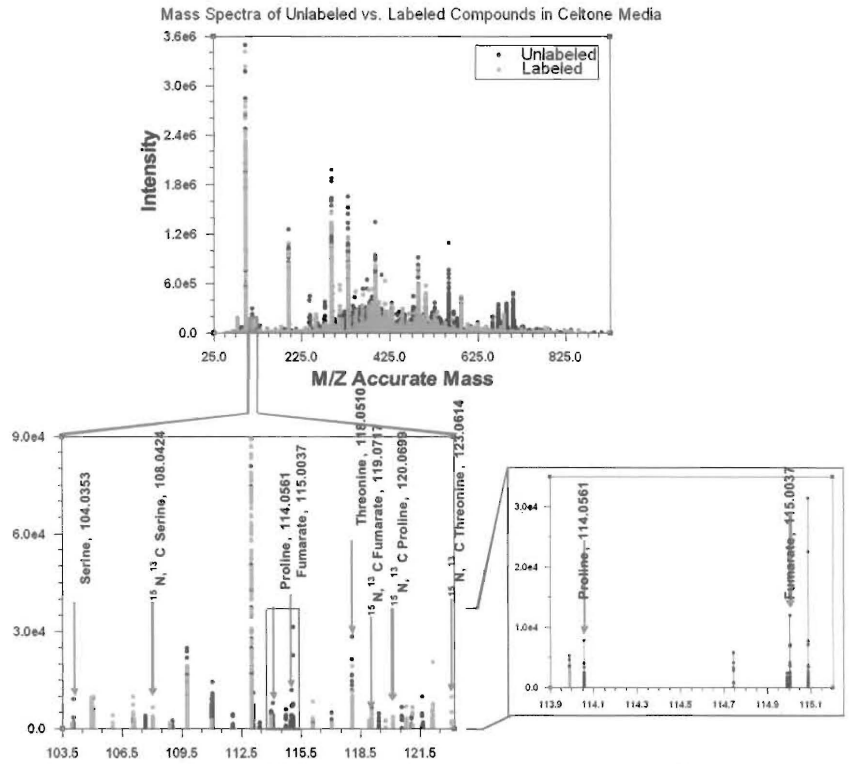

Figure 5. Tracking of complex mixtures with isotopes. The top panel shows the full mass spectrum of labeled vs. unlabeled metabolites. The bottom left panel shows a small section of the mass spectrum (~ 2% of total mass range), with selected compounds and their labeled isotopomers identified. The bottom right panel zooms in still further to show the resolvability of regions that appear overlapped in the bottom left panel. Taken together, these figures illustrate that high-resolution data are critical for label tracking in complex mixtures and that tools for single-compound analysis will quickly be overwhelmed. ABSTRACT (used to extract the data in the figure) is specifically designed to perform whole-system and isotopomer analyses. This figure also shows the power of isotope labels for structural assignment. Uniform ^{13}C- and ^{15}N-labeling shifts the mass of furmarate with four carbons from 115.0037 to 119.0171 while proline with five carbons and a nitrogen shifts from 114.0561 to 120.0699 exactly as predicited. By comparing data from labeled with unlabeled samples, we can easily identify peaks that are not metabolites because they appear at the same mass in both experiments. Peaks that are not isotopically shifted are not cellular metabolites

Isotopomer Analysis - Examination of metabolic pathways, or changes in metabolic profiles, has been used as part of an integrative approach for assessing gene function and relationships to phenotype. For example, stable isotope labeled compounds can be useful in establishing precursor product relationships in the identification of new enzymatic transformations and metabolic pathways. In addition, isotopomer analysis has been used to measure metabolic flux. For example, labeling the pyruvate pool and analyzing the isotopomer distribution in glutamate or aspartate can examine flux through the TCA cycle. Finally, stoichiometric models of enzymatic pathways and the experimentally determined metabolite concentrations are used to estimate the flux of carbon and electrons under varying conditions. Recent metabolomic advances combine isotope labeling with NMR and/or MS analysis, to estimate cellular metabolic fluxes.

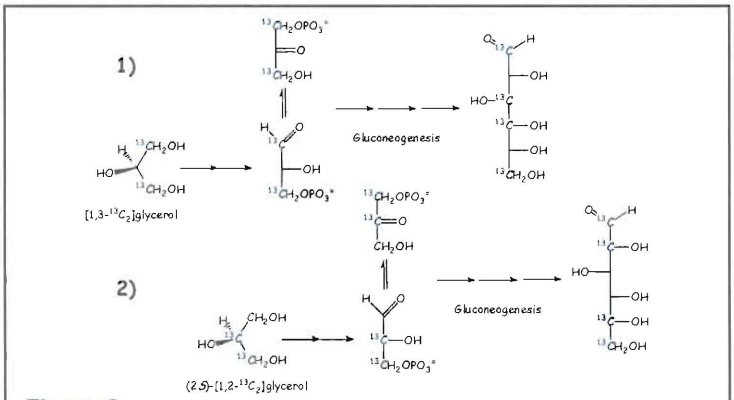

Figure 6 In case 2 it is clear that the proS hydroxymethyl of glycerol of leads via gluconeogenesis to C1 and C6 of glucose

Our general approach involves the addition of ^{13}C-labeled substrates to steady-state chemostat cultures and the identification the labeling patterns in intermediates and products by NMR spectroscopy or MS/MS techniques. This approach has several advantages over more conventional techniques for examining metabolic pathways. Because the resonance frequency of the ^{13}C nucleus is sensitive to its chemical environment, NMR can be used to determine not only the extent of labeling, but also the chemical identity of the labeled products as well as the distribution of label among their carbons. In NMR, the ^{13}C nucleus is sensitive to the spin of its neighbors (spin-spin coupling) so a particular ^{13}C atom will yield a different spectrum if it is chemically bonded to another ^{13}C atom rather than to a ^{12}C atom. Therefore, by double labeling with ^{13}C, one can determine the metabolic fate of a carbon-carbon bond. Similarly, the combination of ^{13}C and ^{2}H labeling allows tracking the metabolic fate of a carbon-hydrogen bond. The disadvantage of this technique is its relative insensitivity. In general, one is limited to observing compounds present in approximately millimolar concentrations. This problem can be circumvented by taking advantage of the fact that the label distribution in amino acids contains a wealth of information about the central metabolic pathways; amino acids will be isolated from crude protein hydrolysates in relatively high concentration. Finally, the sensitivity problem can be solved using a modern HPLC/ES/MS/MS or GC/MS/MS, which in many cases allows isotopomer analysis and the simultaneous analysis of many metabolites.

The ability to synthesize labeled compounds at Los Alamos will give us a distinct advantage in metabolic studies. In addition to unique labeled compounds available via the National Stable Isotope Resource at Los Alamos, we have the synthetic capability to prepare virtually any ^{2}H, ^{13}C, or ^{15}N isotopomers of most biochemicals. In addition, we have developed the stereospecific routes to many prochiral centers, which will allow us to unravel the cryptic stereochemistry of enzymatic transformations. For example, labeling a natural product with the commercially available symmetrical [1,3-$^{13}C_2$]glycerol allows one to identify positions labeled by C1 or C3, but cannot distinguish between the two. We have recently prepared the enantiomers (2*R*)- and (2*S*)-[1,2-$^{13}C_2$]glycerol, which will differentially label positions arising from the pro R or pro S hydroxymethyl carbons of glycerol. As diagrammed in **Figure 6**, double labeling the glycerol at C1 and C2 introduces a correlation in the label. Thus, products arising from intact (2*S*)-[1,2-$^{13}C_2$]glycerol units will be labeled at carbons arising from C1 and C2, but not labeled at carbons arising from C3.