

scientific community. Among the enterprises that have already headed down this road are comparative genomics, with the genomic sequencing and analysis of key model organisms being done by some of the centers that worked on the human genome. Large-scale genomic enterprises now extend well beyond straightforward sequencing, as witnessed by the recent launch from our own funding agencies of other large-scale biology initiatives involving functional genomics (11), structural biology (12), microbial genomics and proteomics (13), and haplotype mapping focusing on human populations (14). Other candidates for the "big science" approach to biology include interagency and international approaches to biological database creation and maintenance, the construction of public small-molecule libraries for use by basic scientists in their efforts to chart biological pathways, and the large-scale application of microarray technologies with the potential for applications in a wide range of biological research settings.

The millions of people around the world who supported our quest to sequence the human genome did so with the expectation that it would benefit humankind. Now, at the dawning of the genome era, it is critical that we encourage the same intensity toward deriving medical benefits from the genome that has characterized the historic effort to obtain the sequence. If research support continues at vigorous levels, we imagine that genome science will soon begin revealing the mysteries of hereditary factors in heart

disease, cancer, diabetes, schizophrenia, and a host of other conditions. Genomics holds the promise of "individualized medicine," tailoring prescribing practices and management of patients to each person's genetic profile. These revelations should also lead us to develop and target drugs in a rational fashion to genes, protein pathways, and networks shown to be involved in primary disease pathogenesis. Furthermore, a better understanding of the genetic factors that influence susceptibility and/or response to various infectious diseases could have an enormous impact on health in the developing world. Genomics also has the potential to help agricultural researchers develop better crops and livestock, environmental scientists create better methods of cleaning up toxic material, production experts streamline industrial processes, and energy researchers work toward sustainable, nonpolluting energy sources.

For this grand vision (15) to come true, however, we in the biologic research community need to pursue the next generation of research projects with the same determination and creativity that the dedicated scientists of the HGP used to spell out the human genetic code. This Herculean challenge is by no means limited to biologists. We call on leaders across science and society, across academia and industry, and across political and geographic boundaries to join us on this exciting voyage to understanding ourselves. If the past 50 years of biology is any indication of the future, the best is certainly yet to come.

VIEWPOINT

Realizing the Potential of the Genome Revolution: The Genomes to Life Program

Marvin E. Frazier,¹ Gary M. Johnson,² David G. Thomassen,^{1*} Carl E. Oliver,² Aristides Patrinos¹

The systems biology revolution is proceeding along multiple pathways as different science agencies and the private sector have adopted strategies suited to their particular needs and cultures. To meet this challenge, the U.S. Department of Energy has developed the Genomes to Life (GTL) program. A central focus of GTL is environmental microbial biology as a way to approach global environmental problems, and its key goal is to achieve, over the next 10 to 20 years, a basic understanding of thousands of microbes and microbial systems in their native environments. This focus demands that we address huge gaps in knowledge, technology, computing, data storage and manipulation, and systems-level integration.

This month we celebrate progress in completing the human genome sequence as well as the 50th anniversary of James Watson and Francis Crick's Nobel Prize-winning description of the DNA double helix (1). Spurred by achievements in genomic research, the 21st century will bring revolutionary solutions to some of our most urgent and

expensive challenges in health, energy, and the environment. In this new era, we will understand the workings of entire biological systems and how they interact with and respond to their environments. The Office of Science of the U.S. Department of Energy (DOE) will use that information to conduct basic research into biological solutions for

References and Notes

1. J. D. Watson, *The New York Times*, 5 June 1990, p. C1.
2. W. H. Lambright, *Managing "Big Science": A Case Study of the Human Genome Project* (The Price Waterhouse Coopers Endowment for the Business of Government, Arlington, VA 2002). Available at www.endowment.pwcglobal.com/pdfs/LambrightReport2.pdf.
3. J. D. Watson, *Science* **248**, 44 (1990).
4. J. Sulston, G. Ferry, *The Common Thread: A Story of Science, Politics, Ethics and the Human Genome* (Joseph Henry Press, Washington, DC, 2002), pp. 73–76.
5. M. V. Olson, *J. Mol. Biol.* **319**, 931 (2002).
6. U.S. Department of Health and Human Services, U.S. Department of Energy, *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years* (Government Printing Office, Washington, DC, 1990).
7. F. S. Collins, D. Galas, *Science* **262**, 43 (1993).
8. F. S. Collins et al., *Science* **282**, 682 (1998).
9. NHGRI Standard for Quality of Human Genomic Sequence [revised 10 February 2002]. Available at www.genome.gov/page.cfm?pageID=10000923.
10. "Sharing Data from Large-scale Biological Research Projects," meeting sponsored by The Wellcome Trust, Fort Lauderdale, FL, January 2003. Available at www.wellcome.ac.uk/en/1/awtpubrepdat.html.
11. The ENCODE Project: *ENCYClopedia Of DNA Elements* [updated February 2003]. Available at www.genome.gov/Pages/Research/ENCODE/.
12. Structural Biology Consortium [updated 19 April 2001]. Available at www.wellcome.ac.uk/en/1/biosfgstr.html.
13. Genomes to Life [updated 26 February 2003]. Available at <http://doegenomestolife.org>.
14. NHGRI, *The Haplotype Map*. Available at www.genome.gov/Pages/Research/HapMap.
15. F. S. Collins et al., *Nature*, in press.
16. We express our profound gratitude to leaders of all of the sequencing centers involved in the International Human Genome Sequencing Consortium, along with all other participants in the HGP, 1990 to 2003.

producing clean energy, removing excess carbon dioxide from the atmosphere, and remediating contaminated environments left as a legacy of the Cold War.

Genomic information is providing the starting point for understanding the instructions for the manufacture of all of life's molecular machines and the systems needed to control and operate them. Through a labyrinth of pathways, networks, chemistry, and

¹U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, SC-70, ²Office of Advanced Scientific Computing Research, SC-30, Germantown Building, 1000 Independence Avenue, S.W., Washington, DC 20585–1290, USA.

*To whom correspondence should be addressed. E-mail: david.thomassen@science.doe.gov

mechanics, this exquisite machinery makes the cell and the organism come alive. Understanding, not “simply” decoding, the operation, function, and coordination of genome information will be the next transforming phase in biology. From experience gained in sequencing genomes and conducting large-scale biology projects, we have learned that the combined capabilities and imagination of biological, physical, and computational scientists will be needed to organize creative new venues for discovery.

Genomes to Life

To address this challenge, over the past 3 years hundreds of scientists from universities, national laboratories, and industry have helped DOE develop a research program called Genomes to Life (GTL) (2). A central goal of this program is to understand the microbes and communities of microbes, and their molecular machines and controls at the molecular level so well that we can use them to address DOE and national needs (3).

The DOE Office of Science has the ability and institutional traditions to bring the biological, physical, and computing sciences together at the scale and complexity required for GTL success. DOE's academic affiliations, national laboratories, and other resources include major facilities for DNA sequencing and molecular-structure characterization, the high-performance computing resources of the Office of Advanced Scientific Computing Research (OASCR), the expertise and infrastructure for technology development, and a tradition of productive multidisciplinary research essential for such an ambitious research program. More than 70 microbes studied in GTL have had their genomes sequenced in the DOE Microbial Genome Program (4). In the effort to understand biological systems, these strong assets and the GTL program will complement and extend the capabilities and research supported by the National Institutes of Health, National Science Foundation, other government agencies and institutions, and industry.

Goals

We know that proteins generally don't work alone within cells but rather carry out their functions together with other cellular components in multiprotein, macromolecular machines. The first goal of GTL is to systematically identify and characterize the complete inventory of molecular machines in key microbes. This is a daunting challenge, partly because the complex intracellular drama is playing out in a space of micrometer dimensions. Nonetheless, we will need to dynamically image and functionally analyze the critical molecular machines within microbes.

An organism's genome also provides the basis for complex regulatory networks that

control and coordinate the assembly and operation of these molecular machines. A second goal of GTL is to deconvolute and model these regulatory networks.

The genome alone (or even the proteins it encodes) does not determine the physiological state of any cell at a given moment. Environmental factors, as well as stochastic and chaotic processes, affect the behavior of individual organisms and communities. This phenotypic, “epigenetic” diversity clearly has important biological consequences, yet conventional averaging techniques are blind to it. For this reason, we ultimately must be able to study individual microbes dynamically within a community often containing many species and grown under well-understood environmental conditions. The third goal of GTL is to understand the organization and functional capabilities of complex microbial communities.

Perhaps the greatest challenge facing GTL, and indeed all of biology, will be the capture, management, and analysis of the torrent of new data and information flooding over us in the next decades. Many different types of data must be integrated into databases and metadata structures that will range well into the petabyte scale—presenting huge analysis, data-mining, and visualization challenges. Making this infor-

mation useful and accessible to all biologists will require faster and more capable computer architectures tied together in advanced networks linked into research laboratories. Substantial improvements in mathematical, statistical, and algorithmic methods clearly will be needed to power tools for analyzing data and developing models for understanding myriad biological processes (Fig. 1). Therefore, the fourth goal of GTL, which underpins the entire program, is to address this critical need by enlisting a large community of mathematical and computational scientists as research partners and by developing an infrastructure that encompasses comprehensive computing, analysis, modeling and simulation, and information management. Indeed, there may already be an applicable body of applied math knowledge, developed on other applications, with which to jump-start a major push on modeling biological systems. To carry out this goal, a partnership has been formed between the DOE Office of Biological and Environmental Research and OASCR.

Why Microbes?

Through billions of years of evolution, microbes and communities of microbes have found homes in every conceivable environment on Earth, from arctic flows to boiling

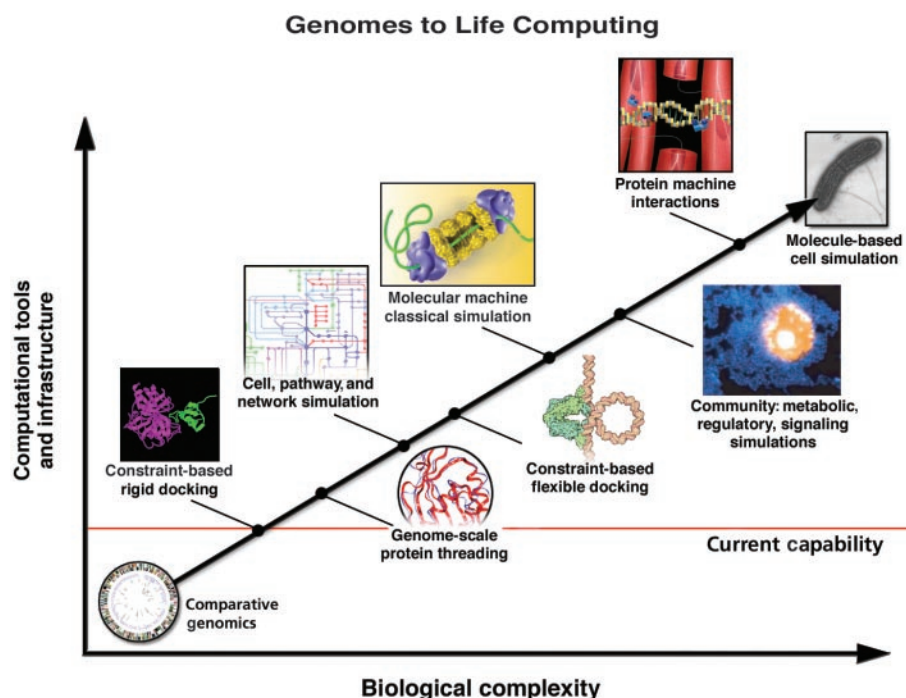


Fig. 1. Concept diagram schematically illustrating a path from basic genome data to a more detailed understanding of complex molecular and cellular systems, and the need to develop new computational analysis, and modeling and simulation capabilities to meet this goal. The points on the plot are very approximate, depending greatly on how each problem is abstracted and represented computationally. Research is under way to create the mathematics, algorithms, and computer architectures required to understand each level of biological complexity. Detailed definitions of computational terms are available as supporting online text. [Images: E. Uberbacher and Y. Xu, Oak Ridge National Laboratory (Constraint-based rigid docking; Genome-scale protein threading); B. Mansfield, Oak Ridge National Laboratory (Cell, pathway, and network simulation; Molecular machine classical simulation); D. S. Goodsell, Scripps Research Institute (Constraint-based flexible docking)]

thermal vents on the ocean floor. They use many different substances as energy sources, even some that we consider toxic wastes, and they produce diverse sources of energy, including biomass, methane, and hydrogen. Microbes make up more than 50% of the Earth's biomass (5) and an even higher percentage of living surfaces, further pointing to the value of microbial activity in maintaining the biosphere. As the foundation of the biosphere, microbes control Earth's biogeochemical cycles and affect soil productivity, water quality, and global climate (6). Indeed, the ability of this planet to sustain life is completely dependent on microbial activity.

Understanding the microbial world thus is key to our energy and climate futures, to the development of sustainable modes of living and advanced industrial technologies, and, in no small part, to understanding how life on Earth functions. Individual microbes and communities of microbes have already "found" solutions for many of our current challenges in energy production and environmental cleanup.

New Technologies

The success of the Human Genome Project was due largely to the development of high-throughput, cost-effective DNA-sequencing tools and methods and new biological resources

that helped to feed the DNA-sequencing "factories." Although GTL will be based on the same model, it will be a more complex endeavor because we also must isolate, identify, and characterize proteins, multiprotein complexes, and regulatory networks in living systems. GTL-defined strategies for implementing and integrating breakthroughs in mass spectrometry (MS), nuclear magnetic resonance, microtechnologies, automation, imaging, and other areas are beginning to reveal a systems view of certain aspects of microbial studies. These new approaches must be used at sufficient scale and with the continual inflow of new technologies that will meet the needs of the current (and next) generation of scientists.

Big Science Versus Small Science: The Need for a Third Model

The genome revolution presents the biology community with the complex, yet critical, challenge of preserving the creativity and entrepreneurial spirit of the single investigator in the face of the increasingly sophisticated and costly resource requirements of leading-edge biological research. Already, only a minority of even large laboratories can afford to be adequately endowed. Individual investigators need the capabilities of big science, and big science needs access to the energy and resourcefulness of individual investigators.

Our strategy is to help build a bridge between large and small laboratories by making the most sophisticated and comprehensive technologies, materials, and information equally available to all scientists. To accomplish this goal, plans for a set of four core research user facilities have been developed (7), and we hope to phase in these facilities over the next 5 to 10 years. Using combinations of new equipment and technologies, automation, data management, and data-analysis tools, the integrated user facilities will build on each other's technological and computational capabilities to provide the GTL program and the scientific community with an unprecedented set of resources for systems biology.

Facility for production and characterization of proteins. The first GTL facility will surmount a principal roadblock to whole-system analysis by implementing high-throughput production and characterization of microbial proteins. It also will generate protein-tagging reagents for identifying, tracking, quantifying, controlling, capturing, and imaging individual proteins and molecular machines in living systems. Over the next 10 years, our goal is to produce 250,000 proteins in milligram quantities; around 1 million molecular tags for those proteins; and multiple biophysical characterizations of each, beginning with an organism's genomic sequence.

Facility for whole-proteome analysis. This facility will be devoted to understanding the range of proteins expressed by microbes under different conditions—the genomic response to environmental cues. Here will be characterized the proteomes and cellular metabolites such as small RNA molecules, glycosylated products, carbohydrates, ethanol, and lactic acid of diverse microbes isolated from different environmental conditions. Advanced MS methods already have been developed and tested for these purposes, but many more will be required.

Facility for characterization and imaging of molecular machines. Working from protein building blocks and information on microbial proteomes, we need to understand the functions of a given microbe's thousands of molecular machines under various environmental conditions. This facility will isolate, identify, and characterize molecular machines, verify their presence and location within living cells, and develop the ability to image component proteins.

Facility for analysis and modeling of cellular systems. Building on the comprehensive knowledge base derived from GTL research, including the three facilities described above, investigators must efficiently elucidate and predict complete metabolic processes and gene regulatory networks in individual microbes and microbial communities. This facility will combine advanced computational, analytical, and experimental capabilities. The goal is to integrate the observation, measurement, analysis, and full-systems modeling and simulation of spatial

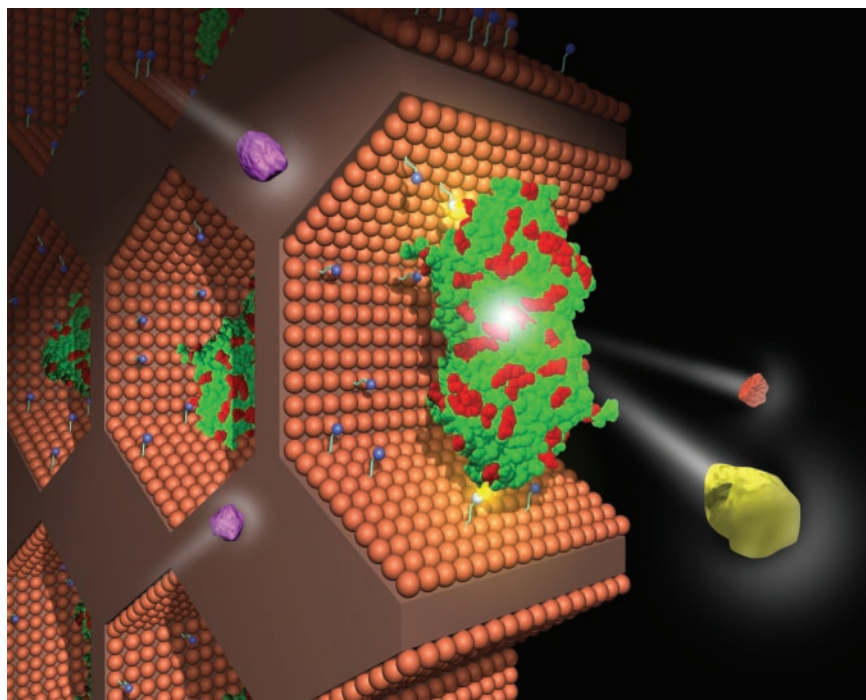


Fig. 2. A biomimetic future. The figure depicts an enzyme, organophosphorous hydrolase (green), embedded in a synthetic membrane that enhances its activity and stability. The enzyme transforms toxic substances (purple molecule at left) to harmless by-products (yellow and red molecules at right). Such nanostructures could eventually be used to develop efficient enzyme-based methods, implemented on an industrial scale, to produce energy, remove or inactivate contaminants, and store carbon to mitigate global climate change. Other potential highly useful applications are food processing, pharmaceuticals, separations, and the production of industrial chemicals. [Image: M. Perkins, Pacific Northwest National Laboratory]

and temporal variations in structures and functions of cellular systems, from individual microbial cells to complex communities and multicellular organisms.

Developing Biotechnology Solutions to National Challenges

On 11 June 2001, President Bush, describing the Administration's approach to addressing global climate change, spoke of the need for science-based solutions. "We all believe technology offers great promise to significantly reduce emissions—especially carbon capture, storage, and sequestration technologies. . ." (8). This theme was further emphasized by guidance from the president's science adviser in preparing the FY 2004 budget, which stated that achieving a molecular-level understanding of life is a national science priority (9). The vast amount of information contained in the hundreds of sequenced genomes and the thousands to come offers an unprecedented starting point for understanding complex biological systems.

Opportunities for clean energy. Petroleum refineries now "crack" crude oil through heat and catalysis to create gasoline and other petroleum products. In the future, biorefineries could use more efficient varieties of microbial cellulases to crack the complex cellulose and hemicellulose in plant walls into simple sugars that would then be fermented, via traditional approaches, into fuel ethanol and other products. GTL basic research can accelerate this vision by improving the understanding of both plant cell-wall construction and microbial enzyme systems necessary to deconstruct the walls into simple sugars.

Hydrogen is another clean fuel source. Within some microorganisms capable of hydrogen production at room temperature and pressure, hydrogenases and nitrogenases are part of the molecular machinery. By understanding the molecular details of pathways and networks involved in hydrogen production, we may be able to create cost-effective and efficient sources of energy and biological generators superior to standard chemical-engineering approaches.

In addition to the potential energy-security benefit of using more bio-based fuels, the resulting reduction in atmospheric CO₂ would help to mitigate global climate change. Knowledge gained from GTL will provide a vital foundation to those responsible for developing new energy sources, including industry, private companies, and the DOE Energy Efficiency and

Renewable Energy programs on biofuels.

Opportunities for reducing climate change. Atmospheric concentrations of greenhouse gases are influenced by microbial activity. Carbon and nitrogen are circulated through the atmosphere, oceans, organisms, and soil largely by microbial metabolic chemistry. Even plant growth and responses to climate change are affected by microbial communities in the root zone.

A more comprehensive understanding of such communities in the terrestrial and oceanic biosphere will enable us not only to understand more about natural cycles but also to predict the future impacts of greenhouse gases on ecosystems and how ecosystems contribute to gas concentrations. Furthermore, we could devise ways for plants and microbes to take up and retain more carbon from the atmosphere, grow faster and in places where they wouldn't normally grow, become easier to harvest, and contain material more easily converted to clean energy.

Future generations of clean energy sources might be based on biological processes such as algal production of hydrogen from water and sunlight or chemical processes such as re-forming fossil fuels and sequestering carbon in the subsurface. We predict that, in the next few decades, the distinction between biological and chemical production in the many and various pathways will be meaningless because biomimetic processes will be used (Fig. 2).

Opportunities for cleaning up the environment. The DOE weapons complex presents some of the most daunting cleanup challenges ever attempted—over 3000 waste sites, 50% of which have soils, sediments, or groundwater contaminated with radionuclides or metals (10). The cost of remediating these sites is prohibitive if today's practice of excavating and treating is used. Estimates range widely but always are above \$50 billion! To make matters worse, this "pump and treat" method frequently turns out to be ineffective.

Neutralizing contaminants in place (in situ technologies), especially by using microbes already living in these sites, could have a huge impact on the cost of cleanup. For several years, bioremediation has been used effectively to break down organic contaminants such as oils or solvents. Metals and radionuclides, however, are a large and unique problem at DOE waste sites, where they present another challenge altogether. They cannot be converted into entirely different chemical forms by microbes or even by harsh chemical treatments. Microbes can, however, change the chemical properties and thus the solubility of metals and radionuclides in water, caus-

ing them to precipitate as solids in their underground locations. Their availability to humans, animals, or plants is thereby reduced, and they are rendered effectively nontoxic.

The revolution in our understanding already has begun. For example, the *Shewanella* Federation (11, 12) consists of teams of scientists from academia, national laboratories, and private industry who are probing in detail this remarkably versatile organism that can precipitate uranium from groundwater. The federation has made progress in preliminary proteome analyses using a combination of MS approaches that yielded nonredundant identification of 3862 proteins out of a predicted 4931 open reading frames representing virtually every functional class, as well as the identification of posttranslational protein modifications (13).

The Time Is Now

Knowledge is power, and we must develop a comprehensive understanding of biological systems if we are to use their abilities effectively to meet daunting societal challenges. With GTL, we are poised to make important contributions to biological solutions for many of the energy, environmental, health, and agricultural challenges facing the nation and the world. Addressing these challenges expeditiously demands that we take bold steps now to achieve a new, faster, and more efficient pace of biological discovery.

References and Notes

1. J. D. Watson, F. W. Crick, *Nature* **171**, 737 (1953).
2. Genomes to Life: <http://DOEGenomesToLife.org>.
3. *Genomes to Life: Accelerating Biological Discovery* (programmatic roadmap); <http://doegenomestolive.org/roadmap/toc.html>.
4. www.ornl.gov/microbialgenomes/organisms.html
5. W. B. Whitman, D. C. Coleman, W. J. Wiebe, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6578 (1998).
6. J. Copley, *Nature* **415**, 572 (2002).
7. *User Facilities for 21st Century Systems Biology: Providing Critical Technologies for the Research Community*, presentation to the Biological and Environmental Research Advisory Committee (3 December 2002); <http://DOEGenomesToLife.org/pubs/GTLfac34BERAC45.pdf>.
8. George W. Bush on Global Climate Change, The White House, 11 June 2001; www.whitehouse.gov/news/releases/2001/06/20010611-2.html.
9. *FY 2004 Interagency Research and Development Priorities*, Office of Science and Technology Policy and Office of Management and Budget (2002).
10. *Estimating the Cold War Mortgage: The 1995 Baseline Environmental Management Report*, DOE/EM-0230 (2 vol., 1995).
11. www.shewanella.org
12. J. M. Tiedje, *Nature Biotechnol.* **20**, 1093 (2002).
13. R. D. Smith, personal communication.

Supporting Online Material

www.sciencemag.org/cgi/content/full/300/5617/290/DC1
SOM Text