Synthetic Genomes: Technologies and Impact

December 2004

Synthetic Genomes: Technologies and Impact

I. Introduction

The successful completion of the Human Genome Project (HGP) was one of the great moments in science. By revealing the genetic program code of many organisms, biomedical science has been transformed, endowing biologists with the ability to analyze whole genomes and giving hope of comprehending the complexity of entire organisms. This "systems biology" will lead to a deeper understanding of life on our planet.

A new class of technical advances in "synthetic biology" promises to advance the field again. Synthetic biology builds on the HGP by enabling researchers to synthesize genes and even whole genomes. These capabilities, now in the early stages of development, could be breakthrough technologies that enable utilization of the wealth of DNA sequence information provided to address key missions of the Department of Energy (DOE). The genome of a virus that infects bacteria was recently synthesized by a research group at the Institute for Biological Energy Alternatives (Smith et al, 2004), funded by the DOE. Though this is not the first genome to be synthesized, and it is approximately 500 times smaller than the microbial genomes of particular interest to the DOE, the technical advance prompts consideration of the future impact of this technology.

The HGP has provided a remarkable catalog of DNA sequences (the ordered strings of A, G, T and C nucleotides) of the genomes of many organisms of particular importance to DOE missions. New genome sequences are being added almost daily, and many more will be forthcoming as we continue to sample branches of the evolutionary tree and the complex microbial communities of the earth and the oceans. This growing catalog provides "parts lists" that opens a new window into the composition of life and opportunities to harness our knowledge of biology for the benefit of society. The genome project also provides tools for understanding and exploiting genome information – the theme of the Genomics:GTL project. Among the most powerful of those tools is the capacity to synthesize DNA molecules of any desired sequence and to put these sequences to use in microorganisms. This offers the exciting prospect of assembling the best "parts" from the vast diversity of life to create organisms with new capabilities for many different purposes.

Synthetic biology is emerging as a major tool - indeed an enabling technology - in the pursuit of this new science and all it promises for the intelligent and informed application of knowledge to key areas of concern to the DOE. Intelligent assembly of genome "parts" coupled with the understanding from "systems biology" may be keys to the future of energy generation and use, keys to remediation of environmental contaminants and keys to dealing with the problem of carbon sequestration. The past support of research by the DOE has been crucial for providing new biological knowledge and technology that have served not just DOE missions but a much broader range of research and development. Past emphasis by the DOE on development of new technologies and key technical advances enabled the success of the HGP. The DOE can play a similar key role in developing synthetic biology capability.

Synthetic biology needs to be pursued thoughtfully and responsibly. The Federal Government has established the National Scientific Advisory Board for Biosecurity (NSABB - http://www.biosecurityboard.gov/) to provide advice to Federal departments and agencies on ways to minimize the possibility that knowledge and technologies emanating from vitally important biological research will be misused to threaten public health or national security. This report describes the benefits and a course of action for synthetic biology research with the full realization that research recommended will be conducted under the guidance provided by the NSABB.

II. State of the Art for Gene Synthesis Technology

The ability to synthesize DNA in the laboratory was developed more than 20 years ago, and steady improvements in the technology have lowered the cost so that the synthesis of whole genomes (the genetic blueprints of living organisms) is now within reach. Short stretches of DNA (50-100 nucleotides in length) can be chemically synthesized automatically, and, in a series of relatively simple steps, joined end to end to make a single long DNA molecule. The challenge to making these long DNA molecules (5,000 to 10,000 nucleotides) is to achieve accuracy with longer and longer molecules, but this challenge is being met through the development of methods for repairing mistakes.

Several companies (BlueHeron, Genosys, GENEART, Entelechon, GenScript, AnaGen, Qiagen) assemble long (up to 18,000 nucleotides) double-stranded DNA sequences with essentially no errors at costs as low as $2.35 per base-pair. Their proprietary procedures typically start with computer designed single-stranded DNA molecules synthesized by conventional methods that have about one error per 160 nucleotides. Errors in the sequence can be detected by automated DNA sequencing and then repaired. Similar procedures are employed in academic laboratories with costs that can be similar.

III. New Technology

A. What is New?

The principle of synthesizing long stretches of double-stranded DNA is not new (Cello et al. 2002 and Stemmer et al. 1995). Incremental improvements in the technology have brought it to the verge of becoming a breakthrough technology allowing very long DNA strands to be synthesized quickly, with great accuracy and at low cost. This opens significant new scientific and commercial opportunities. Smith et al, 2004, were able to assemble up to 130 pieces of synthetic DNA into one long double-helical molecule over 5,000 nucleotides in length. Purification of the initial DNA mixture reduced the error rate and, therefore, also the number of needed repairs to the sequence. The final assembly of 5386 base pairs matched the natural virus genome and was accomplished in only 14 days. Although this genome is 100 times smaller than the smallest genome of a free-living microbe, this success brings the synthetic biology approach into focus.

These particular advances alone are probably not sufficient to make synthetic biology generally practical, but we expect rapid and substantial advances in the near future.

B. Near and Far Future. One to Ten Years

(1) We expect substantial reduction in the cost and improvement in the length and accuracy of synthesis of the DNA starting pieces. This is likely to come from systems that will use existing chemistries to synthesize thousands of custom DNA sequences with 20-200-fold lower costs. However, accuracy of the sequence is critical and additional development of technologies to give more accurate and longer sequences will be required.

2) Computer algorithms for choosing the optimal sequence and avoiding unwanted features (such as folded structures in the starting molecules) are becoming more sophisticated. (However, there is a need for "open source" versions of these algorithms so that the scientific community can benefit from rapid modifications.)

(3) There will likely be significant improvements in methods for correcting errors in the assembled DNA sequence. These could take advantage of mechanisms that have evolved within cells to correct mistakes in their own DNA or this may be achieved by simple physical purification of synthetic DNA molecules.

(4) Progress is also expected in development of methods for assembly of chromosome-sized synthetic DNA in bacterial cells. This will likely come from automation of homologous recombination - the process of precisely joining DNA molecules.

IV. Generic Enabling Technology

The original method of assembling DNA molecules with genes of interest is to pick and choose among genes in existing organisms by breaking up the genomes with reagents that cut DNA at specific places. The thousands of DNA fragments that result are then sorted to find the one of choice. This approach has proven to be very powerful, but it is time consuming and tedious. The development of the polymerase chain reaction (PCR) simplified this process by enabling amplification of only the gene of interest from the whole genome providing an efficient and powerful method to mix and match genes. However, the DNA of the source genome must be available in the laboratory. The newer synthetic biology approach requires *only* the DNA sequence recorded in the DNA databases – the sequence can then be made from scratch. Moreover, the sequence can be modified and customized as desired. As the technology improves and the cost goes down, this synthetic approach is likely to become the method of choice. It brings the democratization of science, since the genome sequences are largely in the public sector and therefore universally accessible and usable.

Development of the ability to synthesize cheaply and efficiently large pieces of DNA (> 10,000 base pairs) and to assemble these into complete microbial genomes will have a wide-ranging impact on biological research. It promises to increase greatly the yield of knowledge from genetic research. The ability to understand microbial biology is currently

severely limited by a lack of understanding of the complex processes of cells. The recent flood of DNA sequences of many genomes provides a rich information resource that has the potential to reveal these complex relationships. Realizing this goal will require testing the numerous hypotheses that are generated from analyses of these data, and this will necessitate experimentation on a scale not yet achieved. Synthetic genome technology promises to fulfill this need by enabling the rapid and easy construction of multiple genomes. Many different genes could be tested in the laboratory in many combinations for their effect on a desired process; many different regulatory elements that control the activity of genes could be altered simultaneously to determine their role in cellular function. The knowledge gained from these experiments is certain to result in significant improvements in our ability to rapidly explore microbial functions and to assemble tailor-made genomes for specific missions.

Currently, constructing and optimizing organisms to carry out one particular process is a laborious process requiring painstaking manipulation of genes one at a time – base pair by base pair. The ability to synthesize large pieces of DNA would enable optimization of gene expression by finding the most efficient coding and the best regulatory information for that gene. The function of the protein product of the gene can be optimized by trying different combinations of amino acid building blocks. Also, various combinations of genes in metabolic pathways can be tested to find the optimal set for the problem at hand. Genes from widely different organisms that might be candidates for executing some needed function can be quickly tested.

The ability to move with ease from sequence to biological tests will bring the full intellect and imagination of scientists to bear on solving problems using biology as a tool.

V. Benefits & Concerns

A. Microbes for Improved Energy Production, Carbon Sequestration and Bioremediation

The astounding diversity of microorganisms in nature has the potential to provide powerful tools for fulfilling DOE missions, but their genetic capacities need to be understood. The new synthetic gene technologies described above are breakthrough tools that promise to advance these goals. These technologies enable efficient use of the best parts of nature's offerings – even providing access to genes from the many organisms that cannot be grown in the laboratory. Genes and gene families from diverse organisms can be assembled to overcome current barriers. For example, genes could be synthesized and combined to optimize conversion of plant biopolymers to high-energy products, to enhance hydrogen production or to improve the fixation of carbon dioxide or nitrogen. This "pathway engineering" could help sequester the fixed carbon in soils and sediments. Optimal pathways for bioremediation could be constructed to detoxify pollutants. For instance, PCBs are too complex for most microbes to metabolize, but new combinations of synthetic genes chosen from among nature's reserve may offer insights towards novel solutions to this and other problems of persistent environmental contaminants.

However, as increasingly complex sets of genes are brought together, the difficulties of achieving the desired result become greater and greater. When extensively modifying the genetic composition of a microbial cell, the changes must be consistent with cell survival in the context of the desired biological niche. The possibility of unanticipated results may increase dramatically with gene number.

B. Genetically Modified Organisms-gene Escape

Concerns about the uncontrolled spread of genetically modified organisms are not new. The synthetic genome work adds a new tool to the kit for genetically modifying an organism and is likely to lead to a more rapid development of organisms with new capabilities. While the crucial issues may not be new, the level of concern may be raised by the rapid pace of technology.

In order for the DOE to employ microbial solutions for alternative energy sources, bioremediation, carbon sequestration, or other applications, additional scientific questions must be addressed. For examples, what will be the ecological impact of releasing a new organism into the environment, how rapidly and in what direction will released organisms mutate, and how can released organisms be controlled or inactivated after release, if desired? Will introduced DNA be exchanged with native organisms, and if so, what is the effect?

At the same time, new solutions may be offered by enhanced synthetic genomics capabilities. This technology will ease the systematic removal or modification of genes for known allergens or the addition of genes for desired attributes, e.g., to make foods that possess vaccine characteristics. It would also enable the ready engineering of tags, forensic ”bar codes,” to allow tracking and attribution of an organism that would assist both ownership verification and regulatory compliance if such organisms were used to generate pharmaceuticals (see JASON report: Block et al. 2004). It may also be possible to use this technology to modify the genetic code of a microorganism so that its genes would not function in other organisms even if they escape. The issues of genetically modified organisms need to be faced by society, including the scientific community. If used wisely, the new techniques of synthetic genomics could be a major benefit.

C. Pharma & Industrial Production: Enzymes, Single-Cell Protein, Protein-drugs

Microbial cells, grown in fermentors, are important factories for the synthesis of many pharmaceuticals, enzymes, vitamins and other high-value products. A fundamental challenge in such processes is mutation and other genetic changes that occur during growth. Accumulation of mutations is a common property of all biological populations, but in these fermentor populations they cause the production to be unstable or fail. Synthetic genome technology would enable rapid modifications of the initial microbes to minimize mutations, allowing longer, more reliable production runs. Engineering cells to attain and maintain lower mutation rates can be a combination of systems modeling and directed laboratory evolution.

For new pharmaceuticals whose biosynthesis depends on novel gene sequences, direct gene synthesis will be very useful in building the drug of choice, and will greatly aid the combinatorial syntheses of many different gene versions to find the most efficacious form. The new synthetic genomics would greatly improve pharmaceutical practices in this regard.

D. Implications for Vaccine Development and Diagnostics

Some benefits of this evolving technology reach well outside DOE missions and have significant implications for missions of other agencies.

Approaches to vaccine development and diagnostics are being revolutionized by the availability of genome sequences and DNA technologies. A key contribution of gene synthesis technologies is the ability to assemble and vary sequences rapidly, even without access to the original infectious organisms. Taking SARS as an example, the rapid determination of the genome sequence of the SARS coronavirus immediately enabled any laboratory in any country to develop a vaccine via gene synthesis. Another example is influenza A for which current vaccine strategies to combat annual epidemics could be enhanced by rapid gene synthesis coupled with current vaccine production methods. DNA can be used for vaccination, and gene synthesis would advance this technology by removing the steps requiring passage or amplification in bacteria. Gene synthesis could be used to make many more variants than is currently possible, which would achieve a major goal of generating a broadly protective immune response for vaccines against highly diverse and variable viruses like human immunodeficiency virus (HIV) and hepatitis C virus. Rapid gene synthesis would also facilitate optimization of other parameters to enhance antigen presentation such as expression of elements that would stimulate the immune response.

Rapid gene synthesis could be an important element in the ability to very rapidly develop vaccines for emerging infectious diseases whether they arise from nature or are the product of human intervention, e.g., bioterrorism. In principle, large-scale DNA sequencing, gene synthesis and DNA vaccines could be combined to allow the development of a new vaccine in response to an ongoing epidemic.

Genome synthesis technology could also provide high-quality nucleic acid standards of defined sequence needed for diagnostics. As the constellation of microbial and viral genome sequences grows, the diagnostic field is in a continual state of flux. DNA and RNA diagnostic standards could be efficiently synthesized and easily modified as needed.

E. Bioterrorism

Synthetic genome technology could also enable more efficient construction of bioweapons. A known pathogen, e.g., smallpox, could be rapidly synthesized from published DNA sequence or an existing organism could be more readily modified to make it pathogenic. Current technology can already be used to these ends, and while the recent synthesis of a viral genome does not provide fundamentally new technology, the new methods used could lower some cost or time barriers. However, genome synthesis technology could provide

means for developing vaccines against pathogens, methods for pathogen detection and new ways to identify therapeutic targets.

The National Research Council's (NRC) Committee on Research Standards and Practices to Prevent the Destructive Application has recently made policy recommendations for scientists and the U.S. Government that pertain to the conduct, review and publication of a broad scope of biotechnology-related research. We agree with the Committee's recommendations, and believe that synthetic genome technology falls completely within the purview of this report. The Committee identified a number of "experiments of concern" such as research that would demonstrate how to render a vaccine ineffective, or experiments that would confer resistance to therapeutically useful antibiotics or antiviral agents. As a tool or method, the synthesis of segments of DNA that contain one or more genes would not fall directly into the category of "experiments of concern," but it does add another enabling technology. We agree with the recommendations put forward by the Committee for scientific review of proposed research and publication for the specific types of research identified by it.

Acknowledging the potential for misuse of synthetic genome technology before adequate defenses can be mounted, it would be prudent for scientists to work together with experts in national security to explore and develop practical strategies to prevent [and/or detect] its misuse, as recommended by the NRC Committee. The Committee recommended additional protection measures such as containment and safeguarding of biological material, but synthetic genome technology would eliminate the need to obtain actual DNA specimens if the desired DNA sequences were available in publications or databases. Biological containment alone may be insufficient to minimize misuse. Strategies such as monitoring DNA sequences shipped from DNA synthesis facilities capable of producing large segments of DNA might be in order much like the Drug Enforcement Agency attempts to monitor chemical purchases to detect drug-making activity.

An additional threat to society is not from actual deployment of a bioweapon but from a loss of public trust in scientists if the scientific community is perceived as not taking the potential threats seriously. Furthermore, to be effective, prevention strategies will have to include researchers in the private sector as well as those supported by federal funding.

F. Ownership and Access

The ability to synthesize stretches of DNA long enough to contain a full gene or even several genes has significant implications for the research process. The ability to synthesize genes will potentially lower barriers to patenting DNA sequences and the products that result by facilitating demonstration of utility and raising questions about the barrier of "product of nature" doctrine. Both beneficial and harmful implications of patenting DNA for the conduct of research and development have already been identified, and enhanced synthetic capabilities will likely contribute to both ends of the spectrum.

The ability to synthesize functional genes and groups of genes should increase access to genetic materials for all scientists because exchange of actual genetic material will not be

necessary. Scientists will be able to synthesize genes from published DNA sequences alone. A positive consequence of this is that a greater number of scientists can have access to genes once their sequences are published. This will impact the use of material transfer agreements and contracts. It may also reduce the need for transfer of live pathogens among laboratories and thus have implications for the regulation of the handling of pathogens for research purposes.

G. Ethical Considerations

The technology of synthetic biology provides a new set of tools. Any ethical challenges come from the use of the tools, not from the tools themselves. Ethical and moral issues related to synthesizing genomic DNA sequences have been considered previously (Cho et al. 1999) with regard to experiments aimed at constructing an organism with a minimal number of genes. The issues raised by this synthetic biology technology are similar, and we do not foresee any ethical concerns unique to this technology. We believe that this technology does not fundamentally change the philosophical and moral considerations discussed previously.

VI. Recommendations

Synthetic biology technology will likely become a standard method for exploring the capabilities and uses of the growing database of genome sequences. Further technology developments should be encouraged and supported. More efficient and accurate synthesis of starting DNA sequences will likely be driven by commercial interests. However, automation of the steps of error correction and reintroduction of new sequences into cells will almost certainly require funding by federal agencies.

This technology fits well with the DOE Microbial Genome Project and with the Genomics:GTL project and is likely to be the most efficient way to use nature's parts list to meet DOE missions.

Research funded by DOE, should be conducted in a manner consistent with the NRC's Committee on Research Standards and Practices to Prevent the Destructive Application provided by the NSABB. Research with the intent of creating organisms for environmental release should include careful evaluation of potential environmental and health impacts and be designed to minimize foreseeable harms.

We believe that synthetic genome technology promises great opportunities. While issues raised by this technology are not new, additional studies of the potential concerns are warranted since the scope and volume of these issues are likely to increase.

We urge the DOE to seize this opportunity and promote further technology development to realize the full benefit of synthetic biology for its missions. Nature has provided a wealth of biological "parts," honed over hundreds of thousands of years of evolution. Maturity of this technology will open this largely untapped reserve.

References

Block S, Donoho D, Hwa T, Joyce G, Nelson D, Stearns T, Weinberger P, Williams E. DNA Barcodes and Watermarks. Jason Report JSR-03-305, 2004; The Mitre Corporation, McLean, VA.

Cello J, Paul AV, Wimmer E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science. 2002 Aug 9; 297(5583): 1016-8.

Cho MK, Magnus D, Caplan AL, McGee D. Policy forum: genetics. Ethical considerations in synthesizing a minimal genome. Science. 1999 Dec 10; 286(5447): 2087, 2089-90.

Forster AC, Tan Z, Nalam MN, Lin H, Qu H, Cornish VW, Blacklow SC. Programming peptidomimetic syntheses by translating genetic codes designed de novo. Proc Natl Acad Sci U S A. 2003 May 27; 100(11): 6353-7.

Smith HO, Hutchison CA 3rd, Pfannkoch C, Venter JC. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci U S A. 2003 Dec 23; 100(26): 15440-5.

Zhu,J, Shendure,J, Mitra, RD, Church, GM (2003) Single Molecule Profiling of Alternative Pre-mRNA Splicing. Science. 2003 Aug 8;301(5634):836-8.

Fink, GR et al, (2004), Biotechnology Research in an Age of Terrorism: Confronting the Dual Use Dilemma. Committee on Research Standards and Practices to Prevent the Destructive Application of Biotechnology. National Research Council of the National Academies. Prepublication proofs, www.nap.edu, The National Academies Press, Washington D.C.

Chair:
Dr. Raymond Gesteland, University of Utah

Participants:
Dr. Mildred Cho, Stanford University
Dr. George Church, Harvard Medical School
Dr. David Galas, Keck Graduate Institute
Dr. Charles Rice, Rockefeller University
Dr. Mark Johnston, Washington University Medical School
Dr. Gerald Joyce, Scripps Research Institute
Dr. James Tiedje, Michigan State University