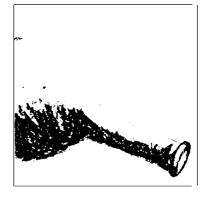# Trends in Science, Technology and Drug Discovery[1]

5

S cience undergirds the growth of the pharmaceutical industry over the past century. The industry has long been highly research intensive. Its roots trace to the advances in physiology and organic chemistry at the end of the last century. In recent years, however, the investment in research and development (R&D) has reached new heights. Drug treatments have transformed medical practice; biomedical research has transformed the discovery and development of drugs.

The expanding science base for drug discovery has implications for the cost of developing new agents. Two implications of the rapid advance of science for the overall costs of discovery are the need to keep abreast of the expanding base of knowledge about disease mechanisms and the need to keep abreast of other competitors in the industry. Pharmaceutical firms compete on several fronts. Science and technology form one basis for competition. This chapter describes the process of pharmaceutical research, particularly the discovery of new drugs. It centers on research more than development and assesses how science and technology might change the process and its costs over the next decade. This chapter explains how a new agent is discovered, particularly how drug discovery is changing in the face of new developments in science and technology.

The focus is on changes in the discovery of new therapeutic drugs brought about by the explosion of knowledge about molecular biology and human genetics. The potential effects of new methods, materials, and instruments for *diagnosis are* not discussed here. The first fruits of many lines of research, including protein analysis and study of DNA (deoxyribonucleic acid) emphasized in this chapter, are likely to be new diagnostic tests.

Improved diagnosis generally advances before treatment, often leading by a decade or more. Advances in body imaging, for example, have obvious implications not only for diagnosis, but also for measuring the efficacy of treatment (281).

Also not discussed in this chapter are new methods of drug delivery currently subject to extensive research efforts (218,333) and research on new assays to predict how the body responds to new agents. These and other topics are central to future directions in pharmaceutical R&D. Each has a rich literature. The discussion emphasizes molecular biology, however, because this best illustrates the shifting scientific foundations of drug discovery. This tack was chosen not because other topics are less important, or will have less impact, on future directions in pharmaceutical R&D. Rather, it was chosen as the most promising way to highlight the widening gulf between current and traditional research methods.

Finally, the emphasis on human genetics is only illustrative. Molecular biology has come to dominate other areas of research equally relevant to drug discovery. Molecular approaches to combat viral infections, to control the immune response, to understand cancer, to penetrate the brain and spinal cord, and to modulate responses to hormones and growth factors are all extremely active fields. A chapter of equal or greater length could be written for each field. The focus on human genetics was chosen in part because it touches on all these other fields, and because it also more starkly contrasts with more traditional approaches and thus emphasizes how the scientific foundations are shifting beneath biomedical research.

This chapter begins with a description of drugs and drug receptors. Next it considers proteins, the workhorses of biology that form structural elements within cells and that mediate biochemical reactions. The next section deals with DNA, the chemical basis of inheritance.

The study of DNA and proteins marks the advance of molecular biology. Pharmaceutical research increasingly relies on molecular biology, and it is here that science most directly joins drug discovery. Protein pharmaceuticals constitute an important new therapeutic class. Several protein drugs have had a dramatic impact on clinical practice in recent years, and many more are being prepared for the market. Novel protein drugs have, for example, constituted roughly half the new major introductions in the last few years; indeed the wealth of new prospects has raised concern that the regulatory pipeline may be overwhelmed by the flow of new biotechnology products (146,341). The direct use of DNA as a treatment constitutes the next logical step. This is currently a frontier of biomedical research, and this section is necessarily more speculative, and illustrates the uncertainty regarding a new treatment modality. A final section draws lessons from the various case histories, focusing on those elements most likely to influence the future costs of drug discovery.

DRUGS AND RECEPTORS

The notion of specific drug action is most often described by a lock-and- key analogy. The drug is seen as the key that acts specifically on a narrow range of locks. The lock is a drug receptor. A drug is a substance that causes a physiological response; the receptor is the molecule through which it works.

Paul Ehrlich formulated the idea of drug-receptor interactions in the first decade of this century. A.V. Hill modified the interaction as the binding of a drug to its receptor in mathematical terms early in this century. This was embellished in the late 1920s by A.J. Clark who forwarded the idea of a drug "occupying" a receptor site, competing with natural agents that bound the receptor at the same site. Drugs were thereafter termed "agonists," those that caused the normal response (simulating the body's own action), or "antagonists, ' those that blocked normal action (105).

The model, in its simplest form, is that the drug binds to a receptor. The receptor has a binding domain that is specific for the drug, typically the same site sought by compounds normally found in the cell. Binding causes the receptor to respond,

causing a cellular change. The receptor may open a gate for the influx or efflux of `charged` molecules through the cell membrane, for example, or it may catalyze a biochemical reaction. Many receptors act by causing the formation of ''second messengers, ' compounds that provoke cellular events such as the secretion of insulin or the synthesis of specific proteins. The site causing a cellular change is separate from, but signaled by, the binding site.

Many enzymes are drug receptors. Enzymes are proteins that mediate biochemical reactions.

They build cellular structures, degrade cellular byproducts, metabolize sugars, and perform the myriad chemical tasks that transpire in cells. Many drugs discovered in the 1970s and 1980s originated from understanding a metabolic pathway, finding the enzyme associated with a particular reaction, and studying how to inhibit its activity. The drug captopril, for example, was developed by tracing the biochemical pathway to produce angiotensin, a short protein (a polypeptide) that raised blood pressure. By slowing production of angiotensin, captopril helped treat high blood pressure (see box 5-A) (300).

Box 5-A--The Discovery of Captopril

Hypertension, or high blood pressure, is a highly prevalent condition. It predisposes to heart attack and stroke, and is a major cause of death and disability. A "hypertensive principle' was discovered in the kidney late in the last century. A theoretical model of hypertension was elaborated in the 1930s and 1940s. In this model, the kidney was a central actor. Blood volume and the resistance to blood flow, caused by contraction of smooth muscles in small arteries, were the main determinants of blood pressure. The kidney controlled blood volume by regulating how much water was retained in the body, in turn determined by how much salt was excreted.

The kidney also produced substances that directly caused contraction of smooth muscles in small arteries. These findings led to several treatment strategies. One long-standing approach entailed diuretics to reduce blood volume. Another involved relaxing smooth muscles in blood vessel walls by blocking the action of epinephrine and other chemicals present in blood. Drug researchers at Squibb discovered a wholly new class of antihypertensive agents through a logical set of steps that began early in the 1960s with the elucidation of one physiological pathway leading to high blood pressure. This line of research culminated in the discovery of captopril. The process was a prototype of rational drug design.

The physiological pathway began when a short protein, or peptide, angiotensin II was formed in the blood stream. Angiotensin II stimulatedthe release of a hormone, aldosterone, from the adrenal gland, which in turn caused the kidney to retain sodium. With more sodium in the blood, the body retained more water. Angiotensin also directly stimulated contraction of smooth muscles in blood vessels. By blocking the fen-nation of angiotensin, one could pursue both paths to antihypertensive therapy in a single stroke.

Angiotensin II, the most active natural agent, consists of a short peptide--a chain of eight amino-acids. These 8 are cleaved from angiotensin I, which has 10 amino-acids. The enzyme that cleaves angiotensin I to II is Angiotensin Converting Enzyme (ACE). The path to captopril started by trying to inhibit the action of ACE.

Researchers at Squibb took several approaches. They started from the natural angiotensin I molecule and fashioned molecules that might have similar shapes but were more difficult to cleave. They also worked with a snake venom that inhibited ACE. The active components in the venom included several peptides, including one chain of 9 amino acids and another of 5 amino acids. These natural peptides were ACE inhibitors, but they had to be injected. The effort centered on finding an agent that could be administered

---

Box 5-A–The Discovery of Captopril--(Continued)

orally, and would find its way into the bloodstream after surviving the brutal transit through stomach and intestines. The team screened a myriad of chemical compounds similar to the natural precursor and to the snake venom. They devised modifications of the venom peptides and the natural precursor. They altered the backbone of the five amino-acid venom peptide thinking that it might retain activity while resisting degradation. Of 2,000 compounds screened, only a few inhibited ACE and only one was specific to it. That one was a metal-binding molecule that promised to be toxic. Another approach was more successful.

The Squibb team also worked back from their conception of the receptor's shape, based on their knowledge of details of the cleavage reaction it performed. Since the ACE protein had not itself been structurally defined, they developed a hypothetical model based on the active site of another enzyme that performed a similar cleavage reaction, and whose structure was known from x-ray crystallography. The research team crafted compounds to fit into the hypothetical active site of ACE and discovered a compound with inhibitory activity that was similar in shape to two amino-acids in tandem. They performed further chemical modifications of this molecule and eventually found a chemical amenable to oral administration that was 1,000 times more potent than their initial "lead" molecule. The synthetic drug had greater activity than the nine-amino-acid venom peptide.

This new agent was named captopril. Squibb submitted a new drug application for captopril in 1979. Two years later, captopril was approved by the U.S. Food and Drug Administration for marketing under the brand-name of "Capoten." Captopril soon became a standard drug to treat hypertension, and also found clinical use in combating heart failure and other cardiovascular conditions. Its sales in 1988 exceeded $1.1 billion, joining only three other compounds with sales over $1 billion in a year. It also became the starting point for a round of new antihypertensive agents pursued by Squibb and other firms.

SOURCES: M.A. Ondetti, D.W. Cushman, and B. Rubin, "Captoril," *Chronicles of Drug Discovery,* J.S. Bindra and D. Ledniger (eds.)(New York: Wiley, 1983). M.R. Ziai and B. Beer, "Making Business Sense of Science with Rational Drug Design," *Pharmaceutical Executive* 10:40-46, October 1990.

▋Identifying a Receptor Expedites Drug Design

The first drugs were typically found by looking for chemicals that caused a particular clinical reaction. Extracts from the foxglove plant were long used to treat "dropsy" (congestive heart failure) and other ills. Digitalis was discovered by purifying chemicals out of those extracts and looking for the compounds that produced clinical improvement. Aspirin (acetylsalicylate) was found by testing chemical modifications of salicylate, a traditional remedy for fever and pain. Although aspirin had been synthesized in 1853, its clinical effects were not discovered until 40 years later. It has become one of the most commonly used drugs since the firm Bayer began making it on a large scale in the late 1890s (396). Screening compounds chemically similar to salicylate was an early example of a strategy that prevailed for many years, starting from chance observations of clinical effects of a known compound, with little or no knowledge of mechanism. This approach was refined until the synthesis of new compounds became a high art and screening tests a centerpiece of pharmacology. One pharmaceutical researcher described drug research in the early 1960s thus:

> There was a prevalent attitude in many places that the conduct of research should be of the man-and-a-boy type in which chemists would create molecules by the pound and send them to pharmacologists who would screen them for their activity in the hopes that luck would strike. It's what I characterize as 'research untouched by the human brain' (394).

Photo credit: NATIONAL CANCER INSTITUTE

The first drugs were typically found by extracting chemicals from natural products and screening them for pharmaceutical activity. Though other methods of drug discovery have evolved, extraction remains an important ingredient of drug discovery even today.

Finding new drugs was not ''irrational, ' but the rationale was not based in chemistry or knowledge of biological function. The purification and screening of natural products for antibiotics, for example, followed a reasonable strategy, but did not rest on knowledge of how the drugs worked. As attention turned to chronic diseases such as atherosclerosis, cancer, and neurological and psychiatric conditions, the development of drugs depended more on knowledge of the disease process. Rational drug design reversed the traditional process, working backward from a known receptor target. Rather than screening agents to serve specific functions, the molecular mechanism underlying a biological function was used to direct a search strategy.

Rational drug design is a general term that covers a broad range of approaches, but the underlying theme is a reliance on structural analysis of target molecules and deliberate design of agents to affect their function. One of the prototype successes was the development of cimetidine to treat peptic ulcer disease (see box 5-B). In this case, the first critical step was to define a class of receptors, The next step was to search for a compound that could at least partially block the action of histamine on these receptors. Once an initial "lead" antagonist compound was found, then chemical modifications of nearby atoms led to more potent and less toxic new compounds that could be tested for clinical effect. One of these proved effective in blocking acid secretion, dramatically tipping the scales in favor of medical management, as opposed to surgery, for most cases of peptic ulcer disease (140,394). Characterizing the H-2 receptor was a critical first step.

■ Structure-Activity Relationships

Having found a receptor, and compounds that act on it, the next step is to search for more potent chemical analogs, drugs that have a similar effect at lower concentrations. The standard way to do this is to synthesize compounds chemically similar to the ' 'lead compound, ' altering one or a few key sites on the molecule by adding or taking away chemical groups, or by deforming its shape. Newly synthesized compounds are then screened for drug effect. The underlying premise is that the structure of the drug affects its activity. The process of chemical modification and searching for functional effects is called structure-activity relationships, or SAR (190).

In recent years, quantitative methods augmented SAR, and earned the name quantitative SAR, or QSAR. The refinements grew in part

---

Box 5-B-The Discovery of Cimetidine

In 1964, histamine was well studied for its role in allergic reactions, whose effects could be partially blocked by a group of antihistamine compounds still used in many over-the-counter cold remedies. James Black worked at the British pharmaceutical and chemical firm ICI, where he focused on drugs that affect the response to epinephrine (adrenaline), itself a large class of drugs now used to treat high blood pressure, disturbances of heart rhythms, diabetes, and many other conditions. The target molecule, the beta-receptor molecule, responded to epinephrine specifically, although its function varied in heart cells, blood vessels, and pancreatic cells, and other tissues.

Black moved to Welwyn laboratories, the British research arm of Smith, Kline& French. He worked to show the existence of a different class of histamine receptors, dubbed H-2 receptors, that caused stomach acid secretion. The task began by refining ways to measure H-2 receptor effects so that compounds could be screened rapidly. Initial tests were insufficiently sensitive; only improved screening procedures and higher doses of histamine enabled work to progress. Once the screening tests were in place, the search focused on finding antagonists, compounds that could block the action of histamine. The compound buriamide blocked H-2 activity.

In the late 1960s, administrators at Welwyn were ordered to stop working on the project to block acid secretion, as it duplicated work going on near corporate headquarters in Philadelphia. The British group persisted by finding a new name for their project, renamed the H-2 receptor program. Research administrator William Duncan resorted to subterfuge, adopting an "arm's length, isolationist policy in relation to headquarters R&D. " Early in 1968, the new president of the company, Thomas Rauch, directly ordered that the project be preserved from drastic budget cuts "one more year," just long enough for success to squeeze through the door. He later reflected, "It's terrifying to think about it today. "

In 1972, Black and other colleagues at Smith Kline& French published their data demonstrating the existence of H-2 receptors. They then searched for more potent antagonists, synthesizing and screening chemical modifications of buriamide, their 'lead' compound. This led to metiamide and then to cimetidine. Metiamide was used in initial clinical trials, and showed promising results. In two patients, however, it suppressed production of neutrophils, white blood cells involved in inflammation. Cimetidine was already in early clinical testing, and work on it intensified. Cimetidine was approved for the market in November 1976, in the United Kingdom, and in August 1977 in the United States. Within a year, it was distributed to 90 countries, becoming king of the "blockbuster" drugs of the 1970s, and revolutionizing the treatment of peptic ulcer disease. Smith Kline & French's moribund stepchild became a robust prince.

SOURCES: C.R. Ganellin, ''Cimetidine,'' *Chronicles of Drug Discovery,* J.S. Bindra and D. Ledniger (eds.) (New York, NY: Wiley, 1982). SmithKline & French International, *The Discovery of Histamine H2-Receptors and Their Antagonists* (Philadelphia, PA: SmithKline & French, 1982).

---

from the increased power and reduced cost of computers. Faster and better computers encouraged use of more complex but more accurate computational methods to predict the three-dimensional structure of small molecules, to display the shape of receptors and ligands using computer graphics, and to study the structure of large molecules including drug receptors. Quantitative methods also enabled simulations of chemical synthesis pathways, the design of agents that might fit into receptor sites, and high-resolution analysis of protein shape—all inputs to the QSAR process.

When multiple agents all work on the same receptor, their common features can serve as a core for seeking new agents to perform the same function. The process of defining entirely new drug agents, however, has no assured computational solution. A generic approach involving structural chemistry and biochemistry has emerged

as a usually successful strategy. The process is far from a simple matter of defining the structure of a receptor, drawing a chemical that fits into it, and testing the result in a laboratory. Indeed, the process has become more complex rather than simpler. As one practitioner of QSAR noted:

> The incredibly rapid advances in biochemistry, molecular biology, theoretical chemistry, and computers along with the accumulated experience of the past 100 years must give drug research directors sleepless nights (179).

Indeed, even the technical advances are far from providing a cookbook to produce new drugs. The process remains too qualitative for precise predictions, and is better at finding agents similar to those discovered by others than in producing real innovation in the form of entirely new classes of agents. At a 1984 symposium called ''Drug Design: Fact or Fantasy, '' G. Jones noted:

> The critical area of pharmaceutical innovation, the de novo lead generation, is unfortunately the area in which fantasy remains preponderant (217).
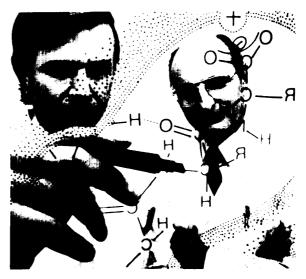
▊ Frontiers of Rational Drug Design

Discovering drugs by design is a relatively new phenomenon, but it dominates the pursuit of new agents today. The logical strategy used to develop captopril is held up as a prototype of the new approach. The approach in its most refined form can be applied only to enzymes, however, and then only when the chemical reaction is well understood. The structure of the enzyme itself, or that of some enzyme with similar function, must first be determined by x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy (explained below).

Drug research is pushing back the frontiers of rational drug design in several directions. One direction is the structural analysis of receptor molecules. Most receptors are large proteins with multiple regions of interest. Determining their shape can be a first step to homing in on parts of the molecule that bind prospective drugs. Until recent years, the only hope of defining the shape of a molecule was x-ray crystallography. Crystal-lography reconstructs the shape of molecules by analyzing how they deflect x-rays. This entails shooting x-rays through a crystal of a substance, and analyzing where they come out, in a process much like determining the shape of a sculpture by shooting balls off it and backing where they bounce. One limitation of the technique is the necessity for a crystal.

Crystallizing large molecules, such as proteins, is a difficult art, and many molecules have proven refractory to crystallography for this reason. Even for the molecules that can be crystallized, the analysis of x-ray deflection patterns requires massive calculations. Defining a crystal structure can take years. The development of very high-intensity synchrotrons radiation has given x-ray methods a new boost, with a particularly high-technology twist (183).

In recent years, a new technique of NMR has been applied to increasingly large molecules, permitting spatial resolution of small molecules, short peptides (small proteins), and regions of macromolecules (58,82). NMR spectroscopy reconstructs shape by analyzing the effects of very

Photo credit: BRISTOL-MYERS SQUIBB COMPANY

Dr. David W. Cushman (left) and Dr. Miguel Ondetti (right) developed captopril using a new approach to drug discovery. The captopril team was able to design a molecule with the structure needed to block the active site of the converting angiotensin enzyme (see box 5-A).

high magnetic fields on the nuclei of atoms in the molecule.

Both NMR and x-ray crystallography require large investments in analytical instruments and computers to analyze the data, entailing multimillion dollar investments just to analyze the structure of potential drug receptors. The justification for such investments, made by many firms in recent years, is that defining the three-dimensional structure of receptor sites will improve the prospects for developing drugs that fit into those sites. The wisdom of the investments will only become clear in several years, if this approach bears fruit.

A second frontier is the analysis of molecules that are joined to proteins to give them distinctive shape, adhesive properties, or other attributes. Chains of sugars, called polysaccharides, project out from the surface of many proteins, for example, acting as antennae for chemical signals or providing structural stability on the outer surface of the cell. These polysaccharides often confer specificity to the receptor. The process of altering sugars, or attaching them to proteins, provides a possible mechanism to affect receptor function, but the chemistry and the biochemical pathways are far from being fully described, and the process of describing them is tedious and difficult.

Progress is nonetheless forthcoming. Modifying polysaccharide structure is a promising avenue to treating inflammation, autoimmune disease, and other disorders. It seems especially promising for diseases such as arthritis, diabetes, multiple sclerosis, and others that involve an inflammatory response that induces tissue destruction by the bodies' own defense mechanisms. As one of many possible examples, the molecule that attracts neutrophils (common white blood cells involved early in an inflammatory response to injury and some tumors) was recently shown to be a sugar chain (329,506). Drugs to control the production of the sugar chain, or to break it down, might temper the destruction by neutrophils, suggesting a path for drug development.

Cell surface molecules are the points at which cell-to-cell communication takes place. The process starts with binding of a molecule to the surface receptor. This entails specific binding, or recognition, that transmits a signal to alter cellular processes. The binding of a hormone from the bloodstream triggers the proliferation of cells, or the binding of a neurotransmitter causes an electrical signal to propagate along a nerve cell. A muscle contracts in response to a transmitter, or the immune system primes itself to fight a nascent infection. Viruses and other infectious agents also take advantage of cell surface receptors. The rabies virus homes in on nerve cells by binding to specific neurotransmitter receptors, for example, and the human immunodeficiency virus attacks cells bearing CD4, a molecule that projects from the surface of certain white blood cells. Cell surface receptors thus mediate many of the most physiologically important functions in the body and are extremely promising targets for future drug development. One serious problem is that there are so many cell surface receptors, and thus too many paths to follow for scientists to understand the structure and function of each.

Another major obstacle is that the methods of predicting three-dimensional structures are simply not up to the task for large molecules. New techniques to deduce the structure of large molecules are powerful but slow and not always successful. The ultimate solution would be predictive tools of sufficient power to predict shape from knowing the order of the building blocks, the amino acids in proteins and nucleotides in DNA. How long strings of amino acids emerge into shapely proteins is termed the ''folding problem, ' and constitutes one of the most vexing and important fields in structural biochemistry today. Computational algorithms and theoretical chemistry are not yet powerful enough to make such robust predictions. All is not lost, however, as the structures determined for one protein can serve as a first approximation of the shape of another with similar function, as in the development of captopril described in box 5-A.

Cell surface receptors are clearly of great importance in understanding the function of the nervous system, the endocrine system, the immune system, cell proliferation, and targets for infection. The major chronic diseases such as cancer, diabetes, heart disease, arthritis, autoimmune disease, necrologic and psychiatric disorders, and endocrine dysfunction involve interactions with cell surface receptors. Understanding the structure and function of surface receptors has grown into one of the great thrusts of pharmaceutical R&D, pursued in the belief it will uncover as yet unknown approaches to treatment of diseases hitherto intractable.

▌ Organizing to Discover New Drugs

The wealth of new knowledge derived from basic biomedical research and the new power of rational drug design recast the drug discovery process, but there is no prototypical process that can be simply described. Pharmaceutical firms track research by talking to scientists and reading the scientific literature. They maintain in-house research teams, often doing research parallel to that performed in academic centers. New ideas may originate in corporate research efforts or in the academic ones; examples of both abound. This industrial research base is a source of new leads, and the stalking grounds for corporate ''champions' of new drug ideas. Each firm has a somewhat different process to decide which leads to pursue. Some organize according to treatment category, assembling teams to focus on finding drugs to treat a disease or organ system; for example, entire firms are dedicated to research on cardiovascular drugs. Some firms make research-targeting decisions in committees, others delegate great authority to research directors who informally circulate in the firm and among academic groups, and still others have strong central direction. All attempt to manage innovation by balancing an endless supply of possible research directions against a need to produce salable products.

PROTEIN ANALYSIS AND PROTEINS AS PHARMACEUTICAL AGENTS

▌ Essentials of Protein Structure and Function

Proteins form structural elements in cells and perform a wealth of functions, including the catalysis of most biochemical reactions. Proteins are composed of linear chains of amino acids. There are 20 common amino acids in cells, each with a common chemical unit called the peptide group. The peptide group enables each amino acid to form a chemical bond with any other. Each amino acid in a protein serves as a chemical module: the peptide units confer structural stability and link the modules together, while chemical groups attached to the peptide backbone confer the distinctive structural and fictional properties. The 20 amino acids have different chemical groups attached to the peptide core. The amino acid proline, for example, introduces turns into the protein backbone, while cysteine can form bridges with other cysteines located on different protein chains or remote parts of the same chain. Some amino acids prefer lipid (fatty) environments, providing stability in cell membranes, while others are highly soluble in water. Some are acidic and others basic.

Some proteins can be transported outside cells to lay down a matrix, such as the collagen that constitutes the bulk of bone and cartilage. Others act as enzymes to carry out the biochemical reactions taking place in cells at every moment. The enormous diversity that is possible when hundreds of amino acids are strung together produces a correspondingly large range of functions in the proteins they form, despite the simplicity of the constituent amino acid components.

▌ Proteins as Receptors

The vast majority of drug receptors in the body are proteins. Many are enzymes or cell surface receptors. Proteins are involved in virtually every cellular process and are components in most

cellular structures. Protein chemistry is thus a dominant theme in current drug research.

Proteins have been a major preoccupation of biochemistry and molecular biology since the inception of those scientific fields. There are several ways to study proteins. Analysis of three-dimensional structure has been briefly described above. Biochemists have devised many ways to study enzyme function. One classic technique is to isolate an enzyme to near purity, then to attempt to study the chemical reaction it mediates in great detail. This is a key strategy in dissecting the biochemical pathways of energy metabolism, biosynthesis, and degradation. Another technique is made possible by modem molecular genetics.

The order of amino acids in a protein chain is specified by the genetic code laid down in DNA. DNA is composed of very long chains of four chemical bases linked through a sugar-phosphate-backbone. The information from DNA is "transcribed' and spliced in the cell's nucleus into a similar polymer, but in a somewhat less stable form called ribonucleic acid (RNA). RNA is transported from the nucleus to the cytoplasm, where it is translated to amino acids. The order of bases in DNA specifies the order of bases in RNA which in turn specifies the order of amino acids in the protein. A gene consists of a stretch of DNA that produces a functional product, either RNA alone or RNA that is subsequently translated into protein. The path from code to product thus typically involves three major steps: DNA, RNA and protein, with the possibility of modifications at each step. The linear order of DNA bases in a gene, modified by splicing out stretches of RNA and adding caps and tails to the message, thus directly determines the order of amino acids in the corresponding protein, In its simplest formulation, the cell translates a linear DNA code into a linear string of amino acids in proteins. The diverse shapes and chemical constituents that result from the chain of amino acids become proteins that form the structural supports and perform the biological functions for cells and tissues.

Through recombinant DNA technology, genes that produce a specific protein can be spliced into bacterial DNA. Large amounts of the protein can be produced in the bacteria, yielding enough DNA to precisely define the DNA sequence for the expressed part of the gene, from which the amino acid sequence for the resulting protein can be derived. Producing large amounts of the protein in bacterial cells is a boon to prospects of crystallizing proteins for crystallographic analysis; it also produces ample supplies for further biochemical analysis. One disadvantage of the process is that bacterial cells may not process the protein in exactly the same way as the cells in which the gene normally resides. Polysaccharides may not be added appropriately, for example, or the protein may fold somewhat differently because processing enzymes are not present in bacteria. The strategy nonetheless works for many proteins.

Manipulating the DNA code alters the order of amino acids in a protein. Scientists can exploit this effect by introducing a gene into bacteria or yeast by recombinant DNA with exquisite precision. They may thus introduce changes (mutations) in the native gene into the whole animal, where the mutation's effect can be observed. By introducing such mutations at specified sites, one amino acid can substitute for another at specific points in the protein peptide chain. To study the binding site of a receptor molecule, for example, the amino acids at that site can be replaced by those with similar, or with vastly different, chemical properties to assess the impact on ligand binding. The active sites for enzyme activity can be similarly studied by targeting just those amino acids thought to be important.

▮ Proteins as Therapeutic Agents

Proteins also serve as chemical signals in the body. At the level of organ systems, proteins regulate immune responses, cell growth cycles, hormone responses, and many other functions. Blood proteins are involved in coagulation and dissolution of blood clots. Several hormones are short proteins, or polypeptides. Many polypep-

tides appear to be involved in modulating diges-
tion, regulation of blood pressure, and other
functions involved in many normal metabolic and
disease processes. These proteins involved in
cellular communication are prominent targets as
drug receptors in many cases, but they can also be
therapeutic agents in their own right. Proteins are
thus of great interest to drug researchers, not only
as drug receptors but also as drugs themselves.

As the tools to study proteins and protein-
protein interactions have advanced, the impor-
tance of proteins and polypeptides has become
increasingly clear. The production of large quan-
tities of proteins became feasible with the advent
of recombinant DNA techniques. Many compa-
nies were founded to exploit the potential of
biotechnology and most established pharmaceuti-
cal firms have many projects underway to de-

velop and market protein drugs. Michael Venuti
of Genentech lists over 100 peptide products
under development in 1991 (498). The first Food
and Drug Administration approval to market a
protein derived from recombinant DNA was
granted on October 29, 1982, for human insulin
(see box 5-C) (171). In 1985, the second approval
was granted for human growth hormone, fol-
lowed by various interferon beginning in 1986,
tissue plasminogen activator in 1987, and erythro-
poietin in 1989. A recombinant DNA-derived
hepatitis B vaccine was approved in 1986, and for
Hemophilus B in 1988. A large number of protein
pharmaceuticals have been recently approved, or
are under active review (146,148).

Many of the products already on the market,
and many yet to be approved, perform functions
not achieved by other drugs. They represent

Box 5-C--Insulin

Charles Edouard Brown-Sequard postulated the existence of circulating factors whose absence caused
disease in the late 1880s. William Bayliss and Ernest Starling worked on pancreatic secretions during the
first decade of this century. In 1905, Starling proposed the term "hormone" for substances secreted into the
blood by one organ to produce a response in another. In the period 1889 to 1920, many groups, most of them
European, established the connection between cells in the pancreas, in the Islets of Langerhans, and the
clinical presence of diabetes. The seminal work leading to the discovery of insulin began under Frederick
Banting and his colleagues at the University of Toronto in May 1921.

Banting worked with physiologist John Macleod and Macleod's student Charles Best to tie off
pancreatic ducts in dogs, to remove the pancreas, and to treat the resultant diabetes with pancreatic extracts.
They quickly found a substance they called "isletin," because it was derived from cells in the Islets of
Langerhans. The name was later changed to insulin. James Collip, a biochemist, joined the team to improve
the extraction procedures, improving the potency and consistency of isletin. The University of Toronto's
Connaught laboratories were brought in to scale up production of the substance, which met with strong
clinical demand soon after the promising initial results were known. G.H.A. Clowes of Eli Lilly & Co. of
Indianapolis then got involved. Lilly was already producing other glandular extracts, and Clowes had an
excellent scientific reputation. Clowes met with Banting and Best after they presented preliminary results
at a meeting in December 1921, and expressed interest in their work. He offered Lilly's services to scale up
production.

In May 1922, the University of Toronto filed a patent on insulin for no other reason "than to prevent
the taking out of a patent by other persons. When the details of the method of preparation are published
anyone would be free to prepare the extract, but no one could secure a profitable monopoly. " Soon
thereafter, the University of Toronto and Lilly began a collaboration, sometimes rocky, to produce insulin
for wider clinical use. George Walden, a chemist at Lilly, developed novel production methods that
simplified production and further improved lot-to-lot consistency by the Fall of 1922. Lilly began to market

---

Box 5-C–Insulin--(Continued)

its insulin as "Iletin,'" hearkening back to the original Banting and Best coinage, but with the dropping of a silent "s."

Insulin remains the main treatment for the most severe form of diabetes. Indeed, in current terminology this form is called insulin-dependent diabetes mellitus. Lilly remains the largest U.S. producer of insulin, and shares the U.S. market with Squibb. Outside North America, Novo of Copenhagen is the dominant producer. But the insulin story did not end with the production of insulin extracts. Insulin also became the target for assaults by molecular biology.

The importance of insulin made it the focus of Frederick Sanger's work in structural chemistry. Sanger applied chemist Emil Fisher's philosophy, using chemical principles to explicate protein structure and function. Sanger eventually established the order of amino acids that made up insulin, the first protein for which the amino acid sequence was determined. Knowing the sequence enabled comparison between insulins from different species, and discovery of the slight differences between them.

Recombinant DNA ushered in new hope for production of peptide hormones. As soon as recombinant DNA was discovered in the mid-1970s, the insulin gene became an early target. The gene for insulin was the first ever cloned from a mammal. Work on insulin emerged as the focus for product development at Biogen and Genentech, two prominent new biotechnology companies established during the late 1970s. The human insulin gene was introduced into bacteria by recombinant DNA, and the bacteria produced insulin. The City of Hope Medical Center synthesized DNA sequences for the gene that Genentech used to produce insulin in bacteria, Genentech licensed insulin produced through recombinant DNA to Lilly for large-scale manufacturing, marketing, and regulatory approval. In 1980, insulin was the first recombinant DNA product tested in humans. In October 1982, Humulin, trade name of the human insulin marketed by Lilly, became the first recombinant DNA drug approved for marketing.

SOURCES: S.S. Hall, Invisible *Frontiers: The Race to Synthesize a Human Gene* (New York, NY: Atlantic Monthly Press, 1987). F. Sanger, "Sequences, Sequences, and Sequences,' *Annual Reviews of Biochemistry 57: 1-28, 1988.* J.P. Swann, *Academic Scientists and the Pharmaceutical Industry: Cooperative Research in Twentieth Century America* (Baltimore, MD: The Johns Hopkins University Press, 1988). U.S. Congress, Office of Technology Assessment, *Commercial Biotechnology: An International Analysis* (OTA-BA-218) (Washington, D.C.: U.S. Government Printing Office, 1984).

---

entirely new classes of therapeutic agents. Erythropoietin, for example, stimulates the replenishment of red blood cells. It was approved in 1989 to treat the anemia that attends long-term renal dialysis (451). In the future, it could be used for many other purposes as well, and many are under investigation. A series of growth factors stimulates the proliferation of different kinds of white blood cells, and could be used to replenish them after cancer chemotherapy or in response to adverse drug reactions. A class of blood enzymes dissolve blood clots after heart attacks. Interferon and interleukins are promising in the treatment of cancer and infections, particularly viral infections for which there are relatively few effective treatments.

Proteins pose several problems as drugs, however. The necessity for parenteral, as opposed to oral, drug delivery is foremost among the limitations. Parenteral administration refers to techniques that break the skin surface, by injecting into veins, into muscles, under the skin, or elsewhere. These are more invasive and require sterile preparations. Proteins are readily degraded in the stomach and intestines, so that simple oral administration is impractical-the agent is destroyed before it can be absorbed. Insulin, the prototype of protein drugs, is still injected or administered by pumps installed in the body. It has been used as a mainstay of diabetes treatment since the 1920s, and has never been replaced by a smaller molecule. Decades of research have

failed to uncover an oral formulation as effective as injectable insulin. The need for parenteral administration is nonetheless an immense impediment to widespread use of a protein drug.

The problems with parenteral administration include the pain of injections, immune reactions against the agent, local inflammation and scarring from repeated injections, the need for sterile preparations, and strong resistance from patients, Parenteral administration includes intravenous injection, intramuscular injection, subcutaneous injection, implantable pumps, and slow- release formulations placed just under the skin. Research along each of these avenues is proceeding apace. Research to introduce drugs through less invasive routes is also moving forward.

Coating peptide drugs with a shell to prevent digestion in the stomach, adding carrier molecules to permit rapid absorption, or formulating drugs with surfactants to promote absorption are all being pursued. Some peptide drugs can already be administered as suspensions for nasal inhalation for absorption by the rich network of blood vessels that can transport agents directly into the brain. Yet another line of research attempts to redesign peptide drugs into small organic molecules that can be administered orally rather than parenterally, making them much more attractive for general use.

For some applications, parenteral administration is not a major obstacle. Clot-dissolving agents, for example, are used mainly in the period immediately after a heart attack. The patient is typically in the hospital receiving intravenous fluids in any event, so that intravenous delivery is not a problem. This same feature will be shared by treatments for anemia from dialysis, cancer chemotherapy, and a few other uses. The clinical benefits of parenteral drugs must be significantly higher than oral drugs to overcome the inconvenience of parenteral administration for long periods.

Many potential drugs must be highly targeted to be useful. Highly potent cellular poisons, for example, would be useful if they could be induced to attack only cancer cells. One novel treatment

approach is to affix poisons such as ricin, tumor necrosis factor, or diphtheria toxin to proteins recognized primarily by cancer cells. The recognition function can be served by antibodies raised against the target cells, or ligands that bind specifically to surface receptors on target cells. Antibodies have also been studied experimentally to coat lipid sacs containing anticancer drugs, as a means of delivering the drugs specifically to regions containing cancer cells (428). These developments are novel uses of proteins as means to target specific cells, but they remain limited by the range of specific recognition molecules, the need for parenteral administration, and a relatively narrow range of therapeutic applications.

Pharmaceutical research scientists are divided about the future significance of proteins as therapeutic agents. In interviews, they formed two camps with differing visions of the future of drug therapy. One camp believes that the problems of drug delivery will diminish over time as technological improvements overcome technical obstacles. They note that many of the promising protein agents have no small molecular counterparts-no small organic compounds have been found to serve the same function, Insulin is cited as a prototype. It is still used as a drug almost 70 years after its discovery (see box 5-C). The clot-dissolving enzymes, growth factors, and hormones approved for use in recent years are cited as other examples of new life-saving protein drugs that flowed straight out of molecular biological research. This group believes that many of the new proteins are so entirely novel in their action that small molecules may never be found to replace most of them, and at least for the foreseeable future proteins alone will be available. This group believes further that small biotechnology companies can grow into pharmaceutical giants through their successes in manufacturing protein drugs. They see protein pharmaceuticals as the technology that mediates a transformation of drug therapy and the pharmaceutical industry. Firms that fail to move aggressively toward protein drugs will lose out on many

Photocredit: *ELI LILLY AND COMPANY*

Protein drugs accounted for over half of the major new drug introductions in the last few years. Insulin, the prototype for this class, remains in use 70 years after its discovery. Insulin was introduced to the medical profession in December 1922 in a *Journal of the American Medical Association* article illustrated with this photo of a mother holding her 3-year-old diabetic son before the child was treated with insulin. The photo on the right shows the same boy after treatment with insulin.

of the major drug innovations over the next few decades.

The other camp points to the immense difficulties of marketing drugs that must be administered parenterally. They acknowledge the importance of the new protein growth factors, hormones, and blood products, but believe that the markets for many protein drugs will be relatively small, seldom achieving the ''blockbuster' status needed to sustain a pharmaceutical firm over the long run. The confinement to narrow therapeutic niches and a need for high prices because of a limited number of doses will limit the impact of protein drugs. This group maintains that protein drugs will be extremely important in a few therapeutic categories such as cancer or organ transportation, but will rarely be commonly used outside the

hospital setting. In **cases** where proteins are the only available means to achieve a treatment goal, companies will concentrate their efforts to find a small organic molecule with the same function, developing drugs that do not require parenteral administration. This camp cites captopril (box 5-A) as the prototype. Here, a peptide drug was replaced by a small molecule that could be orally administered. While this model is only broadly possible today for enzyme drug receptors, the same principles could prove applicable to the full range of peptide drugs. Morphine itself appears to be a nonpeptide mimic of enkephalin, a naturally occurring peptide that modulates pain perception. Drug firms have recently discovered nonpeptide blocking agents for several other natural peptide hormones, such as cholecystokinin (which regu-

lates function in the gastrointestinal tract and gall bladder), substance P (involved in pain sensation), and others. Drug firms may indeed pursue protein drugs, but mainly as stepping stones toward small molecules. Proteins will often serve merely as a research way station *en route* to more widely marketable drugs. These observers see the future pharmaceutical industry as an incremental extension of today's, with many of the same firms continuing to dominate the pharmaceutical market by incorporating new biological technologies. Biology will provide insights, but organic chemistry and the production of small molecules to substitute for the function of larger proteins will become the norm.

Disagreements about the importance of proteins as therapeutic agents does not extend to research. Here, all the experts agree that studying proteins is now central to developing new drugs. Rather, the disagreement among researchers centers on which problem will be solved first— finding small molecules to do what proteins do or finding ways to formulate proteins drugs for easier drug delivery.

Most large firms are hedging their bets, making substantial research investments and pursuing exploratory projects toward protein drug products. Other fins, generally smaller biotechnology companies, are betting a large proportion of their companies' assets on the success of protein drugs. Both strategies are being carried out, so the question will have an empirical answer in the next decade. Small companies may grow large, large companies may engulf them before they get large, or some instances of each may occur. These differing strategies exemplify the difficulties of making resource allocation decisions in pharmaceutical R&D. Both strategies are reasonable, and their success depends critically on factors that cannot be predicted: how many patients suffering from which conditions can benefit from a growth factor or immune modulator? Is it too soon to invest directly in developing a drug agent, based on current research results? Will it be possible to find small molecules to replace protein drugs? Can problems of drug delivery be solved? "There

are so many expensive avenues to explore that even with budgets in hundreds of millions of dollars it is no easy task placing bets" (179).

GENETICS IN BIOMEDICAL RESEARCH

The study of DNA has been the second main thrust of molecular biology, paralleling analysis of protein structure and function. Proteins make up most cellular components; DNA contains the instruction set for when and how to make them. The other important function of DNA is to serve as the structural basis of inheritance.

▌ DNA as the Structural Basis of Inheritance

The DNA base pairs described earlier in this chapter are linked together in long chains. There are four nucleotide bases linked together in extremely long chains, coiled up and bound with proteins to form chromosomes. The information content of DNA is mainly contained in the order of these base pairs, the DNA sequence. DNA is the structure by which individual traits are transmitted from generation to generation. The linear code of four letters is analogous in some ways to the linear code of O's and 1's in computer software, which also instruct hardware to carry out functions and can be copied faithfully for transmission.

Classical genetics, the study of inheritance, dates back to the 19th century. William Bateson's original definition of genetics presumed genes to be the hereditary "elements" discovered by Gregor Mendel in 1865 (230,261,336). Genes were units of inheritance, transmitting specific traits. The discovery of the structure of DNA by Watson and Crick in 1953 (509), spawned a field called molecular genetics. Classical genetics and molecular genetics came together in the study of DNA function, but this required some wobble in the exact meaning of the term "gene.' The simple idea of a gene, an element of inheritance (in classical genetics) that coded for a protein had to be modified under assault from molecular genetics. Genes were not merely present or absent, but also subject to regulatory control.

Genes were expressed (i.e., produced a product) only at certain times, in certain tissues, and in certain amounts. Uncovering such complexities made it difficult to directly map inheritance, with its discrete particles of" inheritance, to the exquisitely complex processes governing expression of genes. The complexity of molecular genetics was reflected in the changing face of human genetics.

▌ Genetic Approaches to Disease

Human geneticist and physician Victor McKusick maintains a catalog of human genes, called *Mendelian Inheritance in Man.* The frost edition in 1966 listed 574 well-characterized traits, and 913 partially validated. ones; most were genetic diseases identified through inheritance patterns in human family pedigree studies. By the ninth edition in 1990, there were 2,656 well characterized traits and another 2,281 partially validated ones (266). This growth in knowledge of human genetics, already significant, understated the growth of knowledge about the contribution of genes to human disease. During this same period, human genetics moved from the backwater to the cutting edge of biomedical research. (Genetics in other organisms has been a central thrust of biology throughout the century, but its full application to humans awaited technologies developed in the 1970s and 1980s.) The genetic factors underlying the most common diseases, such as heart disease, cancer, Alzheimer's disease, diabetes, hypertension, and many others were becoming better delineated. Genetic approaches to understanding such illnesses emerged as a dominant research strategy. Entering the 1990s, genetics was poised to dissect complex diseases because of the growing power of genetic maps.

▌ The Importance of Genetic Maps

Geneticists have been constructing maps of chromosomes since 191.3, when Alfred Sturtevant found traits that were inherited together and inferred their corresponding genes were therefore located on the same chromosome. Geneticists labored for decades to construct maps of the chromosomes of fruit flies, yeast, bacteria, plants, mice, and other organisms (414). These efforts were greatly aided by controlled matings, an option obviously not available in humans. Constructing similar maps for humans relied on finding genes through indirect methods, a slow, uncertain, and tedious process. A global approach to genetic mapping in humans comparable to that enjoyed by geneticists for other organisms was greatly aided by the emergence of recombinant DNA techniques.

The first step toward isolating genetic factors for human traits is to find informative families. These are typically large families with a well-defined genetic character (trait), such as a genetic disease. The way in which the character is inherited provides a great deal of information about whether it stems from a single gene or many. Even single-gene defects vary in their inheritance pattern. A genetic disease may be dominant, expressed if an affected gene is inherited from either parent, or recessive, requiring affected genes from both parents. If affected fathers never transmit the character to sons, and women are only rarely affected, it is good evidence that the character is a recessive gene on the X chromosome (since males have only one copy of the X chromosome, inherited only from the mother). The inheritance pattern for multigenic diseases is even more complex. Yet another layer of complexity is added by genes that express themselves only in combination with environmental factors.

A genetic linkage map is the bridge from the study of inheritance in a family to establishing a gene's chromosomal location. Different individuals have, on average, over a million" DNA differences in their genomes (a genome is defined as a complete set of chromosomes, one of each pair). Most variations have no clinical significance, but they can be used as markers on the chromosomes. The idea of a genetic linkage map is to find chromosomal sites that frequently vary among individuals, and to verify where on the chromosomes these common variations originate. Once this is done, the inheritance of bits of

chromosomes can be tracked through families by tracing the fate of markers. The markers vary among individuals, so they are likely to differ between parents in each family. But the variations are present only at a specific site on the chromosomes. If markers from the same region of chromosome 7 are consistently inherited along with cystic fibrosis (CF) in different families, then this is strong statistical evidence the gene causing CF comes from that region (see box 5-D).

The way to construct a robust genetic linkage map in humans was first proposed in 1978 and published in 1980 (56). The first genetic linkage map of the human genome was published in 1987 (1 10). Other groups constructed genetic linkage maps of individual chromosomes, and the 1990s should see a major push to refine such maps sufficiently to find almost any gene, once family resources are good enough and the genetic character is well enough defined. (These are, however, significant limitations.)

Once the approximate chromosomal location of a gene is known, the next step is to obtain the DNA from that chromosomal region in hopes of finding the gene itself. A physical map is needed for this purpose. The order of genes or markers will always be the same between genetic linkage and physical maps, but the measure of distance is quite different. Genetic linkage maps measure how often markers stay together or separate during inheritance. This is a measure of the probability of being separated by genetic recombination, a DNA exchange process that occurs in the production of egg and sperm cells. The measure of distances in physical maps is the size of DNA fragments, ultimately translating to the number of DNA base pairs. The most useful physical map is a collection of cloned DNA fragments that span the chromosomal region in question and are arranged in order. With such a map, one can go from one end of the region to the other by picking out different clones of known orientation. Since the clones contain many copies of the DNA from that area, this allows direct study of DNA in search of a gene. Physical mapping was pioneered in viruses and bacteria. Groups

working on yeast (whose chromosomes are over 12 million base pairs in length) and the nematode Caenorhabditis elegans (100 million base pairs) scaled up to more complex organisms with larger genomes (99,299). Similar strategies are now being applied to individual human chromosomes (488).

∎ Ways to Find and Study Genes

Once a region of DNA thought to contain a gene is in hand, then the hunt begins for a gene amidst the long string of DNA base pairs. There are many indirect methods to select DNA regions likely to contain "candidate genes. ' All strategies ultimately entail extensive amounts of DNA sequencing from the region, and comparison of sequence differences among individuals.

In the hunt for the cystic fibrosis gene, for example, consistent sequence differences were found in affected children. The CF gene was located on chromosome 7 by genetic linkage in 1985 (425,504,518). Each parent of every CF child had to have 'normal, ' non-CF gene on one chromosome 7 and a CF mutation on the other, (If both copies were normal, parents would not have a CF child; if both were mutations, the parent would be affected.) Using genetic markers, scientists could trace which copy of chromosome 7 was passed on to the CF child from each parent, and this had to be the one with the CF mutation. Comparing ''normal' to ''mutation' sequences in DNA from this part of chromosome 7, taken from many affected children and their parents, revealed a consistent abnormality, the loss of three base pairs (229,351,353). This constituted the most common mutation causing CF. Once a small piece of a gene was identified, it was a straightforward matter to find the rest of the gene and the protein that it produced. Having found the gene, more than a hundred additional mutations causing CF were identified. The successful research strategy was thus: genetic linkage analysis of many families to find the gene's location, physical mapping of the region, DNA sequence analysis to identify the most common mutation (and thus the gene), and further analysis of

Box 5-D-The Search for the Cystic Fibrosis Gene

The search for the gene causing cystic fibrosis (CF) is a recent triumph of human genetics. The symptoms of CF are poor digestion of foods and recurrent lung infections. The main effects of the disease trace to the accumulation of thick mucus plugs in the duct systems of organs throughout the body. The organs most affected are the lungs and the pancreas. Proper lung function depends on clear airways. In CF, airways become obstructed, and pockets of infection develop. Pancreatic ducts normally drain digestive enzymes into the small intestine. When viscid mucus blocks pancreatic ducts, the enzymes are not delivered, and digestion is less effective. The digestive symptoms can be treated by enzyme supplements. Treatments for lung symptoms are primarily careful monitoring, efforts to physically dislodge mucus plugs, and frequent administration of antibiotics to stave off infections.

CF is the most common seriously disabling single gene defect among Caucasian populations. It is a recessive condition, present only when a child inherits defective copies of a gene on chromosome 7 from both parents. CF affects approximately 1 in 3,500 live births in the United States; 1 in 30 Americans has one copy of a CF mutation gene, but most are unaffected because it takes both copies of the gene to cause disease.

Until 1985, the main facts known about CF were its symptoms and its pattern of inheritance. That year, a group led by Lap-Chee Tsui at the University of Toronto found that CF was genetically linked to, i.e., frequently inherited with, a marker on chromosome 7. The marker was contributed by Collaborative Research, Inc., a small biotechnology company located near Boston. Other groups quickly confirmed the linkage.

Finding the gene itself took another 4 years. The process involved studying cloned DNA fragments from that region of chromosome 7, meticulously assembling a map, and then searching for differences in DNA structure that correlated with the presence of CF in patients. A large group of collaborators at several centers, led by Francis Collins, finally identified a common DNA defect, the loss of three DNA base pairs. Seventy percent of patients with CF had this mutation, which served to identify the gene. Once the DNA sequence surrounding this mutation was known, it could be used to find the rest of the gene, and to find the protein produced by the gene. With the gene in hand, it was possible to confirm that its dysfunction caused CF by introducing the normal gene into cells in tissue culture and reversing a molecular defect.

The molecular defect underlying CF involves the transport of chloride ions across cell membranes. Poor chloride transport leads to thick, sticky mucus. The CF protein now serves as a target for drug development. The fact that the CF protein regulates chloride ion flow raises hopes that a drug might successfully replace its function. Another possibility in the long run is to introduce the normal gene into the cells that line lung ducts (see gene therapy section in text), permitting production of normal mucus. This is technically difficult and will take years at best, although investigators took their first steps down this path in 1992 with the first CF gene therapy protocol. In the meantime, a wealth of other treatment possibilities center on increasing chloride flow by drug treatment, or on modulating the inflammation that actually causes tissue damage. Discovery of the gene has rekindled hope among CF families, and has renewed interest in clinical trials of new agents.

SOURCES: K. Davies, "The Search for the Cystic Fibrosis Gene, "New Scientist 124:54-58, 1989. M. Dean, "Molecular and Genetic Analysis of Cystic Fibrosis,' Genomics 3:93-99, 1988. E.D. Green, and M.V. Olson, "Chromosomal Region of Cystic Fibrosis Gene in Yeast Artificial Chromosomes: A Model for Human Genome Mapping," Science 250:94-98, 1990. B.-S. Kerem, J.M. Rommens, J.A. Buchanan, et al., "Identification of the Cystic Fibrosis Gene: Genetic Analysis,'Science 245:1073-1080, 1990. S.T. Reeders, M.H. Breunig, K.E. Davies, et al., "A Highly Polymorphic DNA Marker Linked to Adult Polycystic Kidney Disease on Chromosome 16," Nature 317:542-544, 1985. D.P. Rich, M.P. Anderson, RJ. Gregory, et al., "Expression of Cystic Fibrosis Transmembrane Conductance Regulator Corrects Defective Chloride Channel Regulation in Cystic Fibrosis Airway Epithelial Cells," Nature 347:358, 1990. J.R. Riordan, J.M. Rommens, B.-S. Kerem, et al., "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complimentary DNA," Science 245:1066-1072, 1989. L. Roberts, "The Race for the Cystic Fibrosis Gene," Science 240:141-144, 1988. J.M. Rommens, M.C. Ianuzzi, B.-S. Kerem, et al., "Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping," Science 245:1059-1065, 1989. L.-C. Tsui, M. Buchwald, D. Barker, et al., "Cystic Fibrosis Locus Defined by a Genetically Linked Polymorphic DNA Marker," Science 230:1054-1057, 1985. BJ. Wainwright, P.J. Scambler, J. Schmidtke, et al., "Localization of Cystic Fibrosis Locus to Human Chromosome 7cenq22," Nature 322:467-470, 1985. R.L. White, S. Woodward, M. Leppert, et al., "A Closely Linked Genetic Marker for Cystic Fibrosis," Nature 318:382-384, 1985.

pedigrees to find additional mutations of the same gene.

The polymerase chain reaction (PCR) is a simple technique that enormously expedites the study of DNA (283,358). It is, in essence, a way to make many copies of short stretches of DNA without having to clone it. PCR can be used to generate a DNA sequence directly, or as a "probe" to identify clones or cells that contain the sequence copied by PCR. It requires only a mix of chemicals, a DNA-synthesizing enzyme extracted from bacteria, and short stretches of DNA that define the starting points for DNA copying. PCR was first described in 1985 by investigators at the biotechnology company Cetus. By 1989, PCR was referenced in 860 publications, used in dozens of ways for a multitude of purposes (316). In searching for genes, and as an adjunct to use genes to study proteins, PCR replaces slower and more expensive processes and opens entirely new avenues for exploration.

The story of CF (box 5-D) is one of the most straightforward that can be expected, but it nonetheless required enormous effort. From 1985 to 1989, laboratories throughout the world labored to find the gene. One reason for the protracted search was that the physical maps and regional sequencing had to be done *de novo. Once* complete maps of the human genome are constructed, similar searches should be much faster and less costly.

Other diseases are much more confusing, and tracking them down may prove far more difficult. The gene causing Huntington's disease, for example, was mapped in 1983, but its gene (and consequently, its protein product) remained elusive after a decade. Even when a gene and a gene product are identified, however, it may not be obvious what the protein product does. To understand function, other methods from cellular biology and physiology are needed.

When a gene is found, it is often not readily apparent what it does. A newly discovered protein may be likewise inscrutable-merely a dot on a piece of filter paper or a product derived from DNA sequence. The most reliable tools of molec-

ular biology and protein chemistry produce information about structure, but knowing structure does not guarantee an understanding of function. Genetics does provide tools to determine the function of a gene or protein, but the tools cannot crack open every lock. Genes can be introduced into bacteria, yeast, or animals to study the effects of a gene. As noted above, mutations can be made in the gene coding for a protein, in hopes that "breaking" the protein will clarify what it does when not broken.

The first step is to compare a new gene or protein to others already known, using databases that store the collected knowledge of researchers from around the world. There are databases for many kinds of structural information—crystallographic structure, protein structure, gene map positions, DNA sequence, and others (100). If a match is found—a protein that has a similar sequence of amino acids, for example-the matching protein may give clues to the function of the newly discovered one. Indeed, the ''new'' gene may not be new at all. Russell Doolittle and colleagues shocked the research community in 1983, for example, when they found an unsuspected similarity between a cancer-associated gene (a so-called oncogene) and a molecule that promoted cell growth (1 11).

To study the function of a disease-associated protein, such as the variant protein of cystic fibrosis, the abnormal gene can be introduced into cells in tissue culture, allowing much more precise experiments to be done quickly. The gene may also be introduced into another animal by recombinant DNA, making a transgenic animal. The effects of a gene mutation can thus be directly observed in the whole animal. This is a direct route to creating an animal model of a human genetic disease, with many of the complexities introduced by multiple organ systems, immune reaction, and other factors that are difficult to study in bacteria, yeast, or tissue culture cells. Until the advent of transgenic animal research, one had to hope that scientists had discovered an animal with an appropriate genetic defect similar to a human disease. Many human genetic dis-

eases, however, had no animal counterpart, and for these, transgenic techniques were a godsend.

The technique has obvious implications for drug research. Most drug receptors are proteins with corresponding genes that can be introduced into animals to explore their functions and dysfunctions. Alterations of the drug receptor can be introduced, and their effects observed in animals, without having to put human subjects at risk.
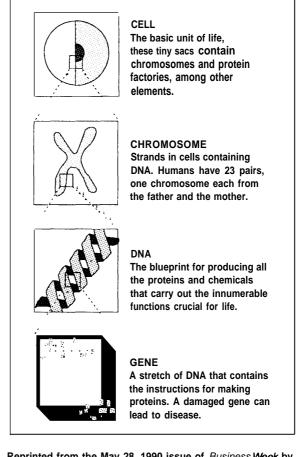
∎ Genetics as a Tool to Dissect Complex Disorders

Genetics may aid drug development by defining specific subpopulations of patients, thus simplifying the process of ascertaining the efficacy of new agents. Alzheimer's disease is an example of how genetics may help advance understanding of a disease by identifying subtypes (see box 5-E). Alzheimer's disease is the most common cause of dementia—loss of thinking ability. Symptoms typically begin only in middle age or later. It affects millions of Americans today, and is expected to afflict tens of millions early in the next century as the population ages (448). Alzheimer's disease is inherited as a single-gene dominant trait in some families, so that children of an affected parent in these families stand a 50-50 chance of developing it if they live long enough. The wide variation in symptoms and in age of onset have puzzled those studying the illness since it was first described in 1907. Indeed, only a small group of specialists were even aware that it could be inherited until recent years. Since 1987, studies of families affected by the familial form of Alzheimer's disease have revealed a genetic heterogeneity obscured by clinical and anatomical diagnosis. Even before the different genes in these families are found, it may be possible to categorize patients into subtypes, thus making it more likely to find effective drugs and other treatments (see **box** 5-E).

Genetics is but one of many approaches to disease. Following the trail down to a mutation in DNA cannot fully explain even most genetic diseases, and clearly genetic factors are only a part of most major diseases. The attraction of genetic approaches to disease, however, is that the tools are becoming so powerful. Most important diseases have been studied for decades. Those that could be easily explained by more traditional approaches have yielded; molecular genetics offers a strategy to crack those that have not.

∎ The Implications of the Human Genome Project

The genetic approach dissects the genetic factors that conspire to cause disease. The basic tools needed to pursue the approach are genetic linkage maps, physical maps, and DNA sequencing capacity. These are complemented by an

CELL
The basic unit of life, these tiny sacs contain chromosomes and protein factories, among other elements.

CHROMOSOME
Strands in cells containing DNA. Humans have 23 pairs, one chromosome each from the father and the mother.

DNA
The blueprint for producing all the proteins and chemicals that carry out the innumerable functions crucial for life.

GENE
A stretch of DNA that contains the instructions for making proteins. A damaged gene can lead to disease.

Box 5-E—The Complex Genetics of Alzheimer's Disease

Alzheimer's disease typically begins insidiously with loss of memory for recent events. It then progresses to more pronounced forgetfulness, loss of cognitive abilities, and frequently to behavioral symptoms such as irritability or depression. The course of the disease can extend for many years, even decades. Disease can begin to show as early as age 40, but most commonly begins after age 70. In 1907, Alois Alzheimer first described a 51-year-old woman with progressive loss of memory and distinctive microscopic changes in her brain. In the 1930s, many groups began to describe families in which many members developed Alzheimer's disease. By 1980, more than 80 such families were in the published literature, but the familial form of Alzheimer's disease was still not widely appreciated.

In some families, the disease travels as an autosomal dominant trait, so that the child of an affected parent has a 50-50 chance of developing it. In such families, a single gene best explains the pattern of inheritance. Getting the Alzheimer's copy of the gene from the affected parent translates to a very high probability of eventually developing Alzheimer's disease. Inheriting the normal copy means that one will not get Alzheimer's disease, or at least not the genetic form, and future progeny likewise will be spared. Several groups throughout the world attempted through the 1970s and 1980s to characterize the genetics of Alzheimer's disease. A picture began to take shape, but it was more complicated than a simple Mendelian disease caused by the same gene in all families.

The first breakthrough came from studies led by Peter Saint George-Hyslop. A large collaborative group studied several families with Alzheimer's disease, and in 1987 linked the disease to a genetic marker on chromosome 21. Another group in Europe confirmed the linkage. Over the next two years, however, other groups found families that seemed to have the same disease but did not show linkage to chromosome 21, even using the same markers. In 1991, a different group of families showed evidence of an Alzheimer's gene on a different chromosome, number 19; in 1992, another international collaboration found most familial Alzheimer families showed linkage to a region on chromosome 14. The story on chromosome 21 also became more complicated. Two groups found patients with a particular DNA mutation on chromosome 21 that correlated with the disease. Other families that showed linkage to chromosome 21 lacked this mutation, making it likely there were two or more causes of chromosome 21-linked Alzheimer's disease. A consensus emerged that more than two genes on chromosome 21 could cause the disease in different families. There was likely another gene on chromosome 19; and a set of families, called the Volga German families for their geographic origin in prerevolutionary Russia, that seemed to have a gene that did not map to either chromosome. There might thus be four or more subtypes: beta-amyloid mutant Alzheimer's disease, a distinct chromosome 21 Alzheimer's disease, chromosome 19 Alzheimer's disease; chromosome 14 Alzheimer's disease, and an unlinked (Volga German) familial Alzheimer's disease.

Time-honored clinical and anatomic classifications obscured the heterogeneity revealed by genetics. The hope was to find the genes, define their protein products, and use these to search for functional clues as to why nerve cells in the brain died prematurely. A similarly complex picture was beginning to emerge in complex diseases such as heart disease, schizophrenia, arthritis, diabetes, various types of cancer, and other disorders. Genetics, with its unique power to break diseases into precise subtypes directly correlated to molecular diversity, was likely to have implications for drug development. If different genes caused clinically similar diseases, different proteins were involved (or, perhaps, RNA products that regulated DNA expression).

All these different gene products are potential targets for drug development. If their function can be restored by drugs, then there is hope for therapy. Finding the gene maybe the first step to finding the right target at which to aim. If the Alzheimer's gene produced a toxic product, for example, it might be possible to inhibit its production. If it resulted from lack of a growth factor, then agents to replace that factor might be synthesized.

Box 5-E—The Complex Genetics of Alzheimer's Disease--(Continued)

The genetic factors involved in Alzheimer's disease underscore how little is known about its causes. The discovery of unsuspected subtypes of disease starkly points out how far medicine remains from a detailed understanding. Genetics is being pursued in hopes of getting a molecular "handle" on conditions for which little is known beyond the fact that they "run in families."

At the very least, being able to distinguish molecular subtypes can direct drug development for defined subpopulations. This could permit advances for diseases such as Alzheimer's that have to date proved refractory. Disease subtyping might also conceivably reduce drug testing costs. If it were possible to select in advance those subpopulations of patients likely to respond to a given drug, then it would be much easier to demonstrate efficacy. Proving efficacy is a major problem in developing drugs for complex chronic diseases. Drug effects apparent in only 20 percent of patients, for example, could easily be lost in the "noise'—random statistical variations in the other 80 percent. If genetic typing could select out only the 20 percent likely to respond, the effect would pop to the surface. Testing a small, molecularly defined population would amplify drug effects, lower testing costs, and speed regulatory approval.

PRINCIPAL SOURCES: A.M. Goate, A.R. Haynes, M.J. Owen, et al., "Predisposing Locus for Alzheimer's Disease on Chromosome 21," *Lancet* I(8634):352-355, 1989. P.H. Saint George-Hyslop, J.L. Haines, L.A. Farrer, et al., "Genetic Linkage Studies Suggest That Alzheimer's Disease Is Not a Single Homogeneous Disorder," *Nature* 347:194-197, 1990. P.H. Saint George-Hyslop, R.E. Tanzi, R.J. Polinsky, et al., "The Genetic Defect Causing Familial Alzheimer's Disease Maps on Chromosome 21," *Science* 235:885-890, 1987. G.D. Schellenberg, T.D. Bird, E.M Wijsman, et al., "Absence of Linkage of Chromosome 21q21 Markers to Familial Alzheimer's Disease," Science 241:1507-1510, 1988. G.D. Schellenberg, T.D. Bird, E.M. Wijman, et. al., "Genetic Linkage Evidence for a Familial Alzheimer's Disease Locus on Chromosome 14," *Science* 258:668-671, October 23, 1992.

ADDITIONAL SOURCES: J. Constantinidis, G. Garrone, R. Tissot, et al., "L'incidence famiale des alterations neurofibrillaires corticales d'Alzheimer," *Psychiatrie et Neurologie* 150:235-247, 1965. W.H. English, "Alzheimer's Disease," *Psychiatric* Quarterly 16:98, 1942. E. Essen-Moller, "A Family with Alzheimer's Disease," *Acta Neurological Scandinavia* 21:233-244, 1946, R.G. Feldman, K.A. Chandler, L.L. Levy, et al., "Familial Alzheimer's Disease," Neurology 13:811-824, 1963. F.E. Flugel, "Zur Diagnostic der Alzheimerschen Krankheit," *Gesamte Psychiatric* 120:783-787, 1929. R.L. Friede, K.R. Magee, "Alzheimer's Disease," Neurology 12:213-222, 1%2. L.L. Heston, D.L.W. Lowther, and C.M. Leventhal, "Alzheimer's Disease," *Archives of Neurology* 15:225-233, 1%6. H. Jacob, "Muscular Twitchings in Alzheimer's Disease, " *Alzheimer's Disease*, G.E.W. Wolstenholme (ed.)(London, England: Churchill, 1970). R. Katz, "The Prevalence and Malignancy of Alzheimer's Disease: A Major Killer," *Archives of Neurology* 33:217-218, 1976. P.J. Landy and B.J. Bain, "Alzheimer's Disease in Siblings," *MedicalJournal of Australia* 2(18):832-834, 1970. H. Lauter, "Genealogische Erhebungen in einer Familie mit Alheimerschen Krankheit," *Archiv der Psychaitrie Gesamte Neurologie* 202:126-139, 1%1. K Lowenberg, and R.W. Waggoner, "Familial Organic Psychosis (Alzheimer's Type)," *Archives of Neurology and Psychiatry* 31:737-754, 1934. T. Luers, "Uber die familire juvenile Form der Alzheimerschen Krankhiet mit nerulogischen Herderscheinungen," *Archiv fur Psychiatrie und Nervenkrankheit* 179:132-179, 1947. R.T.C. Pratt, "The Genetics of Alzheimer's Disease," *Alzheimer's Disease*, G.E.W. Wolstenholme (ed.) (London: Churchill, 1970). T. Sjogren, "A Genetic Study of Morbus Alzheimer and Morbus Pick," *Acta Psychiatrica et Neurological Scandinavica* [supplement] 82:9-63, 1952. L. Wheelan, "Familial Alzheimer's Disease," *Annals of Human Genetics* 23:300-310, 1959. G. Zawuski, "Zur Erblichkeit der Alzheimerschen Krankheit," *Archiv fur Psychiatrie Gesamte Neurologie* 201:123-132, 1960.

enormous diversity of ways to understand the function of genes, by introducing mutations in cultured cells, or by creating yeast models or transgenic animal models of a human disease. The Human Genome Project emerged in 1985 and evolved into a concerted effort to build the infrastructure for large-scale mapping, sequencing, and technology development which, in turn, were intended to lay the foundation for genetic explorations in biomedical research.

The underlying story was a convergence of technologies to analyze DNA, to clone large DNA fragments, to construct genetic linkage and physical maps, and to determine DNA sequences. These developments paralleled developments in computers that lowered costs and added computational power and flexibility. From these technological shifts, several individuals independently struck on the audacious idea of determining the sequence of all 3 billion base pairs in the human genome. The idea provoked considerable controversy, and the genome project was ultimately redefined to include genetic linkage mapping, physical mapping, as well as DNA sequencing. The goal of the Human Genome Project slowly and almost imperceptibly shifted from a complete DNA sequence of the genome to a complete structural catalog of human genes, which may not prove to be the same thing (91).

Work on model organisms is essential to interpret human gene maps, and has been incorporated into the project's goals (488). To accomplish its goals, the human genome project must develop new technologies to make mapping and DNA sequencing faster, less costly, and more accurate. These technologies will themselves be a boon to other investigations, as the analysis of DNA is central to biomedical research of almost every variety.

The objective is ambitious. An estimated 50,000 to 150,000 genes are dispersed through the human chromosomes, of which McKusick's catalog lists just over 2,000 that have been well characterized. The genome project should provide a molecular catalog for tens of thousands of genes that are as yet unknown. Genetic linkage maps with severalfold more markers are needed to locate genes known only by their pattern of inheritance. Most regions of the genome lack physical maps, and less than one percent of the genome has been sequenced as of 1991. The human gene map is only in its infancy (3 13,405). The physical and genetic linkage maps are slated to be near completion by mid-decade, with massive amounts of sequence data to be available in 15 years (488). The elaboration of these various maps, when combined with techniques to catalog the large mass of currently unknown genes, will undoubtedly reveal many genes that influence disease. Even in well studied organisms such as yeast and nematode worms, the direct approach of DNA structural analysis has uncovered many more genes than were known to exist. Each new protein is a potential drug receptor target; many will provide promising new leads for drug development.

DNA AS A THERAPEUTIC AGENT

■ Gene Therapy Is Just Beginning

One promising treatment strategy is deliberately to introduce genes into human cells to compensate for aberrant genes that cause genetic disease. The process is called human gene therapy. Gene therapy as a theoretical possibility was discussed widely for decades. In 1989, the first genes were introduced into the cells of cancer patients in order to *monitor* a novel anticancer treatment (354). The first gene insertion to *treat* **a genetic disease,** *bona fide* gene therapy, was per-formed in 1990.

Gene therapy falls into two major categories. It can be aimed at cells of the body, or somatic cells, so that it affects only that patient. The other, more controversial, alternative is to treat cells of an early embryo, egg cells, sperm cells, or their precursors. Any genes introduced into such cells would not only be present in the individual, but would also be passed on. Treatment of egg cells, sperm, or their precursors would lead to inherited changes in any babies resulting from fertilization. Treating an early embryo would affect not only somatic cells but also those giving rise to eggs and sperm. In each case, some fraction of future generations would carry the altered genes. This variety of gene therapy is termed germ line gene therapy.

The treatments approved to date, and anticipated in the near future, will involve bone marrow cells, white blood cells, skin cells, liver cells, lung cells, pancreatic cells or others that can be extracted from the body, treated, and reintroduced

back into the patient. The first approved gene therapy clinical trial aimed to treat a rare genetic disease caused by a deficiency of the enzyme adenosine deaminase (ADA) known as "Bubble-Boy Syndrome. " This is a recessive disorder, so both copies of the gene coding for the enzyme are abnormal. The result is that the enzyme fails to degrade the chemical adenosine. Adenosine accumulates, most notably in white blood cells, and the white cells responsible for fending off infections consequently function poorly. Untreated patients completely bereft of enzyme function generally die of infection before age 2. In an example of a protein used directly as a drug agent, the ADA enzyme has been chemically linked to polyethylene glycol and injected directly into patients. Patients have improved under this treatment. Another approach is to take white blood cells from such patients, insert the gene that produces ADA, and insert the cells back into the patient. This is the protocol approved as the first instance of human gene therapy (16,49). The frost patient, a four-year-old girl, began treatment in September 1990; by the end of June 1992, there were 14 approved clinical protocols (with 35 patients) involving gene transfer in humans (15,194).

The original notion of human gene therapy was to treat single gene defects. The concept has since broadened considerably. Viewing gene therapy only as a way to compensate for defective genes in the patient's body has given way to seeing it as a way to introduce useful genes into cells that can act as drug delivery devices. This opens a far more diverse set of possibilities. Cells treated with inserted genes could conceivably be used to treat acquired immunodeficiency syndrome (AIDS), heart attacks, diabetes, and cancer (14). Several recent protocols approved or in preparation already illustrate the broader possibilities (305).

Several technical obstacles face gene therapy before it can be used as a standard treatment modality. First, the range of cells that can be targeted for gene insertion must be expanded considerably. Only white blood cells and certain types of bone marrow cells have been successfully treated to date. These cells die off over a period of months, and the treatment expires with them. The next step may be to get genes into the so-called ''stem'' cells that continually divide to produce whole populations of cells. If the gene were inserted into stem cells of the bone marrow, for example, then the treatment might not have to be repeated-there would be a steady stream of new cells expressing the gene, derived from stem cells. Second, the expression of the inserted gene must be sufficient to produce a clinical benefit and not too much as to cause toxicity. The amount of protein produced from artificially inserted genes is, in general, significantly lower than normal amounts. Third, for many applications it will be necessary to "aim' the gene insertion at specific organs or tissues.

For now, this problem is solved by extracting cells from the body before inserting the gene. This severely limits the types of cells that can be treated. It is not practical to remove cells from most organs before treating them. The ability to reliably insert genes only into nerve or muscle cells, for example, would greatly enhance prospects of treatment for intractable neuromuscular diseases. Those hoping to treat Duchenne muscular dystrophy have raised this possibility, and are working to fabricate DNA elements that express genes only in muscle cells (355). Finally, it would be a great boon to gene therapy if genes were not merely inserted, but instead new DNA sequences replaced old ones in the same gene. Current methods of introducing genes into cells insert whole new genes (attached to other genes and regulatory sequences). The chromosome.1 site of insertion is not predictable or controllable. The ideal treatment would instead excise 'bad' DNA sequences while replacing them with "good" ones. This would require that the corrective sequence recognize the gene it was to replace with great specificity. This process is possible in yeast and bacteria, and is a standard tool of genetics. It has even been done in manmals; the problem is that it is successful only very infrequently. Before such techniques were clinically applicable, they would have to be much more reliable.

▌ Prospects for Germ Line Treatments Are Remote

The prospects for germ line gene therapy are quite remote at present, for both ethical and technical reasons. Germ line gene therapy would be directly analogous to transgenic animal methods, in that a heritable gene would be introduced into a human. Changes could be inherited by subsequent generations. The technique is thus technically feasible, but there are extremely important differences between clinical application in humans and transgenic animal research. First, most transgenic animal experiments involve hundreds of animals, only a small fraction of which acquire the desired new gene (a percent or so at best). In many experiments, a fraction of animals become sterile or suffer genetic damage because of the new DNA inserts into critically important sites, disrupting another gene. To be clinically useful, the technique would have to successfully insert a gene almost all of the time and only very rarely cause adverse consequences. Clinical trials of germ line therapy would have to demonstrate, moreover, that the inserted gene had no demonstrable effect during embryonic and fetal development. Even contemplating how to demonstrate this in humans is a major task. Providing evidence of safety in humans without being able to target gene insertion is difficult to imagine.

Germ line gene therapy might be useful for conditions where damage accumulated during embryonic or fetal development, or if multiple organ systems had to be corrected. Germ line therapy would require techniques to insert genes into sperm or egg cells that currently do not exist (or into cells that produce them), or use of in vitro fertilization followed by treatment of early embryos. Prospective parents could more simply and safely choose to implant embryos that would not develop a disease, rather than treating embryos destined to do so. This alternative would not be available in one very unusual situation-if both parents had a recessive genetic disease. Both parents would carry double copies of a defective gene in this case, and so every embryo would

likewise have a double dose of a gene defect and would thus be affected. This clinical situation is not impossible, but it would be quite rare. These technical factors combine with a lack of consensus that germ line therapy is ethically acceptable to make germ line gene therapy unlikely in the foreseeable future, although it might resurface in the more distant future.

▌ Uncertain Prospects for Gene Therapy as a New Treatment Modality

Those surveyed by the Office of Technology Assessment disagreed markedly about whether gene therapy would emerge as a major treatment modality over the next 10 to 20 years, Some saw gene therapy as the coming wave of therapeutics. They cited several advantages. First was that it could attack diseases that other methods could not. Most genetic diseases have only palliative treatments, or only partially effective ones. The optimists foresaw that the problem of getting cells into specific target cells and in specified chromosomal locations would be solved. If so, many single gene defects could be treated.

The most devastating symptoms of CF are due to lung problems. The cells that line lung ducts, for example, are constantly turning over. It might be possible to introduce normal genes into CF cells if the stem cells giving rise to them could be treated by gene therapy. But to treat a sufficient quantity, this would most likely have to be done without removing lung cells from the patient's body. A tamed virus, perhaps inhaled or injected, would have to home in on those stem cells. In addition, the introduction would have to be sufficiently controlled 50 that when many millions of cells were treated, the gene insertion did not induce mutations leading to cancer or cause other unwanted side effects. Gene therapy for CF is nonetheless being vigorously pursued, with the first clinical protocol approved in 1992.

Another advantage of gene therapy is that it might require many fewer treatments. Once stem cells were treated, it might not be necessary to continually administer drugs. In this sense, it would be more akin to organ transplantation or

vaccination than to most drug treatments. One-time treatment is an appealing prospect for diseases such as diabetes, where it would obviate daily injections of insulin. Gene therapy also might restore the body's own feedback controls. Fluctuations in drug levels often fail to synchronize with normal regulatory controls. This is particularly important for hormones and other substances produced in response to environmental changes. Cells treated by gene therapy might be brought back into feedback control, either by including a gene's natural regulatory elements, or by repairing a defective gene in its natural chromosomal site.

Skeptics note the technical difficulties in targeting specific cell types, in getting genes into identified chromosomal sites, and in the high costs of clinical trials and safety testing. The range of diseases that can be attacked with current methods seems narrow. Gene therapy might be useful for cancer therapy, fatal genetic diseases of childhood, and other extremely serious conditions lacking better therapy. Some of those surveyed acknowledged that these revolutionary treatments were imminent, but questioned whether they would prove economically viable on a grand scale. If the number of affected individuals is small and the treatment is expensive to develop and to administer, the clinical advantages of a one-time treatment could prove a commercial disadvantage. The costs of R&D might fall on a small number of patients and single doses, limiting access and boosting the unit price. The range of disorders that can be approached by gene therapy will not broaden appreciably until better cell targeting and chromosome-site targeting are possible, and these make well take a decade of research to develop, if they develop at all. The unknown safety issues also raise concern that liability costs could be high.

Those working to develop gene therapy tend to be optimists. They view current protocols as analogous to Henry Ford's first primitive internal combustion engine, with prospects so revolutionary they cannot be predicted. The Federal Government is, for now, the main investor in gene therapy. The cost of the clinical trials has to date been funded directly by the National Institutes of Health, although this may change as more firms become involved in gene therapy development.

At least three small biotechnology companies were founded with gene therapy as part of their business plan, but these are viewed as long-term investments and the companies are now concentrating on research and new methods. They aspire to turn a profit from marketable therapeutic products in the next several years. These small startup companies have been joined by eight or more larger pharmaceutical firms pursuing gene therapy technology. Even if they cannot sell gene therapy itself, however, they may be able to sell reagents for gene transfer to other researchers. Gene transfer methods are widely applicable beyond gene therapy, and so this market, while not comparable to a major therapeutic agent, might nonetheless sustain a small firm during its formative years. Current work includes collaborative agreements with larger pharmaceutical firms, that fired small exploratory research efforts. For now, gene therapy is in its early exploratory phase.

■ Alternative Uses of DNA and RNA

In the future, DNA itself could serve as a therapeutic agent. There are several possible routes by which DNA (or RNA) could be used as drugs. The transcription of DNA into RNA can be blocked, for example, by proteins designed to bind to DNA. Short stretches of DNA introduced directly into the blood stream can last for hours, are actively taken into cells, and can also block the process of transcribing DNA into RNA. Short stretches of DNA or RNA can also inhibit the process of translating RNA into protein. This strategy of making "antisense" sequences to block the production of proteins is being explored by several pharmaceutical companies for its possible therapeutic value. These efforts closely parallel other drug discovery efforts, the main difference being that the physiological target is DNA or RNA rather than proteins. Like gene therapy, these alternative uses of DNA and RNA

as therapeutic agents are in their preliminary research phase.

DISCOVERY INCREASINGLY DRIVEN BY BIOMEDICAL RESEARCH

An increasing reliance of drug discovery on biomedical research stands out as a salient theme of this chapter. Drugs uncovered by chance clinical observations or systematic refinement of folk remedies early in the century have given way to pharmaceutical firms with thousands of scientific workers pursuing drug research. The teams of organic chemists, specialists in pharmacological screening, and clinical experts have not been abandoned; they are still just as essential as they have always been. Rather, biologists have been added onto the front end of the drug discovery process.

▌ The Foundation of Drug Discovery Is Biomedical Research

The first step along the pathway to discovering a major new drug is the understanding of what causes a disease. Many of the drugs that could be discovered by clinical happenstance have been developed in this first century of the pharmaceutical industry. Screening of microbes and plants is still important, but many of the classes of agents that can be discovered have already been. There are undoubtedly many clinically useful natural products yet to be found, but the process of discovering them has reached a point of diminishing returns. Pharmaceutical firms have turned their attention to understanding the mechanism of disease as a guide to discovering truly novel drugs.

The development of cimetidine and captopril illustrate the new approach. Dozens of other drugs could have served as equally valid examples. These drugs are already well established in the market. Many of the agents under investigation now will only come to market after the turn of the century. The differences of opinion belie an underlying, widely shared philosophy:

"Successful management of industrial research is dependent on rapid access to the latest discoveries in academic laboratories, the ability to recognize the importance of a given discovery, the ability to integrate the information into research programs within an industrial laboratory, and the ability to focus effort to allow maximum chance that the idea will bear practical fruit. It is vital for an industrial laboratory to have its own cutting edge basic research program at early stages of newly evolving fields" (373).

The dominant strategy of modern pharmaceutical firms is to invest heavily in R&D, to form collaborations with academic laboratories and small specialized biotechnology companies, and to pursue the most promising leads with in-house research teams. Drug discovery research thus rests on a broad base of publicly funded basic biomedical research. It picks the fruits of basic research as new ideas emerge about disease mechanisms that suggest potential drug receptor targets.

▌ Biomedical Research Turning to Larger Scale and More Complexity

Biomedical research is itself changing. The scale and complexity of problems are increasing. This is in part due to the massive accumulation of knowledge in the postwar era. Many of the simple problems have been solved, the simplest diseases understood. What remain are the conditions that could not be understood with past methods. Research methods grow more powerful at the same time as harder problems come into view. Structural understanding of proteins and DNA are at the forefront of innovative research methods and pose many of today's most tantalizing problems.

Genetics progressed from the study of viruses whose genomes are thousands of bases in length to those with hundreds of thousands, then to bacteria with millions of bases in their genomes. Only in the 1980s did DNA pioneers venture into genomes of tens or hundreds of millions of base pairs. As the 1990s began, they. were poised to take on the human genome.

▮ The Increasing Role of Instruments and Computers in Drug Research

A person picking up a scientific journal will open its pages to many advertisements aimed at biologists. In these advertisements, the stereotypical biochemist or molecular biologist holds a test tube in one hand and pipette in the other, focusing intently on transferring a fluid containing protein, DNA, or chemicals used to study such molecules. The image of biomedical research conjures up thousands of test tubes and hours spent moving reagents into and out of them, mixing them, and then analyzing the results. The image is accurate, but changing. In many advanced academic and industrial laboratories the test tubes have not gone away, but the person has been replaced by a robot or automated instrument. Sanger's work from 1945 to 1955 to determine the sequence of amino acids in insulin (see box 5-C) took almost a decade of intensive effort by a scientist who ultimately earned two Nobel prizes. Today, this can be done in days by a technician and a machine. The first 24 nucleotides of DNA sequence took many years effort in the early 1970s; today a single automated DNA sequenator can generate a thousand times as many in a day.

Extraordinary leaps in technological capacity open up new approaches to problems of larger scale. Instrumentation and automation thus dramatically increase the efficiency of doing the same experiments. In the 1970s, every well-to-do molecular biology laboratory had its own ultracentrifuge, used to spin test tubes very fast and enabling separation of proteins and DNA fragments according to weight, and a spectrophotometer, used to measure the color of liquids to determine the concentration of chemicals. These instruments cost tens of thousands of dollars each. Now, a laboratory must have not only these instruments, but others even more complex and expensive. To stay on the cutting edge of DNA or protein research, there are robots to do microchemical reactions, instruments to determine the sequence of amino acids in proteins and bases in DNA, machines to synthesize short proteins and

stretches of DNA with specified sequence. Some of these instruments edge over $100,000 in price. The percent of total Federal funding going to equipment has remained relatively stable over the past decade, indeed dropping somewhat from 6.2 percent in 1979 to 5.6 percent in 1988 (481). Within this stable base, however, there has been a shift from centrifuges and tissue culture needs to instruments that synthesize or sequence protein and DNA. Requests for preparative equipment, including centrifuges, dropped from 33 percent of requests in 1984 to 25 percent in 1988, while sequences and synthesizers increased from 11 to 14 percent (480).

Those surveyed by the Office of Technology Assessment agreed that the need for instruments had increased dramatically, and many offered the example of multimillion dollar investments in crystallography and computers as examples. These not only involve investments in instruments but also entail sustained commitments to specialized personnel to run the machines and cadres of support personnel to develop software and to analyze the data. Supercomputers are used and in the process of QSAR to analyze all the data emanating from protein structure studies. Virtually every firm has invested heavily in computers with sophisticated graphics displays for structural chemistry studies. Computer networks maintain and analyze the massive amounts of data flowing out of clinical studies. Many firms lease time on supercomputers. At least one drug company has purchased one outright, representing a several million dollar hardware investment, and has spent millions each year in personnel, software, operating, and maintenance costs.

IMPLICATIONS FOR FUTURE PHARMACEUTICAL R&D COST

▮ Knowledge Better Specifies Targets for Drug Design, but Multiplies the Number

The large public investment in biomedical research since World War II has amply demonstrated how investment in research can translate to knowledge about disease and normal biology.

The mushrooming mass of medical facts has two major impacts on pharmaceutical R&D. Refined models of disease mechanism offer new insights into lines of research that might produce a new drug. Work can concentrate on the molecules that form the links in a causal chain leading to disease. Just as the targets become more precise, however, they proliferate in number.

Drugs to treat heart disease, for example, were until recent decades restricted to a few drugs that strengthened cardiac contraction (such as digitalis), sped heart rate (epinephrine), or constricted or relaxed small arteries (vasodilators and vaso-constrictors). These agents still exist, but drugs that modulate the flow of ions through cell membranes have been added to the list, including agents for calcium channels, potassium channels, sodium channels, and chloride channels. A wealth of drugs inhibit or stimulate cardiac muscle receptors selectively. Understanding diverse mechanisms that influence the contraction of heart muscle and blood vessels has uncovered a plethora of new drug receptor targets.

In general, each project undertaken now is far more precise and the mechanisms are better understood individually, but there are far too many avenues to pursue. This is perhaps the most significant change in drug discovery research over the past decade, and the trend is likely to continue.

▌Uncovering Disease Subtypes May Make Clinical Testing More Precise

As illustrated in box 5-E (on Alzheimer's disease) genetics may prove useful in identifying groups of patients more likely to respond to a given agent. If so, the process of demonstrating a drug's effect would be simplified and the drug approval process expedited. Refinements of drug receptor studies will frequently uncover tests for function that are more precise, capable of improving screening tests used to identifying promising chemical compounds for physiological effects. Narrowing the population that needs to be clinically assessed could dramatically cut costs of clinical trials.

The flip side of this coin is that tests of efficacy and safety could continue to proliferate. The number of tests that regulators could wish to see performed may increase to provide better evidence of safety and efficacy. If more population subtyping leads to an increased demand for studies of more refined groups that could not be distinguished before, then costs could go up rather than down. Here again, scientific advance is a two-edged sword. Each experiment or trial can be more precise, but the number and cost of experiments may go up as well. It is difficult to predict which effect will be greater.

▌Automation and Advances in Analytical Methods Can Make Research More Powerful

Some technologies clearly make essential steps faster and cheaper. The polymerase chain reaction (PCR) described above is a good case in point. PCR will have direct applications for diagnosis, and it also shortens or eliminates many steps in DNA research. This is a technology that clearly saves costs. The growing power of computers similarly makes old procedures practical for a much broader range of experiments. Dropping computer costs and improved instruments have enabled several drug firms to invest in facilities to do their own x-ray crystallography and high-field NMR analysis of proteins, for example.

▌Will New Technologies Speed the Discovery of New Drugs?

The new technologies of biomedical research clearly presage the development of new drugs. Whether the new approaches to drug discovery increase or reduce R&D costs depends in large part on whether the drug development cycle is shorter or longer as a result of the new technologies. A large fraction of the cost of any research-intensive enterprise, and especially drug development, is the cost of capital (see chapter 3). A dollar invested in drug discovery cannot be invested elsewhere, and its return can be known only many years later. The length of time from

first investment to payback is a critical variable, since the costs compound each year.

Developing new therapeutic drugs will likely become more difficult over time, as the easiest to discover give way to more and more difficult tasks. Unless the new technologies expedite drug discovery sufficiently to compensate for the increasing difficulty of finding new agents, costs will rise. Most of the new investments in molecular biology appear to have been added to the ''Iiont end' of drug discovery, and thus represent an increased investment at the earliest stages of research, where the time-cost of capital is greatest.

As drugs move from discovery to clinical testing, the new technologies may make it easier to demonstrate efficacy. If the number of safety an efficacy tests increases at the same time, however, or if the testing process becomes slower, then costs will again rise. The critical factor is again how long trials take, how many new ones are added as a result of new technologies, how much the testing costs, and the duration of regulatory review. Molecular biology clearly promises to dramatically expand the repertoire of drug therapy in the coming decades. The prospects for cost reduction or cost escalation, however, are extremely difficult to gauge. Faster and cheaper methods may be offset by longer product development cycles and a need for more and better clinical trials.