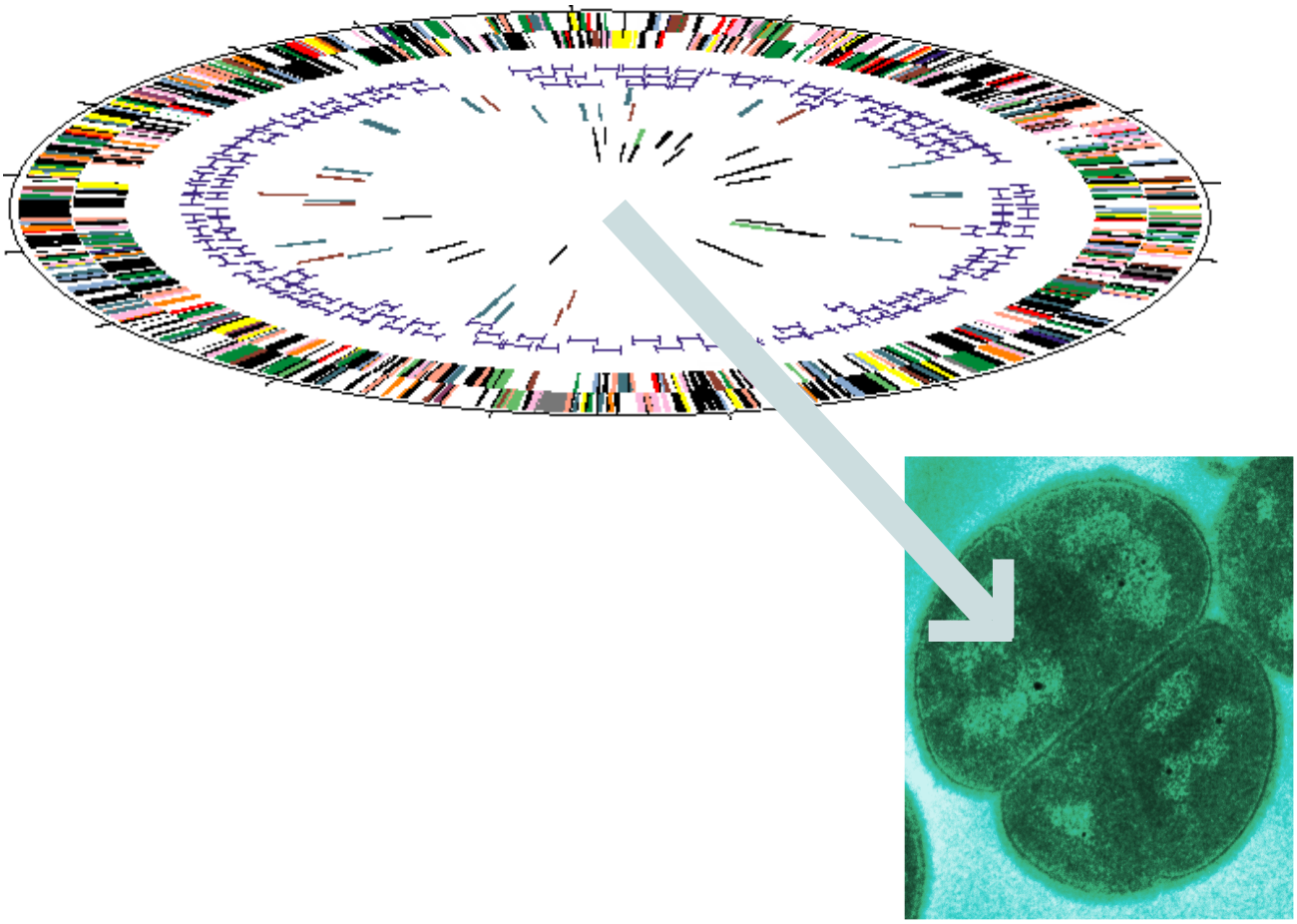
June 2000

Microbial Cell Project



U.S. Department of Energy
Office of Science
Biological and Environmental Research Program

Contact Information

Office of Biological and Environmental Research
U.S. Department of Energy, SC-72
19901 Germantown Road
Germantown, MD 20874

Dr. Daniel Drell, SC-72 (301/903-4742, *daniel.drell@science.doe.gov*)
Dr. Anna Palmisano, SC-74 (301/903-9963, *anna.palmisano@science.doe.gov*)
Dr. Gregory Dilworth, SC-17 (301/903-2873, *greg.dilworth@science.doe.gov*)

Web site for report: *www.ornl.gov/hgmis/publicat/microbial/index.html*

Microbial Genome Program (2000), a report on the DOE program that gave rise to the Microbial Cell Project, is available electronically and in print:

- Web – *www.ornl.gov/hgmis/publicat/microbial*
- Print – 865/576-6669 or *yustln@ornl.gov*

DOE Microbial Cell Project

http://microbialcellproject.org

Executive Summary

During the last decade, scientists have amassed millions of DNA sequences containing the complete genetic instructions for a growing list of microbes and viruses. These DNA sequences offer a virtual "parts list" for life in its simplest form, but scientists do not know what many of the parts do. Furthermore, DNA sequences provide little information on how the parts work together to orchestrate the chemistry of life. (By analogy, a pile of automobile parts would tell us very little about the complex function of an automobile.) In biology, the whole is much greater than the sum of the parts, and understanding this complexity is the exciting challenge science now faces. Revolutionary breakthroughs in genome sequencing, new methods of protein characterization, and access to powerful supercomputers now position scientists to begin to understand the complex pathways that give a microbial cell its life. The Microbial Cell Project (MCP) is an exciting new initiative that will address these challenges. The proposed project builds on previous research sponsored by the Office of Science, including the Microbial Genome Program (MGP), itself a spinoff of the Human Genome Program initiated by the Department of Energy (DOE).

The Microbial Cell Project will support core missions of the Department of Energy and is consistent with several Office of Science strategic goals. One of DOE's missions is to help ensure that the United States continues to have access to sources of affordable and environmentally friendly energy (Goal 1, "Science for Clean and Affordable Energy"). While physical sciences have been the backbone of energy research, new concepts in the biological sciences will shape our energy future by providing ways to use living organisms to produce energy and clean the environment (Goal 2: "Energy Impacts on People and the Biosphere").

Microbes have evolved for 3.8 billion years and have colonized almost every environment on Earth. In the process, they have developed an astonishingly diverse collection of capabilities that will help DOE meet its challenges in toxic waste cleanup, energy production, global climate change, and biotechnology (Goal 3: "Building Blocks from Atoms to Life"). All of this will take place in the context of outstanding science that has characterized DOE since its inception (Goal 5: "Scientific and Operational Excellence").

To embark upon this journey will require the development of new technologies, analytical tools, and modeling capabilities. In addition to working with academic, nonprofit, and industrial partners, DOE will take advantage of the scientific talents available in its national laboratories. These talents include high-throughput genomic DNA sequencing, microbial biochemistry and physiology, imaging, and structural biology. National user facilities such as synchrotrons will play important roles, as will capabilities in high-performance computing. Interdisciplinary collaborations among biologists, chemists, physicists, engineers, and computer experts will be critical to this effort.

In the MCP, scientists will begin to write a comprehensive "owner's manual" for a microbial cell. Microbial cells have internal organization and complex control systems that allow them to respond to their environment. They can work as miniature chemistry laboratories, making unique products and carrying out specialized functions. Ultimately, understanding the complex functioning of a single microbial cell will enable science to go far beyond just exploiting the beneficial capabilities of microbes to meet DOE's missions. The knowledge gained will apply to cells in all living things. Thus the MCP represents a first step in moving from cataloguing molecular parts to constructing an integrative view of life at the level of a whole organism—microbe, plant, or animal.

Background

By some estimates, microbes make up about 60% of the Earth's biomass, yet less than 1% of microbial species have been identified. Microbes play a critical role in natural environmental cycles and processes. Because most microbes do not cause disease in humans, animals, or plants and are difficult to culture, they have received little attention (especially in comparison to the minority that do cause diseases). Microbes have been found surviving and thriving in an amazing diversity of habitats, in extremes of heat, cold, radiation, pressure, salinity, and acidity, often where no other life forms could exist. Identifying and harnessing their unique capabilities, which have evolved over 3.8 billion years, will offer us new solutions to longstanding challenges in environmental and waste cleanup, energy production and use, medicine, industrial processes, agriculture, and other areas. Scientists also are starting to appreciate the role played by microbes in global climate processes, and we can expect insights about both the biological underpinnings of climate change and the contributions of microbes to Earth's biosphere. Their capabilities soon will be added to the list of traditional commercial uses for microbes in the brewing, baking, dairy, and other industries.

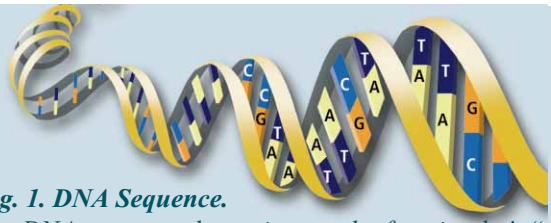


Fig. 1. DNA Sequence.
The DNA sequence determines each of a microbe's "parts" as well as where and when it is expressed.

To explore the possibilities for new applications, in 1994 the U.S. Department of Energy (DOE) established the MGP as a companion to its Human Genome Program (HGP). From the start, the MGP experienced remarkable success, and microbial genomics has become one of the most exciting and high-profile fields in biology today. A principal goal of this spin-off project is to determine the complete DNA sequence—the genome—of a number of non-disease-causing microbes that may be useful to DOE in carrying out its missions . The microbes chosen for genomic sequencing were selected with broad input from the scientific community.

"The microbial diversity of the program is an absolute treasure trove for[research in] biotechnology, ecology, evolution, and bioremediation," notes David Schlessinger (National Institute on Aging). Only a few years ago, scientists could not have imagined having full access to the genetic structure of more than a few such organisms. Today, nearly three dozen complete microbial genomes, eleven supported by DOE's MGP, have been sequenced, and the rate of reported new genome sequences is increasing rapidly. (For a current listing, see *http://www.tigr.org/tdb/mdb/mdb.html*). These DNA sequences, along with those from many viruses and more complex organisms such as the fruitfly, the roundworm, photosynthetic algae and yeast, are freely available in public databases. This information is being used by governmental, academic, medical, and industrial scientists. The number of possible applications of this information is staggering. One of the remarkable surprises that has emerged from the study of a number of microbial sequences is the presence of genetic segments containing entire blocks of genes that appear to have been acquired intact during evolution from other microbes in very distant parts of the tree of life. The bacterium *Thermotoga maritima* is hypothesized to have acquired a quarter of its genome through this process, which is termed "lateral gene transfer." These findings present exciting challenges to our understanding of how microbial species live and evolve. Thus, sequenced genomes provide us not only with a genetic "parts" list, but also insights into the evolution of life on Earth.

Now, the next great challenge is to explore how these parts come together to form a functioning organism, and this is the impetus for the DOE Microbial Cell Project. The MCP is not a genome sequencing program and is thus distinct from the precursor MGP; essentially, the Microbial Cell Project builds on the foundation of the MGP and extends into the realm of complex biology where the next frontier lies.

Microbial Cell Research Thrusts

1. Biological Basics

Goal: Determine and characterize the minimum set of genes and corresponding gene products necessary to sustain a simple free-living microbial cell, express the genes to produce the relevant proteins, and determine their structure.

Challenges: Once the entire genomic complement of a specific microbial cell has been determined, it should be possible to pare down this extensive set of genetic instructions to the minimum framework needed to direct sustainable growth under a defined set of favorable nutrient and environmental conditions. This is analogous to the situation in which a microbe is given all the basic building blocks for proteins; no longer required to synthesize each of these amino acid building blocks, the microbe still must implement the genetic instructions for assembling these building blocks into the appropriate enzymes that will carry out the basic life functions of the cell. Accelerated genomic information and enhanced biotechnological methods will enable scientists to streamline the traditional approach used in classical molecular genetics, in which laborious analyses are used to determine if specific genes are necessary or not under a favorable set of growth conditions.

Once the minimum gene set has been determined, scientists must determine the functions of these essential genes. For the approximately 3 dozen or so microbial genomes completely sequenced to date, an astonishing 45% of the genes do not resemble previously known genes; this means scientists do not have any clues as to the structures or functions of these novel genes. It is like having all the parts to a car spread out on a garage floor, but not knowing what half the parts are (or what they do). So one of the first imperatives is to characterize the unknown genes and their protein products. This can be approached in several ways, including targeted deletions of specific genes (to explore what functions are lost) and solving the physical structure of these gene products using high-resolution crystallographic or spectroscopic techniques. This latter approach is particularly powerful, as structural similarities between two proteins can often suggest functional relationships that scientists subsequently can verify by experimental means. The facilities of the DOE national laboratories (synchrotron sources, computational resources for protein structure prediction and comparisons) will be critically important to this thrust area.

Fig. 2. Whole Genome Sequence. *The complete genome sequence comprises the fundamental "parts list" for a microbe. The genome also contains information about development and responsiveness to the environment.*

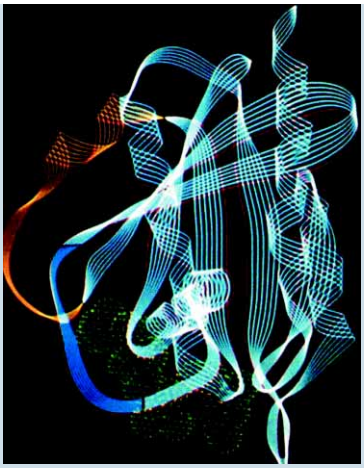


Fig. 3. Gene Products. *Each gene product is a protein and understanding its 3-dimensional shape is key to understanding its functions in the microbial cell and the other proteins with which it interacts.*

2. Functional Foundations

Goal: Determine the physiological and biochemical functions of the gene and specific bioprocesses using standard biochemical techniques and structural/computational biology.

Challenges: Although a given microbe possesses a single set of genes, it is the pattern and timing of how these genes are expressed that enables the organism to adapt and survive. One of the challenges in studying a microbe that has been biologically selected to respond quickly to complex environmental cues (such as changes in nutrients, toxins, temperature,

salt, or acid conditions) is to understand how it expresses a physiologically appropriate subset of its genes. Adaptation by a microbe to a specific environmental cue is not simply the addition of unrelated, independent responses by individual genes, but rather a synergistic and tightly coordinated network of gene expression. A number of technical advances make it feasible, for the first time, to begin to map out these complex regulatory networks in microbial cells. The advent of microarrays of DNA sequence fragments (one for each expressed gene) and high-resolution protein electrophoresis and mass spectrometry can give scientists the gene expression and protein profile of a microbe as a function of time or experimental condition. Among these are responses to external stimuli (e.g., hormones, nutrients, and toxins), or changes in environmental conditions. Similarly, computational algorithms designed to recognize common pattern elements in gene control regions can provide insight into co-regulated genes. Since, as noted above, about half of the genes in a microbial genome have unknown function, these approaches will be especially useful in linking these novel genes and proteins to a particular physiological regulatory network and potentially in elucidating their function. For these data-intensive gene expression profile and sequence comparisons, computational biology will be critical.

3. Modeling Interactions

Goal: Use high-end computing to model gene-gene, gene-protein, and protein-protein interactions as well as the internal biochemistry of the cell.

Challenges: A microbial cell is not simply a "bag of dilute salt water" within which gene products freely diffuse. There is much internal organization due to structural cytoskeletal components, partitioning of gene products in different parts of the cell so that they can efficiently participate in their appropriate pathways, concentration gradients across the volume of the cell, and physical effects caused by the cell membrane and intracellular constituents (in fact, the cell interior is thought to resemble Jello). Very little is known of this internal "milieu" of any microbial cell. Much greater understanding is required so that the results from biochemical, structural, and expression studies can be properly integrated with knowledge of the physical character of the cell's internal environment. All of DOE's capabilities will be needed: imaging the internal components will be a major task pushing the limits of current imaging technologies and new microscopic techniques. Mass spectrometry techniques also will be used to image and identify protein-protein interactions and characterize structures using internal linking compounds. The computational demands will be enormous because the aim is nothing less than to model not only the physical distribution (over time and under different circumstances) of all the gene products but to incorporate into this model the spectrum of gene regulatory, gene-protein, and protein-protein interactions.

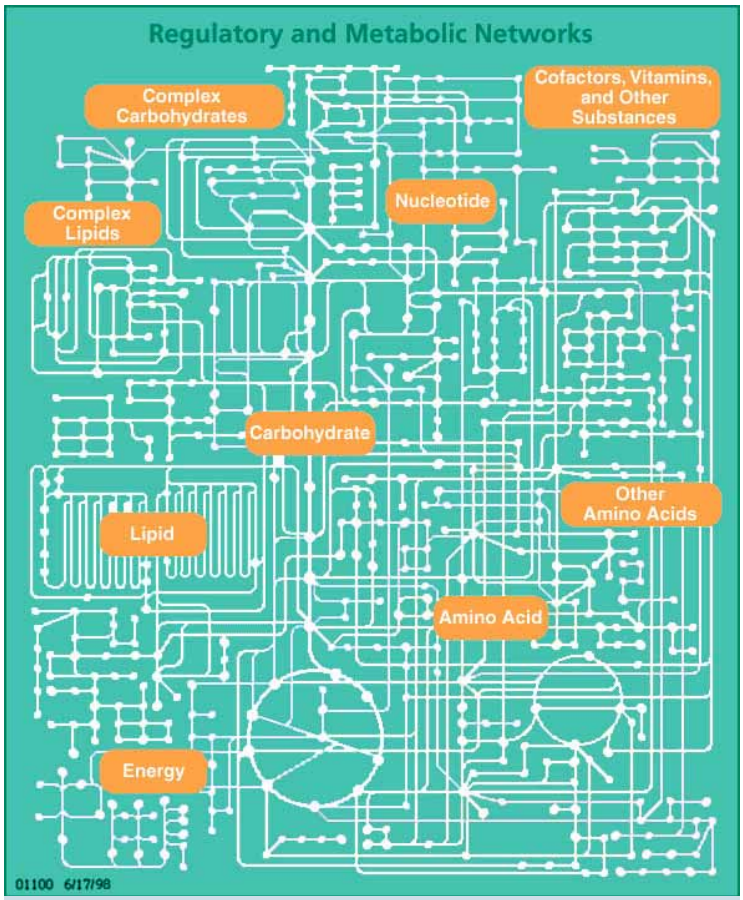


Fig. 4. Metabolic Pathway Diagram. *The microbial proteins comprise a diverse set of reaction pathways that carry out the cell's activities. Modeling this complexity is crucial to understanding and utilizing the cell's capacities.*

4. Regulation and Manipulation

Goal: Use gene-protein manipulation to enhance or suppress various cell functions.

Challenges: The ultimate aim of this undertaking, and the rationale for DOE's concerted effort and focus, is to achieve the necessary level of understanding of cellular functions so that they can be manipulated intelligently (e.g., beneficial functions enhanced) and harmful functions suppressed. Once the minimum gene set is defined, scientists will be able to specifically introduce a gene or gene subset to gain insight into what metabolic flexibility or advantage these added genes confer. These reverse genetic techniques will prove especially powerful in elucidating the role of master regulatory genes on metabolic networks that have been mapped out for a specific set of environmental or growth parameters. Given the complexity of the metabolic networks discussed previously for even the simplest cell, current gene manipulations or interventions are analogous to throwing the proverbial monkey wrench into a complex machine; mostly, the outcome will be negative and uninformative. With greater knowledge about the parts, their interactions, the pathways to which they belong, their partners in biochemistry, and their temporal and spatial distribution within the cell, our interventions can be more precise and predictable.

5. Functional Expression

Goal: Focus on functions that are relevant to DOE goals (e.g., bioremediation, carbon sequestration, and sustainable energy production).

Challenges: DOE has longstanding biological missions focused on genomics; environmental remediation; sustainable energy production via biofuels, carbon sequestration, cycling, and conversion; biotechnology; and other applications. Combining both genomic sequencing information with fundamental insight into the metabolism and biochemistry of the microbial cell offers a unique opportunity to utilize microbes for energy production and renewal. The 50-year legacy of waste from both the military's atomic weapons programs and the civilian nuclear power industry has created a vast challenge for environmental remediation. Exploring biological approaches has acquired even more importance with escalating costs to store and secure these wastes. The first microbes that DOE sequenced were producers of methane; the more recent ones express abilities for carbon fixation, nitrogen fixation, biomass conversion, and toxic chemical degradation. Most recently, DOE supported the sequencing, at The Institute for Genomic Research (TIGR), of *Deinococcus radiodurans*, a ubiquitous microbe that can withstand more than 1.5 million Rads of direct radiation (some 1500 to

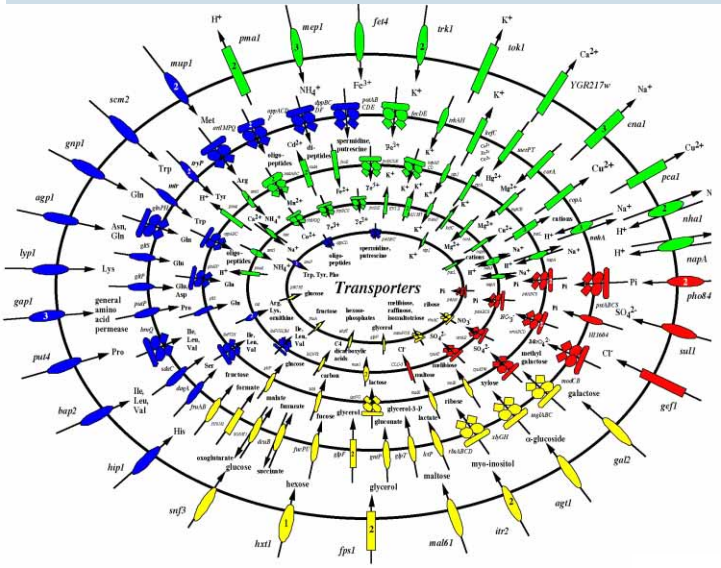


Fig. 5. Transporter Protein Categories. *Biological evolution is conservative; if a protein worked well on a simpler microbe (inner circle), it is often copied and adapted in more complex microbes (outer circles). This is a powerful aid to understanding protein functions. Membrane transport functions are identified by analyzing coding regions in five complete genomes. Circular representations from the center to the outer ring:* Mycoplasma genitalium, Methanococcus jannaschii, Synechocystis PCC6803, Haemophilus influenzae, *and* Saccharomyces cerevisiae. *Colors represent the four role categories: (1) amino acids, peptides, and amines (blue); (2) carbohydrates, organic alcohols, and acids (yellow) ; (3) cations (green); and (4) anions (red). Shapes designate transporter type: ion-coupled permeases (ovals), ABC transporters (composites of circles, diamonds, and ovals), and all other transporters (rectangles). Arrows pointing outward indicate efflux from the cell; those pointing inward designate solute uptake from the environment.* [Reprinted with permission from *Nature*© 1997 (**387**, 461; *www.nature.com*) and Karen A. Ketchum (Celera Genomics Inc).

2000 times the dose that is lethal for humans). This microbe is being engineered to degrade toluene and reduce toxic mercury, serious problems at several DOE waste sites. As the MCP grows, DOE will sustain its focus on environmental remediation, energy production, biotechnology, and carbon management.

Future Directions

While the DOE MGP and MCP focus primarily on environmental, energy, and biotechnological areas, it is worth noting here that there will be spinoff applications in medicine as well. Potential biomedical benefits from comparative genomics studies include insights into the specialized, shared systems used by disease-causing organisms (pathogens) to disable or destroy human cells. Comparing these genomic data with those of other microbes may help scientists understand a diverse range of pathogens that have remarkably similar methods for infiltrating organisms with protein-coding genes capable of sneaking past human defense systems. These protein structures may provide ideal targets for developing completely new types of antibiotics. Thus the MCP will have great value to other agencies and academia, as well as to the private sector. While the MGP will continue to sequence microbial genomes of mission relevance to DOE, its objectives are to mine genomic information from sequenced microbes, improve tools for annotation and analysis of sequence data, develop high-throughput methods for determining gene function and gene expression, and develop methods for examining protein-protein and protein-nucleic acid interaction. The MCP, arising from the MGP, has a bolder goal, to understand a microbial cell in its entirety. Information about the organization and functions of the parts of a simple cell, in its limited complexity, will relate directly to more complex cells and multicellular organisms. We also recognize that perhaps the most significant impact of these programs will come from the delivery of new science, new insights, and new approaches to the difficult challenges that DOE faces in carrying out its varied and demanding missions.

The future promises many exciting developments as the fruits of the DOE–initiated microbial genome and microbial cell research programs mature. Already, we have become more appreciative of the extent of the microbial world's effect on Earth, realizing how little we know about microorganisms and wondering at their potential benefits to our world—if only we are wise enough to discover them. Much work lies ahead, from defining a roadmap for the MCP, with clearly delineated deliverables and benchmarks, to assembling the teams from DOE labs and academia, as well as the private sector, to move towards the goals of this ambitious program. This effort will begin in the near future with a series of workshops to define the specific components of the MCP and build the required roadmap.

DOE already is planning to coordinate and partner with other interested federal agencies, and this will be important to ultimate success. But the ultimate goal, critically dependent on DOE facilities and technologies, as well as DOE's most valuable asset, her interdisciplinary teams of first-rate scientists, is to figure out just how a cell, in its entirety, actually works. Then biologists will have a set of genetic building blocks, and the real adventure can begin—to use them intelligently for assembling useful and specific microbial tools to address the nation's needs.

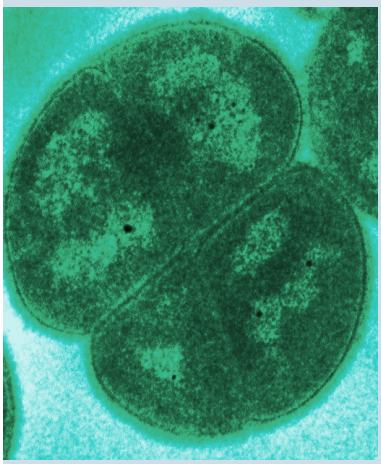


Fig. 6. Forming a Living System. Deinococcus radiodurans *is a completely sequenced microbe capable of cleaning up environmental pollutants in high-radiation environments. Understanding how it (and other microbes) work from their roster of "parts" is the challenge of the Microbial Cell Project.*

DOE Microbial Research and Related Activities

Office of Biological and Environmental Research (OBER), Office of Science

Overview of OBER Research: *http://www.sc.doe.gov/production/ober/ober_top.htm*
Programs and Staff Contacts: *http://www.sc.doe.gov/production/ober/restaff.html*
Funding: *http://www.sc.doe.gov/production/grants/grants.html*

Carbon Management Science Program

http://www.sc.doe.gov/production/ober/carbseq.html

Genome Programs

Microbial Genome Program

http://www.sc.doe.gov/production/ober/microbial.html
http://www.ornl.gov/hgmis/microbialgenomes

Human Genome Program

http://www.sc.doe.gov/production/ober/hug_top.html
http://www.ornl.gov/hgmis

Global Change Research Program

http://www.sc.doe.gov/production/ober/esdrestopic.html

Natural and Accelerated Bioremediation Research (NABIR) Program

http://www.lbl.gov/NABIR

Bioremediation of Metals and Radionuclides: A NABIR Primer

http://www.lbl.gov/NABIR/primer/primer.html

Structural Biology Research Program

http://www.sc.doe.gov/production/ober/msd_struct_bio.html

Environmental Molecular Science Laboratory

http://www.emsl.pnl.gov:2080

Other DOE Offices Supporting Microbial Research

DOE is a major funder of nonmedical microbiology in the U.S. government. Some programmatic areas of microbial study are listed below (for more details, see *www.DOE.gov/people/peoppo.htm*).

Office of Basic Energy Sciences, Office of Science

Renewable energy, carbon sequestration

Office of Energy Efficiency and Renewable Energy

Renewable energy, hydrogen production, ethanol production, organic acid synthesis, cellulose and lignin degradation

Office of Environmental Management

Bioremediation research (organics)

Office of Nonproliferation and National Security

Characterization and detection of potential biological warfare agents