

## STORIES OF DISCOVERY

### **How Sweet It Is!**

The sense of taste plays a critical role in the ingestion of caloric nutrients and in the rejection of toxic, typically bitter-tasting, food compounds. Abnormalities in taste perception can contribute to several diseases or unhealthy conditions, including poor nutrition and diet, obesity, diabetes, hypertension (i.e., excess sodium intake), and dental caries. Consequently, an understanding of the cellular events that mediate the early stages of taste perception at the level of the taste receptor cell will provide important insight into the regulation of balanced food intake and proper nutrition.

A variety of distinct signalling mechanisms are activated by the basic taste qualities of salty, sour (acid taste), sweet, and bitter. Salty- and sour-tasting compounds activate gated ion channels that are located at the apical pore of taste receptor cells. Bitter and sweet taste are mediated by G protein-coupled receptors that are similarly located. Considerable attention has been given to the identification of bitter taste receptor genes, bitter taste receptors, and the signal transduction cascade underlying bitter taste. In contrast, the genes and G protein-coupled receptors involved in sweet taste have been more difficult to identify.

Recently, four NIH-funded laboratories independently identified a putative sweet taste receptor gene, T1R3, at or very near the mouse *Sac* locus. Differences in sweetener intake among inbred strains of mice are partially determined by allelic variation of one or more genes in the saccharin preference (*Sac*) locus. As predicted, the encoded receptor, T1R3, differs in amino acid sequence in "sweet preferring" versus "sweet indifferent" mouse strains and may represent one type of sweet receptor. Since sweet tastants vary considerably in terms of structure, including sugars, amino acids, proteins, and artificial compounds, it is reasonable to suspect that there are multiple types of sweet receptors. T1R3 appears to be one member of the T1R family that is involved in sweet taste. Both human and mouse T1R3 are members of a family of G protein-coupled receptors that contain large extracellular domains. Candidate vomeronasal pheromone receptors, metabotropic glutamate receptors and two taste receptors (T1R1 and T1R2) of unknown ligand specificity are also members of this family. T1R1, T1R2 and T1R3 share about 30 percent amino acid sequence identity and are selectively expressed in subsets of taste receptor cells. However, the T1R1 family is distinct from the T2R family, which is known to respond to bitter tastants, and it is more likely that T1R1 and T1R2 are responsive to sweet tastants.

These results indicate that bitter and sweet taste receptor genes and the corresponding receptors are expressed in different parts of the genome. The mechanisms underlying bitter- and sweet-tasting quality coding are being studied in genetically altered mice where key components of the signal transduction cascade (such as candidate receptors and G-protein subunits) are modified or deleted in an attempt to determine the molecular basis for normal taste function.

## **The Declining Disability of Older Americans**

As recently as 20 years ago, scientists predicted that advances in medicine and technology would increase the number of people surviving to older ages, but would also cause an exponential increase in health care services and costs. In the past few years, however, it has become clear that increased disability is not an inevitable consequence of aging, and that people are living longer and healthier lives, with the potential for working longer and postponing medical care needs. These trends are occurring in both the U.S. and in Europe.

The ability to make reasonably accurate projections with regard to both life expectancy and disability has obvious implications for assessing the solvency of such large national programs as Social Security and the Medicare Trust Fund. In the mid-1970s, the Social Security Administration made life expectancy projections based on the assumption that mortality would be static, and consequently life expectancy would not improve over time. It soon became clear that such an approach was not accurate. In 1982, the National Commission on Social Security Reform was established to determine how to modify Social Security funding so it would remain solvent in the face of expected life expectancy increases at ages 65 and over. The commission mandated an increase in the normal retirement age from 65 to 67 years in increments from 2000 to 2022. This was done without clear evidence that life expectancy increases at ages 65 and over were associated with significant improvements in health at later ages.

In 1993, the National Academy of Sciences found evidence of chronic disability declines from the 1982 to 1989 National Long Term Care Study (NLTCS). Confirmation of this trend required a longer time series. In 1994, new NLTCS data confirmed that a decline in chronic disability was occurring. The chronic disability rates declined 1.1 percent per year from 1982 to 1989, and 1.5 percent per year from 1989 to 1994 – suggesting an accelerating trend. The recent completion of the 1999 NLTCS confirmed that the decline in disability had accelerated (2.6 percent per annum 1994 to 1999) and that there were 2.1 million fewer chronically disabled persons in 1999 than there would have been if the 1982 rates had not changed.

These findings are of great potential significance in terms of identifying and addressing causes of disability, as well as informing national health care policy. It is therefore important that confirmation of the trend toward decreased disability has subsequently come from independent studies and investigators. For example, an analysis of data from the Survey of Income and Program Participation (SIPP) in 1998 showed a reduction in disability rates from 1984 to 1993 in every age group of elderly Americans 50 years and over. The SIPP per year declines, at 0.9 to 2.3 percent, were even larger than in the NLTCS, and were relatively more rapid above age 80. Analysis of the 1991 to 1996 Medicare Current Beneficiary Survey (MCBS) also showed disability declines more rapid than in the 1982 to 1994 NLTCS. Declines of 0.9 percent per annum were found in the 1983-1994 National Health Information Survey (NHIS). A more recent examination of the NHIS data through 1996 shows that improvements did not accelerate through the entire period. However, the latter study also evaluated the consistency of results across five national surveys (NHIS, NLTCS, SIPP, MCBS and 1984 and 1994 Supplements on Aging (SOA) to the NHIS) and found estimates of decline in disability prevalence to be fairly

consistent. All of these corroborating findings, as well as analyses of data from other developed countries, have helped establish that the disability decline may reflect real improvements in underlying physiological health, as well as better therapies or coping strategies.

The evidence suggests that larger declines are found for disabilities due to difficulties with routine care activities or Instrumental Activities of Daily Living (IADLs – e.g., everyday household chores, doing necessary business, shopping, getting around), whereas more severe personal care disability (e.g., difficulty with bathing, dressing, and eating) shows little or no improvement. Recognizing that strong links exist between functioning and cognition, researchers are beginning to document an improvement in cognitive functioning among older Americans in recent decades that parallels declines reported for physical disability and functional limitations. Analyses of data from the Health and Retirement Study show that only 4 percent of older Americans living in the community were identified as severely cognitively impaired in 1998, down from 6 percent in 1993. Also consistent are findings from the NLTCs that show large declines in severe cognitive impairment from 5.7 percent in 1982 to 3.8 percent in 1994.

The long-term implications of disability decline depend in large part on whether the trend continues and at what pace. On the basis of current disability rates, demographers estimate far fewer working age people in the future per disabled person, but if disability rates continue to decline by 1.5 percent per year, there would actually be more working age people in the future per disabled older person. It has been projected that between now and the year 2030, expected increases in educational attainment among the elderly will continue to contribute to improvements in functioning. Related trends that are likely to stimulate continued disability decline include improvements in health related behaviors, the availability and effectiveness of assistive devices, and the treatment of conditions that lead to disability. At this time, there is no consensus about the likely pace of future disability decline.

Research has begun to focus on plausible models to explain the recent declines in disability in order to identify specific interventions, behavioral changes, and survival attributes that can accelerate the trend toward decreased cumulative disability, postponed onset of disability, regaining of function, and improved quality of life. Since the molecular and genetic bases of many chronic diseases are only beginning to be identified, systematic reductions in chronic degenerative diseases have not yet occurred.

A broad research effort is also under way on the long-term economic consequences of the disability decline and the adoption of new medical technologies. Since people are physically capable of working longer, more may defer retirement and continue working until older ages. This has implications for Social Security, the Medicare Trust Fund, and for the overall productive capacity of the economy. It has been argued that if a 1.5 percent annual decline in chronic disability were maintained for the next 40 years, the Medicare Trust Fund would remain solvent. Some also contend that biotechnological innovation has and will continue to stimulate health improvements that may ultimately lower health care costs. Declining rates of chronic disability may also moderate the burden of caregiving, including the informal care provided

within families, the care provided through home health services, and the care provided in long-term care institutions.

## The Human Genome, Chapter Two

News stories appearing over the last year and a half would have everyone believe that scientists have – once and for all – cracked the human genetic code. Indeed, two teams of researchers have already published a draft sequence of our 3-billion-piece jumble of DNA "letters," chemical units called nucleotides. Yet how our bodies interpret, or "express," our cells' genetic code plays a prominent role in establishing our everyday health. Thoroughly deciphering the code's protein-making instructions – something our bodies do with ease all the time – remains a complicated puzzle. Scientists already know that certain spelling differences in DNA cause disease. But other inheritable factors, aside from DNA, can influence how likely a person is to develop a particular disease. A compelling tale of discovery recounts researchers' quest to tease apart these so-called "epigenetic" factors that, along with diet and other environmental influences, profoundly affect our health.

The correct packaging of DNA is essential to the proper functioning of the cells that make up our bodies. Cells contain and protect their precious cargo, genes, in protein-rich complexes called chromatin, which consists of long, stringy DNA spooled around an orderly, ball-like core of proteins called histones. In a sense, chromatin acts as a gatekeeper for our genes, regulating access to DNA by cellular equipment that decodes the genetic instructions. Among other things, this arrangement permits embryos to develop the right way, and it directs precursor cells to form organs and tissues. Conversely, if access to the genes in chromatin is not stringently controlled, cancer and a variety of other diseases can be the terrible consequence.

**Beads on a String.** The chromatin story begins over a hundred years ago. In the late 1800s, researchers first discovered the molecules now known as histones, and there was widespread belief among scientists that these proteins – not DNA, as determined over 50 years later – were the source of heredity. In 1973, cell biologists first obtained electron microscopic pictures of chromatin fibers as "particles on a string." Soon thereafter, a seminal article was published that proposed a model of chromatin structure as repeating units of approximately 200 nucleotide pairs of DNA and 8 histone molecules – the string and beadlike particles, respectively. Virtually every college biology student now knows this description of chromatin as "beads on a string." Beginning with these early studies, NIH has funded a quarter-century of groundbreaking research on chromatin – what it is, how it works, and more recently, how it is tied to cancer and other rarer diseases.

In the mid-1960s, even though researchers did not know precisely what histones did inside cells, scientists suspected that natural chemical tags on histones could control genes by turning them on or off. Throughout the 1970s, more scientists began to appreciate and gradually accept the connection between DNA's physical environment (chromatin) and the activity of genes. Genes are turned on ("transcribed"), for the most part, by proteins called transcriptional activators that must touch DNA to exert their effects. The basic structure of chromatin, in which DNA is spooled and compacted, would appear to be a major obstacle to the transcription process by blocking access of activator proteins to genes. Indeed, researchers have reported many examples of how chromatin can prevent genes from being read. In the late 1980s and early 1990s, basic

researchers discovered that some transcriptional activator proteins (and often combinations of them) can bind to DNA in chromatin, displacing histones. This observation proved to be a major step in researchers' understanding of how genes can be read through a veil of protective chromatin.

**Another Code.** Groundbreaking molecular biology studies have begun to reveal that a key step in how cells interpret their genetic code involves actually finding genes tucked away inside chromatin. Part of a cell's gene-decoding machinery is drawn to the histone proteins in chromatin. In recent years, scientists have defined several cellular systems that carefully balance how histones are "marked" with a variety of natural chemical tags – called acetyl, phosphate, and methyl groups – in a specifically timed order. Putting on these tags and taking them off – something some researchers have called the "histone code" – turns out to be a critical aspect of targeting a cell's gene-reading activities.

The past 10 years in particular have witnessed an explosion of major discoveries that are paving the way toward a better understanding of how genes are controlled and how certain diseases result when gene access is either too lenient or too stringent. In many cancer cells, for instance, inappropriate control of certain growth genes can fuel unchecked cell division. Scientists are finding a prevalence of telltale marks on chromatin in certain cancer cells, leaving growth genes bare and prone to near constant activation. When histone-marking enzymes are revved up in cancer cells, these molecules become important potential targets for developing future cancer drugs.

In recent years, researchers studying model systems such as yeast and worms have also made links between chromatin and normal biological processes like aging. One key discovery revealed that a gene-silencing protein called Sir2 removes the chromatin-marking chemical tags called acetyl groups from genes. Scientists found that Sir2 is intimately dependent on a molecule central to the body's process of metabolizing food into energy. The finding has far-reaching implications, providing a tantalizing potential explanation for the recent observation by other researchers that low-calorie diets in model organisms such as yeast and worms can extend lifespan.

## **Environmental Agent Gives Clues to Arthritis, Depression, and Cancer**

Scientists are using fundamental discoveries in the environmental toxicology of dioxins to understand a number of important human diseases such as arthritis, mental illnesses, and cancer.

The unique nature of this relationship lies in the observation that the mechanism by which these pollutants signal in a cell is very similar to mechanisms used in a number of important physiological processes. These physiological processes are central for normal development and the maintenance of health throughout the life of the organism.

At the center of these diverse areas of biology is a large group of regulatory proteins that control how cells communicate with each other, the PAS proteins. With the completion of the human genome, the PAS superfamily is known to be comprised of twenty-one members, each with a unique role in a particular aspect of human biology. The PAS domain is a region of approximately 250 amino acids that appears to have its origins in light and oxygen sensors found in a number of primitive organisms such as bacteria.

One of the key steps in understanding how PAS proteins work came from studies of the environmental pollutant 2,3,7,8-tetrachlordibenzo-p-dioxin, also known as TCDD or simply dioxin. This highly toxic environmental contaminant is of considerable interest due to its potency as a carcinogen (causes cancer) and teratogen (causes birth defects). A series of elegant genetic and biochemical studies in the late 1970s and early 1980s demonstrated that essentially all of dioxin's toxicity could be attributed to its interaction with a soluble protein known as the Ah receptor. This was interesting for two reasons. First, it was not clear why such a receptor should exist. That is, why do humans and most organisms have a receptor that recognizes environmental pollutants? Second, the dioxin-Ah receptor system became a paradigm for the prediction of health risk based upon a receptor-mediated model. Such models would be put to great use with the later recognition that a large number of environmental chemicals could act through other receptors.

A first look at the Ah-receptor's sequence revealed that this protein was unlike any receptor that had ever been cloned before. The most informative feature of the Ah receptor was the PAS domain. At that time, the PAS domain had only been observed in one other protein, a protein known as the Ah receptor nuclear translocator or ARNT. This similarity was not a coincidence and was related to the fact that the Ah receptor and ARNT were binding partners. When the two proteins bind together, they form a dimer (2-protein complex) that is the actual chemical structure capable of eliciting biological responses. A series of papers would reveal that it was this pair of proteins that was required for dioxin activity and cellular disturbance.

The recognition of the importance of the Ah receptor ARNT dimer led to a watershed in dioxin research and ultimately a better understanding of this novel receptor and its ability to control cell communication through a process known as signal transduction. In addition, elucidation of this partnership led researchers to ask if this was a novel mechanism or if this pair of proteins was a prototype for additional signaling systems in mammalian cells. The answer to this question came quickly as a number of novel PAS proteins were identified. What was most interesting

about these proteins was that each of them required a heterodimer (2-protein complex where each protein is different) of two distinct PAS proteins and each pair regulated gene expression in very similar ways.

An understanding of the Ah receptor-ARNT provided a model to help explain other PAS proteins encoded by the human genome. The next mammalian PAS system to be understood was regulated by the hypoxia inducible factor (HIF). Biochemical and genetic evidence clearly demonstrated that HIF is made up of two PAS subunits, HIF1a and ARNT. This hypoxia-responsive transcription factor regulates the expression of batteries of genes that are turned on when oxygen levels are low, a condition known as hypoxia. A number of important gene products are regulated by this system including erythropoietin (Epo) and vascular endothelial growth factor (VEGF). Epo regulates the increased production of red blood cells and VEGF is an important factor in blood vessel formation both during development and in tumors.

More recently, the mechanism that underlies circadian behavior has also been shown to be regulated by PAS protein interactions. Circadian rhythm is the name given to the daily fluctuations in body enzymes and the processes they regulate that occur in response to the presence or absence of light. At the heart of the mammalian rhythm is a dimer of two PAS proteins known as CLOCK and MOP3. Mutations at either of these loci have been shown to have significant impact on biological rhythms in mice. Given that dozens if not hundreds of biological processes are regulated in a circadian manner, this system may be influencing a broad-spectrum of mammalian biology. In fact, recent evidence indicates that MOP3 may also be an important factor in the development of arthritic disease. Given that biological rhythms are responsive to external light, this system is also under study for its relationship to human depressive illness.

It is interesting to note that to date, all of the PAS signaling pathways described in mammalian systems can be thought of as mediating responses to environmental change. These environmental stimuli include chemicals like dioxins, changes in oxygen availability, as well as changes in day length and light. Such an observation may provide direction as scientists endeavor to understand those remaining PAS proteins of which there is little biological understanding. The future of PAS protein research is promising. At the present time, genetically engineered mouse models that lack one or more of the genes coding for these proteins have been generated at almost half of all known gene loci. These mice are only now being exploited as models for human disease. With only the models currently in hand, there exist powerful new systems to understand tumor promotion and growth, mental illness, and arthritis.

## How Ovaries Fail

The awareness of reproductive disorders resulting from exposure to toxic chemicals in the environment is increasing; however, exactly how these agents damage reproductive organs remains unclear. The ovaries are unique because at birth they house a finite and irreplaceable stockpile of germ cells (oocytes) enclosed within support structures termed follicles. Normal growth and maturation of these follicles are required for both fertility and the maintenance of a proper endocrine environment in the body. Indeed, menopause, driven by near-exhaustion of the oocyte reserve, heralds a major change in women characterized by the end of ovarian secretion of key hormones such as estrogen. As a direct consequence of ovarian aging, post-menopausal women face increased risks for developing cardiovascular disease, the bone-thinning disease osteoporosis, blindness and neurodegenerative disorders. Because a woman's supply of oocytes is limited, environmental agents that accelerate the destruction of ovarian germ cells are of great concern. Early loss of these germ cells can irreversibly hasten the time at which women experience problems with fertility and overall health due to premature menopause.

Very early life events (during fetal development and infancy) play a role in mediating the supply of oocytes through a process called programmed cell death (PCD or apoptosis). Using modern genetic techniques, scientists have established that apoptosis indeed is responsible for female germ cell loss during prenatal (before birth) ovarian development and throughout postnatal (after birth) life. Furthermore, in laboratory animals genetic manipulations of the PCD pathway that lead to altered oocyte apoptosis result in increased numbers of ovarian germ cells at birth, leading to a dramatic extension of ovarian life span and even delay of menopause into a later age.

Ovarian damaging chemicals have been examined for their potential to perturb the basic PCD machinery in germ cells, leading to inappropriate activation of apoptosis. Polycyclic aromatic hydrocarbons (PAHs) are toxic chemicals released into the environment by fossil fuel combustion. A primary route of human exposure to PAHs is tobacco smoke. For over 20 years it has been known that oocyte destruction and ovarian failure occur in PAH-treated mice, and that cigarette smoking causes early menopause in women. However, efforts to define the mechanisms by which PAHs destroy oocytes have been minimal. In many cells, PAHs activate the aryl hydrocarbon receptor (AHR), a member of an important family of proteins that control cell-to-cell communication (the Per-Arnt-Sim family of transcription factors). The AHR is also activated by dioxin, one of the most intensively studied environmental contaminants. Assuming that PAHs kill oocytes via the induction of apoptosis, one might assume that genes involved in the regulation of apoptosis would be prime targets for transcriptional regulation by the PAH-activated AHR. Later experiments have shown that a single injection of 9,10-dimethylbenz[*a*]anthracene (DMBA), a prototypical PAH, rapidly increased an ovarian pro-apoptotic gene known as Bax. Since both the AHR and its binding partner, the AHR nuclear translocator are expressed in oocytes, these results raised the possibility that Bax is directly regulated by the PAH-activated AHR.

Other studies reported that the ovary is capable of breaking down or metabolizing parent PAH compounds, such as 7,12-dimethylbenz[*a*]anthracene (7,12-DMBA), to biologically active intermediates. To determine whether a 7,12-DMBA metabolite, as opposed to or in addition to 7,12-DMBA, triggered AHR activation, experiments were carried out on 7,12-DMBA and its metabolite 7,12-DMBA-3,4-dihydrodiol (7,12-DMBA-DHD). Ovaries cultured with either 7,12-DMBA or 7,12-DMBA-DHD induced Bax accumulation and apoptosis in primordial and primary oocytes, with the metabolite being more potent in initiating germ cell destruction. Co-treatment with a compound that stopped the action of AHR stimulation (an AHR antagonist) inhibited both Bax expression and oocyte death resulting from 7,12-DMBA-DHD exposure. These results were reinforced by findings that the toxic responses of normal ovaries to 7,12-DMBA-DHD were similarly absent in ovaries of mice lacking the gene for AHR that were treated in parallel. Since 7,12-DMBA-DHD is known to bind with and activate the AHR, these data collectively indicate that the AHR is functionally required for PAHs to induce Bax expression and apoptosis in oocytes.

By grafting human ovarian tissue into mice, scientists were able to determine that Bax expression was greatly increased, followed closely by oocyte death in response to PAH exposure. The destruction of human oocytes by PAHs was found to be time- and dose-dependent. In fact, six days following injection of the highest dose of DMBA, approximately 70 percent of the primordial and primary oocytes had degenerated in response to PAHs. These findings provided the first direct evidence that PAHs induce human oocyte death in the living animal, and that this oocyte loss, like that in the mouse, is associated with an induction of Bax expression.

These studies have identified a new intracellular cell death-signaling pathway and have shown that activation of this pathway by DMBA, a member of a ubiquitous class of toxic environmental chemicals, can cause oocyte loss and ovarian failure. Furthermore, mice housing grafted human ovarian tissue provide a long sought-after model to conduct mechanistic studies and risk assessments of suspected chemical hazards and new drugs for human reproductive toxicity. Indeed, data derived from use of this model strongly support the contention that early onset of menopause in women smokers is due, at least in part, to the pro-apoptotic actions of tobacco smoke-derived PAHs in human primordial and primary oocytes.

## Determining the Genetic Cause of Drug-Resistant Malaria

Since ancient times, malaria has taken a devastating toll on people and societies. Today, after decades of eradication efforts, malaria still strikes an estimated 300 to 500 million people and causes more than 1 million deaths each year, with the heaviest toll among young children in sub-Saharan Africa. Countries with a high rate of malaria infection are among the poorest in the world, partly as a result of the economic impact of the disease. In the absence of a preventive vaccine, drugs to treat and prevent malaria are a mainstay of public health efforts to reduce the burden of disease. However, these efforts have been rendered less effective in many parts of the world by the emergence of drug-resistant strains of *Plasmodium*, the one-cell parasite that causes malaria. After more than a decade of intensive research, NIH researchers have found a single *Plasmodium* gene that plays a major role in resistance to chloroquine, a frontline malaria drug. This discovery offers hope that researchers can find ways to stem the tide of antimalarial drug resistance.

**Falling Behind in the Battle with Drug.** Since the discovery of the malaria parasite and its transmission by mosquitoes at the end of the 19th century, public health officials have focused on eradicating malaria with a combination of drug treatment and mosquito control strategies. The introduction of the drug chloroquine and the insecticide DDT in the 1940s set the stage for a worldwide effort to eliminate malaria. But early successes in this endeavor were soon overshadowed by the phenomenon known as resistance – the ability of organisms to develop strains that are impervious to specific threats to their existence. Not only did strains of mosquitoes become resistant to DDT, but chloroquine-resistant strains of *Plasmodium falciparum*, the most deadly species of malaria parasite, also emerged.

Chloroquine resistance first appeared in Southeast Asia and South America in the late 1950s, and emerged in Africa by the late 1970s. The rates and geographical distribution of resistance as well as the rates of infection and death from *P. falciparum*, the most virulent species of the parasite, continue to increase. Today, nearly all regions where *P. falciparum* malaria is endemic (continuously present) have chloroquine-resistant strains. In many areas, this resistance has forced the replacement of chloroquine with newer agents that are more expensive and have more severe side effects. In countries with no affordable alternatives, however, chloroquine remains the frontline antimalarial agent.

**Tracking the Resistance Gene in the Laboratory.** Since the mid-1980s, NIH investigators have been leading research efforts to find the *P. falciparum* gene or genes responsible for resistance to chloroquine and to develop methods to genetically manipulate *P. falciparum*. Researchers then identified a region of DNA on the parasite's seventh chromosome that appeared to be involved in chloroquine resistance. In 2000, after years of searching this particular stretch of DNA, the researchers reported a gene, known as *pfcr*, that contained tiny mutations in several chloroquine-resistant strains of *P. falciparum*. One specific mutation, known as *pfcr* T76, was present in all resistant strains. The team was then able for the first time to convert chloroquine-sensitive parasites to resistant ones by introducing a mutated *pfcr* gene. Although NIH investigators were confident that the *pfcr* gene plays a central role in chloroquine

resistance, they needed field tests to determine whether the same mutations responsible for resistance in the laboratory also were responsible for the failure of patients to respond to chloroquine treatment.

**Confirming Findings in the Field.** Recently, collaborating teams of laboratory and clinical researchers in the U.S. and the West African country, Mali, conducted studies that provided this confirmation as well as other vital information. Clinical investigators in Mali used an ultrasensitive technique called polymerase chain reaction (PCR) to determine the frequency of a gene mutation in a population. Investigators tested blood samples from patients in two villages in Mali before and after chloroquine treatment. The first of several ongoing field studies found that the *pfcr* T76 mutation was present in all samples from patients who remained infected after treatment or who were resistant to treatment, confirming the laboratory finding that this *pfcr* mutation in *P. falciparum* confers resistance to chloroquine. The investigators also found, however, that some patients whose samples carried the *pfcr* T76 mutation cleared the *P. falciparum* organism after treatment. Further, other investigators discovered a link between chloroquine resistance and the mutation of another *P. falciparum* gene, known as *pfmdr* 1. But this link is much weaker than that between resistance and the *pfcr* T76 mutation.

These additional findings suggest that although the *pfcr* mutation is responsible for chloroquine resistance, other factors, such as partial immunity to malaria and other genetic mutations in the parasite, may play a role in modulating responsiveness to the drug.

**Using New Knowledge for Public Health.** The 100 percent correlation between the presence of the *pfcr* T76 mutation and post-treatment infection has many implications for public health efforts against malaria. The finding suggests that PCR-based tests for the *pfcr* T76 mutation could be used to recognize the frequency of the gene in a population and guide public health policy. Moreover, knowledge about how genetic mutations in parasites confer resistance may help researchers restructure chloroquine to renew its effectiveness. In addition, further research on the role that partial immunity plays in eliminating chloroquine-resistant *P. falciparum* infection may prove useful to investigators who are trying to develop a malaria vaccine.

## **Accessing the Human Genome: Public Information Resources as Discovery Tools**

The human genome contains the DNA code for building the body and maintaining the health of a human being. Included in the code, which consists of four chemical constituents, called bases, represented by the letters, "A", "T", "G", and "C", are the particulars for the creation of every protein we use along with control signals governing when and where each protein will be produced. Since our proteins perform the work of our cells, and our cells build and maintain our bodies, the DNA sequence comprising the human genome contains vital clues for better understanding human health and disease.

The multi-national Human Genome project (HGP) was begun in 1990 with the simple but ambitious goal of determining the sequence of the human genome. The idea was to distribute the effort of sequencing all 3 billion bases to many genome centers located around the world. From the outset, one of the primary tenets of the project was that it was important to get the incoming data into a publicly accessible and useable form as quickly as possible so that researchers could begin their analysis of the genome without delay. NIH has also played an important role in providing public access to the genome data and designing tools for its analysis.

The data from the HGP has been deposited in NIH's GenBank DNA sequence database in roughly three overlapping phases, corresponding to the three phases of the Human Genome Project. For the past twenty years, NIH has produced and provided public access to GenBank, a seminal database of DNA sequences founded by NIH in 1982, and to an accompanying suite of search and analysis tools. Since the nature and quantity of the incoming data has varied as the HGP has progressed, NIH has developed new databases, data submission systems, display programs and analysis tools during each phase so as to provide timely public access to the data as the details of the human genome gradually emerged.

In the first phase, running roughly from 1990 into 1999, the human genome was mapped using a technique called radiation hybrid mapping, to establish a dense set of sequence landmarks along the genome acting as a framework to orient the bulk sequence data expected during the subsequent phases. These landmarks were integrated with existing genetic maps and made accessible to the scientific public via an NIH-developed web resource known as GeneMap96. The main practical value of having a dense and integrated genetic-physical map of genes is to accelerate the discovery of disease-related genes. GeneMap96 and its successors, GenMap98 and GenMap99, have allowed researchers to zero-in on candidate pools of disease genes within a region defined by a set of genetic markers. A related tool, called Electronic PCR, allows researchers to map segments of human DNA that they had sequenced by searching for sequence landmarks that were known to be present only in unique locations in the genome.

In the next phase of the HGP, running from 1996 to December 2000, the first of bulk quantities of "draft" quality sequence data began to be deposited in GenBank. Draft sequence consists of large packets of sequence derived from a particular location in the human genome but broken up into dozens of smaller sequence units by gaps for which sequence was not obtained. The challenge at this point was to make this huge amount of data accessible to scientists wanting to

compare it to other public data as well as to their own research data. To accommodate this new influx of sequence data, NIH created, in 1996, a new GenBank division called the "High Throughput Genomic" sequence division. Along with the new GenBank division, NIH established a streamlined procedure for the submission of bulk sequence data generated by the HGP sequencing centers. The data was made searchable by BLAST, an NIH sequence similarity search program used by scientists over 80,000 times a day, by formatting the HTG data as a special "HTG" BLAST database. By August 2000, the HTG database contained over 4 billion base pairs of DNA sequence. By accessing the data in the GenBank HTG division, researchers are able to determine whether a sequence of interest had been deposited by a sequencing center within 24 hours of its submission. In addition, the HTG division affords researchers the opportunity of assembling parts of the genome in their areas of interest without the need to wait for a large-scale assembly. The ready availability of raw sequence data has led to an acceleration of disease gene discovery, as illustrated by the assembly of the Fanconi syndrome gene locus from HTG sequence in advance of a large-scale genomic assembly.

The final phase of the HGP was marked by the first assembly of the raw public sequence data in December 2000, with about 98 percent of the genome deposited in GenBank. The sequencing centers are filling in the remaining gaps in the sequence, a process called "finishing." It has now become feasible to build the sequence data into an assembly of the entire genome and to update this assembly regularly as new DNA pieces are added to GenBank. Having the hundreds of thousands of pieces of human DNA sequence assembled into a single entity is advantageous as it allows researchers to view each piece in its proper relationship to all the others. In order to build and maintain this assembly, NIH has developed a semi-automated procedure for compiling the human genomic pieces generated by the HGP into complete chromosomes by using a variety of techniques including searches for overlapping portions of sequence from two different pieces. In addition, NIH has developed a genome browsing tool, called the Human Genome Map Viewer, to allow researchers to view the assembly and to download selected portions of it for their own research.

As a result of the wide dissemination of the human genome data to the scientific community, many significant discoveries have been made. The genome has turned out to be quite different from our expectations, a result which confirms the importance of the undertaking in the first place. Despite the fact that the primary purpose of the genome is to encode instructions for making proteins, only about 2 percent of the genome has been shown to code for protein amino acid sequences. At least half of the genome is comprised of highly repetitive sequence that appears to be involved in genomic evolution, gene regulation, and, in some cases, disease. These two findings came as quite a surprise to the scientific community because the estimated number of human genes had to be revised from a previous estimate of 100,000, based on the relative genome sizes of human and other organisms such as the fly and the worm, to only 40,000, based on knowledge of the actual human genomic sequence. Analysis of the publicly available human genomic sequence has made it clear that a much larger part of the genome than expected is involved in the regulation and evolution of the small part that actually codes for proteins. Now, the research community is making effective use of the public data provided by NIH to analyze the structure of the genome in order to identify those portions of the genome important

in gene regulation. For example, large segments of repetitive DNA called "CpG Islands" can be seen throughout the genome adjacent to gene-rich areas, and are believed to help regulate gene activity. The genomic sequence is also fueling research into human genetic variation. Scientists have already identified about 1.4 million locations where DNA variations occur in humans. Many of these variations are associated with disease. As the human genomic sequence data continues to accumulate in GenBank, researchers accessing it through NIH web interface will make many of the discoveries that will promise to lead to future medical breakthroughs.

## Providing Sight to Dog Born Blind

The genetic disorder known as Leber's Congenital Amaurosis (LCA) is one of several incurable forms of blindness collectively known as retinitis pigmentosa. It was first described in 1869 by Theodor Leber, who studied this condition in children under the age of one year. No progress in understanding the cause of this childhood blindness was made for over 100 years. Currently, there is no treatment for LCA or related early onset retinal degenerative diseases. As a result of the Human Genome Project and the enormous progress made in the field of molecular genetics, there is now hope that the disease may be treatable soon.

Mutations in LCA account for about 11 percent of patients with early onset retinal degeneration. In 1997 investigators located an area on chromosome 1 that appeared to be involved in early retinal degeneration. This area on chromosome 1 contained a gene called RPE65. Also in that year, a group of investigators showed that the RPE65 gene product was located in the retinal pigment epithelium (RPE), a single layer of cells in close contact with the retinal photoreceptor outer segments. The function of the RPE65 protein was unknown, although it appeared to be involved in vitamin A metabolism in the retina. Assuming that a genetic defect in the RPE might cause an early retinal degeneration, the RPE65 gene was used to screen for mutations in patients with LCA. This resulted in the identification of mutations that cause LCA, when a child inherits two defective copies, one from each parent.

Research on the treatment of LCA has advanced enormously through recent studies of a naturally occurring congenital blindness in Swedish Briard dogs. In the course of long-term breeding by humans, these animals have acquired an RPE65 mutation that is identical to one that causes about 20 percent of the human LCA cases, although mutations in any of a dozen or so other genes are also known to cause LCA. Histopathology studies in homozygous dogs, which carry two defective copies of the gene, show abnormal rod photoreceptor cell morphology early in life, with slowly progressive photoreceptor degeneration and blindness.

Normal versions of the RPE65 gene were genetically engineered into an adeno-associated virus vector. Thousands of copies were directly injected behind the retina, close to the RPE cells, of the right eye of three blind Briards who were between the ages of two and four months old. Ninety-five days after injection, the animals had vision in the treated right eye, judged by pupil response, electroretinogram, and behavioral testing. Tissue from a subretinally injected eye demonstrated persistence of the transferred DNA. The left, control eye, of these animals remained blind. All three dogs were seeing well nine months after treatment, with no ill effects.

This study provides a "proof of concept" to show that gene therapy can restore vision in a large-animal model of a human retinal degeneration. Although previous studies had demonstrated that gene therapy could delay retinal degeneration, the present study demonstrates definitive recovery of function. Function was not restored after intravitreal injection of the genetically engineered vector, a route that targets ganglion cells but not RPE/photoreceptor cells. The treated animals will continue to be studied in order to validate the effectiveness, duration, and safety of the treatment. Future studies will include applying gene therapy to both eyes of a dog.

A logical question leading from these results is whether subretinal injection of the engineered construct would also correct the functional defects found in humans with LCA caused by RPE65 mutations. Researchers are proposing to begin the long-term safety and efficacy studies that must be performed before beginning such trials in humans.

## Free Radicals: A Link Between Alcohol and Liver Damage

Free radicals are molecules that perform important beneficial functions in our bodies, but that also can wreak biological havoc in us. Under certain conditions, they contribute to a host of illnesses. Among them is alcoholic liver disease (ALD), which accounts for about half of all deaths from cirrhosis, the 10<sup>th</sup> leading cause of mortality in the U.S. Alcohol researchers have built on decades of work on free radicals (and other mechanisms that contribute to ALD) to reach a better understanding of how to prevent alcohol's devastating effects on the liver.

Our own bodies make most free radicals from oxygen, as a normal byproduct of metabolism. Fortunately, our cells also make antioxidants, substances that neutralize free radicals. When free radicals form in excess of our antioxidants' ability to neutralize them, however, tissue damage ensues.

Alcohol is among the substances that triggers formation of free radicals, which are molecules that have an extra, unpaired electron. The unpaired electron is very reactive and can disrupt chemical bonds that hold other molecules together, changing their structures and functions. The changes in these altered molecules can lead to unpaired electrons in them, too, leading to disruption of yet more molecules, and so on. This chain reaction is beneficial to some biological systems, but destructive to others, some of them critical.

Antioxidants counteract these activities by taking unpaired electrons from free radicals. Certain enzymes (proteins that regulate the timing and placement of chemical reactions) serve this protective antioxidant function, but other enzymes generate free radicals, instead. Alcohol damages the liver not only by stimulating production of free radicals, but also by inhibiting liver cell's antioxidant production.

***Building on the Background.*** The discovery that free radicals were potent damagers of body tissue came about 50 years ago, from scientists in the field of radiation chemistry. Once chemists had made this discovery, other scientists, in the biological arena, asked the next logical questions: Do biological systems *produce* free radicals, and *how* do they damage biological tissues? In the 1960s, scientists began studying free radicals' effects on the fatty part of the protective membrane that surrounds cells. The importance of this membrane cannot be overstated; it regulates what can enter or exit the cell, so that the cell functions properly. Scientists found that chain reactions of free radicals in the membrane can damage the cell to the point of causing its death.

Other scientists would add to the picture of the devastating damage free radicals can cause. They soon found that free radicals damage DNA, leading to genetic mutations, and, later, that free radicals damage proteins that regulate hundreds of crucial activities in cells, causing the cells to function abnormally.

Scientists had found that free radicals cause tissue damage, and where, but a critical question remained: How to prevent it? In the 1970s, scientists made a discovery that would lay the

groundwork for answering this question, and research on free radicals surged. They found that an enzyme, superoxide dismutase (SOD), removed a specific free radical from biological systems. The researchers now had a tool enabling them to conduct experiments to find out if they could remove free radicals to protect tissues from damage.

In the 1980s, scientists made a major leap with the discovery that blood-starved tissues – the kind of damage seen in heart attacks – sustain additional damage when blood flow is restored to them. The flush of restored blood brings with it a new burst of oxygen, whose metabolism causes a surge of free radicals to form. Scientists found that giving antioxidants attenuated the damage.

***Focus on Alcohol.*** In the alcohol field, researchers knew that some type of association existed between alcoholic liver disease (ALD) and free radicals. What they did not know was how alcohol increased free-radical production and whether free radicals were a cause or an effect of ALD.

Evidence that free radicals play a major role in *causing* ALD is mounting. In the 1980s, alcohol researchers developed animal models to show that, at the whole-organ level, the association between ALD and free radicals existed. Other research suggested that one of the systems through which alcohol increased free radicals might involve enzymes, including one called CYP2E1 that metabolizes alcohol. Further investigations showed that when CYP2E1 metabolizes alcohol, free radicals do form and that cells make more CYP2E1 on exposure to alcohol. These findings were very suggestive, but scientists needed to link them more directly to strengthen the case for a cause-and-effect association between free radicals and ALD.

That study came about 10 years ago, when alcohol researchers injected into the main cells of the liver, *in vitro*, the gene that produces CYP2E1, causing the cells to make the enzyme. When the scientists incubated these CYP2E1-producing liver cells with alcohol, the cells produced more free radicals and sustained damage. The investigators thus more directly linked alcohol with free-radical production and liver damage at the cellular level. They went even further: Adding antioxidants to the liver cells attenuated injury and reduced levels of free radicals.

***Strengthening the Connection.*** In recent years, investigators conducted four studies that greatly strengthened the evidence that free radicals are a cause, not an effect, of ALD. The collective strength of the studies is that they used four different approaches that independently reached the same conclusion: Alcohol-induced free-radical production activates biological mechanisms that lead to liver injury.

In each of the studies, researchers gave animals alcohol for extended periods, which normally triggers free-radical production. This time, however, the scientists either blocked enzymes that generate free radicals in the liver, boosted the animals' ability to produce antioxidant enzymes, or restored depleted levels of the animals' own naturally occurring antioxidants by giving them antioxidant precursors. The scientists used four different genetic or pharmacological techniques to achieve these effects. Each of these interventions reduced free-radical production, attenuated

mechanisms in the biological chain of events through which free radicals damage the liver, and reduced liver injury itself.

***The Final Word?*** This one mechanism of alcohol-induced liver injury, free-radical formation, is not the sole vehicle for alcohol's liver damage, and antioxidants might not be a magic bullet, especially if used indiscriminately, without medical supervision. Neither free radicals nor antioxidants operate in a vacuum; they interact with other biological systems, and antioxidants have their own negative effects, under certain circumstances.

What these decades of research have produced is not the final word on how to prevent and treat ALD, but major strides in getting there. Through this body of work, particularly the research of the last decade, free radicals are emerging as a principal vehicle through which alcohol damages the liver, and antioxidants look increasingly promising as a potential treatment.

## Understanding Disease-Environment Interactions in Global Amphibian Decline

A wide range of observations over the past 15 years have led scientists worldwide to examine the relationships between environmental changes and infectious diseases in humans and other animals. For example, outbreaks of lethal hantavirus in the southwestern U.S. are linked to increases in local field mouse populations that follow *El Nino* weather cycles. Expansion of Lyme disease in the northeastern U.S. has been linked to changes in rodent and deer populations and altered forest cover. Excessive mortality in seal die-offs off the coast of California appears to reflect both environmental-mediated cancers and viral diseases.

Amphibians have emerged as a particularly important class of biological indicator of environment change. Over the past ten years, we have witnessed dramatic declines and even extinctions of frog and salamander species around the world. Whereas some studies point to increased ultra-violet radiation and drying of habitats as culprits, other research implicates one or more pathogenic fungi.

Building upon years of field observations and experimental research supported by NIH and the National Science Foundation (NSF), scientists at Pennsylvania State University have clarified the relationship of several environmental factors and a deadly pathogen. NIH-funded scientists have recently demonstrated that complex interactions between climate, ultraviolet-B radiation, and an infectious agent, *Saprolegnia ferax*, interact in ways that are harmful and sometimes fatal for frog populations<sup>1</sup>.

Previous research conducted by atmospheric scientists worldwide has shown that penetrating ultraviolet-B radiation is increasing due to loss of ozone. Studies have also concluded that the Southern Oscillation weather cycles, commonly known as *El Nino* have a profound effect on annual rainfall and water levels in several parts of the world, including the Pacific coast of the Northwestern U.S. Data obtained from breeding *Bufo boreas* frogs shows that a significant percentage of frog embryos become infected by fungal pathogens and later died. Furthermore, experiments showed that these deaths were more common when UV-B radiation is high and water levels are low. While the exact relationship of the pathogen to frog embryo mortality requires further study, it is clear from this seminal study that decreased rainfall, ultraviolet radiation, and presence of the disease agent combine to produce significant declines in *Bufo boreas* frogs and probably other species of amphibians.

These important discoveries highlight the growing pressures on global biodiversity and serve as sentinel measures of the destructive implications of environmental degradation. These studies do not have direct implications for human disease. However, they demonstrate the complex interactions of environmental change with ecological niches and subsequent emergence of new infectious diseases. Modeling such events may help scientists to develop new insights into the factors that influence the emergence and re-emergence of infectious diseases in humans.

## **A Fifteen Year Investment in Caregiving Research**

By 2030, approximately 20 percent of the American population will be 65 years of age or older. With the aging of the population, issues of providing care to them will become increasingly significant. While some of the care for older people is provided by home health agencies or nursing homes, much of the care is provided informally by family and friends. Current estimates place the number of informal caregivers in the U.S. at 24 to 27.6 million people. Each provides a mean of 17.9 hours of unpaid caregiving per week assisting with activities such as bathing, feeding, dressing and toileting. If the care recipient is cognitively impaired, he or she may need assistance with activities such as paying bills and balancing the checkbook. Assuring safety of the care recipient, who may have impaired mobility, adds to the amount of time spent in caregiving activities.

The informal caregiving activities of family and friends were estimated in 1997 to have an economic impact of \$196 billion. That figure dwarfs the Nation's total expenditure for paid home health care (\$32 billion) and nursing home care (\$83 billion). Informal caregiving also provides less tangible savings through delaying or avoiding institutionalization so fewer nursing home beds are needed.

Caregiving has been a focal point for NIH funding since its inception fifteen years ago. Early studies called attention to the considerable effort care providers spent coordinating care across paid and informal caregivers within multiple, complex care situations. These efforts by the informal care providers helped their loved one attain their goal of an optimal quality of life while living at home and emphasized the importance of the informal caregiving role.

One specific population where informal caregiving is of great importance is with a care recipient who is cognitively impaired. The Alzheimer's type of dementia is used as an example. Approximately 70 percent of the 4 million people with Alzheimer's are cared for in the home by family and friends without payment. Since the duration of Alzheimer's from diagnosis to death averages nearly ten years, caregivers of people with dementia face a very long road of increasing needs to fulfill. NIH-sponsored research has focused not only on the cognitively impaired population, but also among families with other diseases and conditions, as well as the special issues surrounding care at the end of life.

Many people who have chronic diseases such as severe congestive heart failure, respiratory disease, and cancers and those who are at the end of life, for any reason, choose to live at home rather than in an institution. These care recipients may have complex care requirements that fully engage the caregiver in tasks such as managing troublesome symptoms, observing for adverse drug effects, preventing complications and restoring function. Over time, science has refined our understanding of the multifaceted outcomes of caregiving, its stressors, and predictors of bad outcomes. One such predictor is the dynamics within the home. Research has identified predictors of family dynamics that are responsive to interventions.

The consequences of providing informal caregiving on the care provider is another focal point of NIH-sponsored research. The specific health risks and situations for caregivers of cognitively intact recipients are beginning to be described at this time. Maintaining the health of the care providers as well as the care recipient is crucial to successful care within the home.

Studies of families who were successful at providing informal care were used as models to identify and test interventions to sustain caregiving efforts. One approach identified was caregiver support groups that were found to influence the family's decision to place their loved one in a nursing home. Nursing home placement of a loved one places extreme stress on the family, but how this degree of stress compares to long-term caregiving in the home is not known. We have also learned that interventions need to be tailored for specific caregiver and care recipient situations to prevent bad outcomes. There is an emerging body of information about the experiences of caregiving among racial and ethnically diverse caregivers that underscores the need to individualize interventions to the caregiving situation. Studies have also demonstrated that technologies that facilitate communication and access to information are not only feasible but promote positive outcomes.

Currently funded work is investigating how to encourage caregivers to reach out and obtain services and support that might be helpful to them. The role of hospital health care professionals in training and supporting caregivers for their in-home role once the care recipient is discharged is also being evaluated. The interactions among the family caregivers and the hospital personnel when the care recipient is hospitalized is another area of emphasis. A continued focus on health disparities extends both to caregivers and care recipients.

Because of NIH's fifteen year investment in caregiving research, scientists can now begin to map out the breadth and depth of skills needed to provide caregiving. These research results and future directions will enhance the existing body of knowledge about caregiving. Application of this knowledge to family situations tends to occur fairly rapidly by means of a network of advocacy groups and senior citizen organizations. Once the knowledge is developed, applied, and evaluated, families recognize their improved quality of life.

## **Bacterial Pili – Molecular Initiators of Bladder and Kidney Infections**

It all starts with a “handshake” between two molecules of almost infinitesimally small size: a tiny, hairlike projection from the surface of a bacterial cell and a small cluster of sugar molecules on the surface of a human epithelial cell. From this molecular interaction arise over 8 million doctor visits each year. A urinary tract infection (UTI) – which may involve the bladder, kidneys, or both – begins when a molecule on the surface of a bacterial cell recognizes and binds to a molecule on the surface of its target urinary tract epithelial cell. Bladder and kidney infections are important public health concerns because many individuals suffer recurrent infections. Because of the wide prevalence of UTIs, their tendency to recur, and their potential for bladder infections to progress to more serious kidney infections, researchers have long sought to understand the first critical steps in infection. The identification and characterization of the mechanism by which bacteria adhere to epithelial cells is the key to designing therapies to block this process and prevent these infections. Recently, research from the field of structural biology – the study of how molecules interact with one another – has yielded important insights into the earliest stages of UTIs and provided fresh evidence that strategies to block the attachment of bacteria may represent an effective way of treating the initial infection as well as a viable approach to preventing their recurrence.

The adhesion of bacteria to the cells of the urinary tract is mediated by proteins present on the surface of the bacterial cells. Over twenty years ago, studies implicated these proteins in this process. Since that time, families of proteins have been discovered, the genes that encode them have been cloned, and they have been the subject of much research. Although the genetics of these bacteria are well-characterized, studies of bacterial genetics illustrate a paradox about modern molecular biology: as more is learned about the genes of organisms, the more researchers appreciate that genes tell only part of the story. A gene represents a set of instructions for assembling a chain of amino acids in a specific sequence to form a protein. In order for this to occur, a gene must first be turned “on” so that the information within the gene can be read by the cell. This blueprint must then be faithfully translated into a mature protein. However, functional proteins do not exist as linear chains of amino acids; rather, they fold back upon themselves, assuming intricate shapes with highly complex topography. To truly understand how a protein functions, and how it interacts with other proteins, it is often necessary to understand how these proteins are assembled from their genetic instructions, how this process is regulated, and what form these proteins ultimately assume.

Among the people looking for this knowledge are structural biologists – scientists interested in how molecules come together to form critical components of cells. These researchers have long studied protein assembly using bacteria as a model system. Pili are tiny rod-shaped projections from the surface of many bacteria that can act as adhesion molecules, facilitating cell-to-cell contact and communication. Structural biologists have determined that these hairlike pili are formed through a complex series of interactions that begin within assembly of the pilus base within the bacterial cell and the transport of the elongated pilus across the outer cell membrane. As a model system to study protein assembly, pili have proven to be a very rich source of

information about the mechanics of protein assembly and transport within the cell. But pili are important for reasons beyond their value as models of protein assembly.

In UTIs, attachment of bacteria to the surface of the epithelial cells that line the urinary tract is a key event, and a better understanding of this process might reveal novel approaches to preventing infection. Not surprisingly, this initial interaction is mediated in large part through proteins on the surface of the bacterium that recognize and bind specifically to molecules on the surface of the epithelial cells – including pili. Pili consist of multiple protein subunits each encoded by a specific gene. It has long been known that the presence of certain kinds of pili on a cell's surface increases the number of *E. coli* capable of infecting the urinary tract and enhances the persistence of the infection. Although all strains of uropathogenic *E. coli* possess the genes necessary to give rise to pili, the expression of a particularly critical one is controlled by a small segment of DNA located nearby. This small invertible element is capable of “flipping” its relative orientation within the chromosome and thereby controlling gene expression: when it faces one direction, expression of the gene is turned “on” and pili are present on the cell surface; conversely, when the element faces the opposite direction, the gene is turned “off” and the protein is absent. Relatively little has been known about why this element flips, how its orientation might change during an actual infection, and what the implications of these changes could be.

To answer these questions, scientists recently isolated different strains of *E. coli* from cystitis – bladder infection – or pyelonephritis – kidney infection – and then examined the ability of the bacteria to change the orientation of this invertible element. After culturing both bacterial strains under conditions that maintained the element in the “off” position, the strains were introduced into mice. Subsequently, the bladders were removed and the orientation of the invertible element within the bacteria analyzed. The element in the bacterial strain that was originally isolated from the cystitic strain quickly reverted to the “on” orientation; in contrast, it remained largely “off” in bacteria from the pyelonephritis-causing strain. This observation suggests that cystitis-causing strains have an enhanced ability, relative to pyelonephritis-causing strains, to change the orientation of the invertible element to the “on” position and thereby alter pilus gene expression. This finding suggests that bacteria that infect the bladder may rely more heavily on pilus-mediated attachment during infection than bacteria that infect the kidney. It also illustrates how changes in expression of a particular gene – and not its presence or absence – can influence the site of infection in UTIs.

Once bacteria find themselves in the kidney, pili again play a role in adhesion and infection. A protein that sits atop the pilus – PapG – is involved in recognizing a specific molecule on the surface of the kidney cell and thereby mediating binding of the bacteria to the kidney cell. PapG's partner in this molecular interaction is a group of sugar residues on the outer membrane of human kidney cell – globoside. Scientists have deduced the three-dimensional structure of PapG alone and in a complex with globoside. This detailed molecular snapshot – with a resolution finer than one-millionth of a meter – has allowed researchers to identify which areas of the PapG protein mediate binding of the bacterium to the host cells, as well as to pinpoint which regions of the host globoside are important for this interaction. By defining the specific

molecular interaction necessary for binding to occur, this insight not only sheds light on a key event in infection but may also lead to the development of vaccines that target the disease process at its earliest stages.

Even with all of these insights into the molecular events that underlie the earliest stages of infection, the sad fact remains that, even after successful antibiotic therapy, a significant number of women will suffer a relapse of their UTI. Why this is so has remained a mystery for many years. New research has provided fresh insights into why many bladder infections recur – and once again highlights the role of pili. It seems that bacteria use their pili not only to recognize and bind to bladder cells during the initial infection but also to get inside of these cells. Invasion of these bladder cells may provide the bacteria with a relatively safe environment in which they may either replicate or go into “hibernation,” only to emerge later – and cause a subsequent UTI. According to this model, following the initial UTI the bacteria persist at low levels, continuously invading a small number of cells, escaping, and re-infecting neighboring cells. Insights into the role of pili in this process might allow for therapies to be designed that interrupt this cycle of infection and relieve the burden of recurrent UTIs.

The study of pili illustrates how technical and highly specialized research into a seemingly obscure question – how proteins are assembled in bacteria and what shapes they assume – can also have relevance to a common and costly health issue – UTIs. From bacterial models of protein assembly to critical players in infection and possible targets for future therapies, the story of pili shows that there is no such thing as something too small to investigate.

## Genetic Breakthroughs in the Study of Crohn's Disease

In a landmark finding, researchers announced the discovery of the first gene shown to confer susceptibility to Crohn's disease (CD), a debilitating form of inflammatory bowel disease (IBD) that affects an estimated 500,000 Americans. Previous studies had implicated a region of chromosome 16 as the possible location of one or more CD genes. However, it was not until two scientists working in different fields decided to pool their ideas and resources that the gene was finally pinpointed and sequenced. This targeted, interdisciplinary, collaboration revealed that a mutated form of the gene – *Nod2* – significantly increases a person's risk for developing Crohn's disease. This discovery is built upon a solid foundation of research into how genetic and environmental factors combine to initiate an aberrant immune response that cascades into a destructive inflammation of the digestive system.

Crohn's disease typically afflicts young people in their teens and twenties, although it can strike at any age – as President Dwight Eisenhower discovered in his 60s. For decades, only one treatment existed for Crohn's disease: surgical removal of the affected regions of the intestine. Research has made other, less drastic alternatives possible for many patients. The goals of modern therapy for Crohn's disease are to control inflammation, correct nutritional deficiencies, and relieve symptoms such as abdominal pain, diarrhea, and rectal bleeding. Treatment options include oral medication, nutritional supplements, or surgery – alone or in combination. A major medical advance was the development and FDA approval of the drug infliximab (Remicade®), which alleviates symptoms in two-thirds of patients by blocking the action of a pro-inflammatory cytokine known as tumor necrosis factor alpha (TNF- $\alpha$ ). Still, even today, about two-thirds to three-quarters of Crohn's patients undergo surgery at least once, with up to half requiring a second operation. Thus, even though infliximab is an enormous therapeutic advance, a vigorous search continues for the underlying genetic and environmental causes of Crohn's, because future treatments could derive from this critical knowledge.

Though the specific causes of Crohn's disease have long been elusive, substantial evidence points to a complex interplay of genetic and environmental factors in disease initiation. When the immune system of genetically susceptible individuals is exposed to bacterial flora normally present in the gut, the destructive inflammation of CD is unleashed. The existence of strong genetic influences has been reinforced by many research studies showing that the disease occurs in clusters within families and that siblings of CD patients have a 30-fold increased risk of developing CD compared to the general population. In addition, clinical studies conducted with identical twins have reinforced the genetic underpinnings of the disease. However, CD does not seem to be a simple genetic disease caused by a single gene. Rather, it is genetically complex – involving the presence and possible interaction of two or more genes plus environmental factors.

Although the polygenic nature of Crohn's disease has made the hunt for its underlying causes especially challenging, researchers remain undeterred in their quest. A better understanding of the genetic causes of Crohn's disease might identify new targets for therapy, as well as individuals who may be at risk for developing the disease in whom early diagnosis and intervention might prove beneficial.

Important clues about genetic and environmental factors in triggering the immune flare-up of Crohn's disease have emerged from studies in animal models, especially mice that have been raised under special, germ-free conditions and therefore do not have bacteria within their guts. When these "naive" mice are inoculated with the bacterium *Bacteroides thetaiotaomicron*, a prominent component of the normal mouse and human intestinal flora, the cells of their intestine respond by turning on a wide array of genes. These genes include those involved in nutrient absorption, fortification of the intestinal barrier to infection, metabolism of foreign molecules, growth of new blood vessels, and maturation of the intestinal wall. Through the controlled introduction of harmless bacteria, researchers have found an innovative way to examine the cellular and molecular machinery that normally springs into action very early in life. By understanding the normal response of healthy gut tissue to harmless bacteria, scientists have gained important knowledge about how this response may go awry in Crohn's disease – when genetic susceptibility and environmental factors converge to incite the immune system and promote inflammation.

Insights into the relative contributions of genetics and environmental factors to stimulating the immune response in Crohn's disease have also emerged from a strain of mice known as SAMP1/Yit, which are genetically prone to developing Crohn's-like symptoms. When these animals are housed under germ-free conditions they remain disease-free, thus demonstrating that genetic factors alone will not produce the disease. However, under normal environmental conditions – including the presence of bacteria – SAMP1/Yit mice spontaneously develop intestinal inflammation that remarkably mimics human Crohn's disease. This observation suggests that a key component of Crohn's disease is the reaction of a genetically susceptible host to naturally-occurring intestinal bacteria. A close examination of the intestines of animals affected by the disease reveals that the inflammation is centered in the small intestine and that it is not continuous – two cardinal features of human Crohn's. Researchers studying this model have noted that the involved regions have a large number of immune T cells in them, and that these T cells produce interferon gamma (IFN- $\gamma$ ) and TNF- $\alpha$  – two proteins that promote inflammation. When they isolated these T cells and introduced them into an immunocompromised mouse that was otherwise normal, the recipient animal developed intestinal inflammation that was similar to that seen in SAMP1/Yit mice. Importantly, this transfer of disease from one mouse to another could be blocked if the recipient animal first received an injection of an antibody that blocks the action of TNF- $\alpha$ . Together, these results implicate activated T cells – and the TNF- $\alpha$  they produce – as mediators of intestinal inflammation, arising from underlying genetic and environmental causes. This observation provides further validation of the usefulness of research on SAMP1/Yit mice as a model of Crohn's disease, because infliximab, the medication used today to treat many Crohn's patients successfully, is also an inhibitor of TNF- $\alpha$ .

A major advance in unraveling the genetics of Crohn's disease occurred in 1996, when researchers identified a promising region on chromosome 16 believed to include CD genes. In related research, other scientists employed gene profiling techniques to analyze genes thought to be culprits in Crohn's disease. Using cutting-edge genetic technology called DNA microarrays,

they looked for differences in the patterns in which genes are expressed in various tissues of IBD patients. They found “clustering” of genetic expression levels in chromosome areas already linked to Crohn’s disease, thus reinforcing the probability that these regions did indeed harbor the putative gene or genes.

Research took a giant leap forward this past year with the discovery of the first susceptibility gene for CD on chromosome 16. The impetus for this discovery was research on an immune gene called *Nod1*. *Nod1* is a member of a growing family of intracellular proteins with similarities to proteins involved in cell death as well as a class of disease-resistance genes in plants. When a draft sequence of the human genome was released last year, the scientist who led the team that discovered *Nod1* noted a very similar gene – *Nod2* – located in a region of chromosome 16 previously identified as containing a Crohn’s susceptibility gene. He pointed this out to a colleague who was studying Crohn’s disease, and together they used her repository of DNA from 416 families with a history of Crohn’s disease to identify a defective form of *Nod2* in about 15 percent of Crohn’s patients. The mutated gene also is present in about eight percent of healthy people, indicating that it increases the risk of the disorder, but that other factors must also interact for the disease to occur. The discovery of *Nod2* mutations in Crohn’s disease was validated by a second independent study using a completely different approach, known as positional cloning, to find Crohn’s disease genes on chromosome 16. In this second study, additional mutations in *Nod2* were found in patients with Crohn’s disease. Having one flawed copy of the gene doubles a person’s chances of developing CD; having two copies can increase the risk from 15 to 40 fold.

How do mutations in *Nod2* contribute to Crohn’s disease? *Nod2* is normally present in monocytes, a type of white blood cell that is one component of the “innate” immune system. Within the cell, *Nod2* activates molecular factors involved in the response to a common component of the outer lining of bacteria known as lipopolysaccharide (LPS). The mutant *Nod2* protein, which terminates prematurely and is about three percent shorter than it should be, seems less able to respond to LPS. The finding that defects in *Nod2* function might impair the ability of the “innate” immune system to respond to bacteria may seem counterintuitive, given that Crohn’s disease is thought to arise from an *overreaction* of the immune system. However, this impaired response by the “innate” immune system might open the door for the “adaptive” immune system – of which T cells are an important component – to step in. T cell-mediated immunity could possibly overcompensate for the diminished “innate” response and thereby lead to persistent inflammation. This concept is supported by previous observations that certain strains of mice that have other types of defects in their ability to respond to LPS spontaneously develop bowel inflammation similar to IBD. Future studies will attempt to define the role of *Nod2* in contributing to Crohn’s disease.

Extensive research by dedicated scientists and clinicians, coupled with critical advances in scientific technology, have provided the groundwork for extraordinary achievements in genetic research. Scientists studying animal models of Crohn’s disease have provided insights into the intertwined roles of the immune system in genetically susceptible individuals and naturally-occurring bacteria in contributing to inflammation. The availability of the complete sequence of

the human genome played a pivotal role in the identification of the Crohn's gene, as did open communication between researchers working in different fields. Finding this gene is not the first nor final step, but it is a crucial step in conquering this disease of unknown origin.