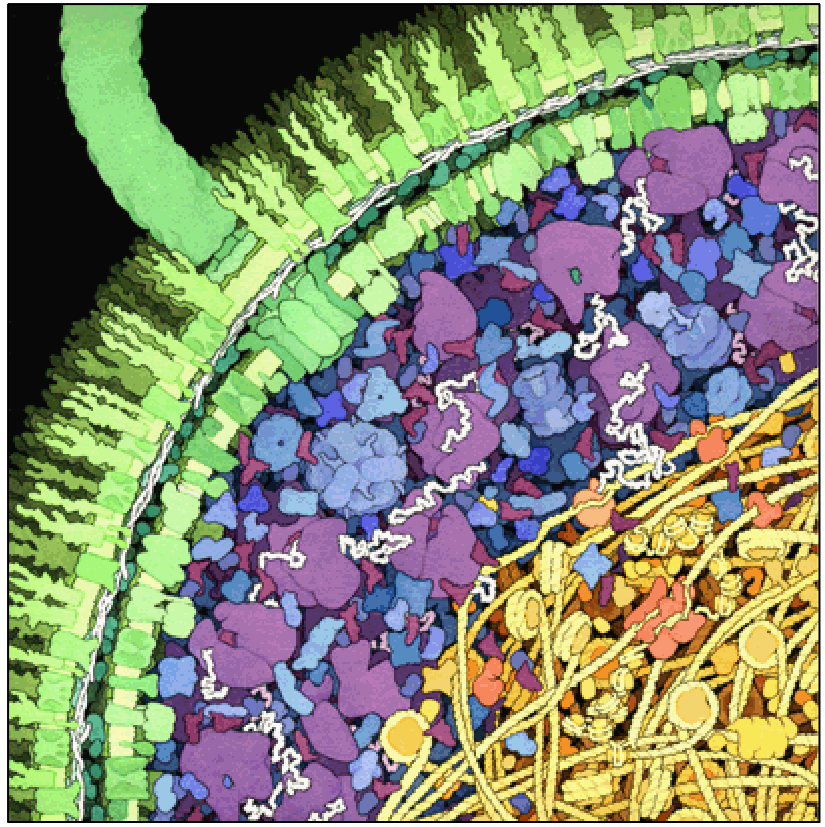
The role of Systems Biology in Bioenergy at ORNL



Brian H. Davison

ORNL

September 2006

President's State of the Union Address January 2006

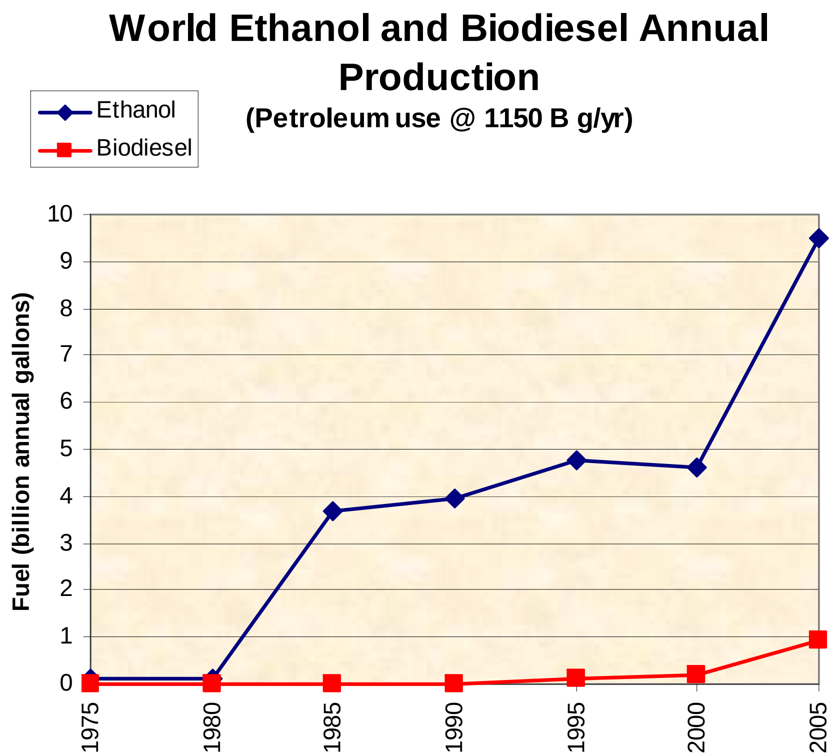
Keeping America competitive requires affordable energy. And here we have a serious problem: America is addicted to oil, which is often imported from unstable parts of the world.

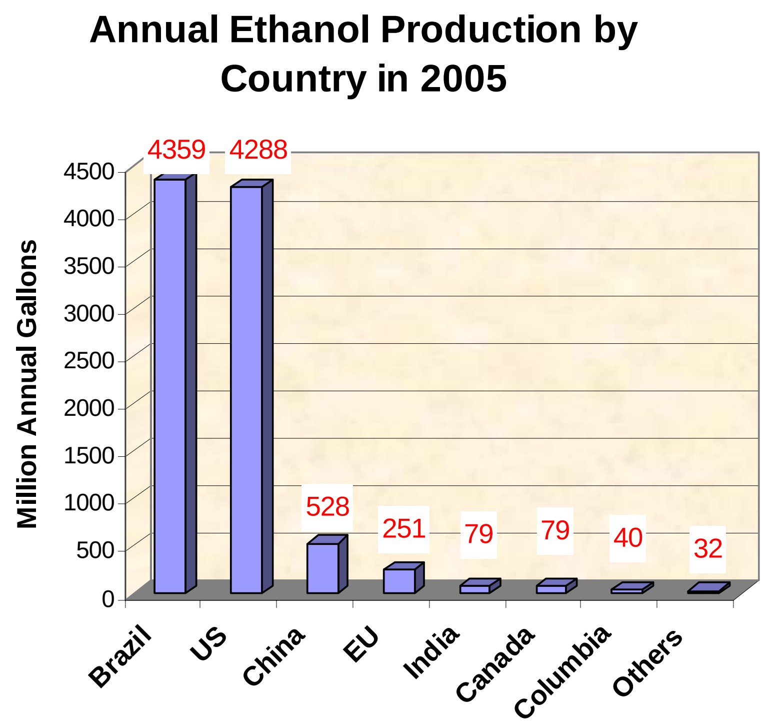
The best way to break this addiction is through technology…. and we are on the threshold of incredible advances…

So tonight I announce…push for breakthroughs in two vital areas…change how we power our homes and offices,…change how we power our automobiles.

Biofuel production

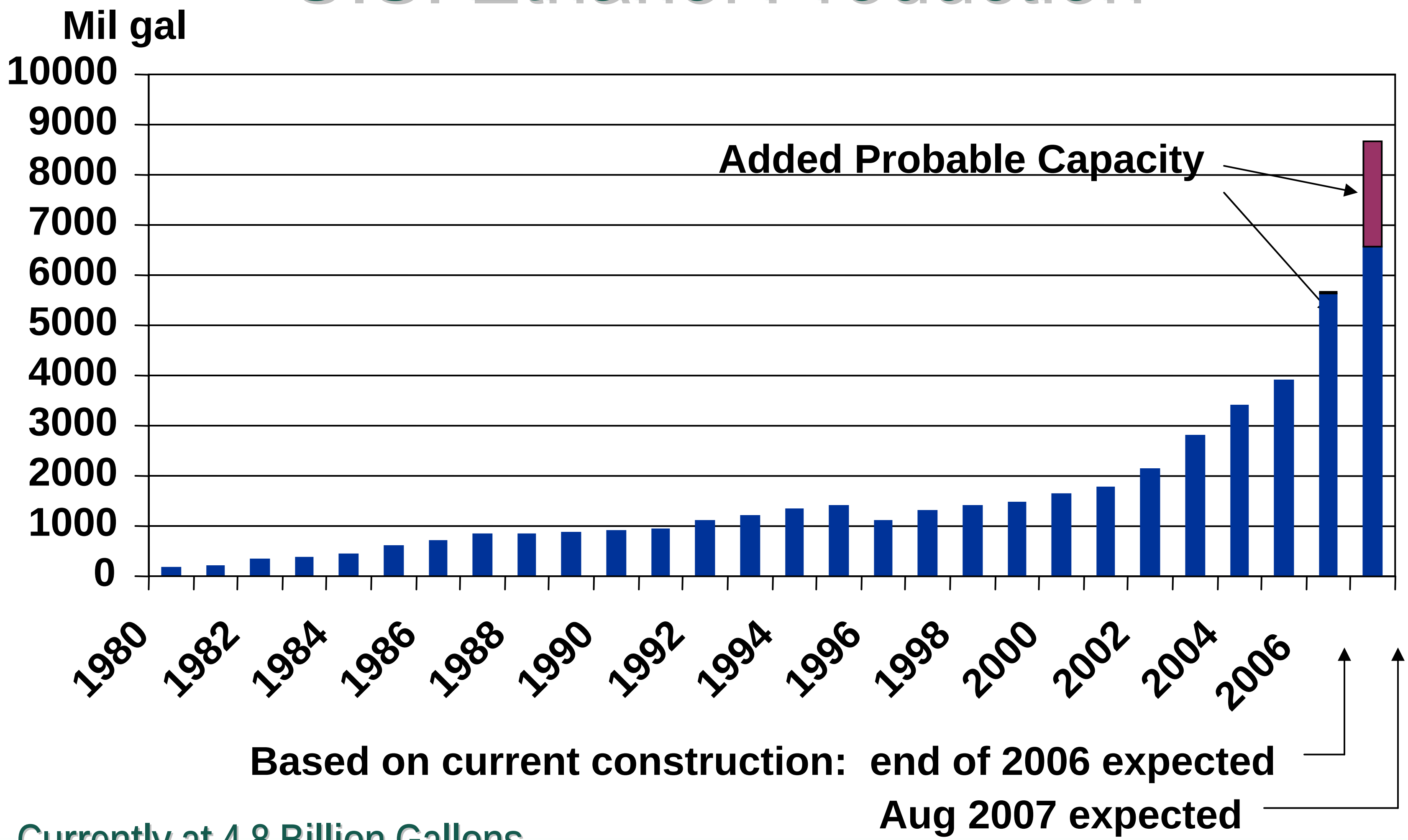
- Biofuel use worldwide is growing but…





- We have a long way to go!

U.S. Ethanol Production

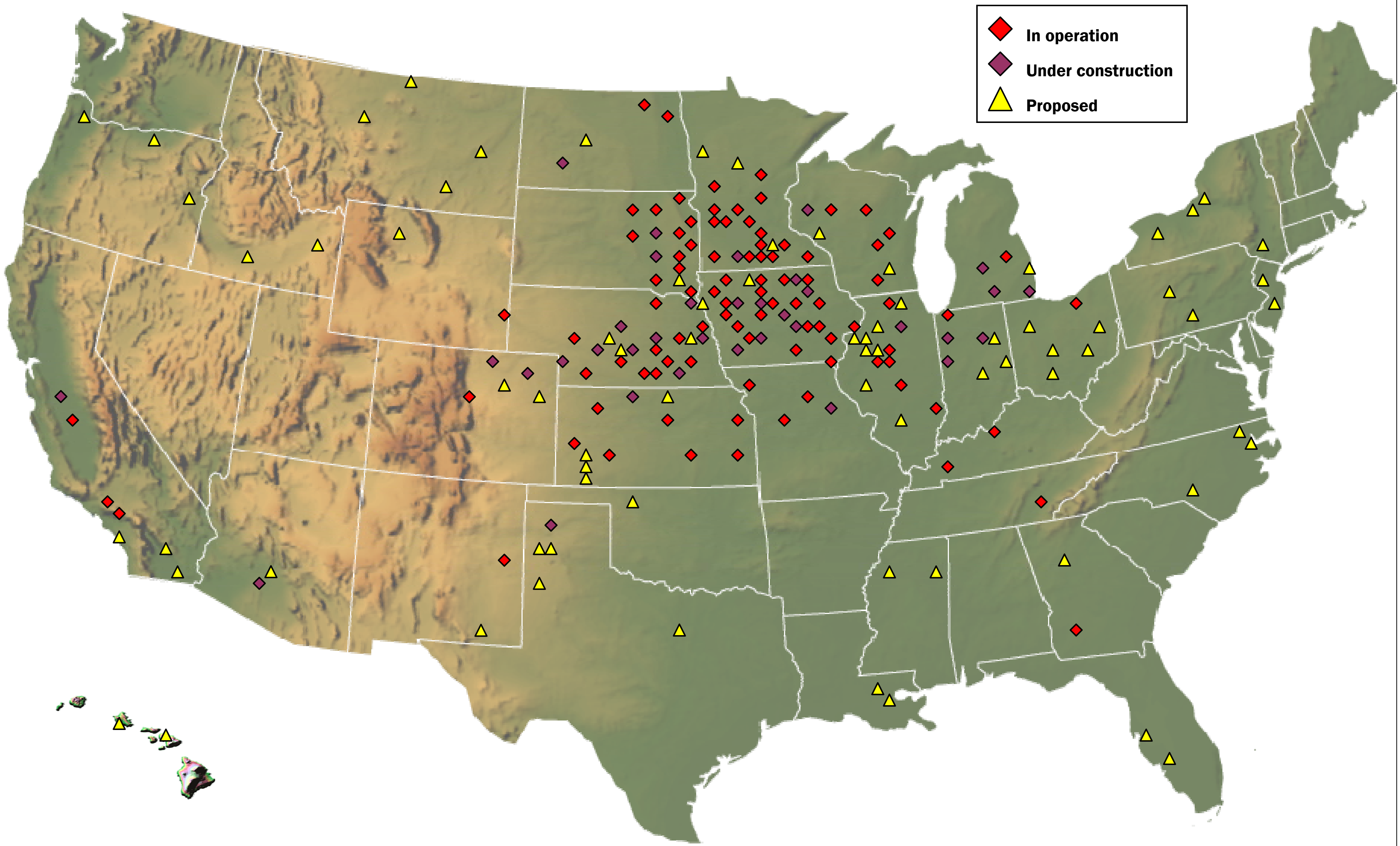


Currently at 4.8 Billion Gallons

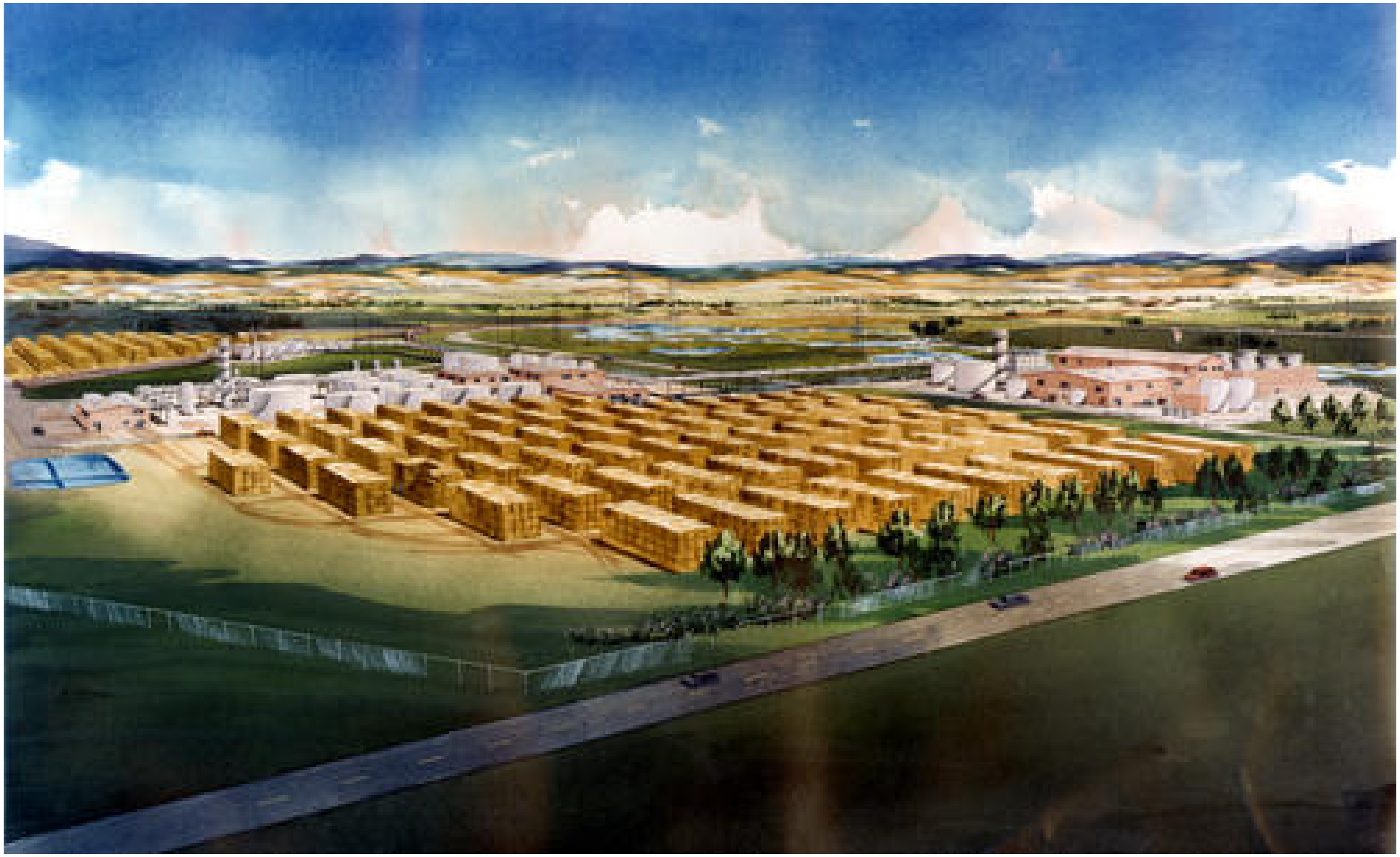
NCGA
www.ncga.com

US Ethanol Plants

AS OF: March 2006



Artist Vision of a Biorefinery with Biomass Storage Adjacent



Source: NREL Pix

Bio-energy – why now?

- Bio-energy is seen as one of the few areas for significant near to mid term impact
  - "Moving the needle"
  - Energy Security and Climate Change benefits
  - Some NGOs are strongly supportive
- It is emphasized by Bush and DOE Secr. Bodman
  - "breaking addiction to oil", "Science to Energy"

Bioenergy Funding Outlook

- Corn Starch ethanol capital investments are rapidly increasing
- OBP has $30-60M in new funds for increased deployment activities
- Private investment and industrial investment increasing
  - Gates, Khosla, BP bioenergy center ($500M)
- BER will establish 2 bioenergy research centers at $25M/y * 5y
- BER has $20-40M other GTL funds that may emphasize bioenergy
- USDA is increasing emphasis
- DoD has BAA for short term and long term needs

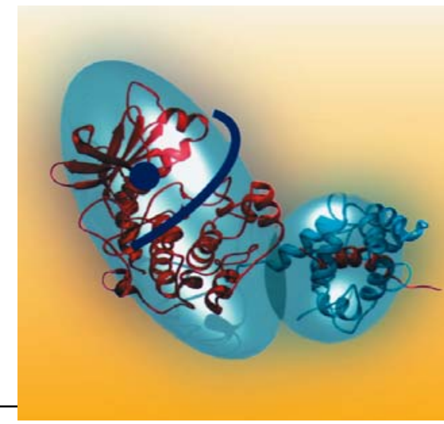
Systems Biology:

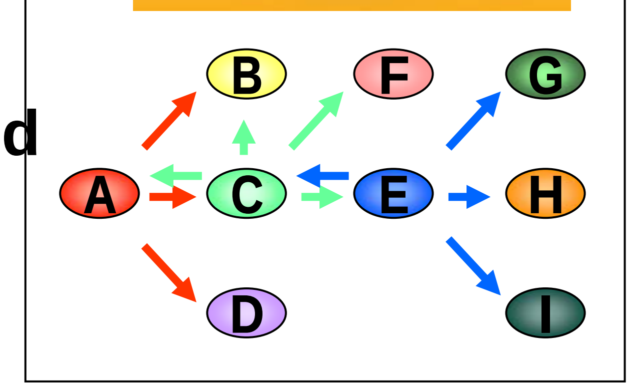
Philosophical approach to consider biology as integrated complex microbial and metazoan systems including:

Molecular complexes & interactions

Molecular networks including cell signaling and gene regulation

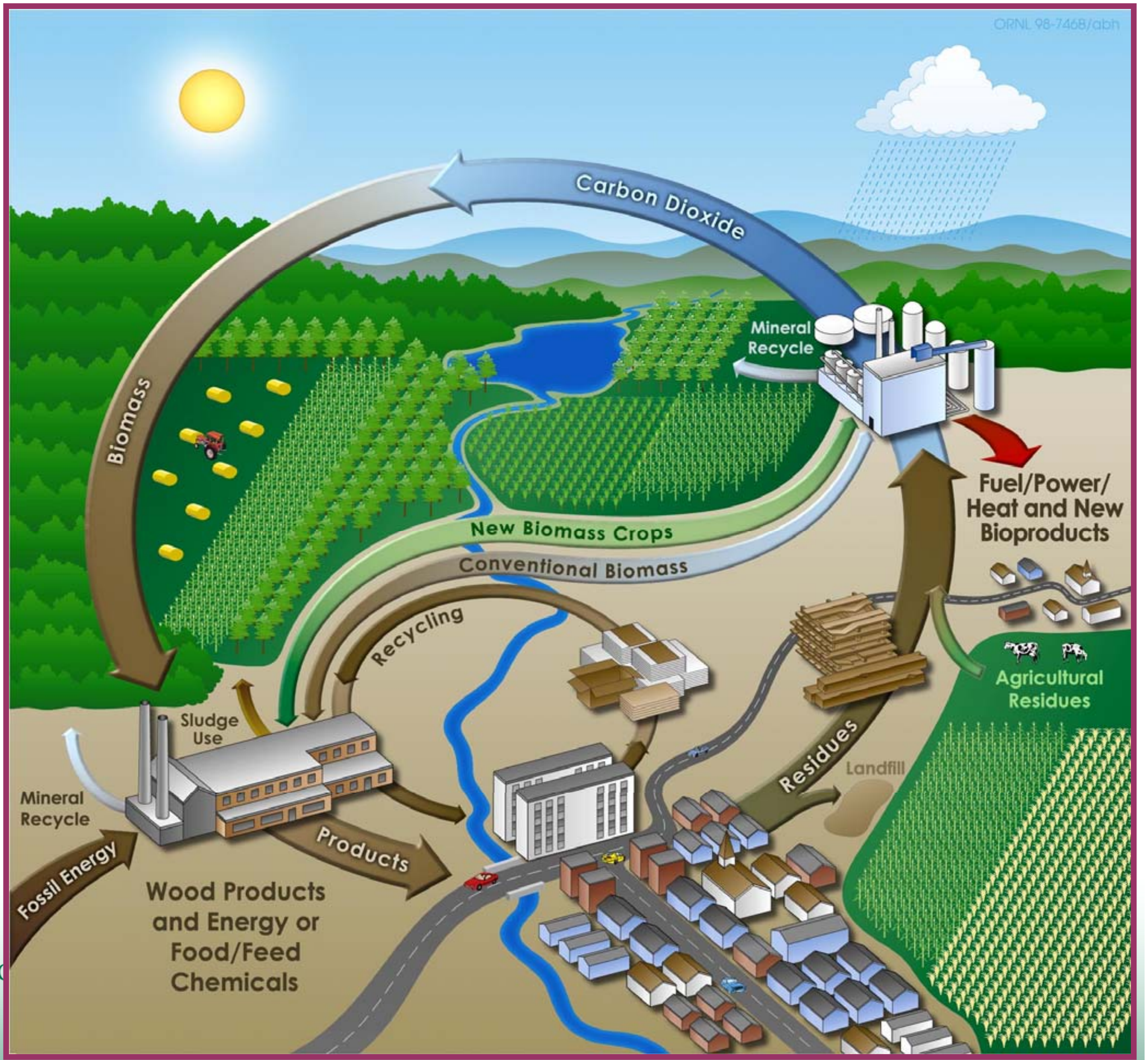
Experimental approach to generate and analyze HTP data



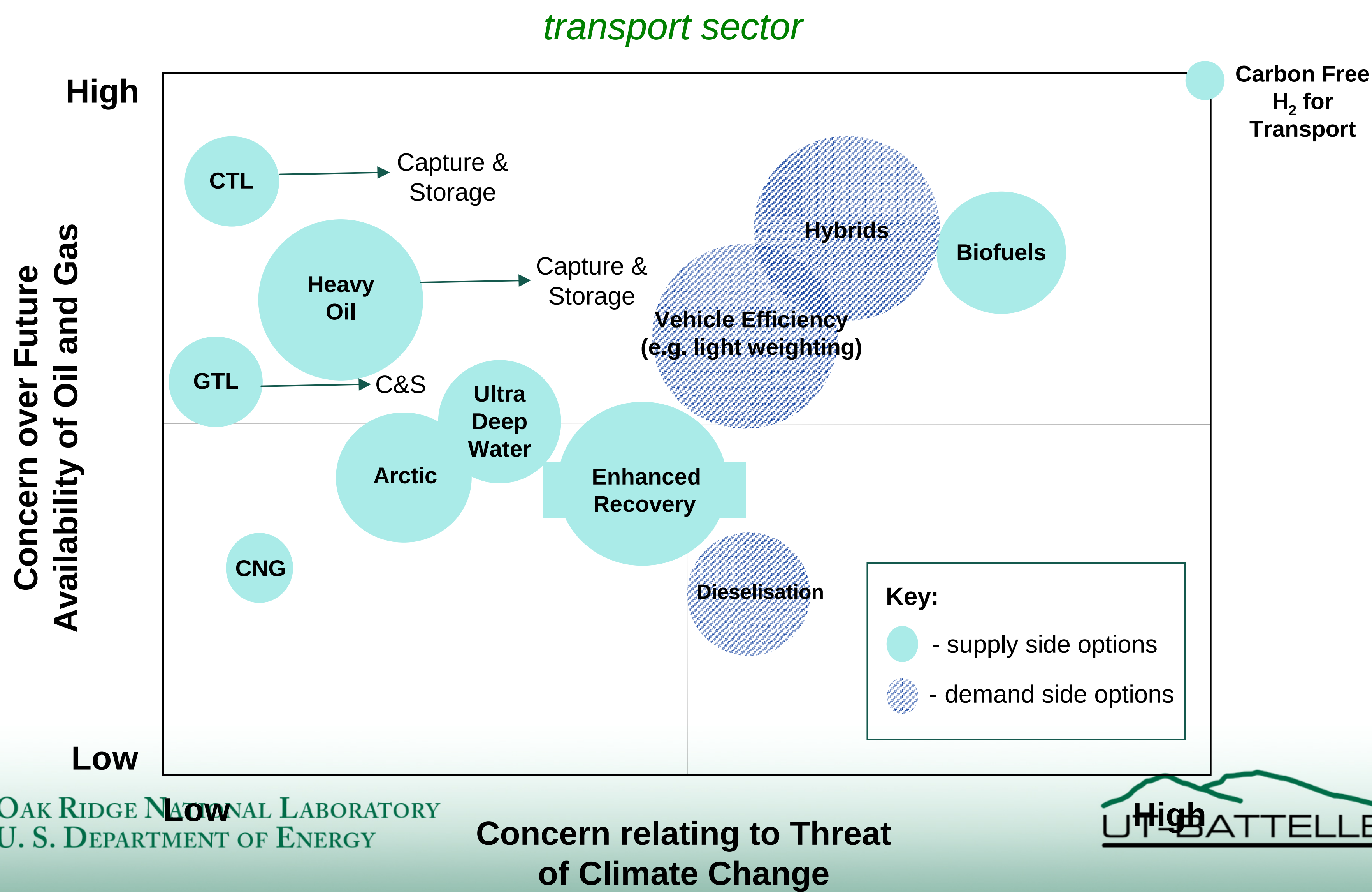


DOE-Office of Biomass Programs Mission And Strategic Goals

- Reduce dependence on foreign petroleum
- Create a new, domestic bioindustry



Technology Options for Transportation – from BP



Issues and facts in corn ethanol (EtOH)

- In the conversion routes chart, we should focus on ethanol, biodiesel, hydrogen
- New corn dry mill plant for ethanol
  - Average cost to build - $100M
  - Produce ~ 70M gl of EtOH, from 26 M bu or 0.7 Mt of corn
  - Process cost ~$0.6/gal
- The one EtOH plant in TN uses corn largely imported from other states
- US production of EtOH from corn is ~4B gal in 2005, another 2.5 B gal capacity planned by 2008
- The RFS demand is 6 B gal to replace MTBE in gasoline
- Capacity to grow corn limits annual EtOH production to less than 15 B gal/yr without impacting food and feed

Corn production in Southeast (2005, MM bu)

Alabama	24
Arkansas	30
Florida	3
Georgia	30
Kentucky	156
Louisiana	45
Mississippi	47
Missouri	330
North Carolina	84
South Carolina	33
Tennessee	77
Virginia	42
West Virginia	3
Total SE	904
Total US	*11,112*

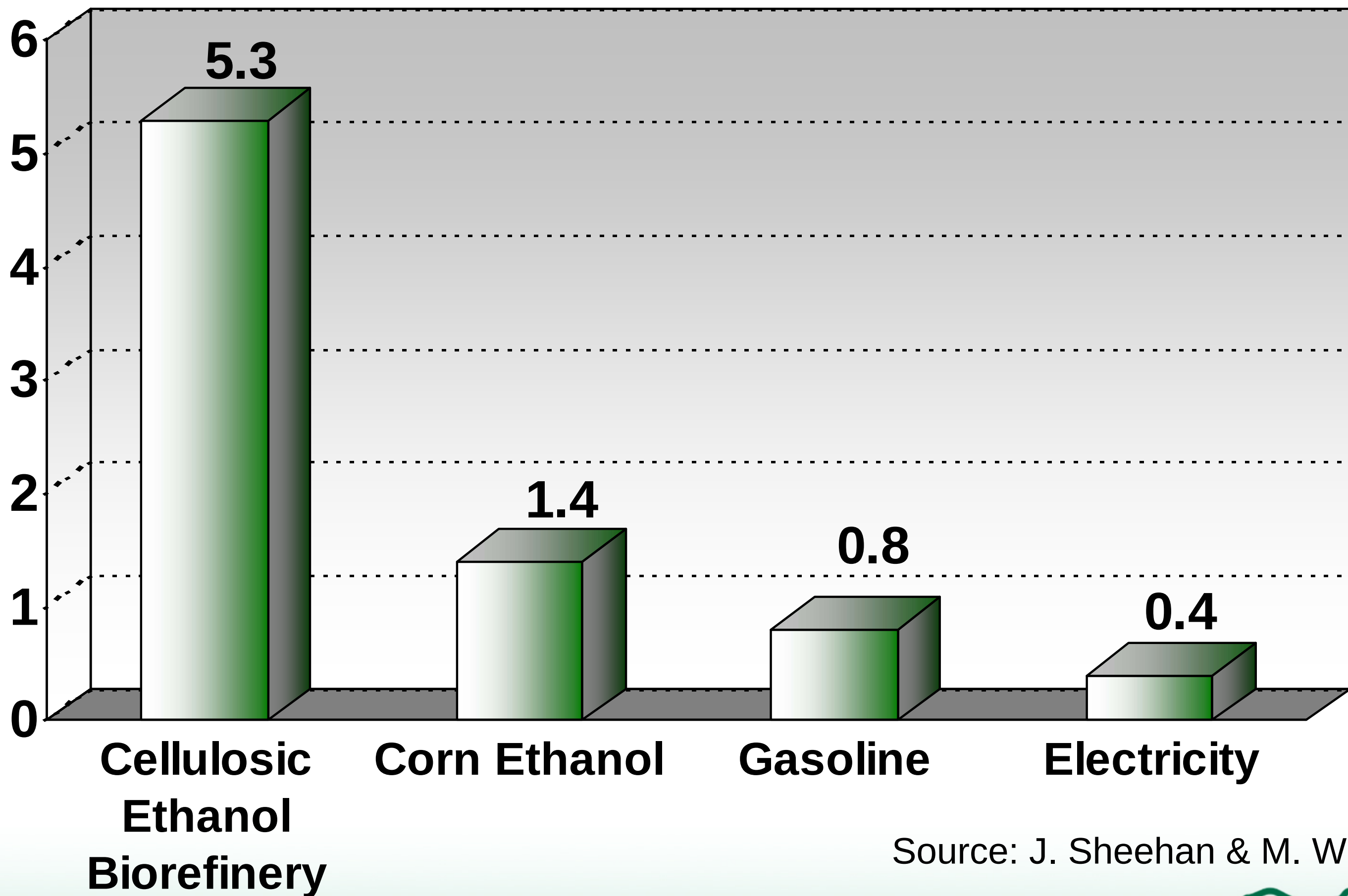
Ref: NCGA World of Corn, 2005

Issues and facts in cellulosic ethanol

- Cellulose to EtOH feasible now but not economical (>$2/gl)
- R&D is expected to bring this down to <$1/gl total price
- Variety of sources of cellulose, more abundant supply that corn, better suited for SE USA
- Almost 5 times more energy delivered per fossil energy unit used for cellulose to EtOH compared to corn (see next chart)
- Estimated construction cost to build cellulose to EtOH production plant with comparable 70 M gal/y capacity is
  - Today's technology estimate is ~$200M (includes both potential $100M reduction from improvements and an added risk cost of $100M)
  - Ultimate future est. construction cost comparable to $100M corn plant
  - Economies of scale make larger better – this 3-4 M ton biomass/d is at lowest end, most designs are 5 to 25 M ton biomass/d
- Huge opportunity in the Southeast to have alternate use for forest products (pulp & paper diminishing)
- Gasoline pipelines not likely usable by ethanol

Fossil Energy Replacement Ratio

Fossil Energy Ratio (FER) = $\frac{\text{Energy Delivered to Customer}}{\text{Fossil Energy Used}}$

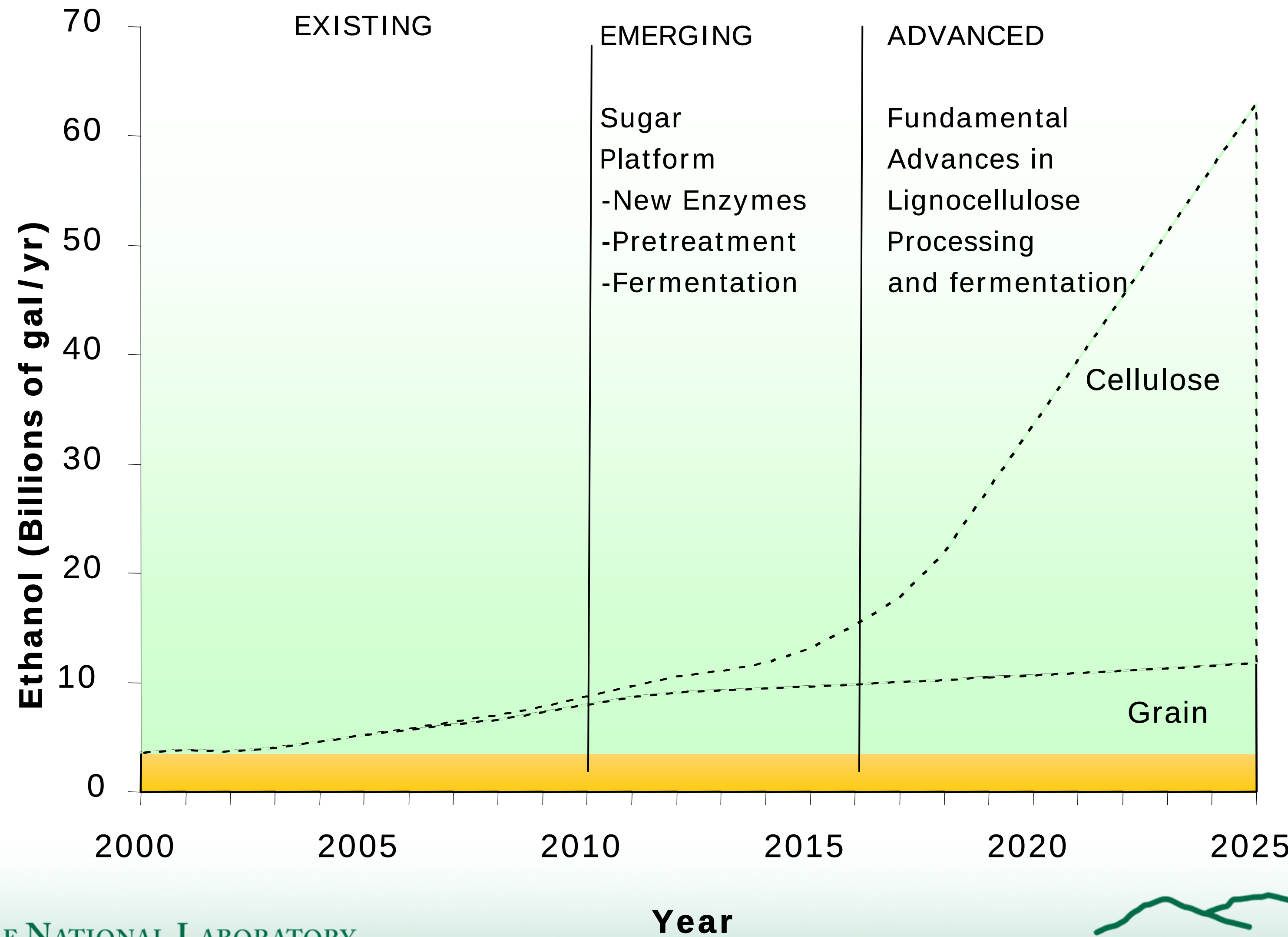


Source: J. Sheehan & M. Wang (2003)

DOE has set a “30 x 30” goal
Biofuels for 30% of current transportation by 2030

- Corn starch is utilized to 10-15 B gal/y
- Cellulosics must be used for remainder
  - First ag residues and wastes
  - Then dedicated bioenergy crops
- Advanced technologies will be required
  - Conversion cost must lower to 20-30% of total cost
  - Feedstock costs (grower payment) will increase as demand rises
  - Risk capital cost is barrier
- Market subsidies suggested for penetration and technology development to about 2020 if goals met
  - Ethanol subsidy continues
  - Biomass suppliers support
  - Loan guarantees for first of kind technology and crops

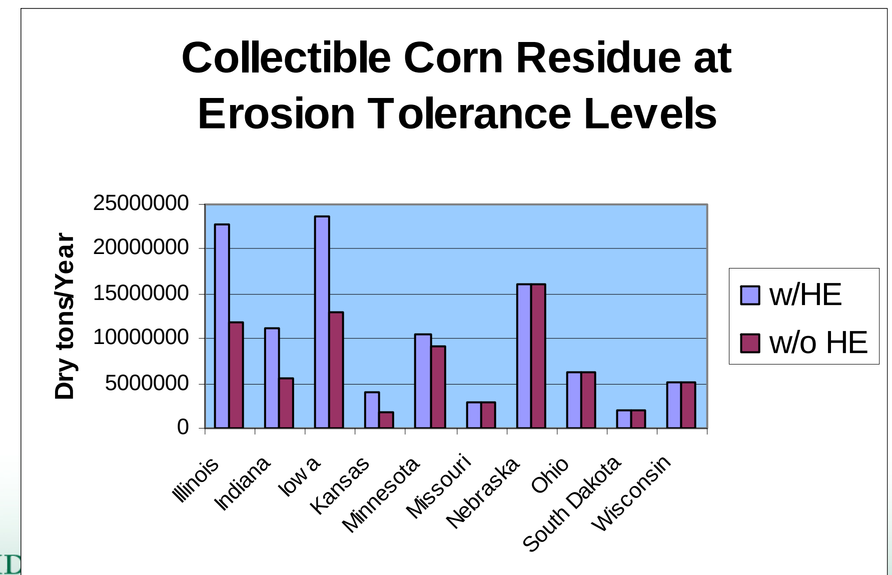
The DOE Biofuels Vision

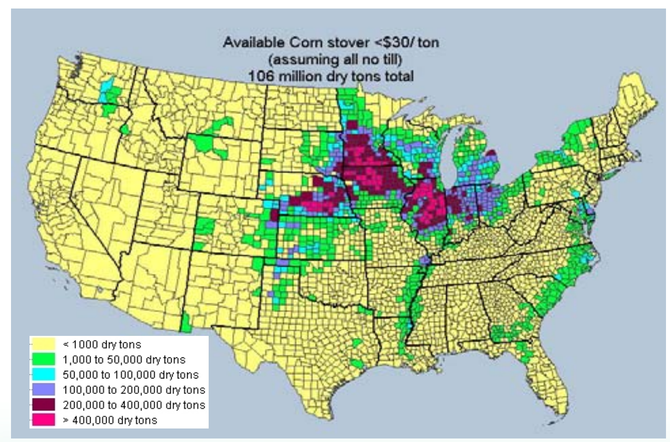


Modified from Richard Bain, NREL

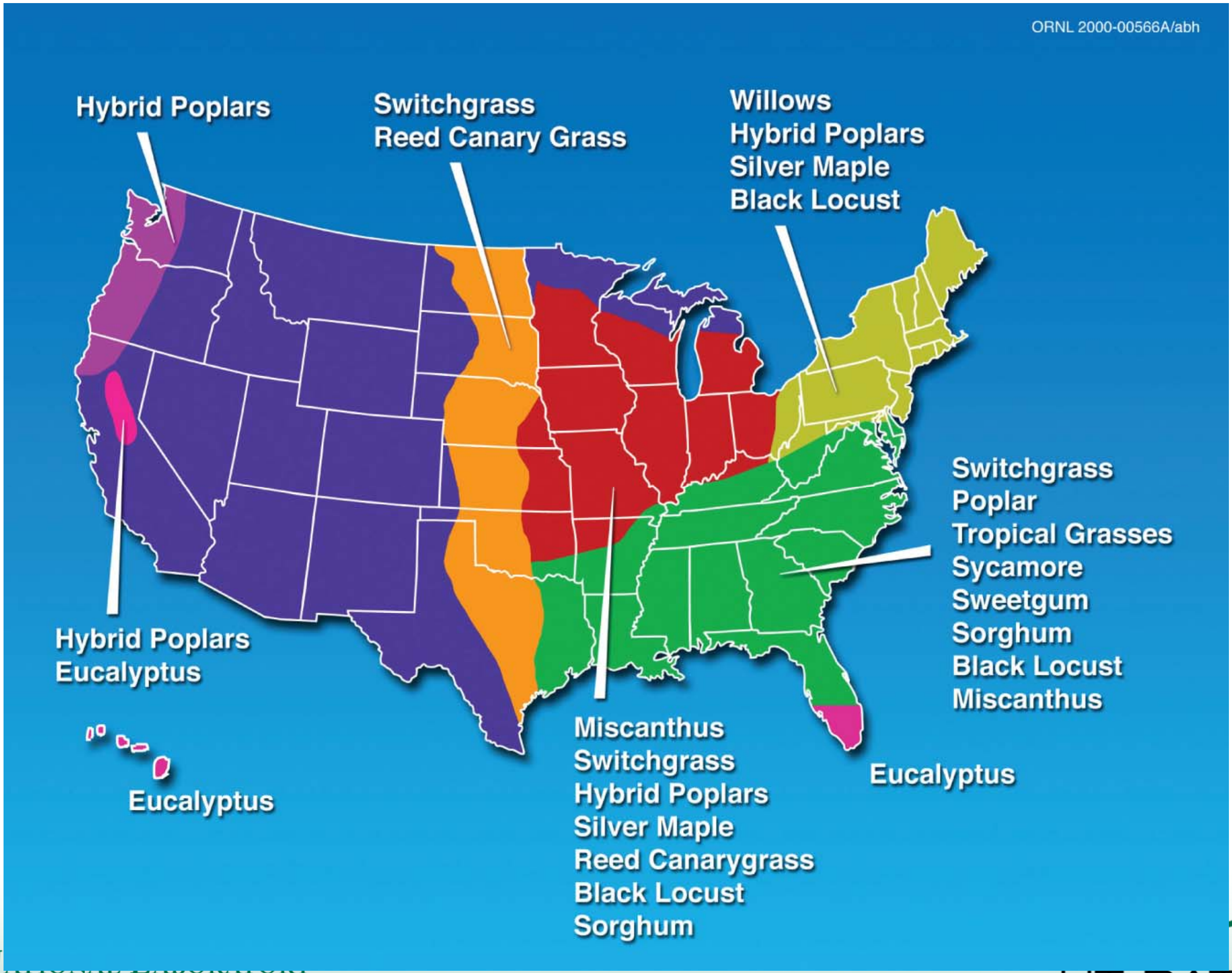
Biomass analysis capabilities at ORNL

- Biofeedstock supply logistics
- Agricultural residue sustainability
- Biomass Supply Forecasting analysis



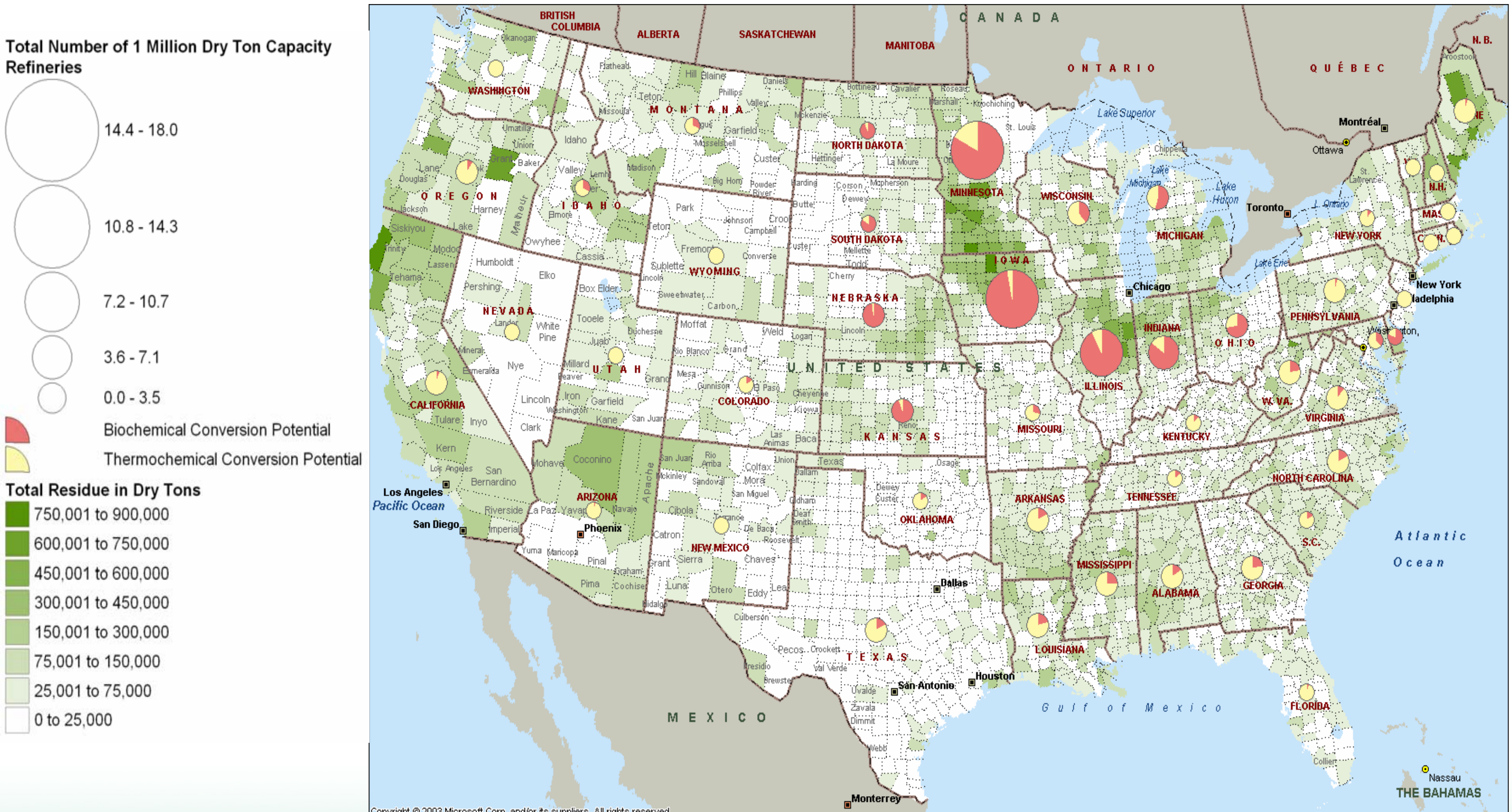


Geographic distribution of biomass crops



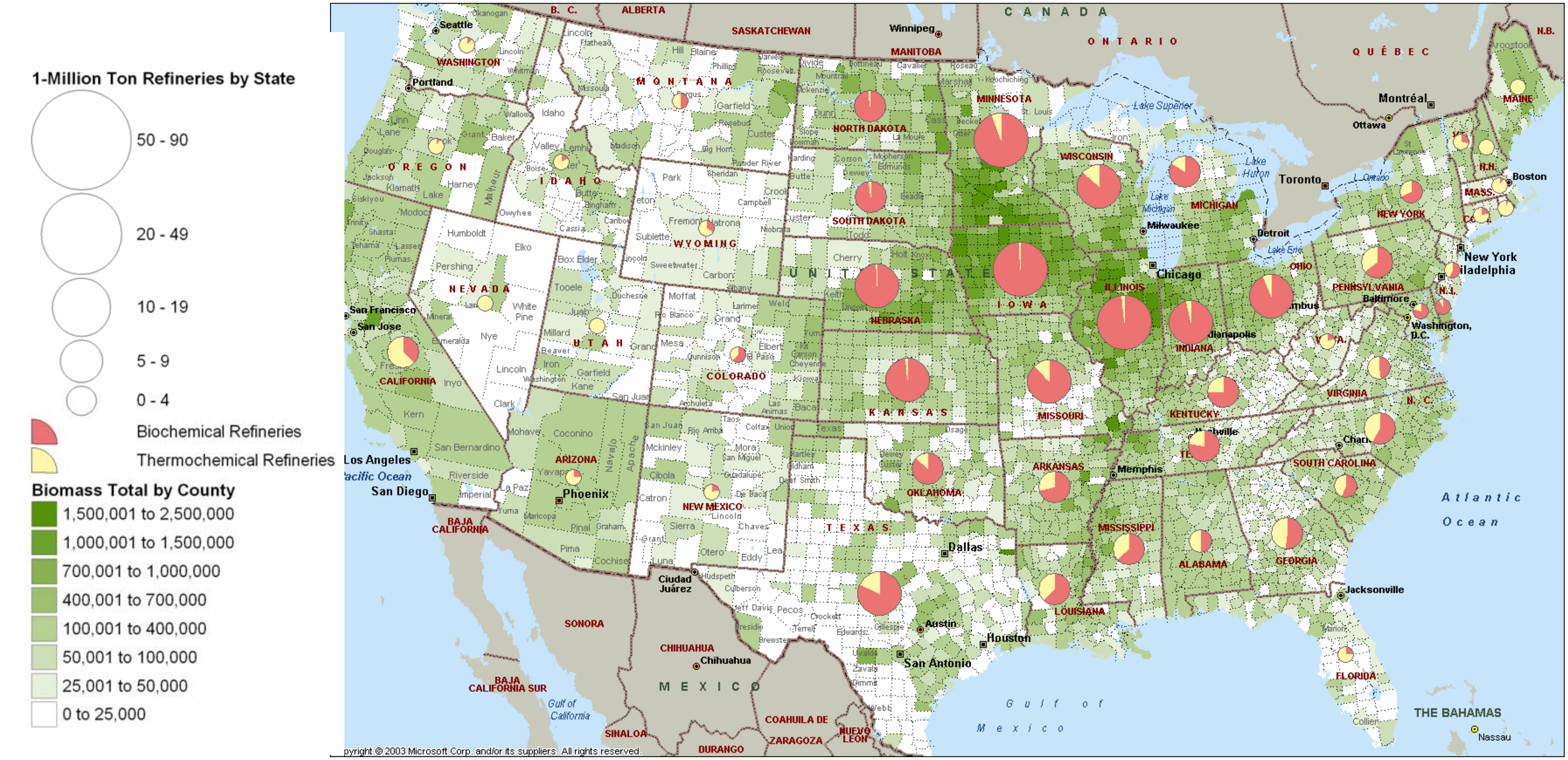
Wright et al DOE-ORNL-EERE

POTENTIAL BIOMASS RESOURCE AND REFINERY CAPACITY IN 2012



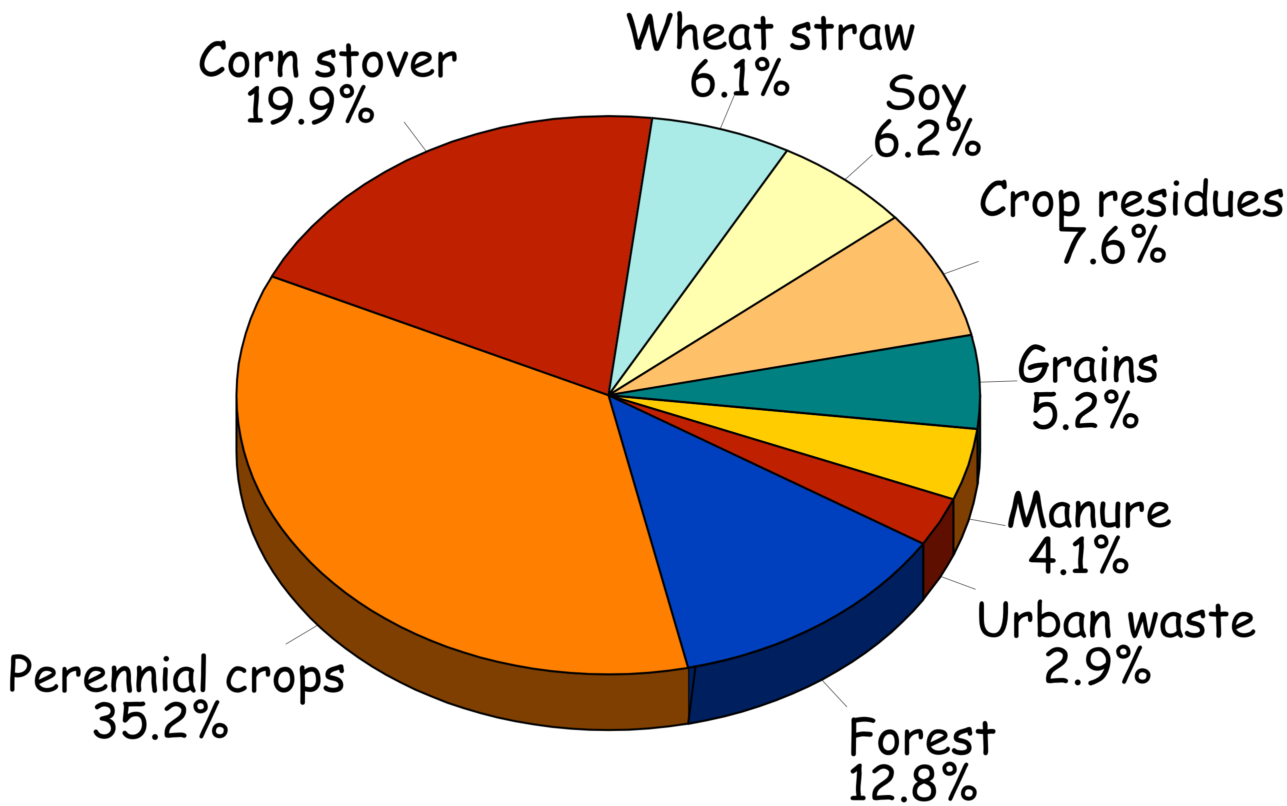
Logging residues, crop residues, switchgrass

POTENTIAL BIOMASS RESOURCE AND REFINERY CAPACITY IN 2030



Forest residues (all), crop residues, switchgrass

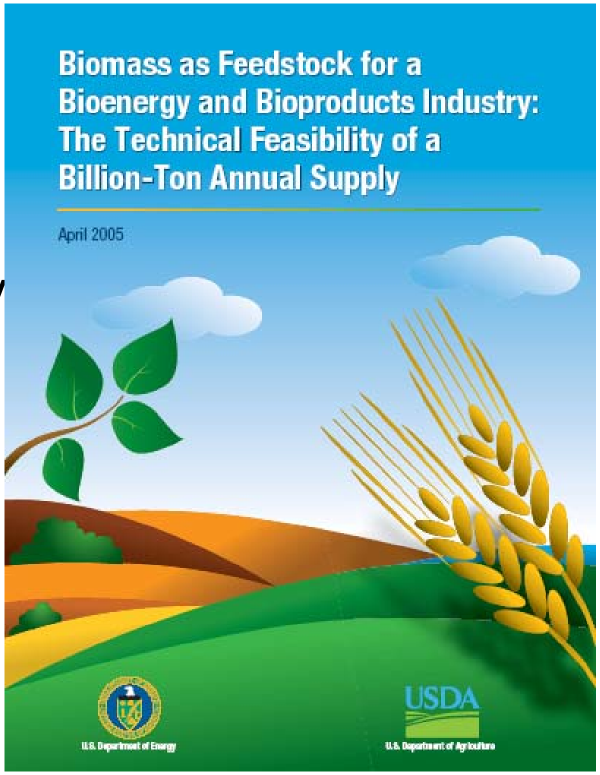
US Biomass inventory = 1.3 billion tons



From: R. D. Perlack *et al*., *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply* (DOE/GO-102005-2135, ORNL, 2005).

Current OBP Projections

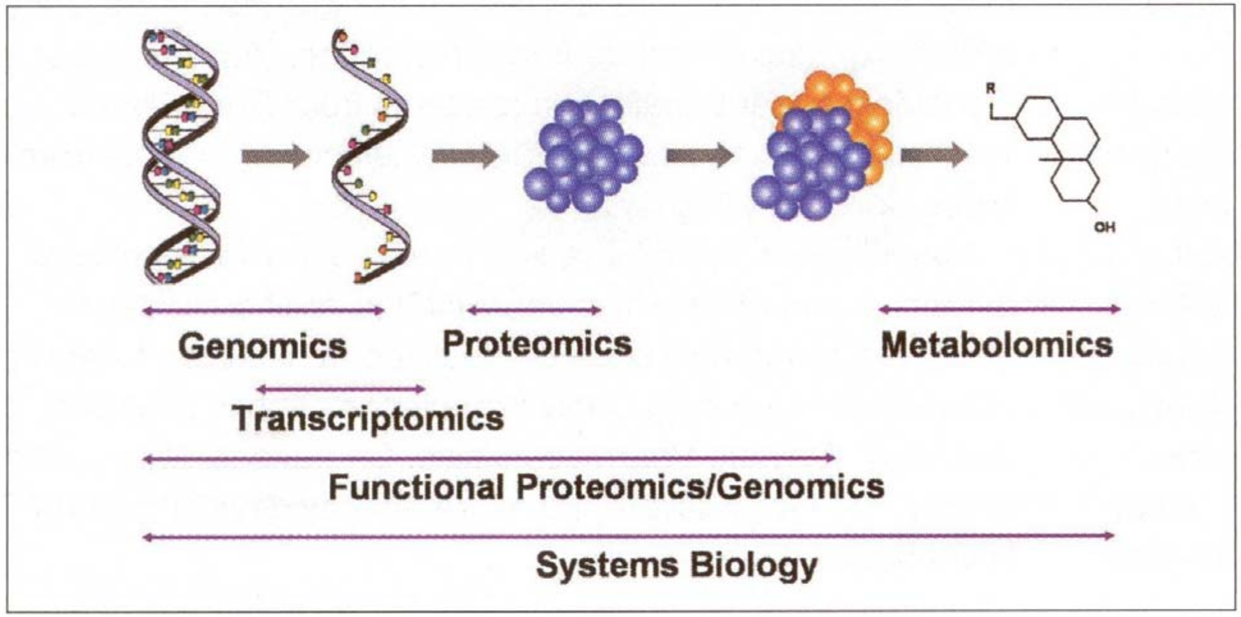
- Biomass Resource Base (ORNL)
  - Land resources of the U.S. can sustainably supply more than 1.3 billion dry tons annually and still continue to meet food, feed, and export demands
  - Realizing this potential will require R&D, policy change, stakeholder involvement
  - Required changes are not unreasonable given current trends
  - <u>Should be sufficient to replace 30% of current US petroleum requirements</u>



USDA/DOE Billion Ton Vision Paper

ORNL (and others) has "omic" capabilities for Systems Biology

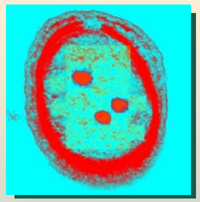
- Genomics
- Transcriptomics
- Proteomics
- Interactomics
- Metabolomics



- Together these can provide a deeper picture of how a microbe is functioning
- This can help identify where improvements need to be made

Components of ORNL Biology Program

Plant Sciences

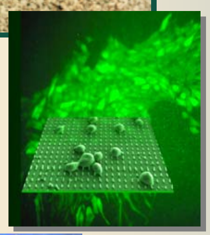


Microbiology

Mass spectrometry

Neutrons for structural biology

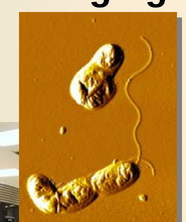
Bioinformatics

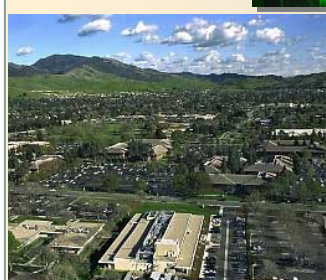


Biotechnology

Understand molecular interactions in cells and communities

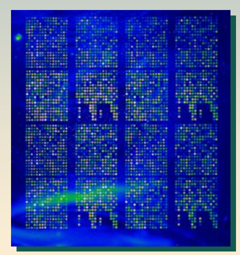
Imaging





Joint Genome Institute

Mouse genetics and comparative genomics

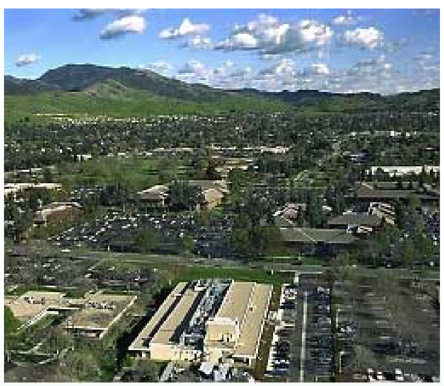


Expression profiling

Computational biology

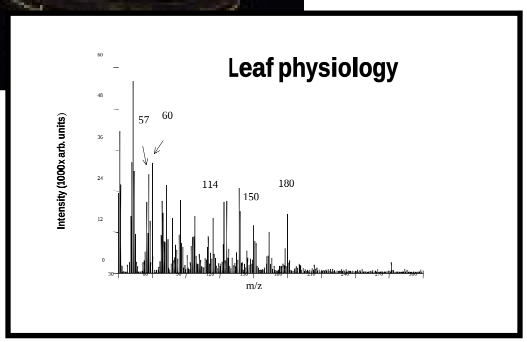
Genomics

- Genome sequence helps provide baseline of what an organisms capabilities are
- ORNL can help provide access to sequencing capability to JGI through off-cycle Lab Science program
- ORNL provides primary annotation to all microbes sequenced at JGI
- This can help OBP's industrial partners

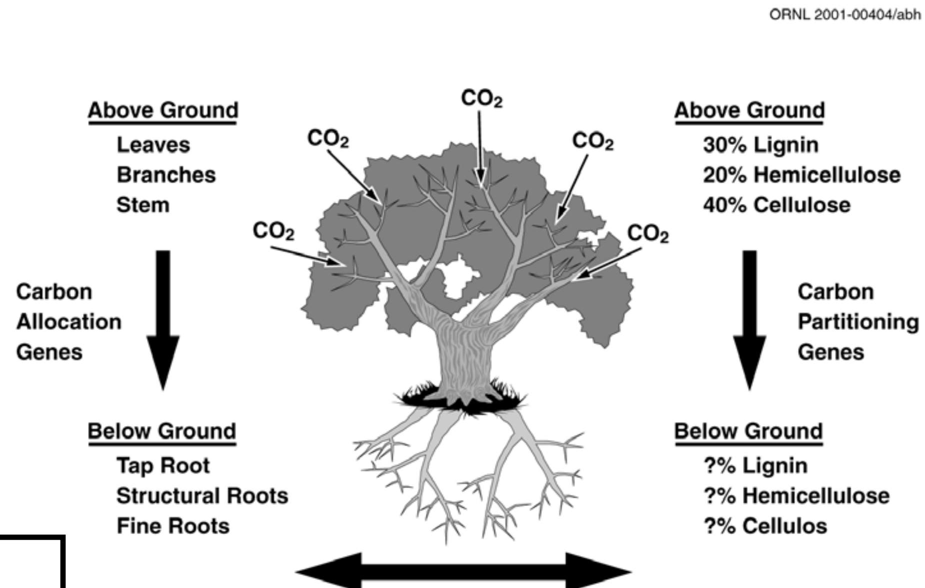


Joint Genome Institute

ORNL is a world leader in woody plant genomics – *designing crops for energy & C sequestration*



High-throughput Phenotying



Plant Traits Needed to Address all Parts of Equation:

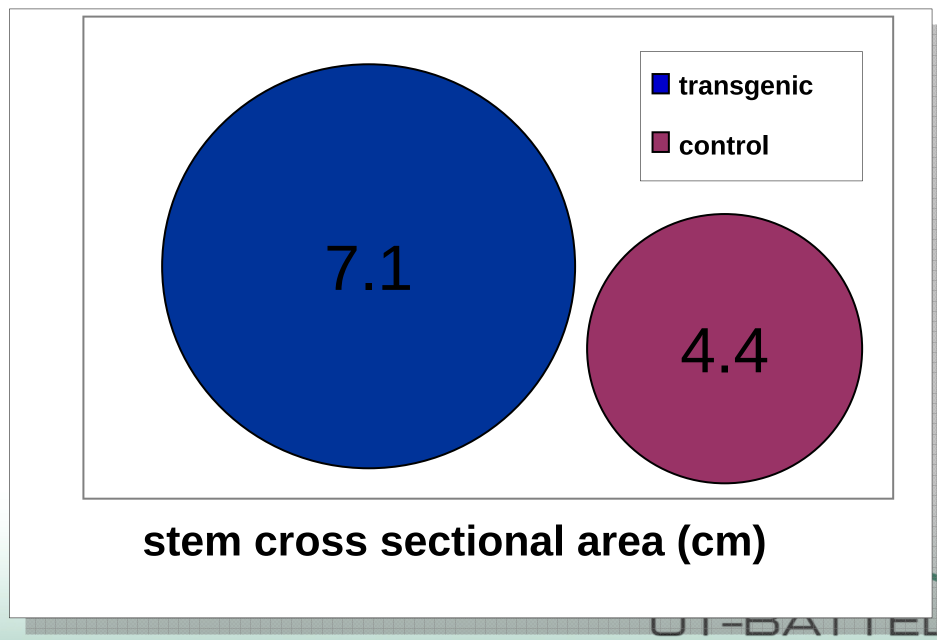
Targeted Goal	Traits & Technologies
Broaden Planting Range	• Tolerance to chronic and acute drought • High salt tolerance • Tolerance to heat shock • improvement in seedling growth under cold conditions
Increase tons per acre	• Increase in biomass in arabidopsis in the greenhouse (5X) • Increase in rice in the field (3X) • Increase in CO2 uptake (30%)
Dollars per acre	• Significant reduction in required nitrogen • Improvement in photosyn efficiency on low nitrogen (20%) • increase in root biomass
Gallons per ton	• Decreased lignin • Increased cellulose/ hemicellulose

Source: Modified from Ceres Corp.

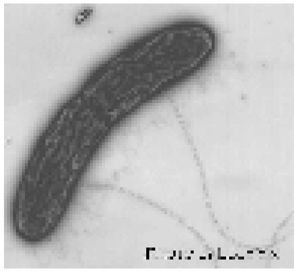
Populus - early results from genome sequence availability (Jerry Tuskan, ORNL)

90-day-old *Populus* cuttings

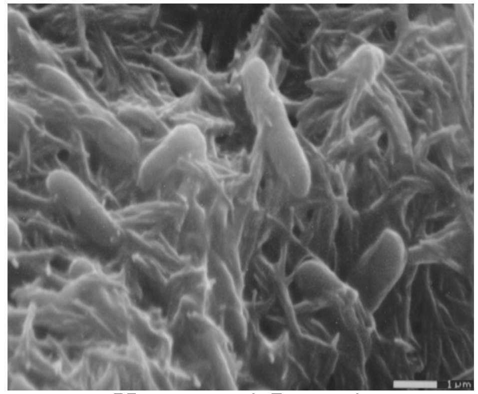
Using Poplar tree genome, the expression of one gene (IAA16.3) was altered. This resulted in enhanced radial growth of IAA16.3 transgenics vs. controls



Whole genome microarrays available



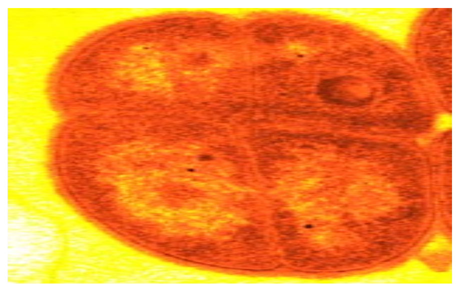
Geobacter metallireducens:
Metal-reducing bacterium



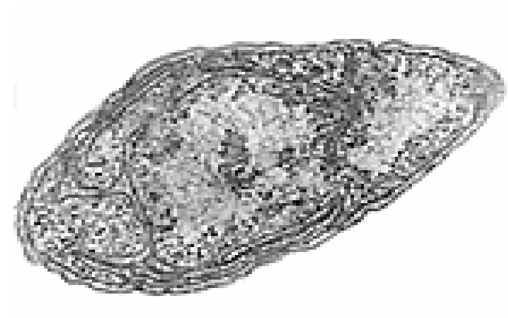
***Shewanella oneidensis* MR-1:**
Metal-reducing bacterium

***Rhodopseudomonas palustris*:**
Photosynthetic bacterium

***Desulfovibrio vulgaris*:**
Sulfate-reducing bacterium

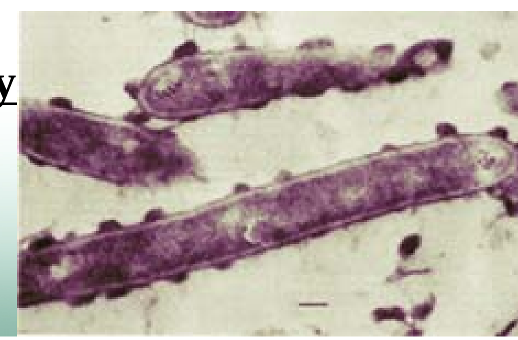


***Deinococcus radiodurans* R1:**
Radiation-resistant bacterium



***Nitrosomonas europaea*:**
Ammonium-oxidizing bacterium

***Clostridium thermocellum* now ready**
cellulolytic fermenting bacterium



Transcriptomics

- Tells what genes the microbe is trying to express under what conditions
- Typically requires a sequenced genome
- ORNL has capability to rapidly prepare new arrays (for quantities to small for commercial interest)
- New approaches (functional genomics) allow testing of mixed cultures and unknowns in a Functional Genome Array (FGA)

Summary of New FGA Probes Designed

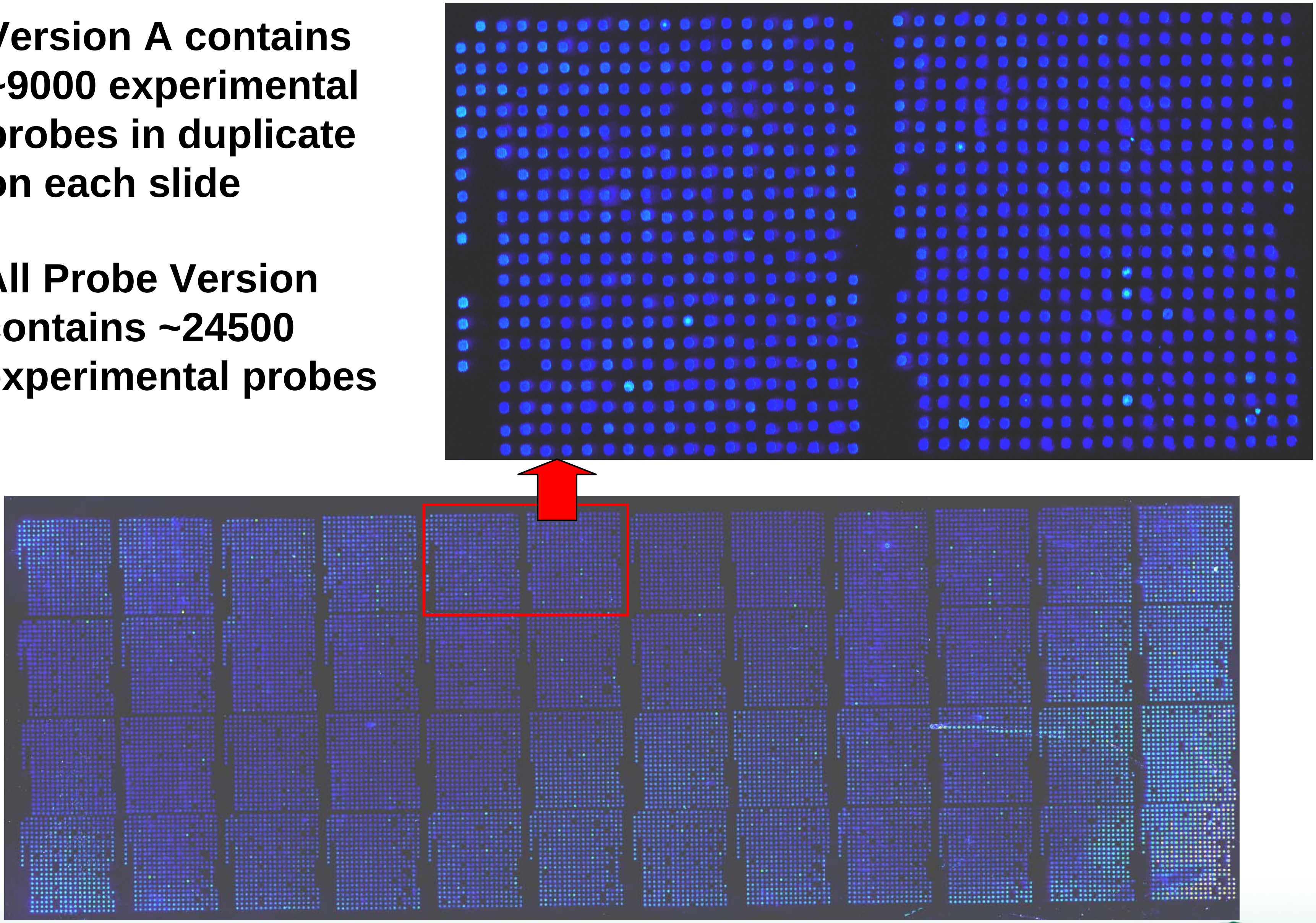
	Number of Probes		
Gene Category	Unique	Group	Total
Carbon Degradation	2,532	276	2,808
Carbon Fixation	584	215	799
Metal Resistance/Reduction	4,039	507	4,546
Methane/Methanogenesis	437	333	770
Nitrogen Fixation	1,225	0	1,225
Nitrogen Metabolism	865	902	1,767
Nitrogen Reduction	1,805	501	2,306
Organic Contaminant	6,920	1,087	8,007
Perchlorate Remediation	21	0	21
Sulfur Reduction	1,286	329	1,615
Total	**19,714**	**4,150**	**23,864**

Details of probe design for C cycle related genes

Carbon polymer degradation	# of input genes	# with Unq probes	# with Grp probes	% inclusion
laccase	190	97	79	92.6
chitinase	413	248	100	84.3
cellulase/endoglucanase	567	417	80	87.7
beta-mannosidase	73	66	4	95.9
polygalacturonase	112	86	25	99.1
glutamate dehydrogenase	112	88	10	87.5
subtotal	**1467**	**1002**	**298**	**88.6**
Carbon Fixation				
carbon monoxide dehydrogenase	10	10	0	100.0
formyltetrahydrofolate synthetase	155	52	87	89.7
rubisco large subunit	470	158	87	52.1
rubisco small subunit	299	16	85	33.8
subtotal	**934**	**236**	**259**	**53.0**

Version A contains ~9000 experimental probes in duplicate on each slide

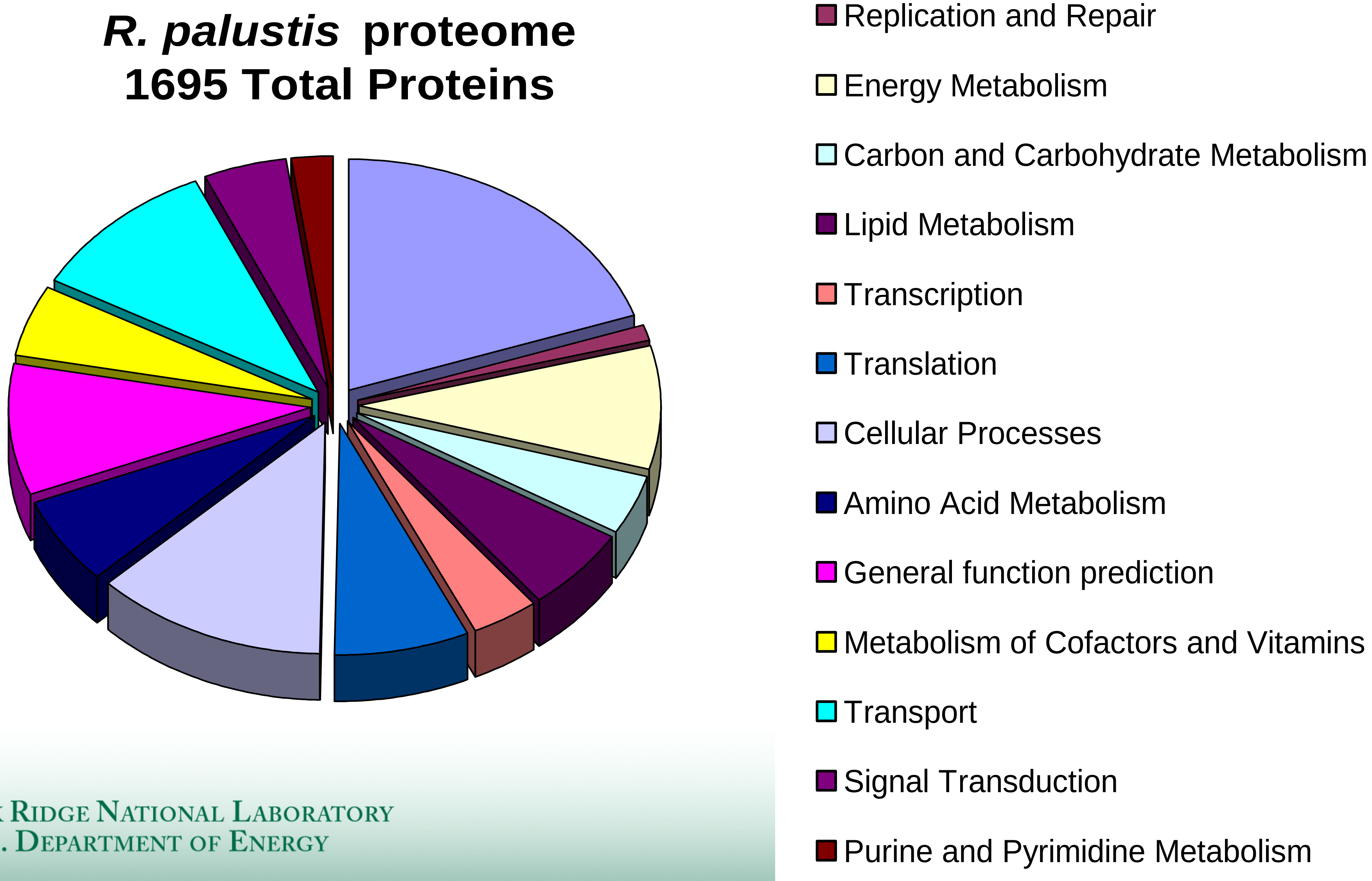
All Probe Version contains ~24500 experimental probes



MS Proteomics

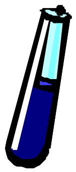
- Can identify what proteins are actually present
- When and where proteins are present
- Detailed analysis can identify how proteins are altered to change activity and location (post translational modifications)
  - Top-down and bottom-up analysis of isolated R. palustris ribosome
  - Identified 52 of known 54 subunits
  - Identified PTMs (methylation, acetylation, glycosylation) and location on many of the subunits
  - Improper PTMs are why many genetically engineered proteins do not work properly the first time

Proteomics: Functional Categories of *R. palustris* Proteins



Proteome Analysis with Mass Spectrometry

Lysate

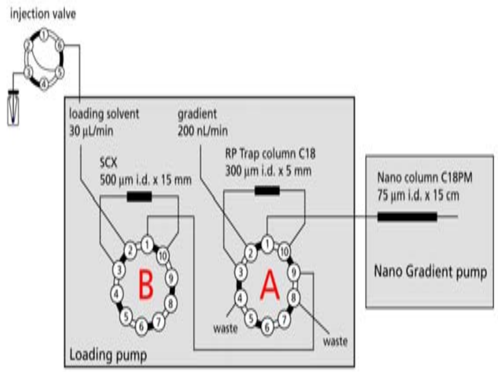


Complex Mixture
Tryptic Peptides

K R R K

Tryptic Digestion of
Mixed Bacterial Lysates

Autosampler makes injections on to the 2D system from fractions in microtiter plates



Automated 2D HPLC
1st Dimension: SCX
2nd Dimension: Nano RP C18

ES-Ion Trap MS
Data-Dependent MS/MS

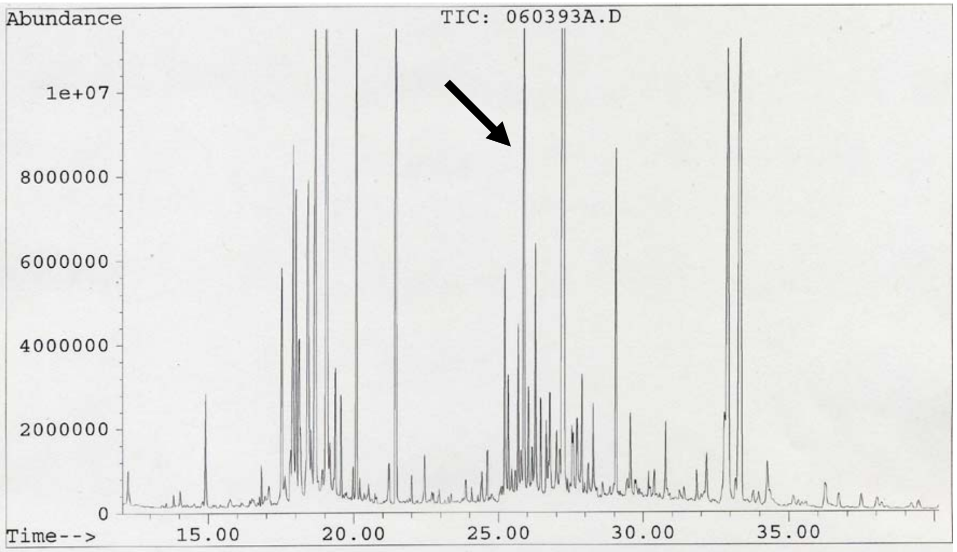
Interactomics

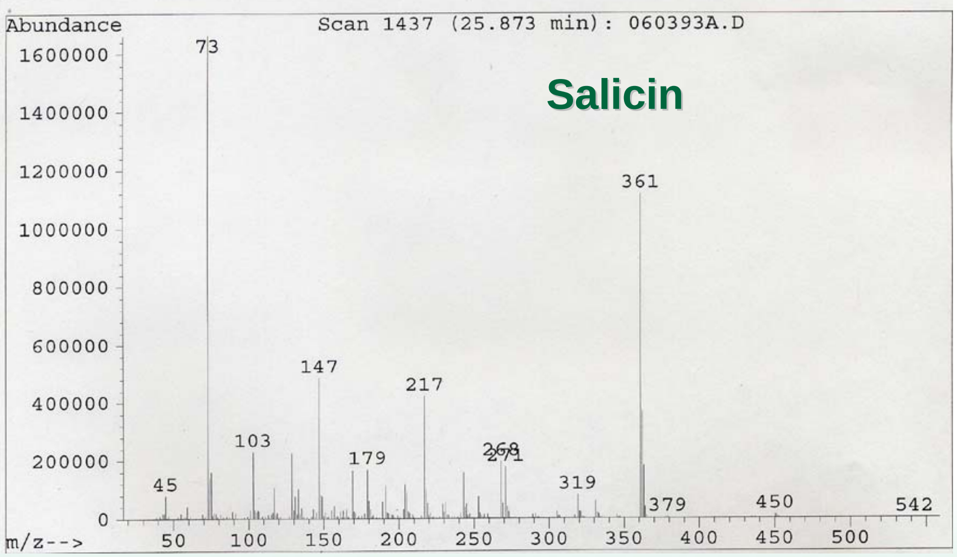
- Identifies required units of a protein complex
  - So that all pieces can be used together
- Identifies transitory protein interactions
  - These may be critical regulators or cofactors

Metabolomics

- LC-MS/MS techniques for plant and microbial metabolites
- Identifies what is actually being made
- Challenge is the “steady state” levels are often low
- Metabolic flux requires modeling

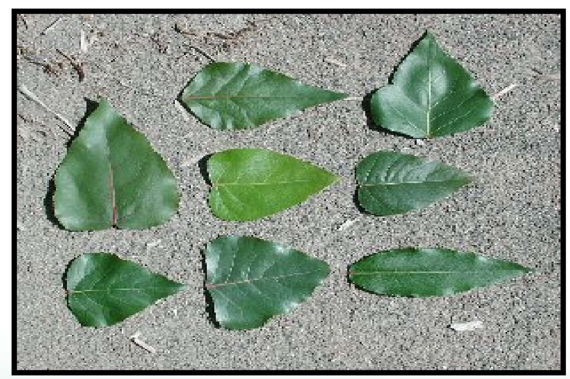
Metabolic Variants





Segregating Population

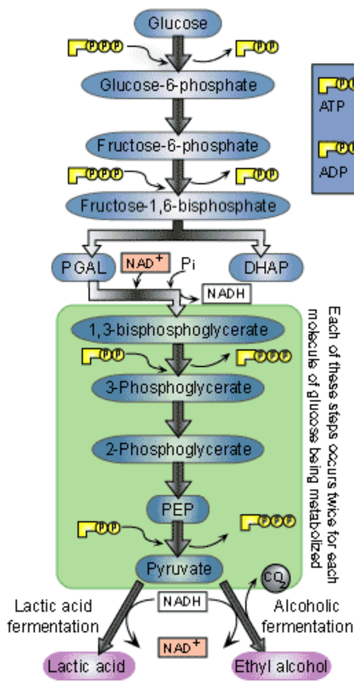
F2 Progeny



Charged metabolites targeted by PACE-ES-MS

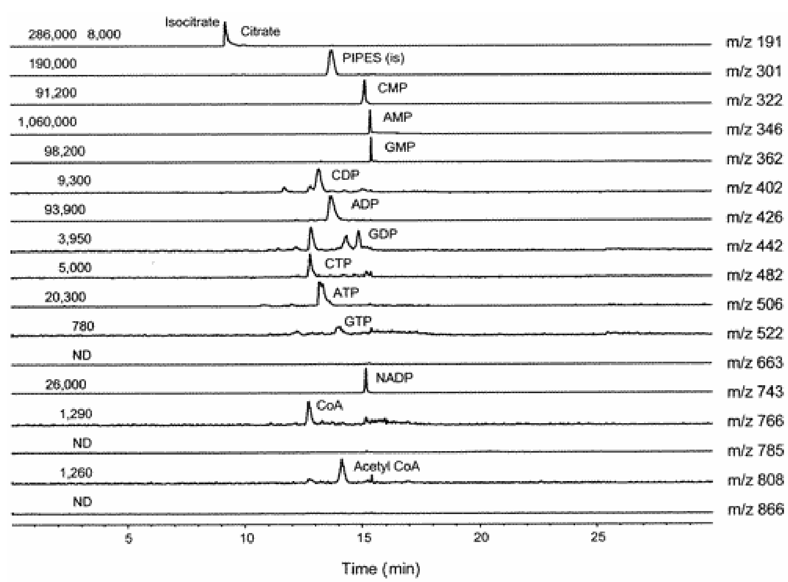
Beckman Coulter PACE system

- acetate
- lactate
- pyruvate
- acetyl phosphate
- acetyl-CoA
- ferredoxin (reduced)
- ferredoxin (oxidized)
- NAD
- NADH
- NADP
- NADPH
- ATP
- ADP
- AMP
- glucose-6P
- glucose-1P
- fructose-6P
- fructose-1P
- fructose-1,6-P2
- glyceraldehyde-3P
- glycerate-1,3P2
- glycerate-3P
- glycerate-2P
- phosphoenolpyruvate
- ribulose-1,5-P2
- inositol phosphates



API Qstar XL

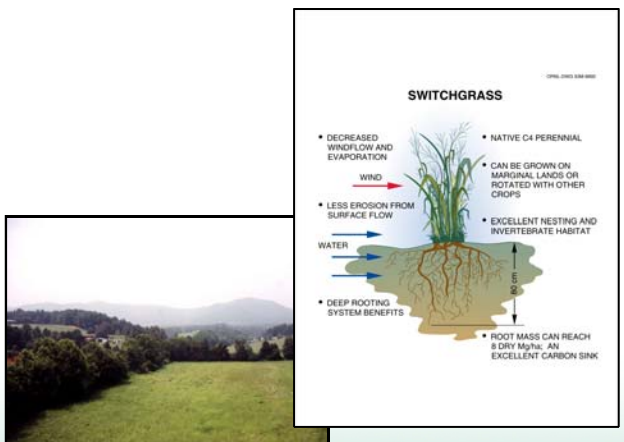
Hybrid Qq-Time-of-Flight-MS System



Bioenergy and Plant Genomics: Expanding the nation's renewable energy resources

Conventional Forestry

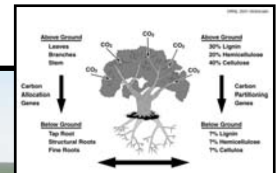
Yesterday



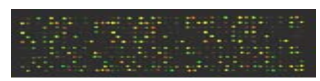
Short rotation hardwoods

Today

Carbon allocation



Metabolic Profiling



Whole Genome Microarrays

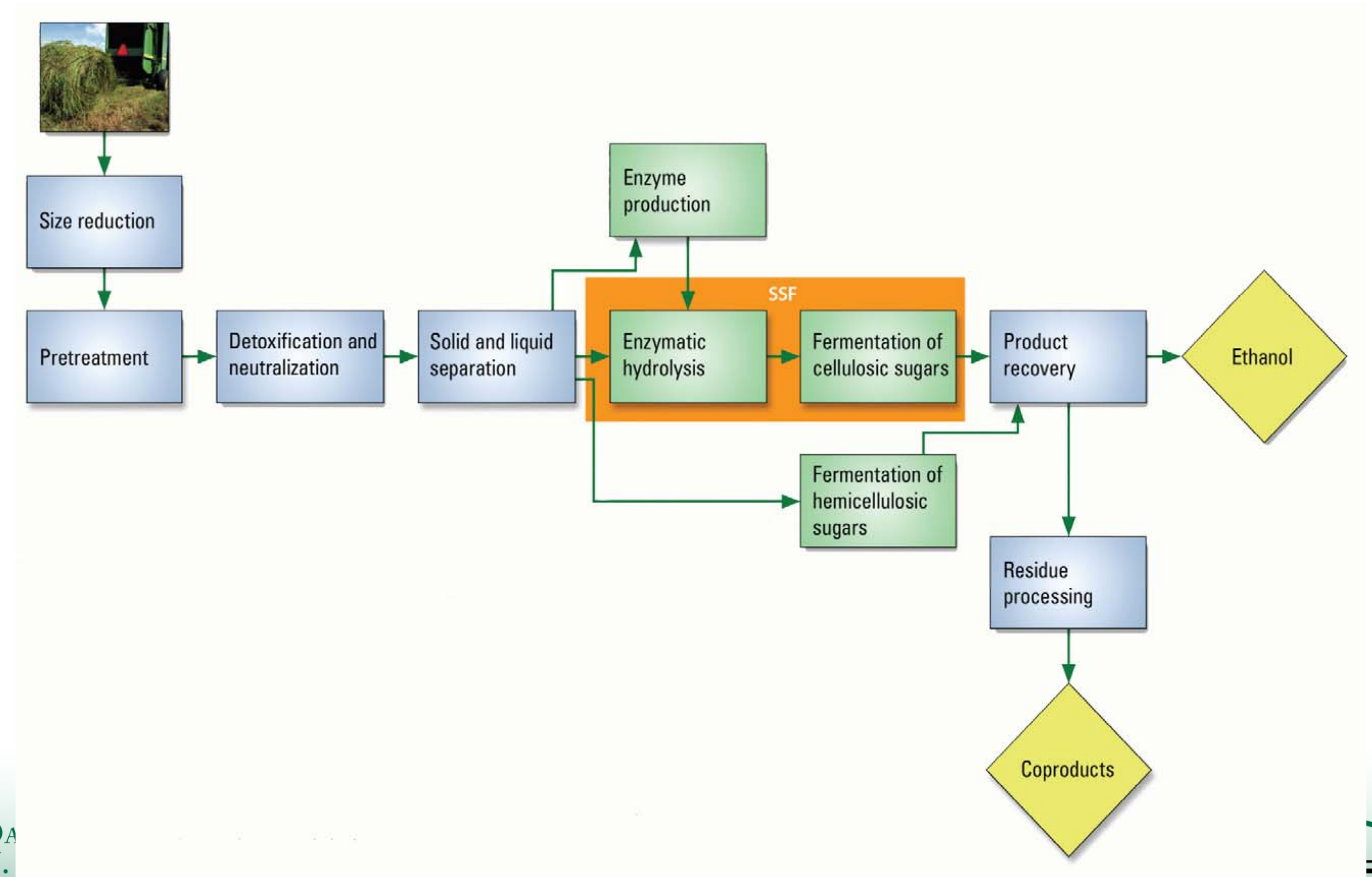
Accelerated Domestication

Tomorrow

High yield wood crops

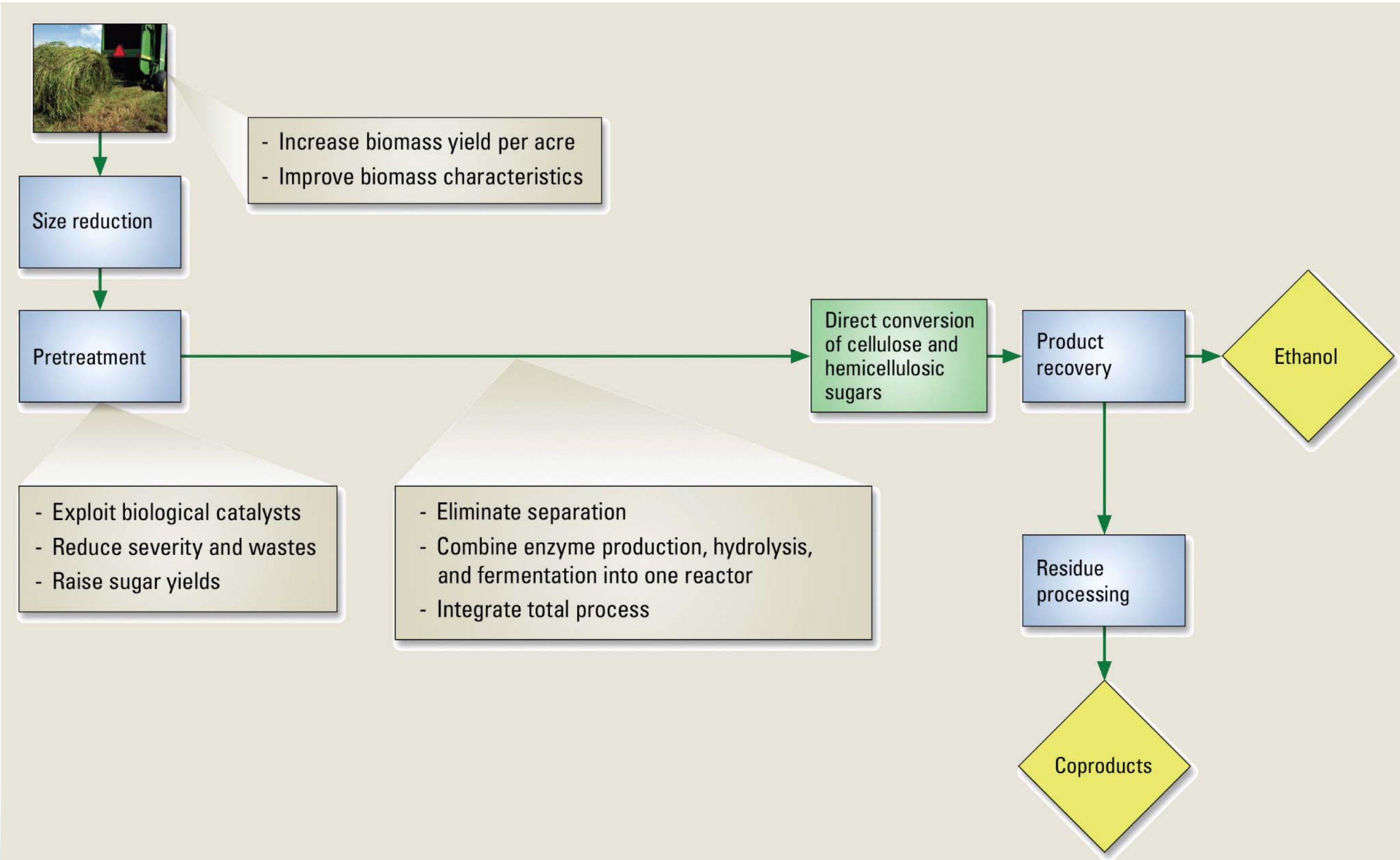
Putting Genomes to work For Energy Security

Steps in cellulosic ethanol production



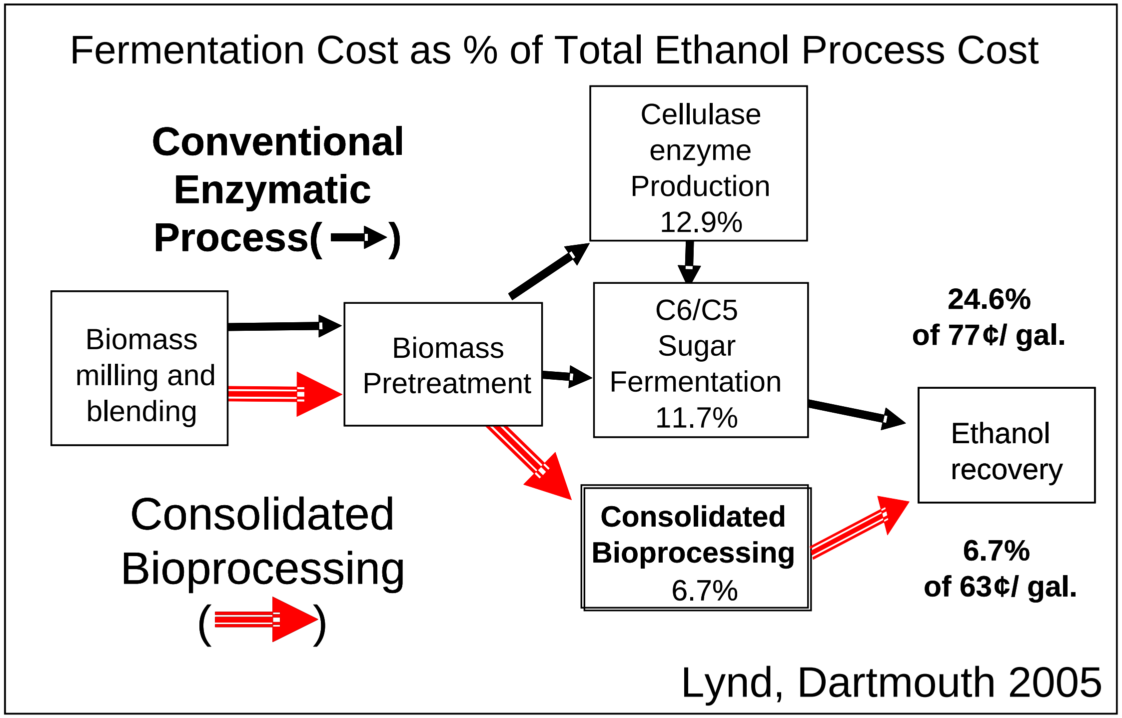
From: Breaking the Biological Barriers to Cellulosic Ethanol

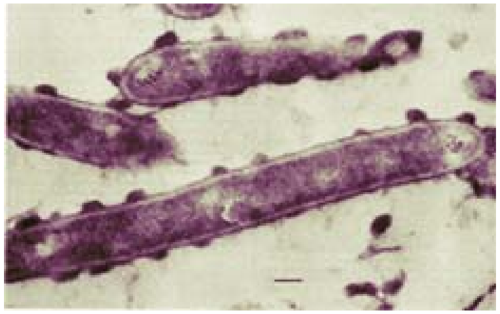
Desirable improvements



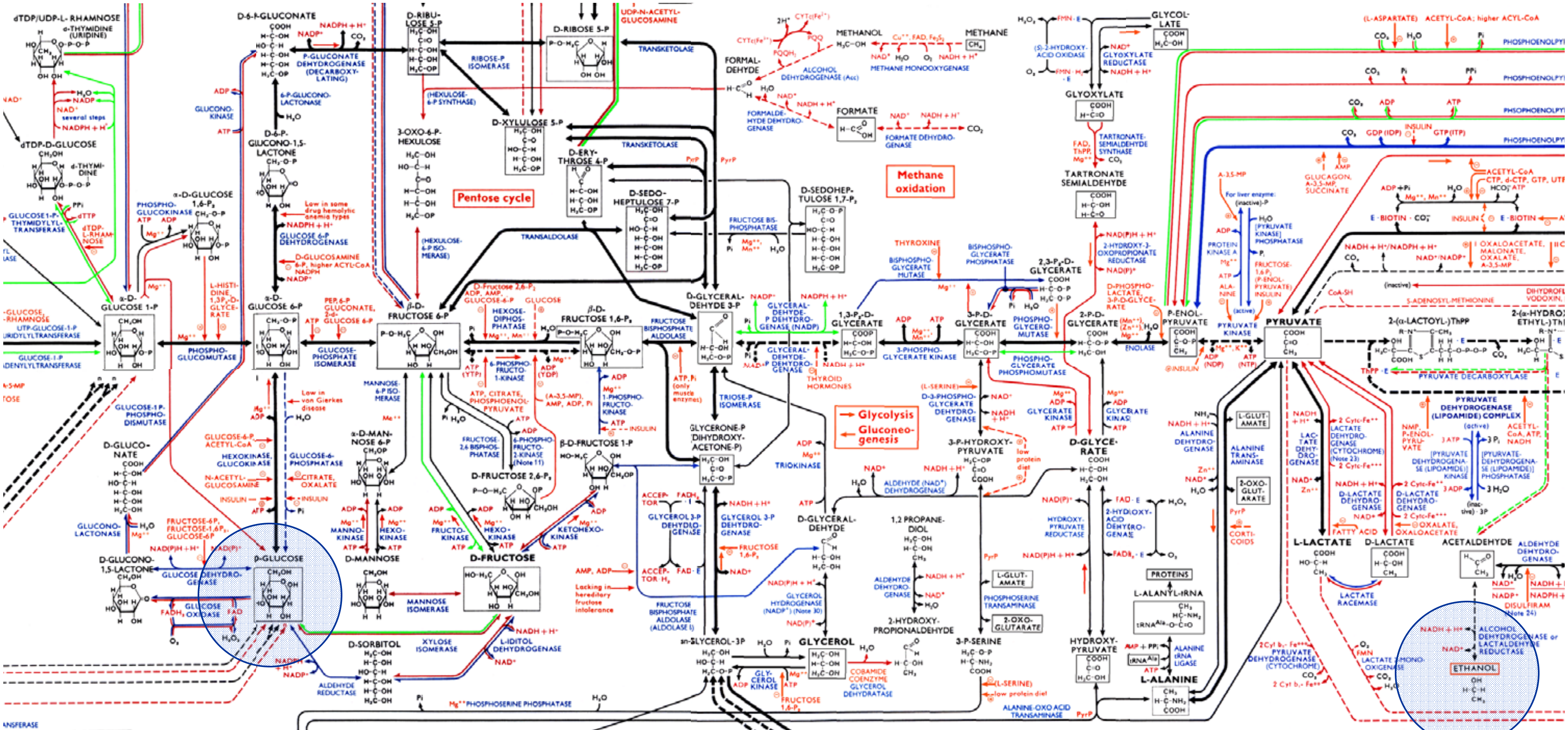
From: Breaking the Biological Barriers to Cellulosic Ethanol

Using Systems Biology to discover regulation in Consolidated Bioprocessing (CBP)





Clostridium thermocellum
Cellulolytic and Ethanologenic bacterium

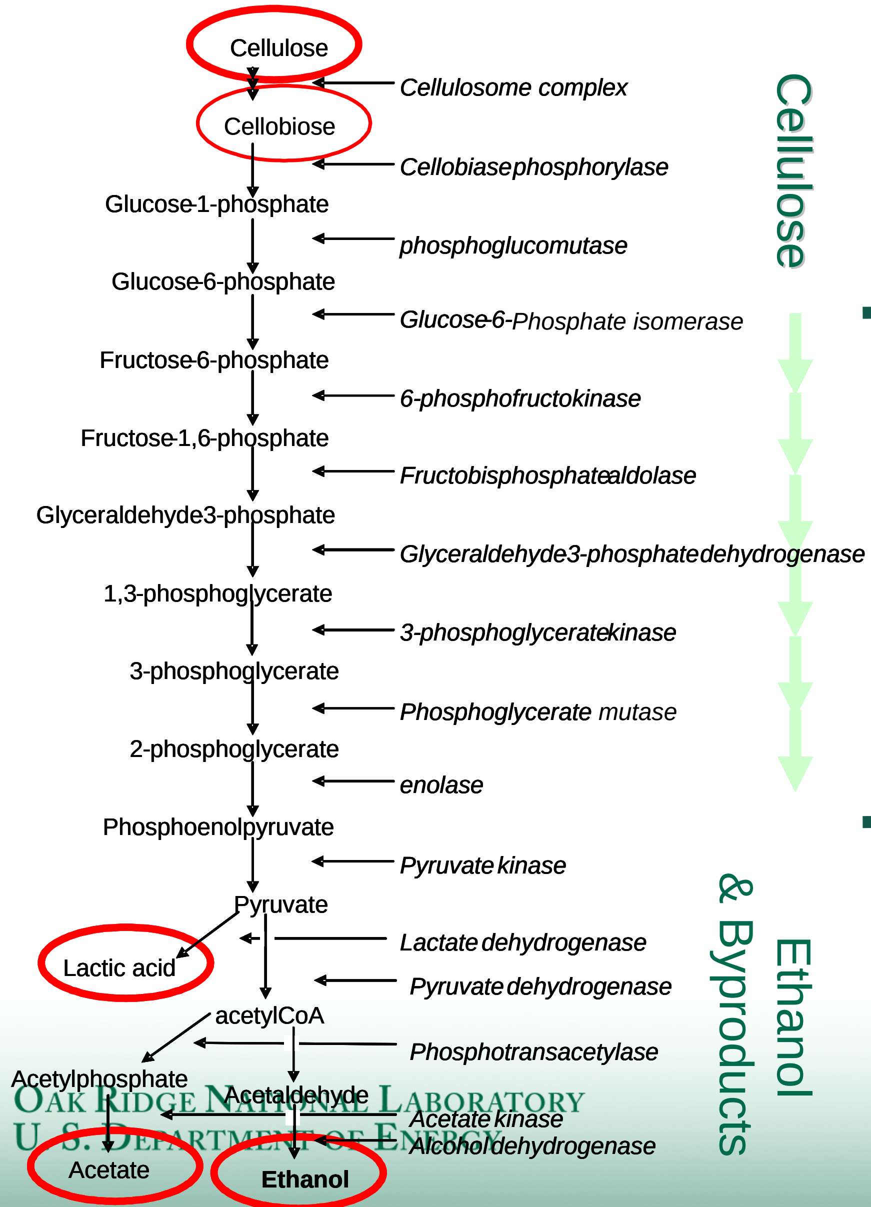


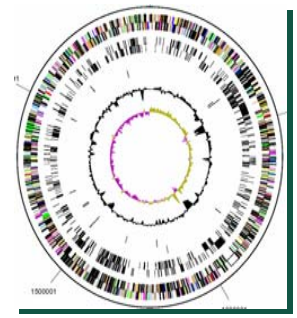
GLUCOSE

ETHANOL

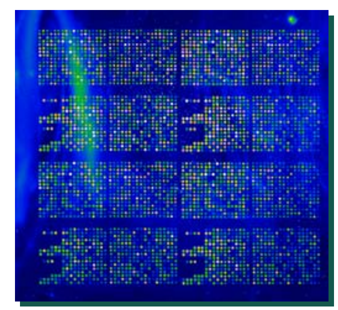
Snapshot of the metabolic pathways involved in central glycolysis and fermentation from BMBP pathway map
http://ca.expasy.org/cgi-bin/show_thumbnails.pl

‘Omics technologies applied to help design better ethanologens - where are the biochemical limitations?





Genome annotation gives possible genes, proteins, etc.



Transcriptome - Gene expression measured by microarray analysis

Proteome- Protein expression measured by mass spectrometry

Metabolome - Cellulose hydrolysis fragments and fermentation products measured by mass spectrometry

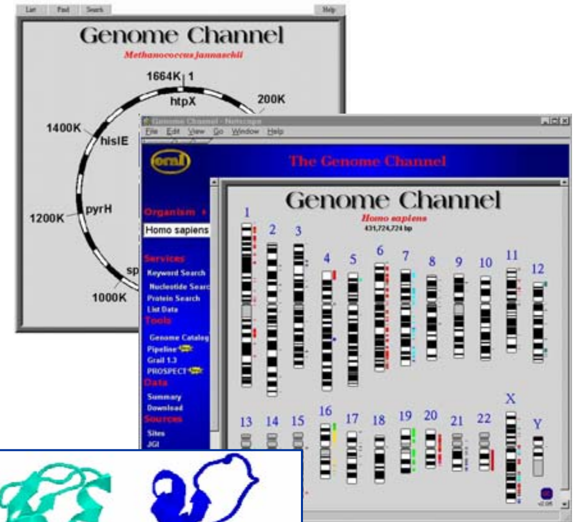
Approaches to create a better ethanologen

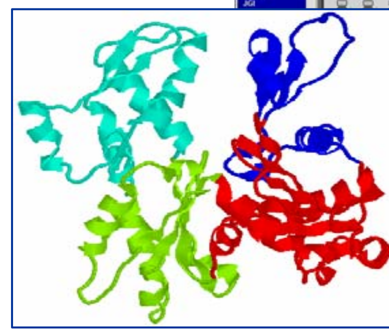
- Take well understood microbes and alter to desired traits
  - (e.g. *E. coli, B. subtilis*, *S. cerevisiae, Z. mobilis*)
  - Works for understood pathways and single gene traits
  - Known genes for sugar metabolism and fermentation
  - Kinetics and regulation appear to be barrier
  - Unknown genes for inhibition, thermal tolerance, etc

- Take novel microbe with complex traits and alter to desired
  - (e.g., *Clostridia*, thermophiles)
  - Exploit biodiversity
  - Lack of tools (genetic and screens)
  - Unknown robustness

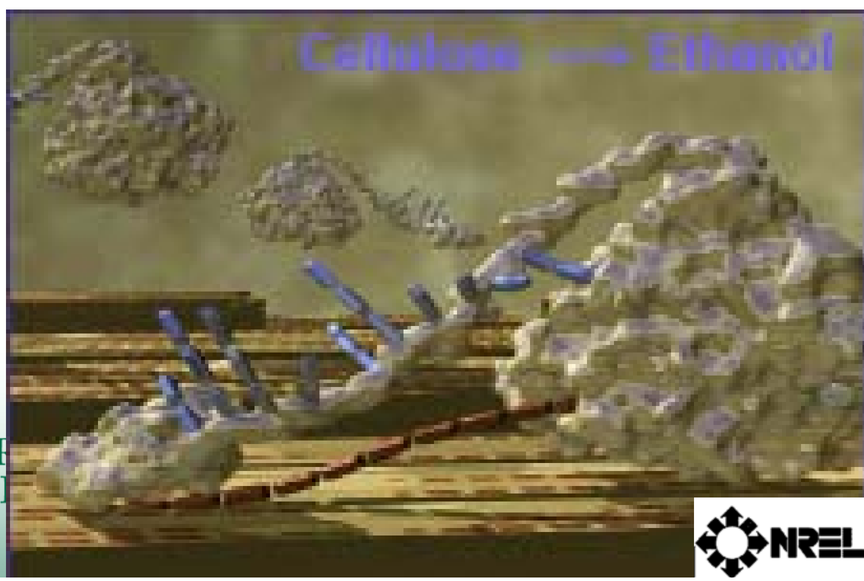
Computational Biology

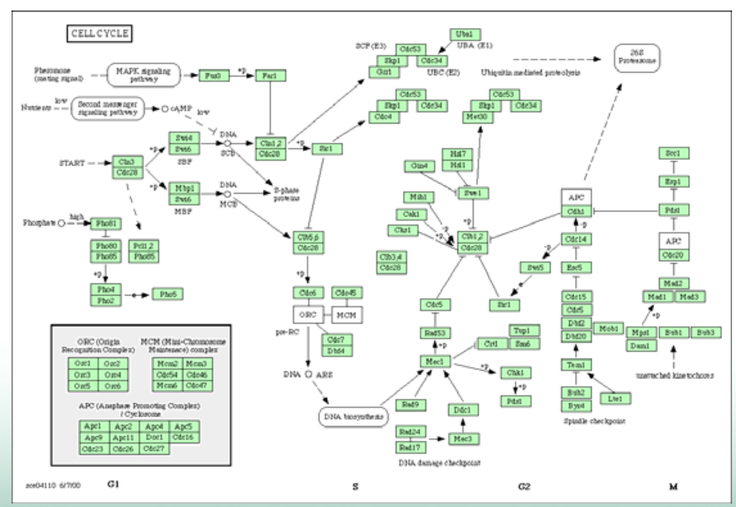
- Annotation DNA sequence analysis, gene finding, regulatory regions, comparative analyses for JGI Microbes
- Systems analysis of mRNA and protein expression patterns
  - Meta-omic databases
- Protein structure prediction
- Protein complexes
- High-performance computing
  - Petaflop computing



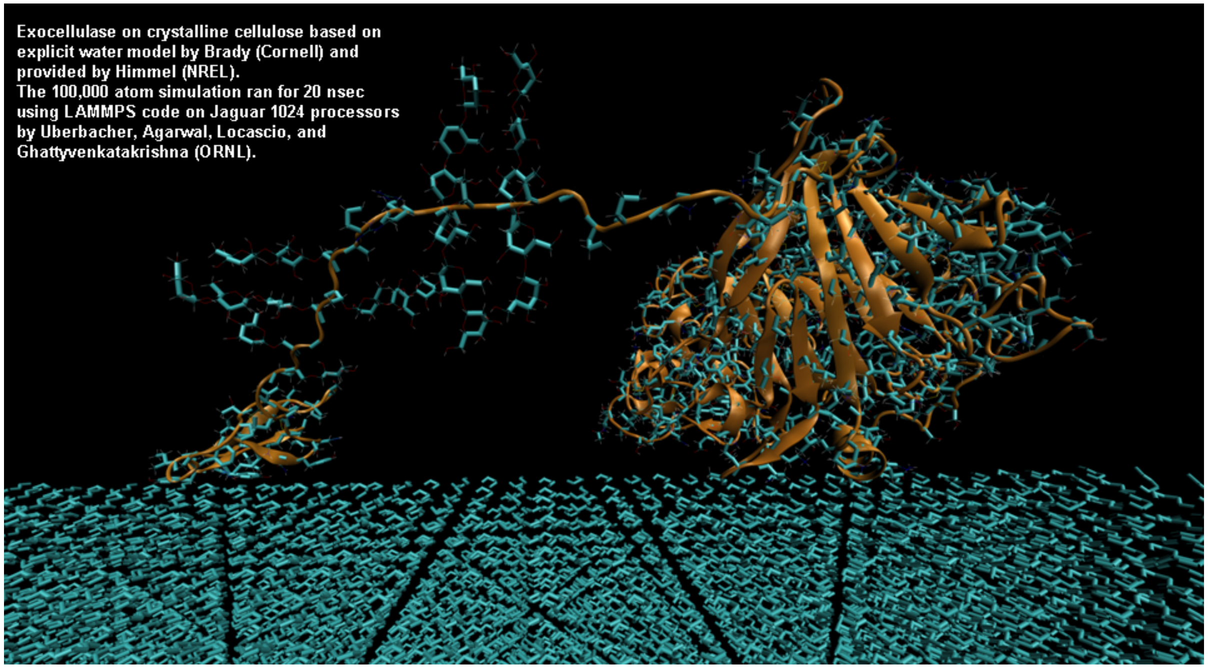


yeast



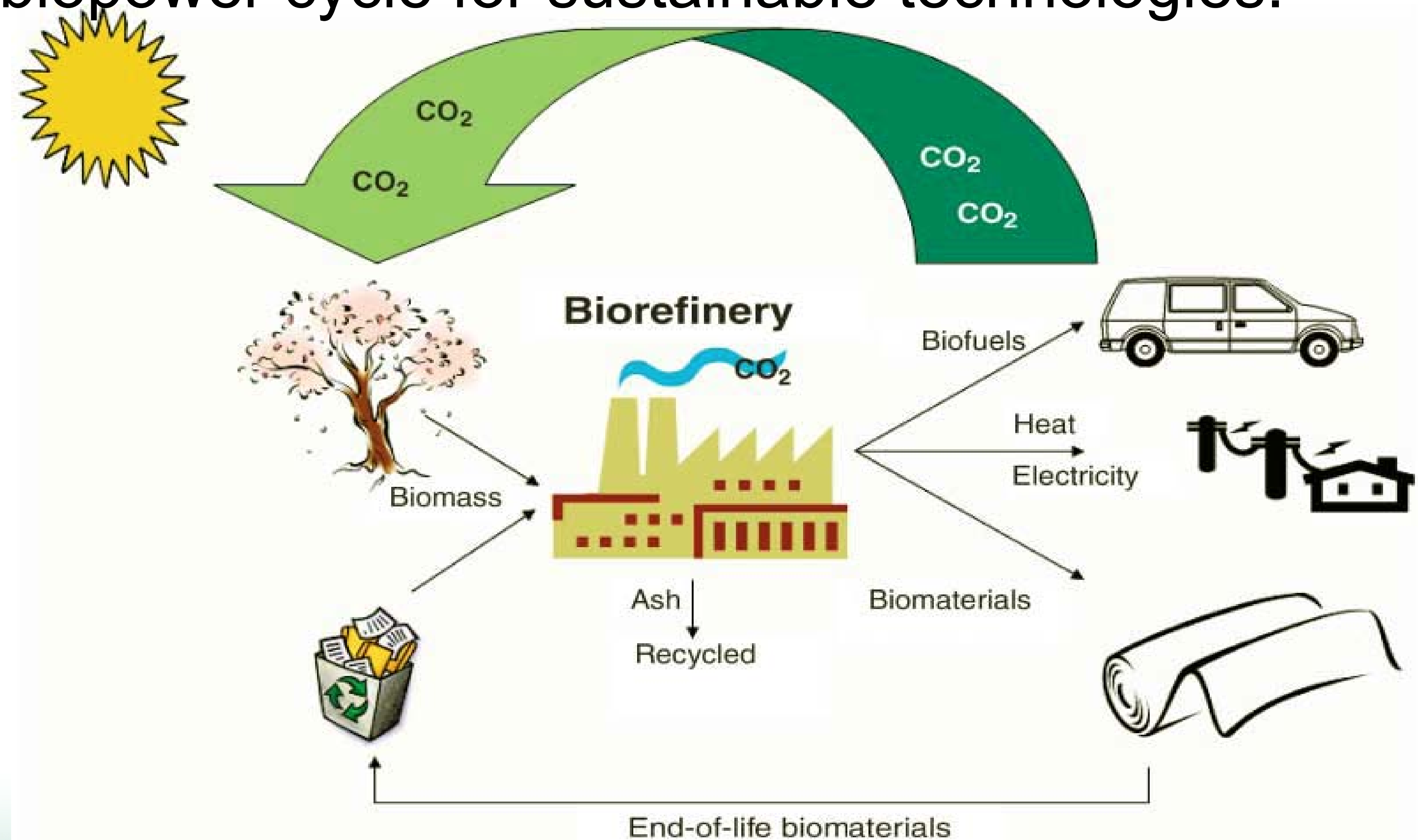


Modeling cellulase on cellulose



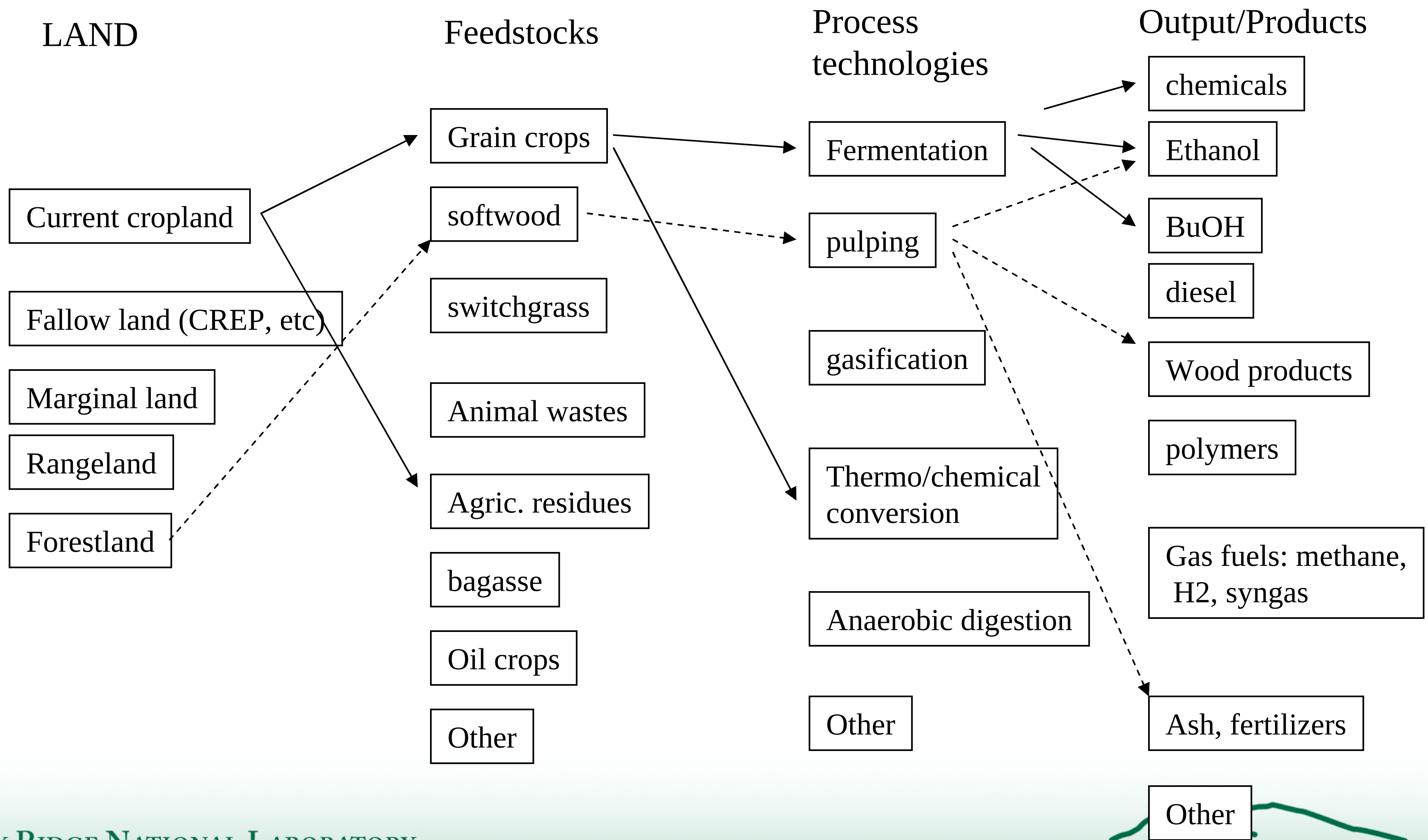
- 100,000 atom model of enzyme cellulase on crystalline cellulose in water
- Computation for 50 nS on Jaguar by Uberbacher et al.
- Based on explicit model of Brady (Cornell) and Himmel (NREL) – leads of new project.
- Shows cellulose binding domain and active site interaction with surface

The fully integrated agro-biofuel-biomaterial-biopower cycle for sustainable technologies.



Ragauskas et al. 2006, *SCIENCE* **311**:484-489

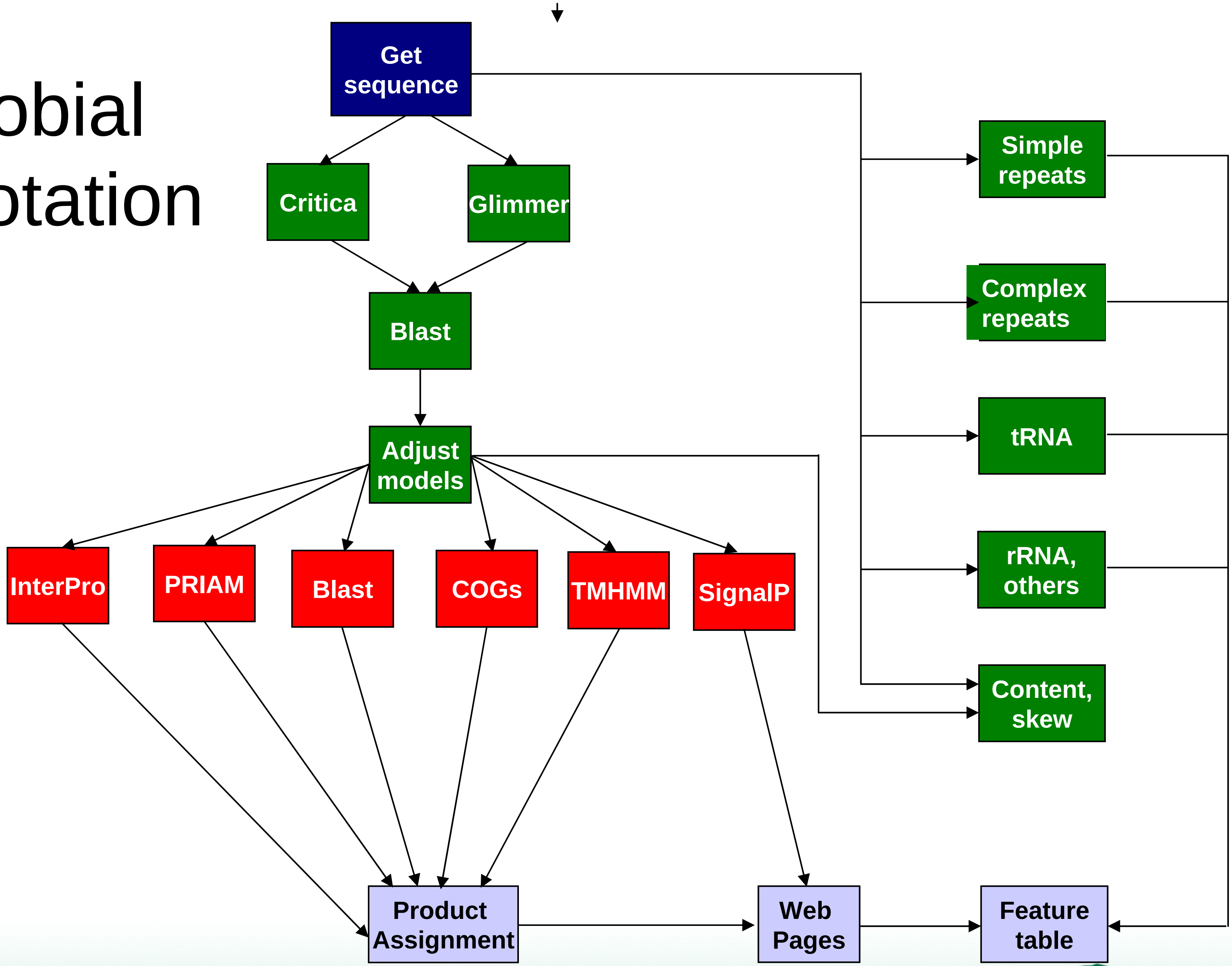
Biomass Utilization is a multi-factorial problem (multiple choice)

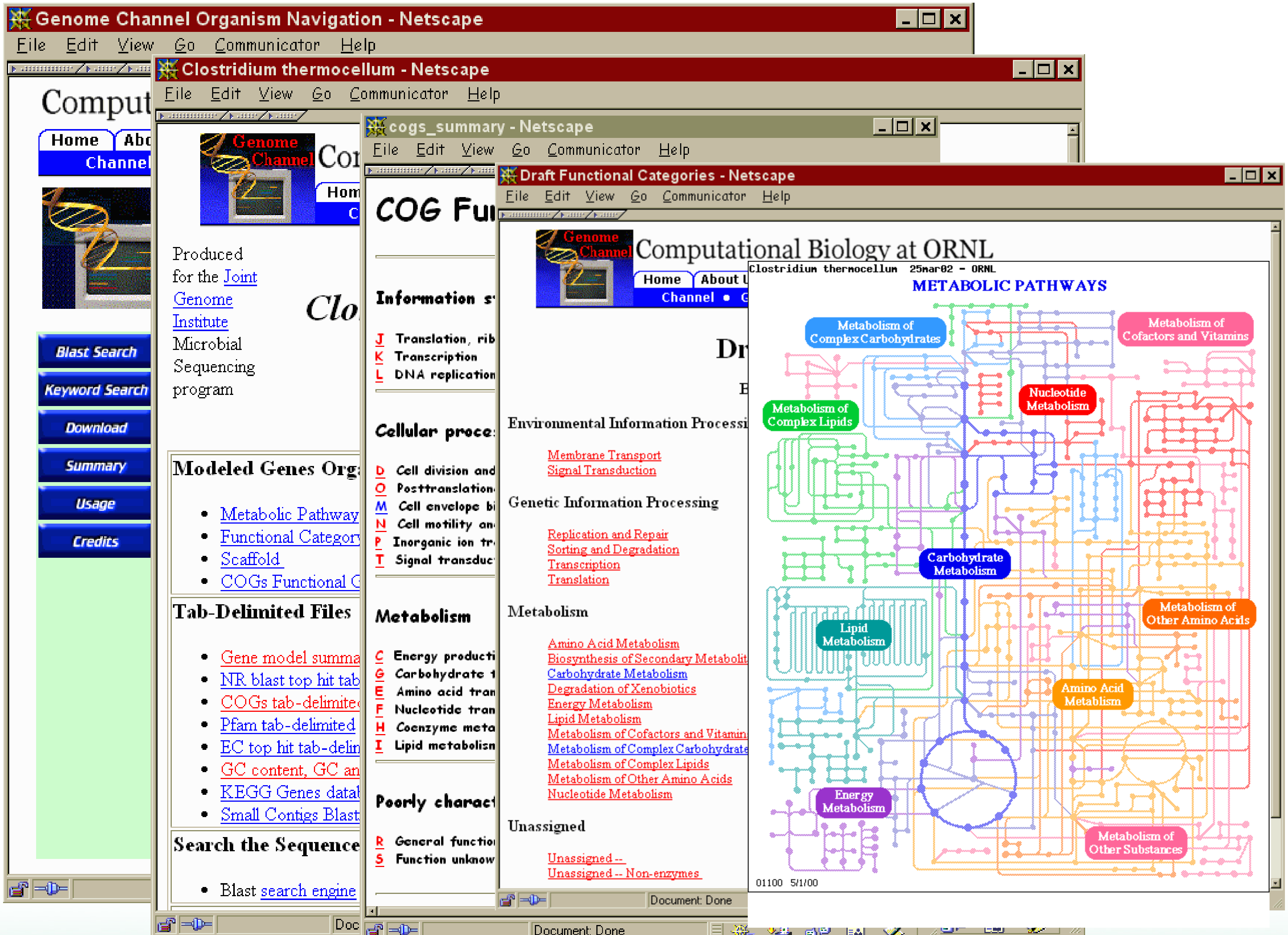


Panel Consensus from 27th Symposium on Biotechnology for Fuels and Chemicals

- We need commercialization and policy support for these current and near-term opportunities to grow the industry from their current base.
- Equally important, we need R&D to increase the impact, the efficiency, the sustainability, and the ability to add value to markets and farmers while preparing for possible long-term limits on land use.
- Corn ethanol is the starting point for this domestic industry and can expand somewhat further; <u>however, lignocellulosic ethanol is ready to be developed in niche markets now and substantially improved for widespread use.</u>

Microbial Annotation





Genome Channel Organism Navigation - Netscape
File Edit View Go Communicator Help
Home
Channel
Blast Search
Keyword Search
Download
Summary
Usage
Credits
Clostridium thermocellum - Netscape
File Edit View Go Communicator Help
Genome Channel
Produced for the Joint Genome Institute Microbial Sequencing program
Modeled Genes Org
Metabolic Pathway
Functional Category
Scaffold
COGs Functional
Tab-Delimited Files
Gene model summa
NR blast top hit tab
COGs tab-delimite
Pfam tab-delimited
EC top hit tab-delin
GC content, GC an
KEGG Genes datal
Small Contigs Blast
Search the Sequence
Blast search engine
cogs_summary - Netscape
File Edit View Go Communicator Help
COG Fu
Information s
J Translation, rib
K Transcription
L DNA replication
Cellular proce
D Cell division and
O Posttranslation
M Cell envelope bi
N Cell motility an
P Inorganic ion tr
T Signal transduc
Metabolism
C Energy producti
G Carbohydrate t
E Amino acid tran
F Nucleotide tran
H Coenzyme meta
I Lipid metabolism
Poorly charact
R General functio
S Function unknow
Document: Done
Draft Functional Categories - Netscape
File Edit View Go Communicator Help
Genome Channel
Computational Biology at ORNL
Home About U
Channel
Environmental Information Processi
Membrane Transport
Signal Transduction
Genetic Information Processing
Replication and Repair
Sorting and Degradation
Transcription
Translation
Metabolism
Amino Acid Metabolism
Biosynthesis of Secondary Metabolit
Carbohydrate Metabolism
Degradation of Xenobiotics
Energy Metabolism
Lipid Metabolism
Metabolism of Cofactors and Vitamin
Metabolism of Complex Carbohydrate
Metabolism of Complex Lipids
Metabolism of Other Amino Acids
Nucleotide Metabolism
Unassigned
Unassigned --
Unassigned -- Non-enzymes
Document: Done
Clostridium thermocellum 25mar02 - ORNL
METABOLIC PATHWAYS
Metabolism of Complex Carbohydrates
Metabolism of Cofactors and Vitamins
Nucleotide Metabolism
Metabolism of Complex Lipids
Carbohydrate Metabolism
Metabolism of Other Amino Acids
Lipid Metabolism
Amino Acid Metabolism
Energy Metabolism
Metabolism of Other Substances
01100 5/1/00

Neutrons and Biology

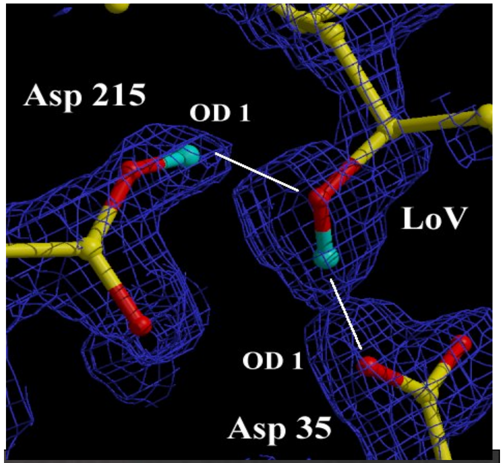
ORNL will provide world-leading instruments for neutron scattering at HFIR and at SNS

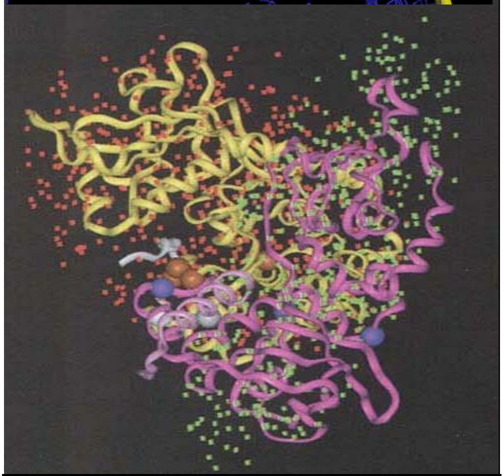
Neutrons are excellent probes for Hydrogen –and can discriminate between hydrogen and deuterium

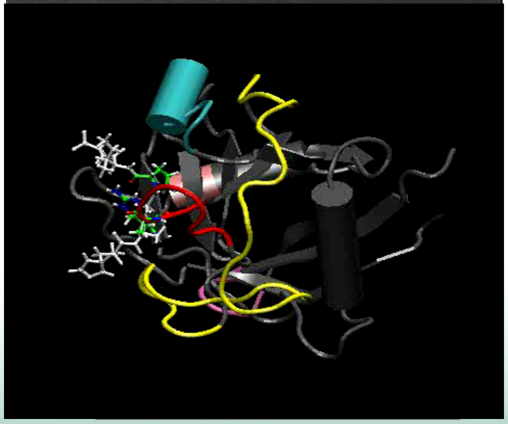
Function: H/D in enzyme mechanism; proton shuttling & transfer

Structure: H/D Labeled protein in complex systems

Dynamics: Specific H-Labeling in deuterated systems

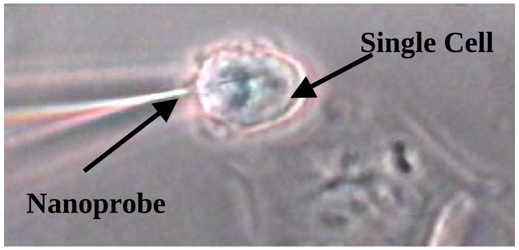






Provide biological interface for nanobiology

Challenges: examine the interface between biological and physical structures with materials, energy, and information using the tools of biology, physical and computational sciences



Examples

- Probes for single cells
- Single cell genetic engineering
- Cell membrane mimics