SANDIA REPORT
SAND2006-7221
Unlimited Release
Printed January 2007

Biological Research Survey for the Efficient Conversion of Biomass to Biofuels

Michael S. Kent
Katherine M. Andrews

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2006-7221
Unlimited Release
Printed January 2007

Biological Research Survey for the Efficient Conversion of Biomass to Biofuels

Michael S. Kent
Biomolecular Analysis and Imaging

Katherine M. Andrews
Computational Biosciences

Abstract

The purpose of this four-week late start LDRD was to assess the current status of science and technology with regard to the production of biofuels. The main focus was on production of biodiesel from nonpetroleum sources, mainly vegetable oils and algae, and production of bioethanol from lignocellulosic biomass. One goal was to assess the major technological hurdles for economic production of biofuels for these two approaches. Another goal was to compare the challenges and potential benefits of the two approaches. A third goal was to determine areas of research where Sandia's unique technical capabilities can have a particularly strong impact in these technologies.

Table of contents

I. Introduction

Four weeks of funding was available to research these topics. The work included the following activities:

-Attended the Society for Industrial Microbiologists annual meeting in Baltimore, Md.

-In-depth review of the DOE/OBER/GTL Biomass to Biofuels workshop report

-Literature search on cellulose synthase rosette complex

-Literature search on cellulosomes and designer cellulosome constructs

-Literature search on optimizating mixtures of free enzymes for degradation of lignocellulosic biomass

-Literature search on sugar transporters

-Literature search on solvent pumps

-Literature search on use of spectroscopic methods (mainly FTIR and Raman) to identify components in complex mixtures derived from plant cell walls

-Literature search on X-ray diffraction studies of cellulose crystalline structures

-Literature search on X-ray and neutron scattering studies of plant cell walls

-In-depth review of NREL report from June 2004 - Biomass Oil Analysis: Research Needs and Recommendations

-In-depth review of July 1998 close-out report on DOE's Aquatic Species Program on Biodiesel from algae

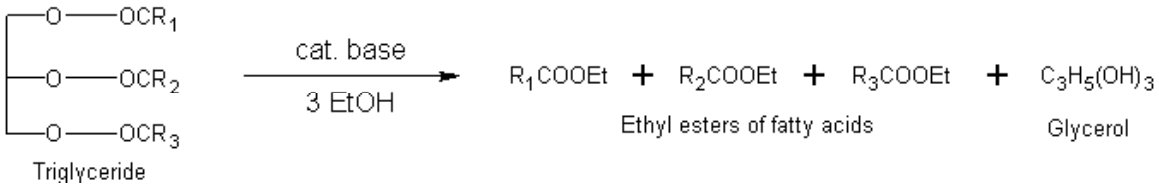
-Patent search on new technologies relating to production of biodiesal fuel (emphasis on biodiesel from algae)

-Discussions with Sandia researchers regarding application of technical expertise to specific technological problems related to biofuels production

II. Biodiesel

General

Biodiesel refers to alkyl esters made from transesterification of triglycerides in vegetable oils or animal and plant fats in which glycerol is replaced with a short chain alcohol such as methanol or ethanol (see below). Currently biodiesel is produced mainly from oilseed crops and from algae. Crude glycerol is produced as a byproduct. Glycerol production could help offset the costs of producing biodiesel fuel. However, no market exists for glycerol unless it is 99.5% pure so better (cheaper) glycerol refining technologies are needed. In general lower production costs and higher productivities are needed to make biodiesel production viable on a large scale.



II.a Biodiesel from Algae

Potential. Algae have much faster growth rates than terrestrial crops. The unit area yield of oil from algae is 7-31 times greater than the next best crop, palm oil. The total oil content in algae can be up to 60-70% of their dried weight. Algae may be the only viable method to produce enough transportation fuel to entirely replace gasoline from petroleum (1). One study claims that enough biodiesel to replace all petroleum transportation fuels could be grown in 15,000 square miles or roughly 9.5 million acres (1). This is far less than the 450 million acres currently used for crop farming in the US. Another important feature is that algae are able to fix CO_2 very efficiently (2). Thus, they have the potential to sequester carbon in waste CO_2 emissions from coal power plants, reducing this greenhouse gas.

History and hurdles. The US and Japan have both sponsored research programs into algae-biodiesel. Different approaches were taken, but neither has been commercially successful to date. From 1978 – 1996 the DOE funded an 18 yr program to produce biodiesel from algae (Aquatic Species Program, ASP). The work resulted in several important steps of progress, in particular progress in genetic engineering to increase the storage of carbon in the form of lipids rather than carbohydrates. However, the work also highlighted major roadblocks. It was believed at that time that economics demanded that the algae be cultivated in large open ponds rather than in enclosed ponds that would necessarily be of more limited size or in expensive bioreactors. However, cultivation of algae in open ponds revealed the inability to maintain selected laboratory algal strains in

large-scale open production systems. Other problems included lower than expected yields and high costs of harvesting the algal biomass and of the entire process due to the need for large agitators, pumps, and centrifuges.

The Japanese program focused on very expensive closed algal photobioreactors utilizing fiber optics for light transmission. In these controlled environments, much higher algal productivities were achieved, but the algal growth rates were not high enough to offset the capital costs of the expensive systems used.

Research opportunities:

The problems involved with maintaining a selected species in large open ponds are reduced substantially by cultivating the algae in bioreactors. Photobioreactors also allow for control of lighting cycles, which can impact productivity. The overall photosynthetic efficiency ranges from 6% for open pond-type reactors to 24% in the most efficient lab scale photobioreactors. However, the photobioreactors to date are still too expensive. Research is needed to improve the productivity and lower the cost of photobioreactors. Some new designs for photobioreactors and associated processes have been reported in the last few years in the patent literature (US20050260553 A1, US20050239182 A1, GB2423525 A). SNL could potentially contribute in this area by using its modeling capabilities to guide design of low-cost, high efficiency bioreactors. SNL has advanced computer codes coupling mass and thermal transport with chemical reaction rates. These codes should be adaptable to the design of photobioreactors.

Increases in productivity can be expected through further strain development. Algal strains are needed with high lipid content, high growth rates, and greater tolerance to fluctuations in operating conditions such as lighting conditions, temperature, salinity, and pH. The final years of the DOE ASP focused mainly on this topic. They succeeded in isolating the enzyme acetyl CoA carboxylase from a diatom. This enzyme catalyzes a key metabolic step in the synthesis of oils in algae. They were able to overexpress this gene in algae, with the hope that this would lead to higher oil production. The experiments did not demonstrate increased oil production in the cells, but these efforts have hardly scratched the surface of what could be done. Further work in metabolic engineering to promote pathways in which carbon is stored in lipids rather than carbohydrates is likely to yield some success. Recent patents have reported progress in genetic engineering of plants for the production of oils. One recent patent reported a method for increasing the storage reserve content in plants by expressing a leghemoglobin (US20060053515 A1). An increase in the oil content of seeds of Arabidopsis plants ranging from 16-23 % was reported. Another recent patent reported the ability to alter the ratios of saturated and unsaturated fats in soybean plants through genetic engineering by inserting thioesterases (US20050262588 A1). However, no patents of this nature have been reported in the last five years for genetic engineering of algae for biodiesel production. Metabolic engineering of algae appears to be an area with potential for high impact on this field.

A related area of research is to develop methods for producing strains that are preconditioned to the specific operating conditions expected in photobioreactors.

Another area of research likely to have an impact on the economics of algal-oil production is the harvesting process. Algae can be harvested using microscreens, centrifugation, or by flocculation. Harvesting by chemical flocculation is typically too expensive for large operations. Interrupting the CO_2 to an algal system can cause algae to flocculate on their own, called autoflocculation. Research is needed to find low cost methods for efficient flocculation.

Another potentially productive research area is oil extraction methods. Current practice involves the use of chemical solvents such as hexane. However, the use of hexane is hazardous, and efficient low-cost alternatives are needed. Possible alternatives are enzymatic extraction (but estimated to be too costly), mechanical pressing of dried algae, osmotic shock, supercritical CO_2 extraction, and ultrasonic-assisted extraction. Sandia has substantial expertise in supercritical CO_2 extraction.

II.b Biodiesel from biomass oils (lipids from animal fats, fish and poultry oils, plant oils, and recycled cooking greases)

Potential. Most biomass oils come from seed crops and are extracted through a crushing process. Compared to soybeans, all other oilseed crops that could expand with demand are minor. Only 19% of soybean mass consists of oil. Soy meal is the main product and soy oil is a by-product. For soy oil supplies to increase, there has to be a major increase in demand for soy meal. It is estimated that biodiesel from biomass oils can only replace a maximum of 10 billion gallons of petroleum by 2030 (Biomass Oils Analysis Rep) compared with a total usage of > 200 billion gallons.

History and hurdles. Biomass oil biorefineries already exist. Biodiesel from rapeseed oil is a substantial commercial enterprise in Europe. Costs of feedstocks and production are too high to replace petroleum unless incentives are given by the Federal government. However, a recent report in Nature described a new catalyst for transesterification that costs1/10 to 1/50 that of conventional catalysts (3).

Japan does not have the land resources to produce large quantities of feedstocks for conversion to biodiesel, so they have developed collaborations with other countries. Japan has supported research on genetic breeding of sunflower plants for propagation and cultivation in many countries with different climates (NEDO). Japan also supports research investigating conversion of palm oil mill effluent in Malaysia, the world's largest palm oil producing country. (NEDO 98EF1, NEDO 01EF4, NEDO 01EF002).

Research opportunities.

Research is needed to increase yield and reduce costs of high oil seed crops (canola or sunflower). One aspect needed for expansion is to optimize varieties in new geographical regions.

Research is needed to increase the oil content of soybeans through metabolic engineering.

Research is needed into oil production from yeasts, fungus, bacteria, and other organisms that can be produced with minimal land, water, or sunlight. Biomass oils can be produced from fungi and yeasts that do not require photosynthesis and have minimal requirements for land usage. These organisms can produce 20% to 40% of their dry weight as oils that are similar to vegetable oils (4). Low cost feedstocks are needed and the conversion efficiency of the feedstocks into oils must be optimized through metabolic engineering. Most testing has involved glucose, which is too expensive. Biomass could be used if the organism could convert a high fraction of it efficiently.

Research is needed into the use of supercritical CO_2 to extract the oils. If supercritical CO_2 could solubilize methanol along with the oil, then oil extraction could be combined with biodiesel production. SNL has considerable expertise in supercritical CO_2 extraction technology, and could potentially have an impact in this area. One company (Crown Iron works) already has a continuous CO_2 extraction technology for soybean crushing suitable for small-scale plants.

If a technology could be developed to extract corn oil from dry ethanol mills in a cost effective manner, ~ 740 million gallons of corn oil could be recovered (4).

III. Ethanol from lignocellulosic biomass

The US has the potential to sustainably harvest 1.3 billion tons of biomass, which could support ethanol production at the level of > 60 billion gallons/year, enough to replace 30% of the nations' current gasoline use. Conversion of lignocellulosic biomass to ethanol has been demonstrated at the lab bench scale, but commercialization has not yet been achieved. However, several pilot scale plants are in the later stages of development or are in operation, indicating that the technology is close to commercial viability. Logistical issues, such as feedstock availability and transport to the site of processing, and several technological hurdles must be overcome to improve the economics for widespread commercialization. Biomass requires extensive processing, and high costs are associated with pretreatments and enzymatic deconstruction. Furthermore, robust organisms are needed for fermentation of a range of sugars in the presence of potentially inhibitory chemicals arising from the deconstruction steps.

III.a The structure of plant cell walls (feedstocks).

Cellulose synthase complex Cellulose fibers are composed of six strands of β-1,4-glucan that crystallize. The strands are synthesized by the "rosette" complex of cellulose synthase (CESA) proteins. The detailed structure of the rosette is not known, but is thought to contain 6 subunits to account for the 6 strands of β-1,4-glucan, where each subunit contains 5-6 cellulose synthase proteins. It has been shown that mutations in the CESA proteins lead to alterations in cell wall structure, presumably by altering the structure of the rosette complex and the nature of the microfibrils. Thus, systematic alteration of the rosette structure is one possible route to tailoring cell wall structure for easier breakdown while still allowing plant growth. Judicious redesign of the rosette complex requires knowledge of the genes involved, characterization of the structure of the complex, and an understanding of the mechanism of operation. Recently the CesA genes that encode protein subunits of the cellulose synthase have been identified (5-7).

Research opportunities

Structure of the cellulose synthase rosette complex. The rosette is known to contain several different types of enzymes that carry out different functions. Establishing the arrangement of proteins within the rosette is critical to understanding the mechanism. Crystal structures of the rosette complex would be invaluable but are not yet available. In fact, the complex may require an intact membrane to form. Other structural studies, distinct from and complementary to efforts to obtain 3-D crystals of the complex, should be pursued. First, attempts should be made to resolve details of the structure of the complex in plant cells using high-resolution electron microcopy and AFM imaging. Second, since the complex is believed to self-assemble into a unique structure, attempts should be made to reconstitute it from the basic protein components on biomimetic lipid membranes that allow for high resolution AFM imaging, grazing incidence X-ray diffraction characterization of the overall 2-D structure, neutron and X-ray reflection characterization of how the proteins interact with and insert into lipid membranes, and TIRF/FRET characterization of lateral associations of specific fluorescently-labeled subunits. Data from these characterization methods could be combined with molecular modeling to understand the effects of specific mutations on the structure of the rosette complex. Finally, attempts should be made to make crystalline cellulose either in vitro using the reconstituted complex and plant extracts or in yeast or bacteria, which has not yet been achieved to date. If successful, this would allow study of many details of the synthetic pathway and the effects of mutations or structural alterations.

Sandia has technical expertise in electron microscopy, AFM imaging, TIRF/FRET, preparation and characterization of biomimetic tethered lipid membranes, X-ray and neutron reflection and grazing incidence X-ray diffraction studies of proteins interacting with lipid membranes.

Characterization of the various structures of crystalline cellulose. Much progress has been made in recent years in using X-ray and neutron diffraction to determine detailed structures of the various forms of crystalline cellulose. Neutron and X-ray diffraction can be used to determine the detailed crystal structure and hydrogen bonding system in cellulose. X-ray diffraction can be used to determine the C and O positions. By using deuterated solvents, crystalline fibers can be prepared in which all the OH groups in the crystalline unit cell are replaced by OD. By collecting neutron diffraction data from both samples, the positions of the hydrogen atoms can be determined.

Two crystalline phases are observed: cellulose phase Iα (triclinic) and phase Iβ (monoclinic). It would be beneficial to understand if one phase is easier to degrade than the other. Similarly, it would be helpful to know if certain cellulose binding domains (discussed below) bind to one phase more readily than the other. Another potentially fruitful area of research would be diffraction studies performed at different stages of hydrolysis (discussed below).

Paul Langan, at the Los Alamos Neutron Scattering Facility, is one of the leading researchers in this field. This could represent an opportunity for Sandia due to the close proximity of Sandia and Los Alamos and the association of several Sandians (Kent, Rodriguez) with LANSCE in general and with Paul in particular. Future work could include development of a collaborative program to assess variations in the structure of cellulose from different sources, or of cellulose synthesized by variations of the rosette complex.

Characterizing the composition in cellulosic biomass. Infrared and Raman vibrational spectroscopies with multivariate analysis can provide quantitative analysis of lignocellulosic biomass, including cellulose structure, lignin content, hemicellulose content, and polysaccharide identification without the use of tags or dyes. Carbohydrates and lignin give both IR adsorption and Raman spectra. FT-IR spectra give information about the various types of polysaccharides present in a complex mixture. The spectra contain main band maxima that are distinctive to the type of polysaccharide present, and also contain features that are dependent upon the axial and equatorial positions of the OH groups. However, care must be taken as FTIR spectra are also sensitive to changes in the conformations of polysaccharides. Polarized IR is sensitive to bond orientation, and therefore can be used to distinguish hydrogen bonding patterns in the different types of cellulose. Visible and NIR Raman can also be used to study changes in polysaccharides, if special measures are taken to reduce fluorescence arising from lignin. Solid state NMR, used in cross polarization-magic angle spinning mode, is sensitive to variations in polysaccharide conformations, but is not sensitive enough to detect the relatively rare covalent crosslinks believed to hold the cell wall together. NMR relaxation experiments can reveal the relative mobilities of various components in a complex mixture, as well as the relative mobilities of main chains and side chain groups. SNL has particular expertise in the abovementioned vibrational spectroscopies, multivariate analysis, and NMR.

Genetic engineering of plants to reduce lignin content will almost certainly be one strategy pursued in the long term to aid the deconstruction process. Therefore, methods

for precise characterization of lignin content will be needed. **Imaging methods, such as electron microscopy, are helpful in qualitative characterization of lignocellulosic materials. Regarding quantitative charcterization,** UV resonance Raman is highly sensitive to lignin, and detection of 0.1 wt % lignin in pulp has been reported. This method can be used to evaluate strategies for reducing the lignin content in plant cell walls or strategies for degrading lignin. However, care must be taken to avoid beam damage from the more highly energetic UV laser radiation.

Characterizing the fine structure of plant cell walls. The micro and nanostructured arrangement of cellulose fibrils, hemicellulose, and lignin in plant cell walls is believed to strongly affect the intransigence of this material to enzymatic breakdown. Characterization of the fine structure of plant cell walls is therefore an important component of studies to re-engineer cell walls, as well as in studies of deconstruction methods. Imaging methods such as electron microscopy, hyperspectral imaging, and AFM have important roles to play in this. In addition, NMR and vibrational spectroscopies hold much promise in this regard. These techniques can distinguish between the different major components, and recent work suggests that there may be ways to probe stress-strain relationships at the molecular level with these methods (8). Such stress-strain data, if resolved for cellulose, hemicellulose, and lignin, can give insight into the nano-and microstructured arrangement of the different components. Neutron and X-ray scattering and diffraction can also contribute in this area. The basic fibrillar structure can be quantified with microfocused wide angle X-ray scattering (WAXS). In addition, load extension curves obtained simultaneously with SAXS/WAXS reveal substantial differences among cellulose composite materials despite the fact that the basic fibrillar structures are very similar. These differences are likely due to variations in connectivity associated with entanglements and crosslinking (9). SNL has strong expertise in all the abovementioned experimental methods.

III.b Deconstruction

It is important to note that pretreatment approaches and cellulase mixtures for deconstruction must be developed in combination, as the method of pretreatment will affect the nature of the material and thus the efficacy of different cellulase mixtures.

Pretreatment approaches. Various methods of pretreating lignocellulosic biomass prior to enzymatic hydrolysis have been explored including physical, biological, and chemical methods. Currently the physical methods are too costly and the biological methods are too slow. Thus, current efforts are focused largely on chemical methods. These include use of dilute acids (HCl, H_2SO_4), alkaline treatments, explosive decompression (steam), ammonia fiber explosion, hot water, or treatments with organic solvents. A better understanding of the physical and chemical changes of different lignocellulosic materials subjected to these pretreatments is needed for selection and optimization of approaches that are most synergistic with enzymatic degradation by cocktails of free enzymes or by cellulosomes. Furthermore, hemicellulose degradation (acetylation) products are toxic to microbes and so must be quantified for the different pretreatments and sources of

lignocellulosic material. As mentioned above, data on physical and chemical changes in lignocellulosic materials can be provided by IR and Raman spectroscopies with multivariate analysis, NMR, AFM and IFM, and X-ray and neutron scattering and diffraction, all areas in which SNL has considerable expertise. Multicomponent characterization of lignin and hemicelluloses in the presence of hydroxycinnamic acids is possible, allowing kinetics of degradation to be followed.

Several methods will be useful to establish the extent of water penetration into crystalline fibers as a function of pretreatment, which may correlate with enzyme binding sites and thus impact hydrolysis rates. IR spectroscopy is one method. By exposing samples to D_2O and measuring the extent of exchange of H with D, the fraction of chains that are exposed to the aqueous solution can be determined. This can be used to provide insights into mechanisms and kinetics. Neutron scattering can provide the detailed profile of D_2O penetration into crystalline fibers as well as changes in the fractal character of the fiber surface. This information could be obtained as a function of time in the presence of the degrading enzymes.

Synergistic effects of mixtures of free enzymes. Understanding the hydrolysis mechanisms and synergistic action of enzymes would be of enormous benefit in leading to knowledge-based design of cellulase cocktails. Cellulases are divided into exoglucanases, which release cellobiose units from the chain ends in crystalline cellulose, and endoglucanases that act randomly on amorphous cellulose regions. Free cellulases are modular, having a catalytic domain for hydrolysis and a cellulose-binding domain (CBD) linked by a flexible spacer. Cellulosomes are large multienzyme complexes that also have CBDs. The CBD have been shown to contribute significantly to the activity of cellulases against native cellulose. More than 180 CBDs have been identified and classified into 13 families, many displaying different substrate binding affinities and specificities (10). **Three general approaches can be envisioned for exploring the large parameter space involved with variations in both CBDs and catalytic domains. First, large combinatorial libraries of CBD/catalytic domain combinations could be constructed and tested individually and also in mixtures of members from the different families. Second, physical insights can be gleaned from study of a small group of representatives from the different families of CBDs and catalytic domains, and design of cellulase cocktails could then be approached** from a knowledge-based perspective. Third, directed evolution approaches could be employed. All three methods should be pursued of adequate funding is available.

With respect to the knowledge-based approach, studies are needed to determine where different CBDs bind, the binding affinity and reversibility, and how they move along crystalline cellulose in an attempt to understand **some of** the mechanisms for synergistic effects. **Studies are also needed that compare different cellulase enzymes linked to the CBDs. It has been shown that the synergistic effects of enzymes are substrate-dependent. Thus, detailed characterization of the lignocellulosic substrate, amount of bound enzyme, location of bound enzyme, changes in substrate during hydrolysis, and the kinetics of adsorption and hydrolysis are all needed for different sources of lignocellulosic material.**

In studies of microbial degradation of lignocellulosic materials, it is important to find methods to distinguish between the mass of adsorbed cells and the mass of adsorbed enzymes for quantitative comparison. Techniques for direct measurement of adsorbed enzymes would be desirable but are seldom employed. One recent approach combined an ELISA assay for enzyme quantification with cell counts by total protein concentration (11). Another related research opportunity is in understanding mechanisms of adhesion of cellulolytic microbes to their substrates.

Imaging methods will be helpful in characterizing lignocellulosic materials at different stages of deconstruction, but it is likely that other techniques will provide important complementary insights as well. Infrared and Raman vibrational spectroscopies can provide additional characterization of substrate material, as cellulose structure, lignin and hemicellulose content, and polysaccharide identification are possible without the use of tags or dyes. Furthermore, IR and Raman can be used to understand how different cellulases, or mixtures of cellulases, alter the physical and chemical nature of the substrate during hydrolysis. **N**eutron and X-ray scattering/diffraction may be useful to establish other characteristics of lignocellulosic materials, including crystalline structures, fibril size, and extent of fibrillation. Neutron scattering can be used to establish the extent of hydration of (profile of water penetration into) crystalline fibers, which may correlate with enzyme binding sites and thus hydrolysis rates. In cases where enzymes can be obtained enriched in deuterium, the quantity of bound enzymes could be determined in the presence of other species by neutron scattering. AFM and EM can be used to identify the characteristics of sites to which each CBD binds. Changes in local mechanical properties due to enzymatic deconstruction can be examined with the interfacial force microscope (IFM), a noncompliant form of atomic force microscopy developed at SNL. This would provide further evidence of how and where different enzymes act. Information from this suite of methods compiled for individual cellulases, as well as for combinations, would provide a much more comprehensive analysis of this highly complex system than exists to date and should be of great benefit in the design of cellulase mixtures.

The rate of hydrolysis typically declines rapidly as the extent of conversion increases, presumably due to product inhibition, thermal denaturation, and formation of enzyme-lignin complexes (12-14). Some evidence indicates that other factors may be involved as well. Understanding the causes of the nonlinear kinetics is a research opportunity with the potential for a large impact on this technology. If mechanisms of inactivation are understood, then it may be possible to design solutions to mitigate this problem. The reactivity of the substrate appears to decrease with conversion. One possibility is that adsorbed enzymes are deactivated/denatured. This possibility could be probed using several methods. If enzymes are obtained enriched in deuterium, then the accumulation of bound enzyme on lignocellulosic materials could be quantified using neutron scattering. This could be followed as a function of time during the hydrolysis process to look for a signature that correlates with a decreased rate. Denaturation of bound enzymes could be probed indirectly by examining the reversibility of binding at different stages of conversion. This can be accomplished by diluting the enzymes in the bulk solution or by

introducing other strongly binding molecules in excess and determining if displacement of the bound enzymes occurs. Model studies could be conducted to test specific hypotheses. In contrast to indirect methods like reversibility studies, neutron reflection is capable of probing the detailed conformations of enzymes bound to flat surfaces. Conformational changes, reversibility, and binding of other species to the adsorbed enzymes can all be examined as a function of time and solution conditions. However, this method requires fabrication of smooth surfaces that present cellulose-like binding sites to the solution. SNL has extensive expertise in studying the structure of proteins adsorbed to surfaces and conformational changes upon binding by this method.

Cellulosomes. Cellulosomes are large protein complexes synthesized by some anaerobic microorganisms that localize suites of synergistic enzymes for efficient degradation of crystalline cellulose (15). Re-engineering of these complexes holds the potential to further improve the efficiency of cell wall deconstruction and cellulose breakdown (16-20). Detailed structural and biophysical characterization of these complexes as they interact with lipid membranes will be needed for modeling and subsequent optimization of designer cellulosome complexes. In particular, understanding the cohesin-dockerin interaction is critical to producing designer-cellulosomes (15-22). Critical issues are binding affinities and reversibility, the origin of inter and intraspecies selectivity, and conformational changes of the enzymes upon docking or as a function of Ca^{2+} concentration that affect their activity. The conformations of cohesin-dockerin complexes in association with lipid membranes can be examined in detail by grazing incidence neutron and X-ray scattering methods combined with high resolution scanning probe microscopies and single-molecule spectroscopies. Crystal structure data are available (21, 22), so modeling the docking of cohesions with dockerins may be possible and could provide helpful insights. SNL has both experimental and modeling expertise to address these areas.

III.c Fermentation

Research opportunities exist in the metabolic engineering for parallel and complete sugar utilization. However, sufficient time was not available in this late-start LDRD to fully explore this research topic. A related topic, the transport of various sugars into the cell, is discussed in detail below.

Sugar Transporters

Research opportunities related to sugar transporters are specific to each organism, and so the discussion below is divided into sections related to E-coli, Z. mobilis, and yeast.

E-coli.

Despite the important role of transport systems in bacterial physiology, relatively few studies to modify transport functions for strain improvement have been reported. One system where this is important is in the modification of E-coli for simultaneous fermentation of hexose and pentose sugars to ethanol. In E-coli the uptake and phosphorylation of glucose occurs mainly through the phosphotransferase system (PTS), which phosphorylates substrates during transport (23). The PTS is a low affinity, high yield transporter system found widely in bacteria but absent in archaea and eukaryotes. In the E-coli chromosome, there 21 different PTS enzyme systems that transport a variety of different carbohydrates including glucose, glucoside, fructose, mannitol, maltose, lactose, mannose, sorbose, glucitol, galactitol, N-acetylglucosamine, galactosamine, N-acetylgalactosamine, trehalose, arbutin, salicin, cellobiose, N,N-diactylchitobiose, pentitol, and pentose. The substrates of several additional permeases have not yet been identified. In addition to facilitating transport of sugars into the cytoplasm, PTS is involved in the regulation of catabolic repression and chemotaxis, and also links to sugar metabolic pathways. Catabolic repression is the regulatory process by which the organism selects from a mixture of sugars the one that provides the highest growth rate. This leads to sequential sugar utilization, which is disadvantageous for industrial processes to produce ethanol from lignocellulosic materials, as hydrolysis of hemicellulose and cellulose produces C5 sugars such as xylose and arabinose along with glucose. In particular, glucose exerts a repressive effect on the consumption of other sugars, both PTS sugars and non-PTS sugars (24). However, the phenomenon is not restricted to glucose, as many PTS carbohydrates are used in preference over other carbon sources. Another consequence of the PTS is that a high rate of glucose uptake by PTS results in high acetate production under oxygen-limiting conditions. The accumulation of acetate is an important problem in an industrial process as it inhibits cell growth.

A general approach to eliminate both sequential sugar utilization and high acetate production is to eliminate or alter the PTS. Strategies to date have included elimination of the entire PTS system though inactivation of the genes for non-sugar specific enzymes in the PTS system, and inactivation of only the integral membrane glucose permease. The PTS is a complex system deeply integrated into the cell's physiology and much work is needed to fully understand the results of these modifications. Nevertheless, significant improvements in mixed sugar utilization have already been demonstrated (24-27).

Elimination of entire PTS system

Elimination of the entire PTS system requires replacement with another glucose transport system. E-coli has a high affinity glucose transport system of the ATP binding cassette (ABC) superfamily that is induced when glucose is present at very low concentrations. However, this system has a high energetic cost of 2 moles of ATP per mole of glucose that is internalized and phosphorylated. This high cost is expected to cause lower growth

rate and cell yield in oxygen-limited growth. Therefore, replacement with another glucose transport system is required.

By contrast, in the absence of the PTS E-coli can grow efficiently on sugars that have alternative transport systems, such as xylose, arabinose, galactose, and lactose (27). In E-coli, xylose and arabinose are transported by two inducible systems: a low affinity high capacity arabinose/H+ (or xylose/H+) symport AraE (or XylE) and a high affinity ABC-type transporter AraFGH (or XylFGH). Arabinose/H+ is the main arabinose transporter in the wild type, and mutants lacking the ABC transporter can grow on minimal media anaerobically at a rate comparable to the parent, whereas mutants lacking AraE have significantly lower growth rate than the wild-type parent. On the other hand, XylE xylose/H+ symport is apparently not the main contributor to xylose transport in E-coli, and xylose is preferentially transported by the high affinity ABC-type XylFGH system. Specific growth rates for mutants lacking XylG and utilizing only XylE is minimal in anaerobic conditions, and is reduced about 50% even under aerobic conditions. Increasing the copy number of XylE does not increase the growth rate either anaerobically or aerobically and deleting the xylE gene has only a minimal effect on the growth rate of the mutant on xylose minimal media. Transport of xylose through the ABC-type system has a high cost in ATP (two ATP equivalents - one each for transport and phosphorylation). Recent work has shown that the arabinose/H+ symporter can transport xylose efficiently if repression in the absence of arabinose is removed by using a synthetic promoter.

Improvements in the ability to import glucose after disabling PTS have followed three strategies. One approach is a continuous culture selection method wherein strains lacking PTS are selected for growth on glucose as the only carbon source (28). Mutants that could grow on glucose arise relatively rapidly, and it was determined that the mutants imported glucose through facilitated diffusion using the galactose permease, a member of the MFS. Thus, these strains are dependent upon a functional galactose permease gene (galP) in the chromosome.

A second approach was developed as a modification of the first. In order to avoid repression of the galP gene in the presence of glucose, the native regulatory region of galP was replaced by another promoter (24). Strains resulted that were able to co-metabolize glucose and arabinose. However, xylose was only utilized after depletion of glucose and arabinose. This strain could transport glucose as rapidly as its wild-type parental strain.

A third approach, entirely distinct from the first two, involves expression of genes for a different functional pathway for glucose uptake and phosphorylation. Work to date has incorporated genes from Z. mobilis, namely the glucose facilitator (glf) and glucokinase (glk) (29, 30). The glf/glk system of sugar uptake does not use metabolic energy in the form of a proton potential or phosphoenolpyruvate. Glucose uptake was restored by the GLF but no studies were reported of mixed sugar utilization. The GLF is known to be very promiscuous, providing transport for D-glucose, D-mannose, D-xylose, D-fructose, D-glucosamine, D-galactose, L-sorbose, and D-arabinose. It is interesting to note that the

glf of Z. mobilis functioned well in E-coli despite the fact that there are substantial differences in plasma membrane lipids between the two organisms. While this finding is very promising, it remains to be seen if insertion and functioning of other transporters are as tolerant of variations in membrane lipid composition.

Elimination of the glucose PTS system only

Several groups have shown that in strains with the gene for the PTS integral membrane permease for glucose (ptsG) inactivated, mutants quickly arise with restored ability to grow on glucose and allow simultaneous uptake and utilization of mixed sugars. This been demonstrated in several strains, including an ethanologenic strain (25,26,31,32). One group showed that a ptsG mutant strain metabolized glucose and arabinose or glucose and xylose simultaneously and the mutant consumed the total amount of sugars in about half the time as that of the wild-type strain (25). The transport mechanism was not determined. It is known that in the absence of the glucose PTS, E-coli can transport glucose by the mannose-PTS system (29), but other possibilities may also exist.

Another study of this type used fosfomycin as a selection agent to screen for PTS mutants of an E-coli strain that are unable to transport glucose or other PTS sugars (27). Fosfomycin is a bacteriocidal analogue of phosphopyruvate. Some of the mutant strains lost the capacity to ferment all PTS sugars, while others retained a functional PTS and the ability to metabolize all sugars transported by the PTS but lost the ability to metabolize glucose-6 phosphate and glycerol 3-phosphate. The specific genetic mutations were not determined in this work. This E-coli strain contained genes encoding the ethanol pathway from Z. mobilis. The study reported mutants of the second category, those that retained the ability to metabolize sugars transported by the PTS but not glucose-6 phosphate and glycerol 3-phosphate, that had the ability to simultaneous ferment mixtures of xylose and glucose, with higher ethanol yields than the parent strain. The mechanism by which catabolite repression was circumvented in this case was not determined.

Research opportunities

These studies show great promise, but have only scratched the surface of the possible avenues to explore.

There are various ways to disable PTS, which will likely have varying affects on the repression of other sugar uptake systems. We will pursue strategies for disabling the entire PTS system and also for disabling only the glucose PTS. Retaining portions of the PTS has the advantage that it is a high flux transport system that requires lower energy than the other systems. Comparisons of sugar fluxes into the cells, cell growth rates, and rates and extents of conversion of sugars to ethanol are needed for the different approaches.

Several components of the PTS are known to have differing regulatory functions. For example, enzyme IIA^{Glc} affects transcription of operons for other sugars through regulation of cAMP concentration. High cAMP concentration is required for RNA polymerases to bind to Lac and other operons, but the presence of glucose lowers cAMP concentration because in the presence of high glucose IIA^{Glc} exists mainly in the nonphosphorylated state and cannot bind to the enzyme adenylate cyclase to activate cAMP synthesis. Methods should be sought to disable glucose PTS while maintaining high cAMP concentration.

Further studies should be pursued in which the glucose PTS in E-coli is disabled and replaced by other glucose permeases of the MFS. The glf from Z. mobilis was mentioned above, but other candidates also exist. In this regard, a series of BLAST and FASTA searches were done in 1997-2000 (33-36) for transporters in prokaryotes. This has led to identification in E-coli of the complete PTS system (23), the ATP cassette binding proteins (37), and the members of the major facilitator superfamily (38). A number of MFS proteins in genomes of bacteria other than E-coli were also identified. However, at that time only 25 genomes were sequenced and 100 additional projects were underway. New searches of this type should be performed for bacterial species to update the list of MFS sugar transporters from the vastly larger database of recently sequenced genomes. Importing MFS permeases should be pursued for both strategies – when the entire PTS is inactivated as well as when only the glucose PTS is disabled. In the latter case, the use of synthetic promoters to avoid repression may be required. This work will include genomic searches to uncover other glucose transporters for replacement of glucose PTS and also more efficient (less ATP-dependent) transporters of xylose.

In the same vein, research is needed to establish if other MFS transporters are tolerant of variations in membrane lipid compositions as was demonstrated for the glf of Z. mobilis. This can be done using proteoliposomes. SNL has extensive expertise in this area.

Further selection experiments should be performed for mutants in strains lacking the glucose PTS permease that can grow on glucose.

Z-mobilis:

Wild-type Z mobilis can utilize only sucrose, glucose, and fructose as sole carbon sources, which are largely converted into ethanol and carbon dioxide. Recombinant Z-mobilis strains have been developed that can efficiently convert glucose, xylose, and arabinose to ethanol, however glucose is used more rapidly and more efficiently. In contrast to E-coli, in Z. mobilis sugars enter primarily through the glucose facilitator (glf). Thus, sequential sugar utilization is not due to catabolite repression by PTS, as in E-coli, and may be due to transport. Recent studies have revealed that glf is capable of transporting a wide range of C5 and C6 sugars with varying levels of efficiency (39). Attempts should be made to determine if transport is limiting the rate of C5 sugar utilization in Z. mobilis. If so, studies to screen for mutants of glf that are more highly

efficient at transporting C5 sugars, and / or research to import other transporters of the MFS to achieve more efficient simultaneous sugar utilization should be pursued.

Yeast.

Sacch. cerevisiae (SC) is the choice organism for ethanol production due to its high productivity, ethanol tolerance, and tolerance for inhibitors present in lignocellulose hydrolysate. However, it is unable to utilize pentose sugars. Genes to metabolize xylose have been inserted, but growth rate and ethanol productivity with respect to xylose are generally low compared to natural xylose utilizing yeasts like Pichia stipitis or Candida shehatae. SC takes up xylose with low affinity by the high-affinity hexose transporters, whereas Pichia stipitis or Candida shehatae possess specific xylose transporters. Recent analysis of metabolic flux control showed that, in strains with low xylose reductase (XR) activity transport has very little control over the specific xylose utilization rate, whereas in strains with high XR activity transport plays a significant role in controlling flux below extracellular concentrations of 1 g/L or 6.7 mM (40). Thus, xylose transport in SC strains currently has limited control over the specific xylose utilization rate, but is expected to become important as the xylose metabolism pathway is improved. Pichia stipitis (PS) has the highest native capacity for xylose fermentation of any known microbe. Its genome has recently been sequenced, and contains 40 putative sugar transporters. A potentially promising avenue to pursue is to express xylose transporters from PS in genetically modified strains of SC that are able to ferment xylose. Batch fermentation of each of the recombinant strains could be performed to investigate xylose transport and product formation. It would be beneficial to express a xylose transporter with higher yield and low energy cost. It would be beneficial to express an active xylose transporter, which is capable of taking up xylose against a concentration gradient, for utilization of xylose below 1 g/l. It would also be desirable to express a transporter that is specific for xylose and not inhibited by glucose, for cofermentation of xylose and glucose.

General

Transport assays:

It is important to determine under what conditions, if any, sugar transport limits the rate of ethanol production. Sugar consumption on the timescale of minutes or hours can be followed by HPLC using RI and UV detectors. Cell mass can be followed using optical density at 550 nm. Ethanol production can be determined using gas chromatography.

The initial sugar uptake rates (within the first few seconds) can be accurately determined using radiolabelled sugars. The cells are brought into contact with the radiolabelled sugars very rapidly, the reaction is allowed to proceed for a given time, and then the reaction is rapidly quenched. The cells are filtered, washed extensively, and then the isotope concentration is counted using a scintillation counter. The initial uptake rates of

various sugars such as glucose, xylose, and arabinose can be determined alone and then also in the presence of the other sugars.

Solute transport can also be assayed using proteoliposomes. This method would be especially useful as a rapid way to determine the degree to which MFS transporters are tolerant of variations in membrane lipid compositions. A large number of such transporters have been identified in prokaryotes and many more are certain to be found in searches of more recently sequenced genomes. These are candidates for import into E-coli or Z-mobilis to optimize uptake of mixed sugars. Proteoliposomes are formed by mixing liposomes with detergent solubilized membrane proteins, and then removing the detergent by dilution or by dialysis. The protein leaves the detergent and becomes incorporated into the phospholipid bilayer. When the proteoliposomes are incubated with solute, they catalyze uptake of the solute into the vesicles, provided that the appropriate transporter has been incorporated. The vesicles can then be isolated by centrifugation or filtration and the amount of solute inside the vesicles can be quantified. Fluorescence measurements can also be used to determine the rate of solute uptake. SNL has expertise in proteoliposomes and could impact this research area.

Cloning and expressing transporters

Methods to clone and express transporters such as galP and glf have been reported previously. For example, plasmid pCLvGalP1 is a pCL1920 derivative containing the galP gene controlled by the trc promoter. Expression of this plasmid has been shown to allow simultaneous utilization of glucose and arabinose in PTS-negative E-coli mutants (24). Plasmid pLOI740 contains the full Z. mobilis glf-zwf-edd-glk operon and has been shown to restore glucose utilization in glucose-negative mutants of E-coli (29).

Biophysical studies

A great deal is known about the structure and mechanism of the lactose permease from work over several decades by the Kaback group (40,41). LacP is a member of the MFS. It may be possible to use this insight to guide studies of other MFS transporters. For instance, this work has led to a proposed mechanism involving binding and dissociation of the substrate on either side of the membrane wherein binding leads to a conformational change (tilting) of the transmembrane helices that exposes the substrate to the opposite side of the membrane. The residues involved in substrate binding and the transmembrane helices involved in the conformational change have also been identified (42). A promising avenue of research is to compare the sequence of LacP with other sugar porters in the MFS and use this information to guide mutational studies to understand the basis for substrate promiscuity. Insight from this previous work should aid one to determine the most important residues for mutational studies. The effects of these mutations on the uptake rates of a range of C5 and C6 sugars should be explored. Rates of efflux or equilibrium exchange across a membrane can be studied by exposing concentrated

suspensions of right side-out membrane vesicles to solutions of radiolabeled sugars. Molecular dynamics simulations could possibly supplement the experimental work.

Other areas of research

Pursuing avenues to increase the tolerance of microorganisms to ethanol concentrations greater than 5% is another area of research with potential to significantly impact this industry. In that regard, understanding the effects of ethanol on lipid membranes and cellular processes will be important. SNL has unique expertise in molecular modeling of lipid membranes and the effects of ethanol on membrane properties (44). Such calculations may reveal strategies to improve the ethanol resistance of microbes. Genetic engineering could prove fruitful as well. Another promising area of research related to this involves ways to sequester/separate and then recover ethanol. Finally, research into solvent pumps could also lead to avenues for enhanced ethanol tolerance as well as tolerance to fermentation inhibitors.

References

1. http://www.unh.edu/p2/biodiesel/article_alge.html

2. NREL close out report on The Aquatic Species Program - http://govdocs.aquake.org/cgi/content/abstract/2004/915/9150010

3. Toda, M., Takagaki, A., Okamura, M., Kondo, J. N., Hayashi, S., Domen, K., Hara, M. Nature 438, 178 (2005).

4. Biomass oil analysis: research needs and recommendations, NREL/TP-510-34796, 44, 2004.

5. Read, S. M., Bacic, T. Science 295, 59 (2002) and references therein

6. Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K., Turner, S. R. PNAS 100, 1450 (2003).

7. Doblin, M. S., Kurek, I., Jacob-Wilk, D., Delmer, D. P. Plant Cell Physiol. 43, 1407 (2002).

8. Jarvis, M. C., McCann, M. C. Plant Physiol. Biochem. 38, 1, 2000.

9. Scattering Studies of Plant Cell Walls. A M Donald and O M Astley. Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK. Received 10th December 2001; www.ccp13.ac.uk/fdr/2001/html/pdf/print/19.pdf

10. Carrard, G., Koivula, A., Soderlund, H., Beguin, P. PNAS 97, 10342 (2000).

11. Zhang, Y., Lynd, L. R. Anal. Chem. 75, 219, (2003).

12. Mansfield, S. D., Mooney, C., Saddler, J. N. Biotechnol. Prog. 15, 804, (1999).

13. Zhang, Y., Lynd, L. R. Biotechnol. & Bioeng. 88, 797 (2004).

14. Zhang, Y., Himmel, M. E., Mielenz, J. R. Biotechnol. Advances 24, 452 (2006).

15. Demain, A. L., Newcomb, M., David Wu, J. H. Microbiol. And Molec. Biol. Rev. 69, 124 (2005).

16. Fierobe, H. P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J. P., Bayer, E. A. J. Biol. Chem. 276, 21257 (2001)

17. Fierobe, H. P., Bayer, E. A., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y., Belaich, J. P. J. Biol. Chem. 277, 49621 (2002)

18. Mingardon, F., Perret, S., Belaich, A., Tardif, C., Belaich, J. P., Fierobe, H. P., Appl. and Env. Microbiol. 71, 1215 (2005).

19. Fierobe, H. P., Mingardon, F., Mechaly, A., Belaich, A., Rincon, M. T., Pages, S., Lamed, R., Tardif, C., Belaich, J. P., Bayer, E. A. J. Biol. Chem. 280, 16325 (2005)

20. Sabathe, F., Soucaille, P. J. Bacteriology 185, 1092 (2003).

21. Carvalho, A. L., Dias, F. M., Prates, J. A. M., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M. A., Romao, M. J., Fontes, C. M. G. A. PNAS 100, 13809 (2003).

22. Adams, J. J., Pal, G., Jia, Z., Smith, S. P. PNAS 103, 305 (2006).

23. Tchieu, J. H., Norris, V., Edwards, J. S., Saier, M. H. Jr. J. Mol. Microbial. Biotechnol. 3, 329 (2001).

24. Hernandez-Montalvo, V., Valle, F., Bolivar, F., Gossett, G. Appl. Microbiol. Biotechnol. 57, 186 (2001).

25. Nichols, N. N., Dien, B. S., Bothast, R. J. Appl. Microbiol. Biotechnol. 56, 120 (2001).

26. Dien, B.S., Nichols, N. N., Bothast, R. J. J. Ind. Microbiol. & Biotechnol. 29, 221 (2002).

27. Lindsay, S. E., Bothast, R. J., Ingram, I. O. Appl. Microbiol. Biotechnol. 43, 70 (1995).

28. Flores, N., Xiao, J., Berry, A., Bolivar, F., Valle, F. Nature Biotechnol. 14, 620 (1996).

29. Snoep, J. L., Arfman, N., Yomano, L. P., Fliege, R. K., Conway, T., Ingram, L. O. J. of Bact. 176, 2133 (1994).

30. Weisser, P., Kramer, R., Sahm, H., Sprenger, G. A. J. of Bact. 177, 3351 (1995).

31. Kimata, K., Takahashi, H., Inada, T., Postma, P. Aiba, H. PNAS 94, 12914 (1997).

32. Chatterjee, R., Millard, C. S., Champion, K., Clark, D. P., Donnelly, M. I. Appl. and Environ. Microbiol. 67, 148 (2001).

33. Paulsen, I. T., Sliwinski, M. K., Saier, M. H. Jr. J. Mol. Biol. 277, 573 (1998).

34. Paulsen, I. T., Sliwinski, M. K., Nelissen, B., Goffeau, A., Saier, M. H. Jr. FEBS Letters 430, 116 (1998).

35. Saier, M. H. Jr., Eng, B. H., Fard, S., Garg, J., Haggerty, D. A., Hutchinson, W. J., Jack, D. L., Lai, E. C., Liu, H. J., Nusinew, D. P., Omar, A. M., Pao, S. S., Paulsen, I. T., Quan, J. A., Sliwinski, M., Tseng, T.T., Wachi, S., Young, G. Biochimica et Biophysica Acta 1422, 1, (1999).

36. Paulson, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R., Saier, M. H. Jr. J. Mol. Biol. 301, 75 (2000).

37. Linton, K. J., Higgins, C. F. Mol. Microbiol. 28, 5, (1998).

38. Pao, S. S., Paulsen, I. T., Saier, M. H. Jr. Microbio. and Molec. Biol. Rev. 62, 1, (1998).

39. Weisser, P., Kramer, R., Sprenger, G. A. Appl. & Environ. Microbiol. 62, 4155 (1996).

40. Gardonyi, M., Jeppsson, M., Liden, G., Gorwa-Grausland, M. F., Hahn-Hagerdal, B. Biotechnol. & Bioeng. 82, 820 (2003).

41. Kaback, H. R., Sahin-Toth, M., Weinglass, A. B. 2, 610 (2001).

42. Kaback, H. R. C. R. Biologies 328, 557, (2005).

43. Vazquez-Ibar, J. L., Guan, L., Wienglass, A. B., Verner, G., Gordillo, R., Kaback, H. R. J. Biol. Chem. 279, 49214 (2004).

44. Frischknecht, A. L., Douglas Frink, L. J. Accepted for publication in Biophys. J. (2006).

Distribution list

3	MS1411	Michael Kent
1	MS1413	Kathe Andrews
1	MS9292	Blake Simmons
1	MS0895	Tony Martino
1	MS1413	Grant Heffelfinger

2	MS9018	Central Technical files, 8944
2	MS0899	Technical Library, 4536