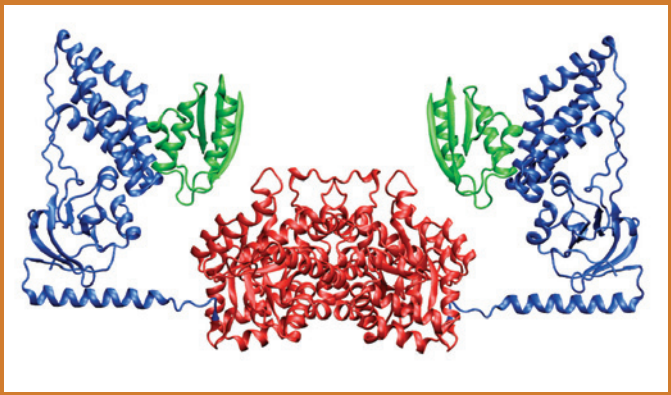
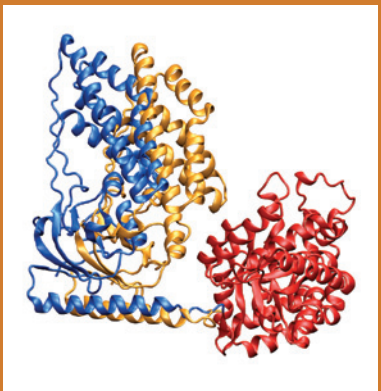




Research highlighted in this chapter describes a novel method for determining the solution structures of large complexes of proteins in an environment that closely mimics that within a cell. An important part of biomedical research is to understand the relationship between a protein's structure and biological function. Scientists from NIDDK's Intramural Research Program used a combination of biophysical techniques (specifically nuclear magnetic resonance spectroscopy and solution X-ray and neutron scattering) and novel computational approaches to determine how different bacterial proteins fit together to form a large molecular complex that regulates the transport of sugar molecules in and out of the bacterial cell. The image on the left illustrates how the proteins—called Enzyme I (blue/red) and HPr (green)—interact to form a complex, and the image on the right reveals changes in the structure of Enzyme I that occur on forming the complex that may be important for the function of this bacterial enzyme. This advance in methodology provides a new approach for researchers to study the structures of large proteins and protein complexes. Understanding the relationship between a protein's structure and function is an important part of biomedical research, as proteins carry out many of the biological functions underlying normal health and disease.

Graphics provided by Dr. G. Marius Clore, NIDDK.

Cross-Cutting Science

Advances in medicine are largely dependent on the accumulation of new knowledge about biologic processes, often at the smallest levels of an organism—its genes, the proteins they encode, the inner workings of cells, and the ways cells communicate with each other. Major strides in fighting disease can be traced to laboratory studies whose immediate relevance to health could not have been fully known or appreciated at the time they were conducted. Opportunities to make exciting new discoveries are arising ever more rapidly with the development of new technologies, new approaches, and even new scientific disciplines as teams of talented, creative researchers join together to pursue increasingly complex challenges. Described in this chapter are several recent studies, each of which spans multiple areas within the NIDDK research mission. The insights gained through this research can be expected to aid progress in many scientific endeavors, for today's research advances may lead to tomorrow's cures.

IMPLEMENTING THE AMERICAN RECOVERY AND REINVESTMENT ACT OF 2009 AT NIDDK

On February 17, 2009, President Barack Obama signed into law the American Recovery and Reinvestment Act of 2009 (Recovery Act; Public Law 111-5). With potential for biomedical research to stimulate the economy, NIH received a generous infusion of funding from the Recovery Act, and this funding was distributed throughout NIH.

Poised to capitalize on this unique investment opportunity, NIDDK developed a funding plan to meet the stimulus goals of the Recovery Act and to advance scientific progress. The funding supported a range of biomedical research efforts across the Institute's research mission. As a priority, the Institute remains committed to funding outstanding science as judged by the NIH peer review process. The Recovery Act provided an opportunity for NIDDK to support many highly meritorious, peer-reviewed research project grants that met program priorities and could be realistically accomplished within a 2-year funding period. In addition, NIDDK was able to provide short-term funding for many peer-reviewed applications to help investigators continue ongoing studies, maintain laboratory personnel, and get exciting projects off the ground.

The NIDDK used Recovery Act funds to accelerate and expand research in two "Signature" programs. The "Novel Cell Therapies in Regenerative Medicine for Diabetes" Signature Program is investigating novel human islet cell replacement therapies for patients with type 1 diabetes. Replacement of insulin-producing pancreatic islet beta cells holds great promise for the treatment and cure of type 1 diabetes. The "Genome-Wide Association Studies and Replication in Diseases of Interest to NIDDK" Signature Program is identifying genetic variations associated with diseases within the NIDDK mission. High priority areas for this program included extending studies to minority populations and to diseases that have not been addressed in previous genome-wide studies. Once new genetic associations are identified, researchers can use the information to develop better strategies to detect, treat, and prevent diseases.

In addition to supporting Signature Programs, NIDDK used Recovery Act funds to supplement ongoing projects that span all aspects of its research mission. Administrative Supplements to existing grants provided resources for additional lab personnel and research equipment. Through the Recovery Act, NIDDK also was able to provide additional funding to advance many of its ongoing major research initiatives, such as the Beta Cell Biology Consortium, the Diabetes Prevention Program Outcomes Study, the Halt Progression of

Polycystic Kidney Disease (Halt-PKD) clinical trial, and the Action for Health in Diabetes (Look AHEAD) clinical trial. In addition to supporting existing programs, NIDDK was able to use the Recovery Act to begin a new program—the NIDDK Consortium Interconnectivity Network, or "dkCOIN"—which serves to connect investigators, resources, and data from various NIDDK research networks.

The Recovery Act also provided funds for investment in new high-impact initiatives that will affect the course of future research. The NIDDK participated in the NIH Challenge Grants in Health and Science research initiative. This grant program was created to support research addressing specific challenges in biomedical and behavioral research that would benefit from significant 2-year jumpstart funds. The NIDDK used the Recovery Act to support high-impact, peer-reviewed Challenge Grants in the areas of enabling technologies, stem cells, clinical research, translational research, and comparative effectiveness research.

In addition to stimulating the U.S. economy, NIDDK expects the Recovery Act to stimulate a number of scientific areas that advance its research mission. The return on this investment is already evident, with a number of scientific advances funded, in part, by the Recovery Act highlighted in this research compendium. It is likely that these and other exciting new findings that result from the Recovery Act investment will answer many pressing research questions and open entirely new areas of biomedical research.

RECENT GENETIC STUDIES: PAVING THE WAY TO IMPROVING PEOPLE'S HEALTH

Genetics, the study of genes—the basic unit of inheritance—is one of the tools in a researcher's arsenal to understand human health and disease. Understanding genetic contributions to the development of disease can lead to clues to biological pathways involved that are driven, at least in part, by variations in genes and their regulatory regions, many of which remain to be identified. Some gene variants directly cause disease, while others confer susceptibility to disease in combination with other genes or environmental factors. Not only will genetics research open new avenues for developing therapies and invigorate the drug development pipeline, it also holds promise to advance development of personalized therapy.

In a breakthrough that revolutionized the field of human genetics, scientists realized that variations in DNA sequences, such as those called "single nucleotide polymorphisms," or SNPs, can be used as genetic markers for mapping disease-associated genes. The Human Genome Project provided a comprehensive map of genetic sequences and the Haplotype Map provided a catalog of genetic variation. Additionally, scientists devised tools and methods to more rapidly and routinely analyze DNA sequences and variations. These advances led to genome-wide association (GWA) studies, in which researchers have compared variants across the entire genome from individuals with and without a particular disease so as to pinpoint genetic regions, called loci, that track with the disease and may thus harbor a disease gene.

These new research tools have enabled an explosion of advances in understanding the genetic basis of complex human diseases and have identified hundreds of disease-associated gene loci, uncovering new knowledge of disease mechanisms. For example, unexpected pathways have been implicated in disease, and previously unknown associations between diseases have been illuminated. Findings from GWA investigations and other advanced genetic studies also indicate that common diseases result from the influence of multiple risk genes with varying contributions and interactions with environmental factors. A more complete understanding of genetic variation contributing to disease will be available as comprehensive sequencing methods are undertaken, which will uncover rarer variants.

The NIDDK has supported many advances in genetics research related to diseases and disorders within its mission. For example, as recently as 2005 only two genes were known to affect type 2 diabetes. Today, there are over 40 new gene loci associated with risk of type 2 diabetes, many of which were found by NIDDK-supported studies. Similarly, in 2003 only three genes were known to affect type 1 diabetes. Today, the NIDDK-led Type 1 Diabetes Genetics Consortium (T1DGC) and its collaborators have identified over 40 gene loci associated with risk of type 1 diabetes, bringing the total number of known

regions to nearly 50. NIDDK-supported consortia like the Family Investigation of Nephropathy and Diabetes, the Genetics of Kidneys in Diabetes Study, and the Epidemiology of Diabetes Interventions and Complications study have collected a wealth of genetic data, which has been deposited in the Database for Genotype and Phenotype (dbGAP), and researchers continue to mine this information to elucidate the genetic contributors to diabetes complications. In addition, novel genetic loci associated with healthy blood glucose levels are being identified, providing new insight into the biology of metabolism. See the Diabetes, Endocrinology, and Metabolism (DEM) chapter for more information on this specific advance.

In another example, NIDDK-supported researchers have discovered genetic regions associated with inflammatory bowel diseases (IBD). As reported in the Digestive Diseases and Nutrition (DDN) chapter of this compendium, the NIDDK Inflammatory Bowel Disease Genetics Consortium recently identified several susceptibility loci for ulcerative colitis, a form of IBD that causes inflammation in the tissues lining the colon and rectum. In this advance, approximately 30 gene loci were implicated in ulcerative colitis. The same chapter also describes the identification of five new genetic variations that predispose children to developing IBD. As other genetic risk factors for IBD typically have been identified in adults, this advance extends these studies to pediatric populations.

Researchers have also made strides in determining the genetic factors that influence immune-mediated digestive diseases. As described in the DDN chapter of this report, NIDDK-supported scientists have now identified disease-associated genetic variants within a cluster of genes that play an important role in the immune system. These variants affect multiple chronic inflammatory diseases with autoimmune features, including Crohn's disease and ulcerative colitis, the two major forms of IBD. New genetic risk factors have also been uncovered in celiac disease, an autoimmune disease in which an aberrant immune response to the gluten protein found in many dietary grains results in chronic inflammation and tissue damage in the intestine. This research is also described in the DDN chapter.

Kidney disease research has also gained new insights from advanced genetics studies. In one study described in the Kidney, Urology, and Hematology (KUH) chapter of this report, NIDDK-supported researchers identified 13 new genetic loci affecting renal function and chronic kidney disease, and seven loci suspected to affect production and secretion of a protein found in urine. Another group of NIDDK-supported researchers reported that variants around the *MYH9* gene are linked to susceptibility to various forms of kidney disease among African Americans. Further research, also described in the KUH chapter, revealed that much of the kidney disease risk is actually due to variants in an adjacent gene, *APOL1*. These variants likely protect against a trypanosomal infection that causes African sleeping sickness, a degenerative and potentially fatal disease affecting tens of thousands of people in sub-Saharan Africa. This finding may lead to the development of better treatments for both chronic kidney disease and African sleeping sickness.

Challenges remain, however, in translating these exciting genetics findings to the development of new therapeutics and improvements in health. First, many disease-associated variants have been identified in populations of European origin. To ensure that all Americans benefit from the fruits of these studies, it will be necessary to conduct similar studies in ethnically and racially diverse populations. The NIDDK is promoting such research; for example, a new initiative, the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multiEthnic Samples (T2D-GENES) Consortium, is investigating and comparing the genetic causes of diabetes in multiple ethnic groups. Second, while scientists have identified an impressive number of loci associated with various diseases, for a number of these loci the causal gene—the gene and precise mutation that influence disease risk—is not yet known. This Consortium will be sequencing within the loci, in multiple populations, to see if different mutations contribute to diabetes in different populations. Third, once the exact gene associated with disease risk has been identified, research is required to determine how and to what extent the gene affects disease risk or disease progression.

Another critical challenge is to bridge the gap from understanding disease development and progression to efficacious prevention and treatment strategies. Prediction of disease risk—based on an individual's genome—will be useful if proven interventions for

specific patient populations exist. Genetic prediction of disease risk, however, may also help advance the development of interventions. For example, scientists can identify, in part from genetics, individuals at heightened risk for type 1 diabetes; this knowledge gives people the opportunity to volunteer for clinical trials to test potential prevention strategies. As part of the NIH Genes, Environment, and Health Initiative, NIDDK leads two initiatives to bring genetic knowledge to clinical utility. One initiative focuses on the translation of significant genetic findings into clinical or public health use. The other initiative supports research to measure the responses of patients and providers to information about genetic determinants of common diseases and to determine how to effectively educate the public to use the information appropriately for clinical care and disease prevention.

Genetic studies are an important component of the research to improve understanding of disease and to develop better predictions of risk, new prevention strategies, and novel treatment approaches. In addition, studies in different ethnic populations may uncover a genetic contribution to health disparities. With continued research, NIDDK can capitalize on exciting genetic findings to improve the health of Americans at risk for and burdened by these diseases and disorders.

THE MOLECULAR LIBRARIES PROGRAM: DRUG DISCOVERY AT THE INTERSECTION OF BIOLOGY, CHEMISTRY, AND TECHNOLOGY

Recent biomedical research, in particular sequencing of the human genome, has produced enormous potential for the identification of novel proteins and biological pathways that could serve as molecular targets for new drugs. To understand the normal functions of these new targets and determine their roles in disease, scientists require molecular tools for research. "Small molecule" probes—a class of chemical compounds that can interact with and help define the functions of biological molecules—are widely used as experimental research tools and for therapeutic drug development. Great effort is needed to identify a suitable small molecule for a given target. Specially-developed experimental tests (known as assays) determine a small molecule's ability to modulate the activity of the target protein or pathway.

To identify just one or a few promising small molecule probes, hundreds of thousands of compounds may be tested using these assays, and thus high-throughput screening instrumentation is often required. The chemical structures of candidate small molecules are then fine-tuned through "medicinal chemistry" to make their activity as potent and specific as possible.

For many years, the pharmaceutical and biotechnology industries have utilized high-throughput screening technologies, but similar resources had been lacking for researchers in the public sector, who are addressing different biomedical questions. In addition, while some academic biomedical scientists include small molecule discovery in their research, most have never engaged in this challenging process and thus lack the requisite specialized knowledge and experience. To bridge the gap between the biologists in academia and scientists with expertise in small molecule development and high-throughput screening, the Molecular Libraries Program (MLP) was launched as an initiative of the NIH Roadmap.

The MLP is a trans-NIH effort that established nine small molecule production and screening facilities located in California, New Mexico, Kansas, Tennessee, Alabama, Maryland, Florida, and Massachusetts. These centers, collectively called the Molecular Libraries Screening Centers Network, work as a consortium to provide a "library" of chemical compounds, drug discovery expertise, and automated high-throughput capabilities to identify previously unavailable probes that researchers can use to study their target proteins of interest.

Identifying a functional small molecule is no easy task; to be useful for research, numerous strict criteria must be met. For example, the compound must effectively modulate the activity of the desired protein or pathway but should also be specific, *i.e.*, it should interact only with its intended target to avoid unwanted side effects. The ability to screen vast numbers of compounds improves the likelihood of successfully identifying a small molecule capable of fulfilling the necessary rigorous requirements.

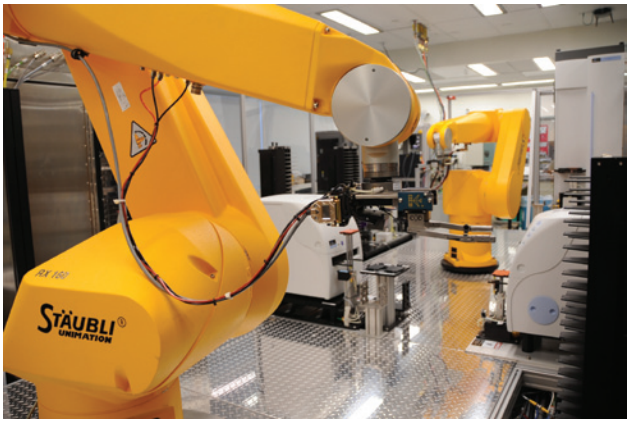
Once discovered, new small molecule probes are added to a growing collection of chemical compounds stored in the Molecular Libraries Small Molecule Repository. Related information (about the chemicals

and assays) is catalogued in a public internet database called PubChem (*http://pubchem.ncbi.nlm.nih.gov/*), which is widely used by biomedical scientists, and to date includes information for more than five million chemical compounds. With vast information from previous research, the PubChem directory provides a valuable starting point for new efforts to identify small molecules that may have useful effects on a broad range of biological processes or diseases.

The NIH Chemical Genomics Center (NCGC), located within the National Human Genome Research Institute, is one of the nine Centers and illustrates the Centers' capabilities. In this comprehensive facility, 75 scientists use state-of-the-art automation to address most aspects of small molecule discovery, including assay design, high-throughput screening, and medicinal chemistry. One major interest of the NCGC is in rare or neglected diseases, which are important areas of academic research and are less likely to be addressed by the private sector. Many projects have been completed by this facility. As just one example, the NCGC developed chemical probes to better understand a rare genetic disorder known as Gaucher's disease. People with this disease have mutations in a protein called glucocerebrosidase. These mutations change the molecule's three-dimensional shape and alter its function, leading to symptoms that include spleen and liver enlargement, neurological disorders, and osteoporosis. The NCGC developed chemicals that interact with this protein and correct its structure. These probes are now being used in the laboratory to better understand the protein's function and are promising candidates for future drug development for possible use in humans.

Chemical probes offer great potential to accelerate the progress of biomedical research. Independently of the MLP, an increasing number of academic scientists are conducting small molecule discovery research in their own laboratories. The MLP complements those efforts by greatly expanding the reach of chemical libraries and high-throughput screening technology to many other biomedical researchers. The chemical probes generated by scientists in the MLP, partnered with academic researchers, have served as useful tools to address challenging research problems, and could potentially lead to new therapies for people suffering from disease.

For more information, please visit: *http://mli.nih.gov/mli*



Automation is a key component of high-throughput screening. Using specially developed assays, robots and machines rapidly screen small molecule libraries, which can contain hundreds of thousands of chemical compounds. A series of steps comprises each assay utilizing different machines—ultimately leading to the identification of small molecules that have a desired function. The particular steps vary depending on the experimental procedure. In this picture, two robots (yellow machines in the foreground and background) operate at different stages of the assay process to transfer individual plates containing large numbers of independent miniature-scale tests stepwise from one automated station to another, until the assay is complete and the results can be analyzed. The machines at these stations can work around the clock to perform these repetitive experimental steps, reducing the time it takes to test the vast collection of small molecule compounds.

STEM CELLS

Induced Pluripotent Stem Cells Find It Difficult To Forget: Researchers have recently discovered why induced pluripotent stem (iPS) cells don't always function as well as embryonic stem (ES) cells. Human ES cells are "pluripotent"—that is, they have the ability to form virtually any cell type in the body and may thus possess the ability to repair human tissues and organs. They also can be propagated indefinitely in the laboratory. However, the use of ES cells is controversial because their isolation entails the destruction of early-stage human embryos.

In recent years, scientists have developed two different ways of reprogramming adult cells, such as those derived from blood or skin, to revert back to an ES cell-like state, with the potential to give rise not only to new cells of their original type but also to more

stem cells and a multitude of different types of cells. One technique involves adding just four genes that convert adult cells into iPS cells. The other technique entails transferring the nucleus from an adult cell into a fertilized egg from which the original nucleus has been removed. The egg then reprograms the adult nucleus, enabling the isolation of pluripotent stem cells with the genetic makeup of the adult cell. Both types of cells could be used to model diseases and potentially to create cells to treat specific diseases, but the use of human pluripotent stem cells generated by nuclear transfer has ethical issues similar to those of traditional ES cells.

The ability of iPS cells to form cells of lineages other than the type from which they were originally derived has been limited, and scientists sought to determine why this was the case. In this study, conducted with mouse cells, the research team examined the role played by DNA methylation. DNA methylation is a type of modification of the genetic material where small chemicals, called methyl groups, are attached to various parts of DNA. Although this process does not alter the sequence of the genetic code, DNA methylation can affect the cell's ability to activate or deactivate genes. Different types of cells have different characteristic methylation patterns. The scientists found that mouse iPS cells retain residual DNA methylation reflecting their tissue of origin. The methylation patterns serve as a kind of "memory," affecting gene expression and restricting the number and kind of ultimate fates of the cells. Moreover, the researchers discovered that compared to iPS cells, nuclear transfer-derived pluripotent cells were more similar to ES cells in their methylation patterns and their ability to form a wider variety of different cell types. The researchers did find that the methylation memories of iPS cells could be reset, at least in part, either by growing the cells longer in the laboratory or by treating them with drugs that affect DNA methylation.

Cautious optimism continues to describe the eventual use of iPS cells for experimental models of disease, as targets in drug screening studies, and as a source for regenerating tissue.

Kim K, Doi A, Wen B, et al. Epigenetic memory in induced pluripotent stem cells. Nature 467: 285-290, 2010.

Reprinted, in slightly modified form, from NIH Research Matters*; original article by Harrison Wein, Ph.D., published on August 2, 2010.*

INSIGHTS INTO PROTEIN STRUCTURE

Novel Approach for Determining the Structure of Large Protein Complexes: Scientists from NIDDK's Intramural Research Program have developed a novel approach for determining the structure of large proteins and protein complexes in an environment that closely mimics that of a cell. Proteins are the molecules that carry out many of the biological functions underlying normal health and disease. In many instances, disease-causing mutations alter a protein's function by disrupting its three-dimensional structure. Thus, an important part of biomedical research is to understand the relationships between the structure and biological function of proteins.

In this study, researchers used a novel combination of two biophysical techniques—X-ray scattering and nuclear magnetic resonance (NMR) spectroscopy—that provide complementary information on the size and shape of the protein complex and the relative orientations of the different components of the complex. By integrating the two different types of data with structural knowledge of the individual components, the researchers were able to determine how two bacterial proteins—Enzyme I and HPr—fit together to form a large molecular complex that regulates the transport of sugar molecules in and out of the cell. In addition, by comparing the structure of the complex with that of Enzyme I alone, the scientists found that the enzyme changes shape when it joins with HPr; this change in shape may be important for the enzyme to function properly. Since X-ray scattering or NMR alone would have been insufficient for characterizing the structure of this enzyme system, the combined use of the techniques provides an important new strategy for determining the size, shape, and orientation of large proteins and protein-protein complexes. The application of this new approach for studying the structural properties of large proteins and protein complexes will provide important information on their biological functions in health and disease.

Schwieters CD, Suh J-Y, Grishaev A, Ghirlando R, Takayama Y, and Clore GM. Solution structure of the 128 kDa enzyme I dimer from Escherichia coli *and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. J Am Chem Soc 132: 13026-13045, 2010.*

Timing How Fast Proteins Fold, One Molecule at a Time: Scientists in NIDDK's Intramural Research Program have applied recently developed statistical methods to determine how fast individual protein molecules "fold" into their correct biological shapes. When a protein is made in a cell, it folds into a uniquely defined three-dimensional structure. This "native" structure, as it is called, is intimately linked to the protein's function. In cases where a protein's native structure is disrupted, such as by genetic mutations that alter its ability to fold properly, the resulting impairment of the protein's normal function can cause disease. Given the importance of protein structure and function in normal health and disease, scientists are interested in understanding the physical process that leads to the formation of a protein's correct native structure.

NIDDK intramural scientists have previously pushed the limits of a technique called single-molecule fluorescence resonance energy transfer (FRET) to observe the transitions between unfolded and native states during the folding process of individual protein molecules. To do this, they labeled proteins with dye molecules that emit light of different colors depending on whether the protein is folded or unfolded. These researchers have now analyzed the FRET data using new statistical methods, developed by another team of scientists in the Intramural Research Program, to determine how much time the protein molecules spend in the unfolded state and the native state. This type of data analysis also provides information on the percentage of molecules in the folded and unfolded states under the experimental conditions, as well as information on the overall shape of the molecules in each state. In addition, the scientists applied a new computational procedure to identify the transitions between the unfolded and native states. As all of the important structural changes associated with going from the unfolded state to the native state occur during these transitions, the application of this new methodology may allow researchers to uncover the fundamental mechanisms that guide the folding process and identify the misfolding steps that are often associated with diseases.

Chung HS, Gopich IV, McHale K, Cellmer T, Louis JM, and Eaton WA. Extracting rate coefficients from single-molecule photon trajectories and FRET efficiency histograms for a fast-folding protein. J Phys Chem A doi: 10.1012/jp1009669, 2010.

THE HUMAN GENOME ARCHITECTURE

Mapping the Three-Dimensional Architecture of the Human Genome: Using a chemical linkage strategy combined with a large-scale DNA sequencing effort, scientists have revealed how DNA is organized within cells. Genomic DNA exists as a set of separate segments that are packaged along with other components into units called chromosomes. Within a cell, the chromosomes are found in a specific compartment, the nucleus. Generally, chromosomes are in a dense, dynamic clump. Scientists have been trying to understand how the chromosomes are organized in the nucleus. Previous studies indicate that chromosomes occupy distinct "territories" within the nucleus. Within territories, chromosomes are flexible such that the DNA portion of one part of a chromosome can interact with the DNA portion of another chromosome, and DNA at one end of a chromosome can interact with DNA at the other end. A technique developed earlier has allowed scientists to determine all the inter- and intra-chromosomal interactions between chosen areas of the genome.

Building on this technique, investigators developed a new method, which they call Hi-C, to enable mapping of inter- and intra-chromosomal interactions throughout the entire genome. Data from this technique provided insight into how the genome is organized in the nucleus. The data confirmed the concept of "chromosome territories" and that small, gene-rich chromosomes tend to be near each other physically. Furthermore, the scientists discovered that the chromosomes could be categorized into two compartments that were physically separated. One compartment contained active genes while the other compartment contained inactive genes.

The scientists propose that their data are consistent with a fractal model—one that has the same pattern no matter how close one zooms in. In this model, the

chromosomes are densely folded in the nucleus and highly compact. In addition, there are no knots in this model. Previously suggested models for chromosome organization contained knots that did not offer the flexibility that might be necessary for genes to turn on and off. These findings and this new technique will allow scientists to look even more closely at genome-wide chromosome interactions and learn more about how chromosomes are organized, including whether genome shape is altered across cell types. Evidence suggests that genome structure affects the turning on and off of genes. Therefore, understanding genome structure will provide insight into how the shape of the genome affects health and disease.

Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326: 289-293, 2009.