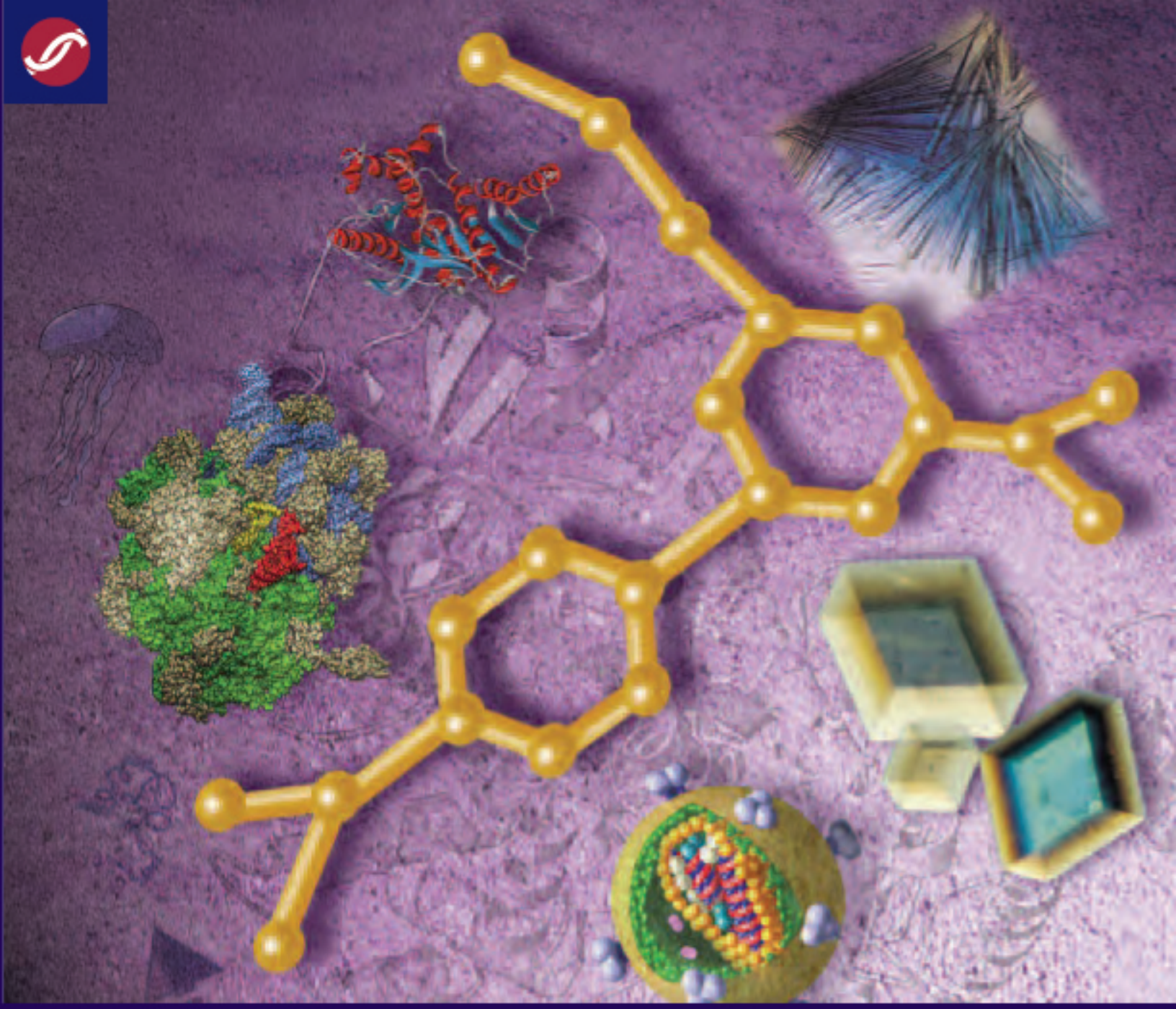


# *The Structures of Life*

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
National Institutes of Health
National Institute of General Medical Sciences

# *The Structures of Life*

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
National Institutes of Health
National Institute of General Medical Sciences

NIH Publication No. 07-2778
Reprinted July 2007
http://www.nigms.nih.gov

# Contents

# *Why Structure?*

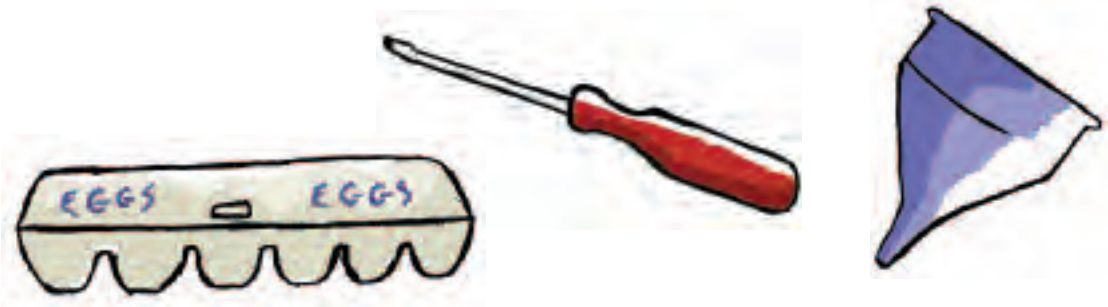
Imagine that you are a scientist probing the secrets of living systems not with a scalpel or microscope, but much deeper—at the level of single molecules, the building blocks of life. You'll focus on the detailed, three-dimensional structure of biological molecules. You'll create intricate models of these molecules using sophisticated computer graphics. You may be the first person to see the shape of a molecule involved in health or disease. You are part of the growing field of structural biology.

The molecules whose shapes most tantalize structural biologists are proteins, because these molecules do much of the work in the body.

Like many everyday objects, proteins are shaped to get their job done. The shape or structure of a protein offers clues about the role it plays in the body. It may also hold the key to developing new medicines, materials, or diagnostic procedures.

In Chapter 1, you'll learn more about these "structures of life" and their role in the structure and function of all living things. In Chapters 2 and 3, you'll learn about the tools—X-ray crystallography and nuclear magnetic resonance spectroscopy—that structural biologists use to study the detailed shapes of proteins and other biological molecules.

*In addition to teaching about our bodies, these "structures of life" may hold the key to developing new medicines, materials, and diagnostic procedures.*



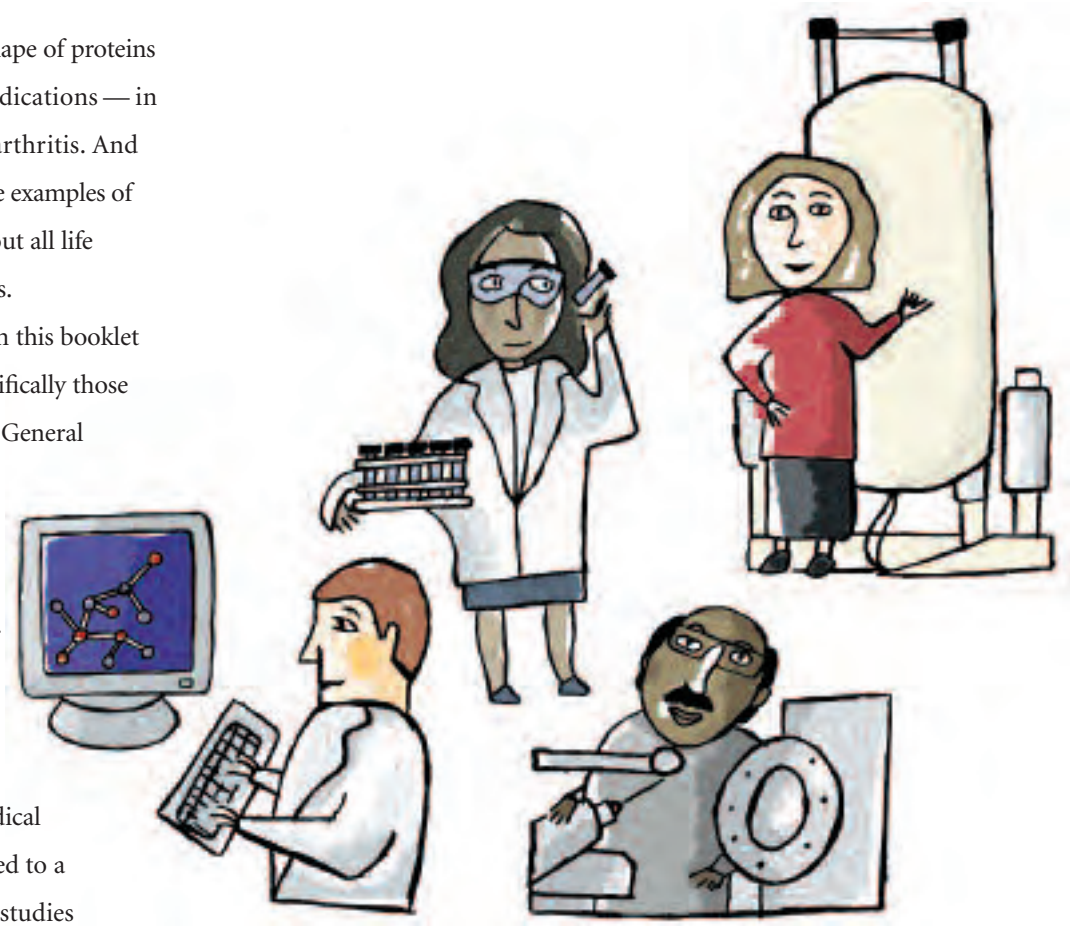

▲ Proteins, like many everyday objects, are shaped to get their job done. The long neck of a screwdriver allows you to tighten screws in holes or pry open lids. The depressions in an egg carton are designed to cradle eggs so they won't break. A funnel's wide brim and narrow neck enable the transfer of liquids into a container with a small opening. The shape of a protein—although much more complicated than the shape of a common object—teaches us about that protein's role in the body.

Chapter 4 will explain how the shape of proteins can be used to help design new medications — in this case, drugs to treat AIDS and arthritis. And finally, Chapter 5 will provide more examples of how structural biology teaches us about all life processes, including those of humans.

Much of the research described in this booklet is supported by U.S. tax dollars, specifically those awarded by the National Institute of General Medical Sciences (NIGMS) to scientists at universities across the nation. NIGMS is one of the world's top supporters of structural biology.

NIGMS is also unique among the components of the National Institutes of Health (NIH) in that its main goal is to support basic biomedical research that at first may not be linked to a specific disease or body part. These studies increase our understanding of life's most fundamental processes — what goes on at the molecular and cellular level — and the diseases that result when these processes malfunction.

Advances in such basic research often lead to many practical applications, including new scientific tools and techniques, and fresh approaches to diagnosing, treating, and preventing disease.



▲ Structural biology requires the cooperation of many different scientists, including biochemists, molecular biologists, X-ray crystallographers, and NMR spectroscopists. Although these researchers use different techniques and may focus on different molecules, they are united by their desire to better understand biology by studying the detailed structure of biological molecules.

Alisa Zapp Machalek
Science Writer and Editor, NIGMS
July 2007

# *Proteins Are the Body's Worker Molecules*

You've probably heard that proteins are important nutrients that help you build muscles. But they are much more than that. Proteins are worker molecules that are necessary for virtually every activity in your body. They circulate in your blood, seep from your tissues, and grow in long strands out of your head. Proteins are also the key components of biological materials ranging from silk fibers to elk antlers.

*Proteins are worker molecules that are necessary for virtually every activity in your body.*


▲ Proteins have many different functions in our bodies. By studying the structures of proteins, we are better able to understand how they function normally and how some proteins with abnormal shapes can cause disease.

## Proteins Are Made From Small Building Blocks

Proteins are like long necklaces with differently shaped beads. Each "bead" is a small molecule called an amino acid. There are 20 standard amino acids, each with its own shape, size, and properties.

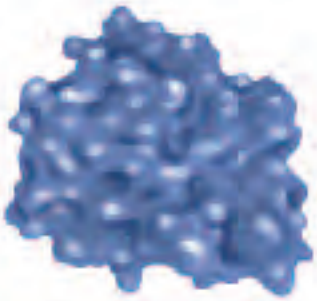
Proteins typically contain from 50 to 2,000 amino acids hooked end-to-end in many combinations. Each protein has its own sequence of amino acids.

These amino acid chains do not remain straight and orderly. They twist and buckle, folding in upon themselves, the knobs of some amino acids nestling into grooves in others.

This process is complete almost immediately after proteins are made. Most proteins fold in less than a second, although the largest and most complex proteins may require several seconds to fold. Most proteins need help from other proteins, called "chaperones," to fold efficiently.

Proteins are made of amino acids hooked end-to-end like beads on a necklace.

To become active, proteins must twist and fold into their final, or "native," conformation.



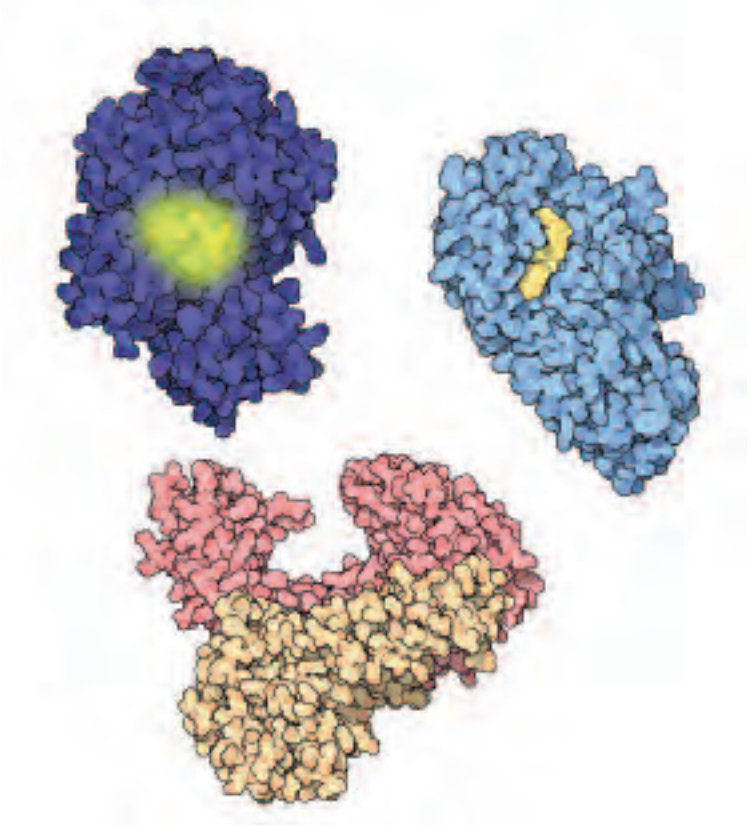
This final shape enables proteins to accomplish their function in your body.

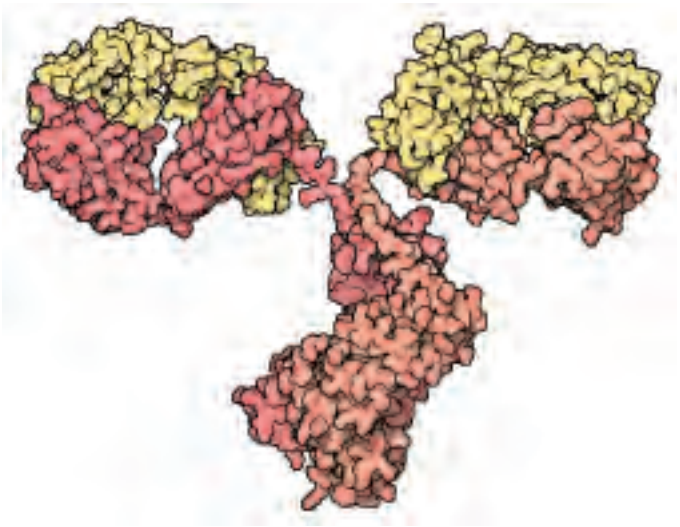
# *Proteins in All Shapes and Sizes*

Because proteins have diverse roles in the body, they come in many shapes and sizes. Studies of these shapes teach us how the proteins function in our bodies and help us understand diseases caused by abnormal proteins.

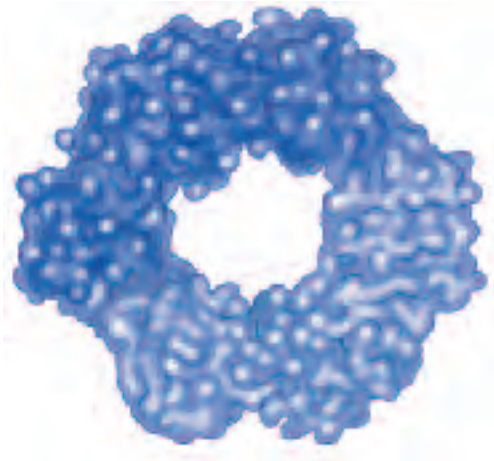
To learn more about the proteins shown here, and many others, check out the Molecule of the Month section of the RCSB Protein Data Bank (http://www.pdb.org).

Molecule of the Month images by David S. Goodsell, The Scripps Research Institute





Antibodies are immune system proteins that rid the body of foreign material, including bacteria and viruses. The two arms of the Y-shaped antibody bind to a foreign molecule. The stem of the antibody sends signals to recruit other members of the immune system.



▲ Some proteins latch onto and regulate the activity of our genetic material, DNA. Some of these proteins are donut shaped, enabling them to form a complete ring around the DNA. Shown here is DNA polymerase III, which cinches around DNA and moves along the strands as it copies the genetic material.

Enzymes, which are proteins that facilitate chemical reactions, often contain a groove or pocket to hold the molecule they act upon. Shown here (clockwise from top) are luciferase, which creates the yellowish light of fireflies; amylase, which helps us digest starch; and reverse transcriptase, which enables HIV and related viruses to enslave infected cells.

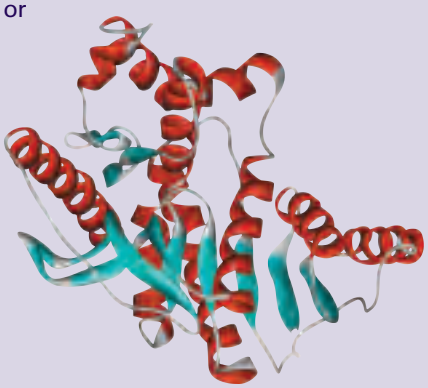
▲ Collagen in our cartilage and tendons gains its strength from its three-stranded, rope-like structure.

## *Computer Graphics Advance Research*

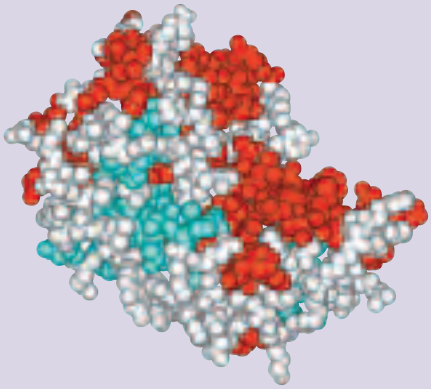
Decades ago, scientists who wanted to study three-dimensional molecular structures spent days, weeks, or longer building models out of rods, balls, and wire scaffolding.

Today, they use computer graphics. Within seconds, scientists can display a molecule in several different ways (like the three representations of a single protein shown here), manipulate it on the computer screen, simulate how it might interact with other molecules, and study how defects in its structure could cause disease.

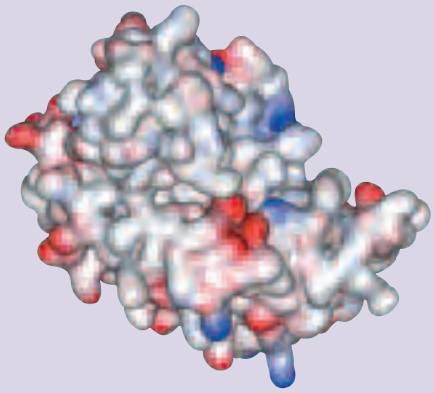
To try one of these computer graphics programs, go to http://www.proteinexplorer.org or http://www.pdb.org.



▲ A ribbon diagram highlights organized regions of the protein (red and light blue).



▲ A space-filling molecular model attempts to show atoms as spheres whose sizes correlate with the amount of space the atoms occupy. The same atoms are colored red and light blue in this model and in the ribbon diagram.

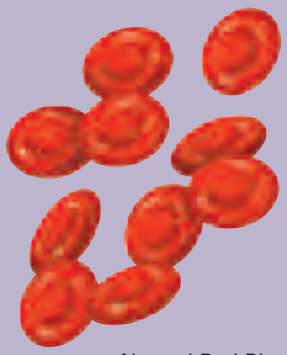


▲ A surface rendering of the same protein shows its overall shape and surface properties. The red and blue coloration indicates the electrical charge of atoms on the protein's surface.

# *Small Errors in Proteins Can Cause Disease*

Sometimes, an error in just one amino acid can cause disease. Sickle cell disease, which most often affects those of African descent, is caused by a single error in the gene for hemoglobin, the oxygen-carrying protein in red blood cells.

This error, or mutation, results in an incorrect amino acid at one position in the molecule. Hemoglobin molecules with this incorrect amino acid stick together and distort the normally smooth, lozenge-shaped red blood cells into jagged sickle shapes.



Normal Red Blood Cells

Sickled Red Blood Cells

The most common symptom of the disease is unpredictable pain in any body organ or joint, caused when the distorted blood cells jam together, unable to pass through small blood vessels. These blockages prevent oxygen-carrying blood from getting to organs and tissues. The frequency, duration, and severity of this pain vary greatly between individuals.

The disease affects about 1 in every 500 African Americans, and 1 in 12 carry the trait and can pass it on to their children, but do not have the disease themselves.

Another disease caused by a defect in one amino acid is cystic fibrosis. This disease is most common in those of northern European descent, affecting about 1 in 2,500 Caucasians in the United States. Another 1 in 25 or 30 are carriers.

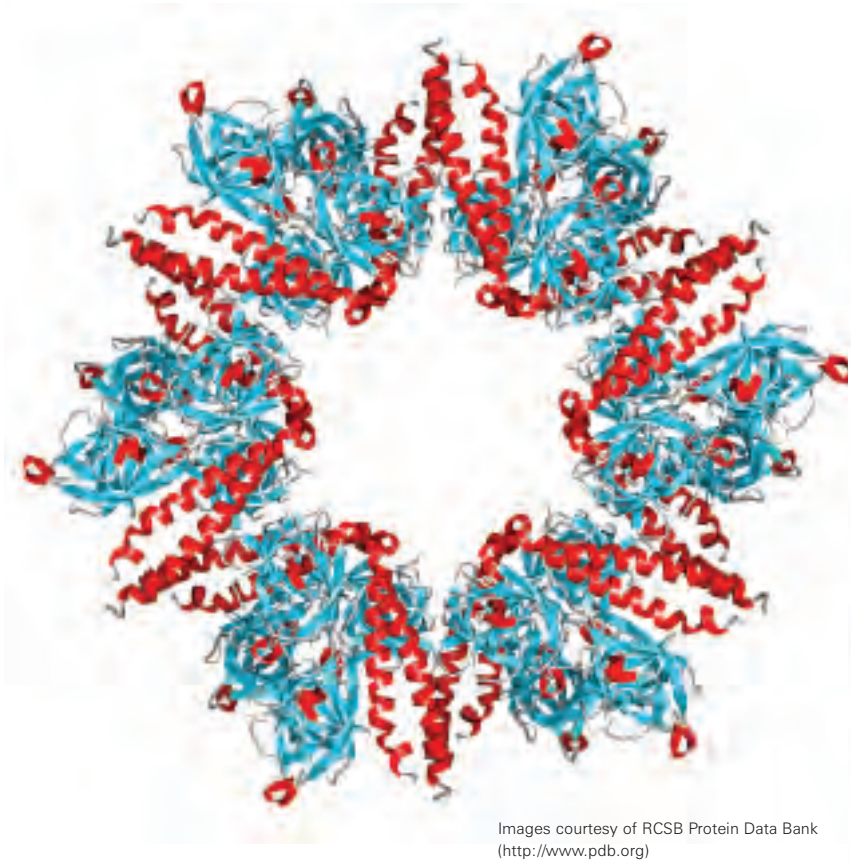
The disease is caused when a protein called CFTR is incorrectly folded. This misfolding is usually caused by the deletion of a single amino acid in CFTR. The function of CFTR, which stands for cystic fibrosis transmembrane conductance regulator, is to allow chloride ions (a component of table salt) to pass through the outer membranes of cells.

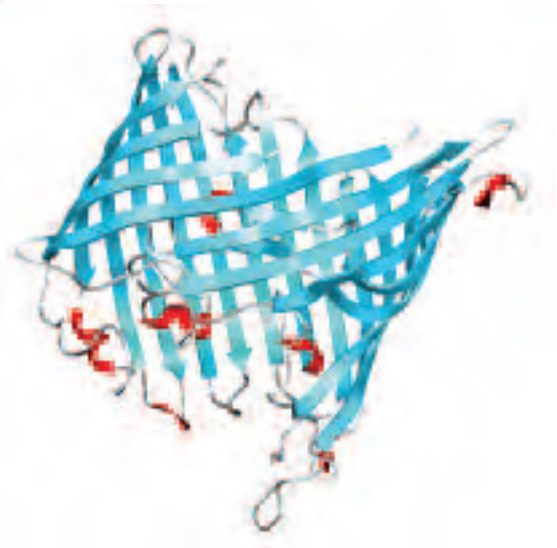
When this function is disrupted in cystic fibrosis, glands that produce sweat and mucus are most affected. A thick, sticky mucus builds up in the lungs and digestive organs, causing malnutrition, poor growth, frequent respiratory infections, and difficulties breathing. Those with the disorder usually die from lung disease around the age of 35.

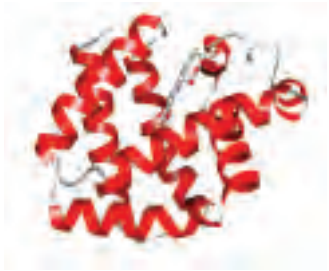
## Parts of Some Proteins Fold Into Corkscrews

When proteins fold, they don't randomly wad up into twisted masses. Often, short sections of proteins form recognizable shapes. Where a protein chain curves into a corkscrew, that section is called an alpha helix. Where it forms a flattened strip, it is a beta sheet.

These organized sections of a protein pack together with each other—or with other, less organized sections—to form the final, folded protein. Some proteins contain mostly alpha helices (red in the ribbon diagrams below). Others contain mostly beta sheets (light blue), or a mix of alpha helices and beta sheets.







Images courtesy of RCSB Protein Data Bank (http://www.pdb.org)

## Mountain Climbing and Computational Modeling

*Many scientists use computers to try to solve the protein folding problem. One example is David Baker, a mountain climber and computational biologist at the University of Washington. He designs software to predict protein structures—and harnesses unused computer power from college dorm rooms to do so. Read about it at http://publications.nigms.nih.gov/findings/sept05/business.html.*

## The Problem of Protein Folding

A given sequence of amino acids almost always folds into a characteristic, three-dimensional structure. So scientists reason that the instructions for folding a protein must be encoded within this sequence. Researchers can easily determine a protein's amino acid sequence. But for more than 50 years they've tried—and failed—to crack the code that governs folding.

Scientists call this the "protein folding problem," and it remains one of the great challenges in structural biology. Although researchers have teased out some general rules and, in some cases, can make rough guesses of a protein's shape, they cannot accurately and reliably predict the position of every atom in the molecule based only on the amino acid sequence.

The medical incentives for cracking the folding code are great. Diseases including Alzheimer's, cystic fibrosis, and "mad cow" disease are thought to result from misfolded proteins. Many scientists believe that if we could decipher the structures of proteins from their sequences, we could better understand how the proteins function and malfunction. Then we could use that knowledge to improve the treatment of these diseases.

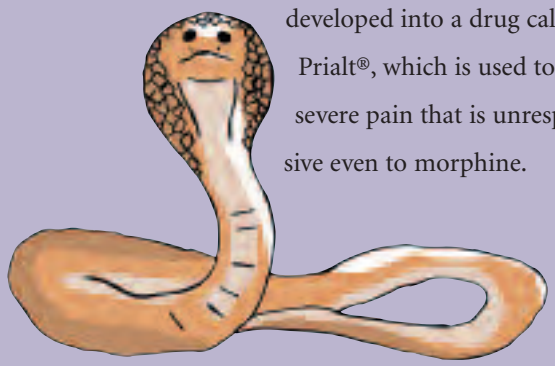
# *Provocative Proteins*

- Each one of us has several hundred thousand different proteins in our body.

- Spider webs and silk fibers are made of the strong, pliable protein fibroin. Spider silk is stronger than a steel rod of the same diameter, yet it is much more elastic, so scientists hope to use it for products as diverse as bulletproof vests and artificial joints. The difficult part is harvesting the silk, because spiders are much less cooperative than silkworms!

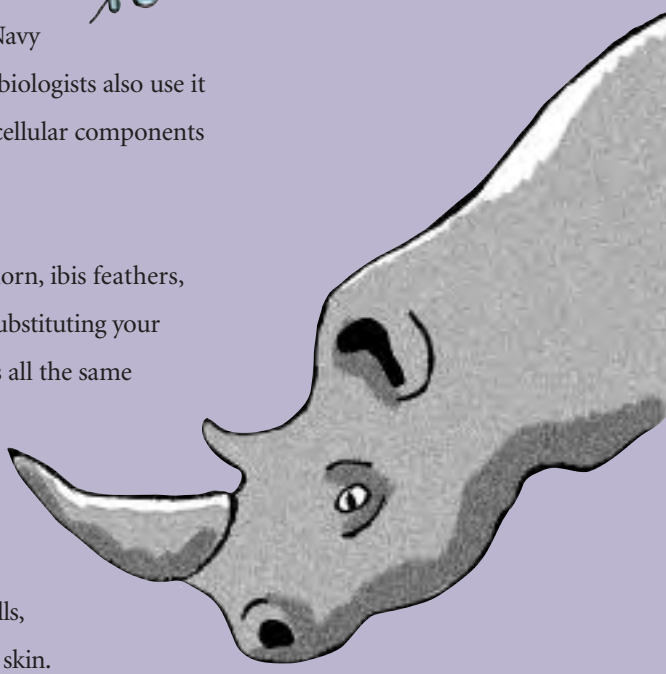
- The light of fireflies (also called lightning bugs) is made possible by a protein called luciferase. Although most predators stay away from the bitter-tasting insects, some frogs eat so many fireflies that they glow!

- The deadly venoms of cobras, scorpions, and puffer fish contain small proteins that act as nerve toxins. Some sea snails stun their prey (and occasionally, unlucky humans) with up to 50 such toxins. One of these toxins has been developed into a drug called Prialt®, which is used to treat severe pain that is unresponsive even to morphine.



- Sometimes ships in the northwest Pacific Ocean leave a trail of eerie green light. The light is produced by a protein in jellyfish when the creatures are jostled by ships. Because the trail traces the path of ships at night, this green fluorescent protein has interested the Navy for many years. Many cell biologists also use it to fluorescently mark the cellular components they are studying.

- If a recipe calls for rhino horn, ibis feathers, and porcupine quills, try substituting your own hair or fingernails. It's all the same stuff — alpha-keratin, a tough, water-resistant protein that is also the main component of wool, scales, hooves, tortoise shells, and the outer layer of your skin.



## Structural Genomics: From Gene to Structure, and Perhaps Function

The potential value of cracking the protein folding code skyrocketed after the launch, in the 1990s, of genome sequencing projects. These ongoing projects give scientists ready access to the complete genetic sequence of hundreds of organisms — including humans.

From these genetic sequences, scientists can easily obtain the corresponding amino acid sequences using the "genetic code" (see page 12).

The availability of complete genome sequences (and amino acid sequences) has opened up new avenues of research, such as studying the structure of all proteins from a single organism or comparing, across many different species, proteins that play a specific biological role.

As part of the Protein Structure Initiative, research teams across the nation have determined thousands of molecular structures, including this structure of a protein from the organism that causes tuberculosis.

Courtesy of the TB Structural Genomics Consortium

The ultimate dream of structural biologists around the globe is to determine directly from genetic sequences not only the three-dimensional structure, but also some aspects of the function of all proteins.

They are partially there: They have identified amino acid sequences that code for certain structural features, such as a cylinder woven from beta sheets.

Researchers have also cataloged structural features that play specific biological roles. For example, a characteristic cluster of alpha helices strongly suggests that the protein binds to DNA.

But that is a long way from accurately determining a protein's structure based only on its genetic or amino acid sequence. Scientists recognized that achieving this long-term goal would require a focused, collaborative effort. So was born a new field called structural genomics.

In 2000, NIGMS launched a project in structural genomics called the Protein Structure Initiative or PSI (http://www.nigms.nih.gov/Initiatives/PSI). This multimillion-dollar project involves hundreds of scientists across the nation.

The PSI scientists are taking a calculated shortcut. Their strategy relies on two facts.

First, proteins can be grouped into families based on their amino acid sequence. Members of the same protein family often have similar structural features, just as members of a human family might all have long legs or high cheek bones.

Second, sophisticated computer programs can use previously solved structures as guides to predict other protein structures.

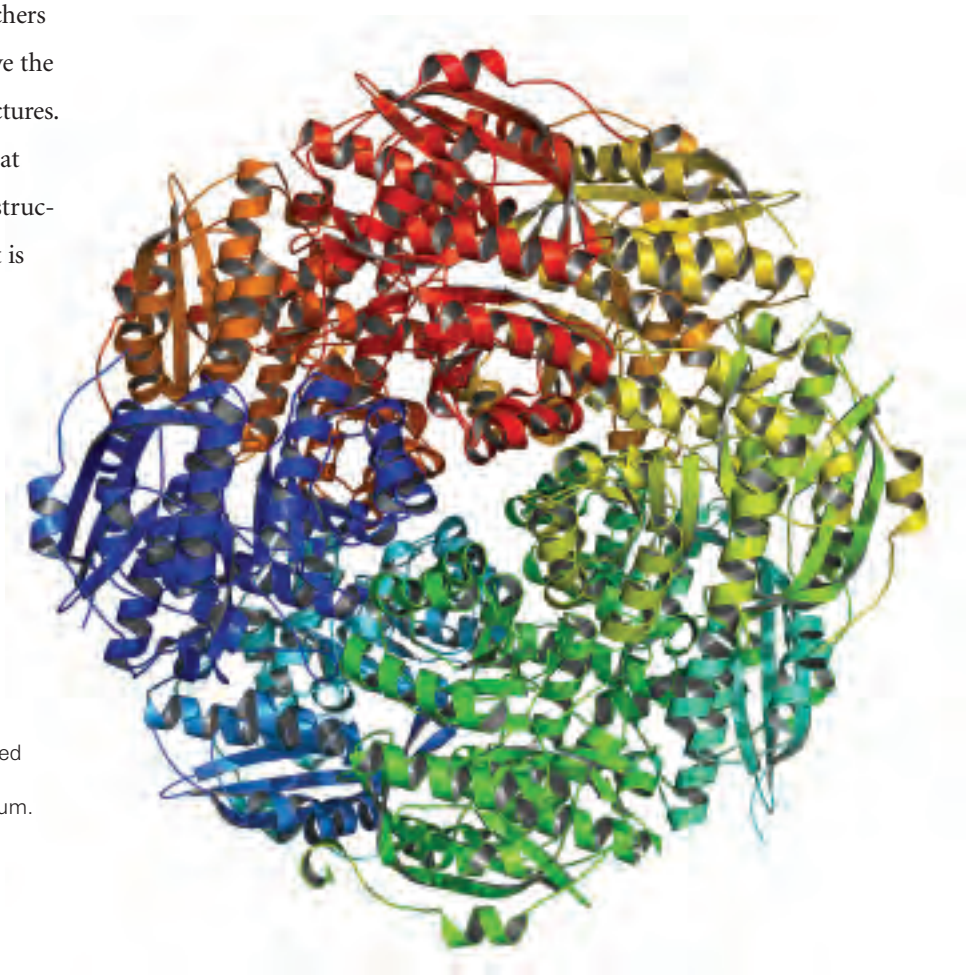
The PSI team expects that, if they solve a few thousand carefully selected protein structures, they can use computer modeling to predict the structures of hundreds of thousands of related proteins.

Already, the PSI team has solved a total of more than 2400 structures. Of these, more than 1600 appear unrelated, suggesting that they might serve as guides for modeling the structures of other proteins in their families.

Perhaps even more significant, PSI researchers have developed new technologies that improve the speed and ease of determining molecular structures. Many of these new technologies are robots that automate previously labor-intensive steps in structure determination. Thanks to these robots, it is possible to solve structures faster than ever before. Besides benefiting the PSI team, these technologies have accelerated research in other fields.

PSI scientists (and structural biologists worldwide) send their findings to the Protein Data Bank at http://www.pdb.org. There, the information is freely available to advance research by the broader scientific community.

To see other structures solved by the PSI team, go to http://publications.nigms.nih.gov/psi/gallery/psi.htm.



Members of the Protein Structure Initiative determined this structure of an enzyme from a common soil bacterium.

Courtesy of the New York Structural GenomiX Consortium

# *The Genetic Code*

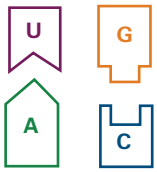
In addition to the protein folding code, which remains unbroken, there is another code, a genetic code, that scientists cracked in the mid-1960s. The genetic code reveals how living organisms use genes as instruction manuals to make proteins.

| 1st mRNA Letter | U | C | A | G |
|---|---|---|---|---|
| U | UUU phenylalanine | CUU leucine | AUU isoleucine | GUU valine |
| | UUC phenylalanine | CUC leucine | AUC isoleucine | GUC valine |
| | UUA leucine | CUA leucine | AUA isoleucine | **GUA valine** |
| | UUG leucine | CUG leucine | **AUG methionine** | GUG valine |

DNA Nucleotides


▲ DNA (deoxyribonucleic acid) is composed of small molecules called nucleotides, which are named for the main unit they contain: adenine (A), thymine (T), cytosine (C), and guanine (G).

RNA Nucleotides




▲ RNA (ribonucleic acid) is chemically very similar to DNA, but uses uracil (U) where DNA uses thymine (T).

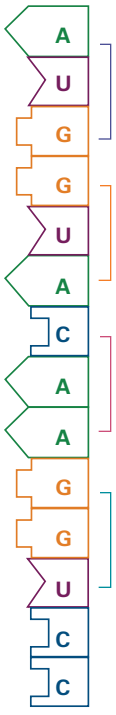
Gene


▲ Genes are long stretches of DNA.

Transcription

▲ Genes are transcribed into complementary strands of messenger RNA (mRNA).

mRNA




Translation

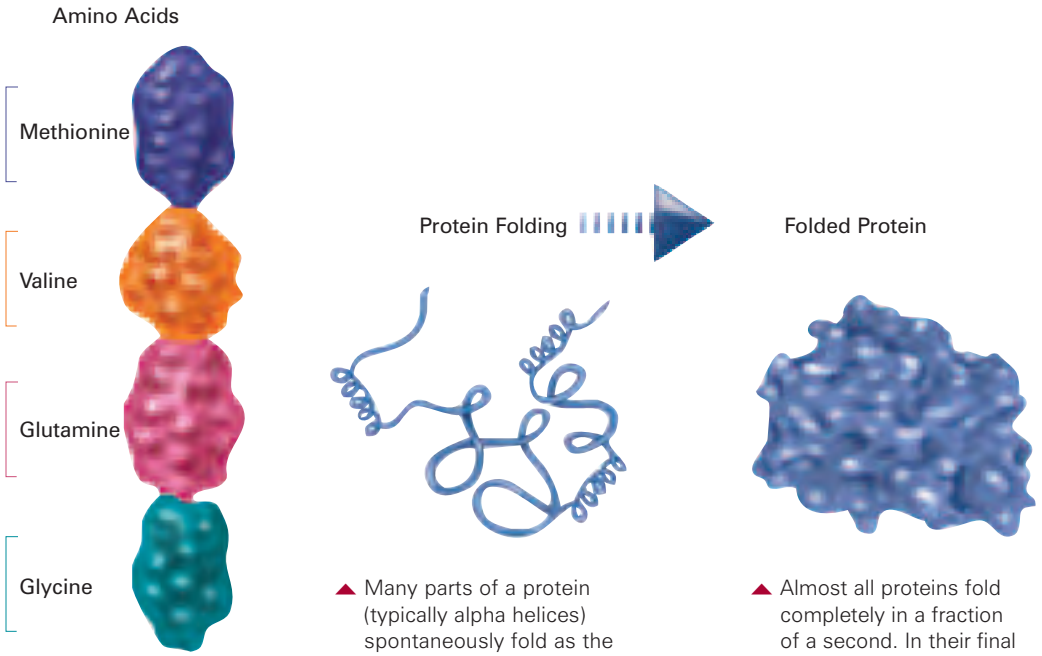
▲ Ribosomes (see p. 23) make proteins by using mRNA instructions and the genetic code to join amino acids together in the right order. Three adjacent mRNA nucleotides (a triplet) encode one amino acid.

## Genetic Code

2nd mRNA Letter

| C | A | G |
|---|---|---|
| UCU serine | UAU tyrosine | UGU cysteine |
| UCC serine | UAC tyrosine | UGC cysteine |
| UCA serine | UAA stop | UGA stop |
| UCG serine | UAG stop | UGG tryptophan |
| CCU proline | CAU histidine | CGU arginine |
| CCC proline | CAC histidine | CGC arginine |
| CCA proline | **CAA glutamine** | CGA arginine |
| CCG proline | CAG glutamine | CGG arginine |
| ACU threonine | AAU asparagine | AGU serine |
| ACC threonine | AAC asparagine | AGC serine |
| ACA threonine | AAA lysine | AGA arginine |
| ACG threonine | AAG lysine | AGG arginine |
| GCU alanine | GAU aspartic acid | **GGU glycine** |
| GCC alanine | GAC aspartic acid | GGC glycine |
| GCA alanine | GAA glutamic acid | GGA glycine |
| GCG alanine | GAG glutamic acid | GGG glycine |

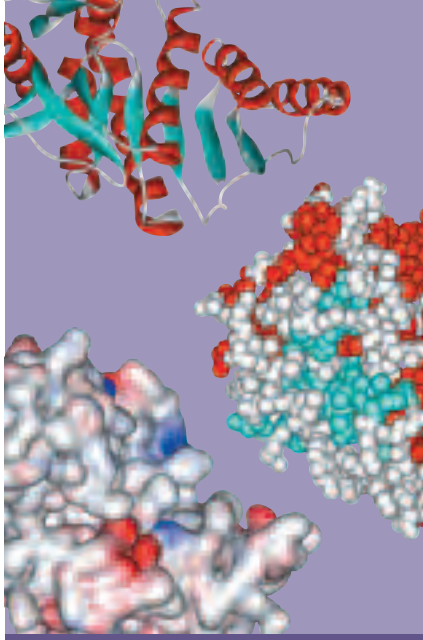
This table shows all possible mRNA triplets and the amino acids they specify. Note that most amino acids may be specified by more than one mRNA triplet. The highlighted entries are shown in the illustration below.




▲ Proteins typically contain from 50 to 2,000 amino acids. Many proteins include two or more strands of amino acids.

▲ Many parts of a protein (typically alpha helices) spontaneously fold as the protein is made. To finish folding, most proteins require the assistance of chaperone proteins.

▲ Almost all proteins fold completely in a fraction of a second. In their final form, some proteins contain metal atoms or other small functional groups.



## Got It?

What is a protein?

Name three proteins in your body and describe what they do.

What do we learn from studying the structures of proteins?

Describe the protein folding problem.

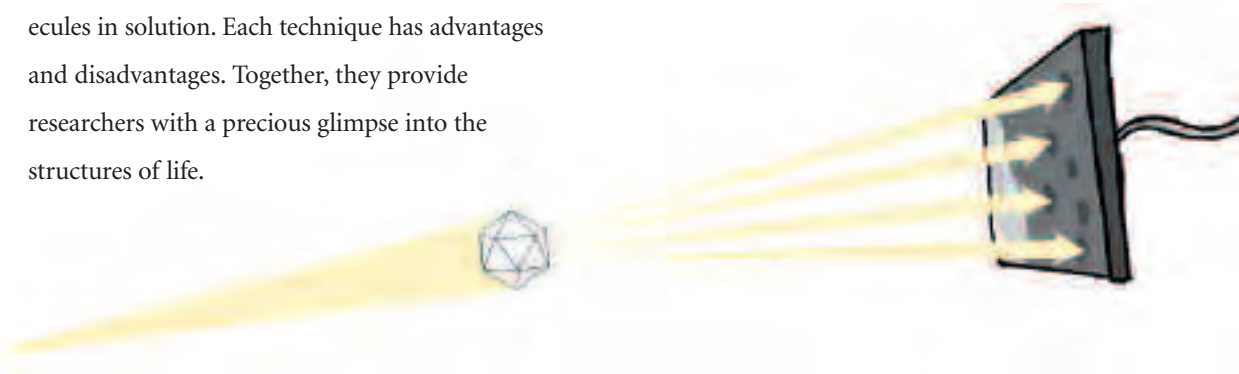
# *X-Ray Crystallography: Art Marries Science*

How would you examine the shape of something too small to see in even the most powerful microscope? Scientists trying to visualize the complex arrangement of atoms within molecules have exactly that problem, so they solve it indirectly. By using a large collection of identical molecules — often proteins — along with specialized equipment and computer modeling techniques, scientists are able to calculate what an isolated molecule would look like.

The two most common methods used to investigate molecular structures are X-ray crystallography (also called X-ray diffraction) and nuclear magnetic resonance (NMR) spectroscopy. Researchers using X-ray crystallography grow solid crystals of the molecules they study. Those using NMR study molecules in solution. Each technique has advantages and disadvantages. Together, they provide researchers with a precious glimpse into the structures of life.

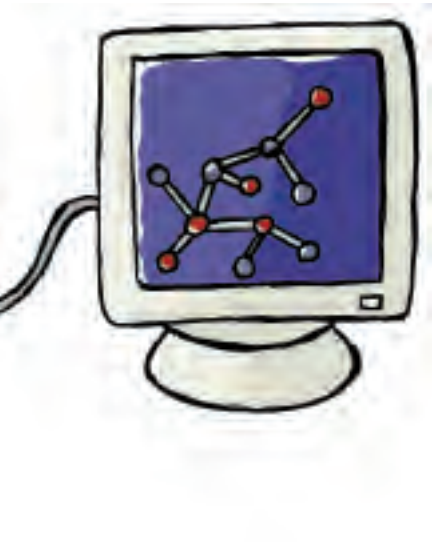
More than 85 percent of the protein structures that are known have been determined using X-ray crystallography. In essence, crystallographers aim high-powered X-rays at a tiny crystal containing trillions of identical molecules. The crystal scatters the X-rays onto an electronic detector like a disco ball spraying light across a dance floor. The electronic detector is the same type used to capture images in a digital camera.

After each blast of X-rays, lasting from a few seconds to several hours, the researchers precisely rotate the crystal by entering its desired orientation into the computer that controls the X-ray apparatus. This enables the scientists to capture in three dimensions how the crystal scatters, or diffracts, X-rays.

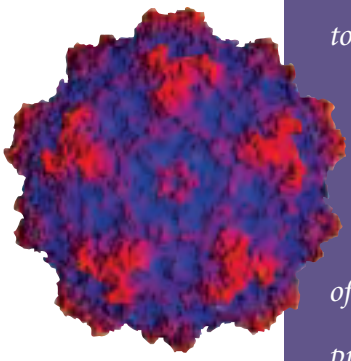


The intensity of each diffracted ray is fed into a computer, which uses a mathematical equation called a Fourier transform to calculate the position of every atom in the crystallized molecule.

The result — the researchers' masterpiece — is a three-dimensional digital image of the molecule. This image represents the physical and chemical properties of the substance and can be studied in intimate, atom-by-atom detail using sophisticated computer graphics software.



Computed Image of Atoms in Crystal



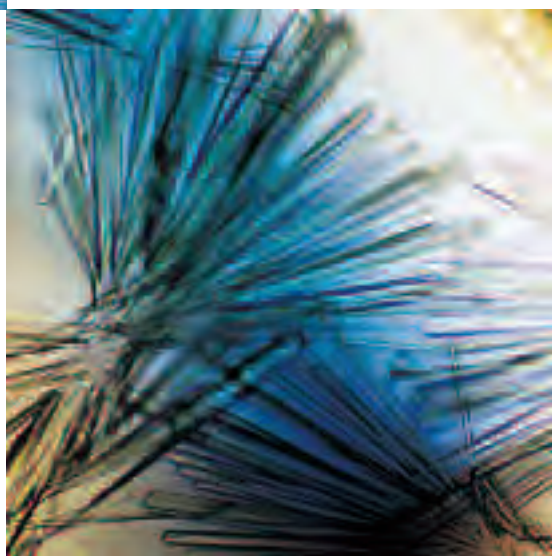
▲ Agbandje-McKenna's three-dimensional structure of a mouse virus shows that it resembles a 20-sided soccer ball.

## Viral Voyages

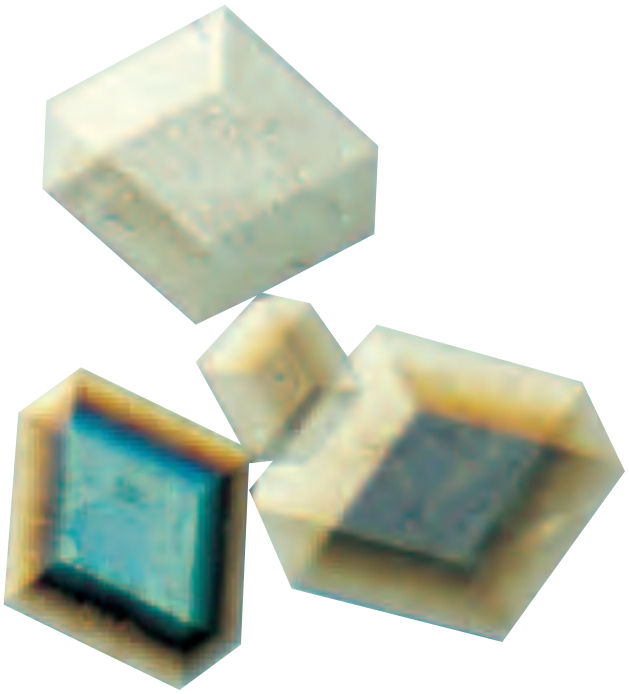
*Using X-ray crystallography, scientists can study enormous viruses that contain several hundred proteins. Mavis Agbandje-McKenna uses the technique to investigate how viruses infect cells. Read about her unusual scientific and personal journey from a rural village in Nigeria to the University of Florida in Gainesville, at http://publications.nigms.nih.gov/findings/mar06/voyages.html.*

## Crystal Cookery

An essential step in X-ray crystallography is growing high-quality crystals. The best crystals are pure, perfectly symmetrical, three-dimensional repeating arrays of precisely packed molecules. They can be different shapes, from perfect cubes to long needles. Most crystals used for these studies are barely visible (less than 1 millimeter on a side). But the larger the crystal, the more accurate the data and the more easily scientists can solve the structure.

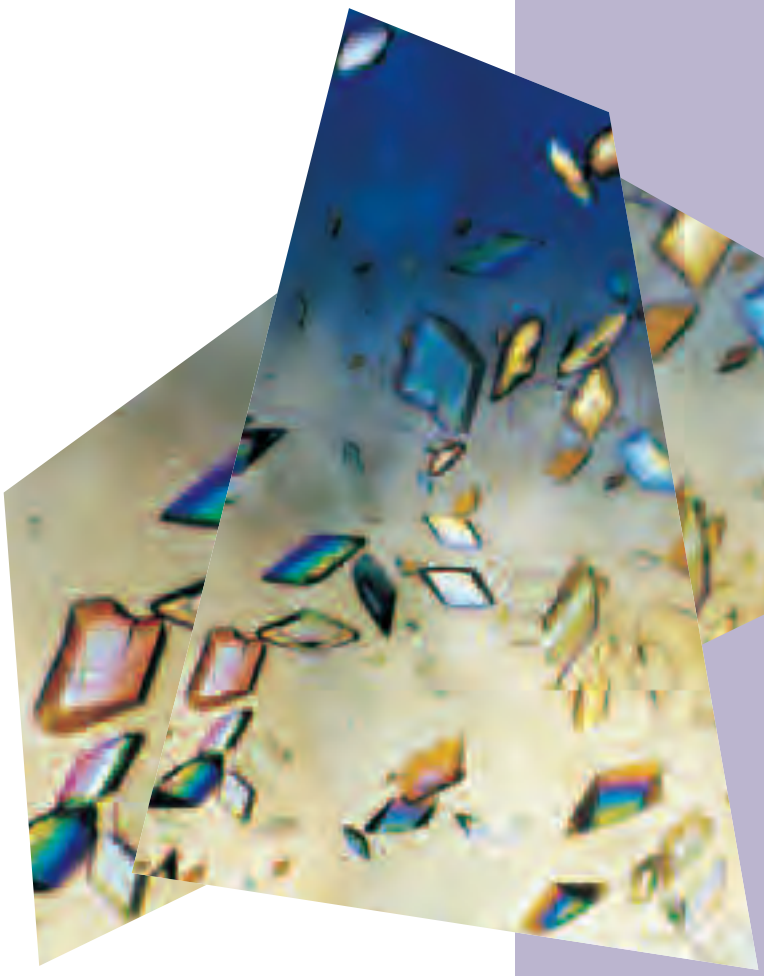


Crystallographers grow their tiny crystals in plastic dishes. They usually start with a highly concentrated solution containing the molecule. They then mix this solution with a variety of specially prepared liquids to form tiny droplets (1-10 microliters). Each droplet is kept in a separate plastic dish or well. As the liquid evaporates, the molecules in the solution become progressively more concentrated. During this process, the molecules arrange into a precise, three-dimensional pattern and eventually into a crystal — if the researcher is lucky.



Sometimes, crystals require months or even years to grow. The conditions — temperature, pH (acidity or alkalinity), and concentration — must be perfect. And each type of molecule is different, requiring scientists to tease out new crystallization conditions for every new sample.

Even then, some molecules just won't cooperate. They may have floppy sections that wriggle around too much to be arranged neatly into a crystal. Or, particularly in the case of proteins that are normally embedded in oily cell membranes, the molecule may fail to completely dissolve in the solution.



# *Calling All Crystals*

Although the crystals used in X-ray crystallography are barely visible to the naked eye, they contain a vast number of precisely ordered, identical molecules. A crystal that is 0.5 millimeters on each side contains around 1,000,000,000,000,000 (or $10^{15}$) medium-sized protein molecules.

When the crystals are fully formed, they are placed in a tiny glass tube or scooped up with a loop made of nylon, glass fiber, or other material depending on the preference of the researcher. The tube or loop is then mounted in the X-ray apparatus, directly in the path of the X-ray beam.

The searing force of powerful X-ray beams can burn holes through a crystal left too long in their path. To minimize radiation damage, researchers flash-freeze their crystals in liquid nitrogen.

Some crystallographers keep their growing crystals in air-locked chambers, to prevent any misdirected breath from disrupting the tiny crystals. Others insist on an environment free of vibrations — in at least one case, from rock-and-roll music. Still others joke about the phases of the moon and supernatural phenomena. As the jesting suggests, growing crystals remains one of the most difficult and least predictable parts of X-ray crystallography. It's what blends art with the science.

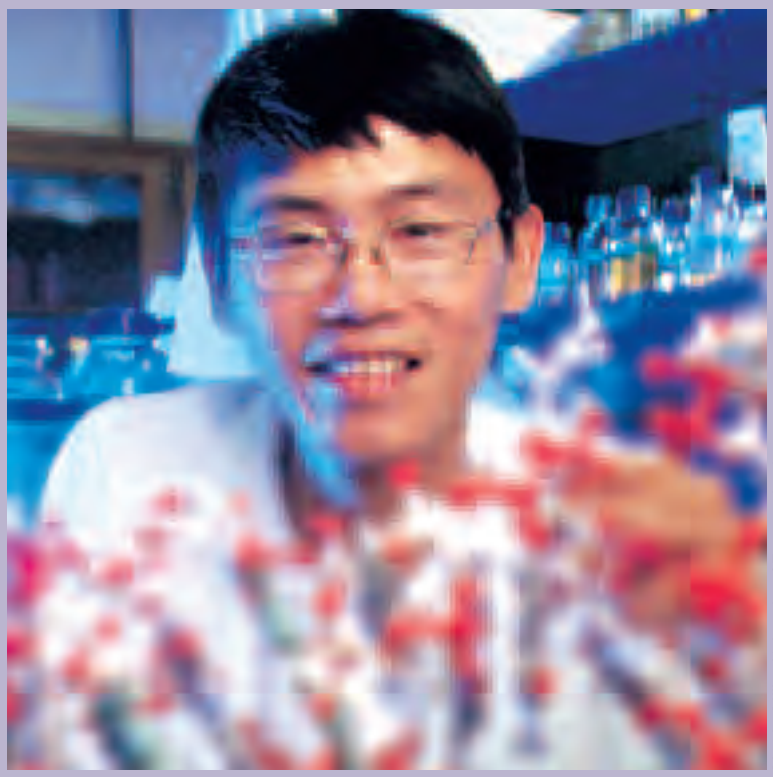
Crystal photos courtesy of Alex McPherson, University of California, Irvine

**STUDENT SNAPSHOT**

# *Science Brought One Student From the Coast of Venezuela to the Heart of Texas*

"Science is like a roller coaster. You start out very excited about what you're doing. But if your experiments don't go well for a while, you get discouraged. Then, out of nowhere, comes this great data and you are up and at it again."

That's how Juan Chang describes the nature of science. He majored in biochemistry and computer science at the University of Texas at Austin. He also worked in the UT-Austin laboratory of X-ray crystallographer Jon Robertus.



Marsha Miller, University of Texas at Austin

Chang studied a protein that prevents cells from committing suicide. As a sculptor chips and shaves off pieces of marble, the body uses cellular suicide, also called "apoptosis," during normal development to shape features like fingers and toes. To protect healthy cells, the body also triggers apoptosis to kill cells that are genetically damaged or infected by viruses.

By understanding proteins involved in causing or preventing apoptosis, scientists hope to control the process in special situations — to help treat tumors and viral infections by promoting the death of damaged cells, and to treat degenerative nerve diseases by preventing apoptosis in nerve cells. A better understanding of apoptosis may even allow researchers to more easily grow tissues for organ transplants.

Chang was part of this process by helping to determine the X-ray crystal structure of a protein

*"Science is like a roller coaster. You start out very excited about what you're doing. But if your experiments don't go well for a while, you get discouraged. Then, out of nowhere, comes this great data and you are up and at it again."*

Juan Chang
Graduate Student
Baylor College of Medicine

that scientists refer to as ch-IAP1. He used biochemical techniques to obtain larger quantities of this purified protein. The next step will be to crystallize the protein, then to use X-ray diffraction to obtain its detailed, three-dimensional structure.

Chang came to Texas from a lakeside town on the northwest tip of Venezuela. He first became interested in biological science in high school. His class took a field trip to an island off the Venezuelan coast to observe the intricate ecological balance of the beach and coral reef. He was impressed at how the plants and animals — crabs, insects, birds, rodents, and seaweed — each adapted to the oceanside wind, waves, and salt.

About the same time, his school held a fund drive to help victims of Huntington's disease, an incurable genetic disease that slowly robs people of their ability to move and think properly. The town in which Chang grew up, Maracaibo, is home to the largest known family with Huntington's disease. Through the fund drive, Chang became interested in the genetic basis of inherited diseases.

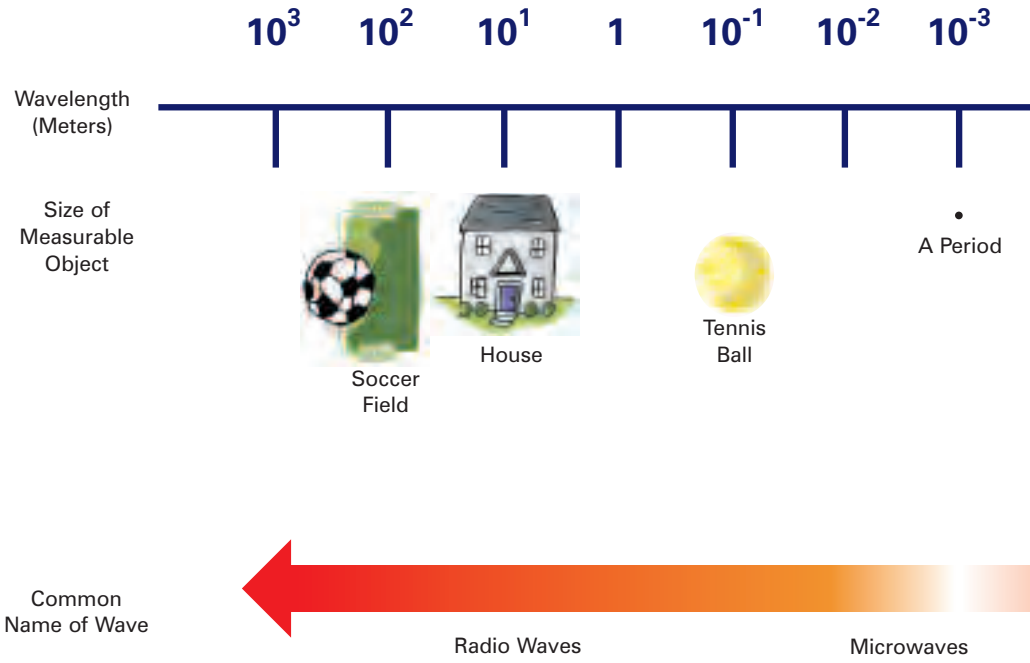
His advice for anyone considering a career in science is to "get your hands into it" and to experiment with work in different fields. He was initially interested in genetics, did biochemistry research, and is now in a graduate program at Baylor College of Medicine. The program combines structural and computational biology with molecular biophysics. He anticipates that after earning a Ph.D., he will become a professor at a university.

## Why X-Rays?

In order to measure something accurately, you need the appropriate ruler. To measure the distance between cities, you would use miles or kilometers. To measure the length of your hand, you would use inches or centimeters.

Crystallographers measure the distances between atoms in angstroms. One angstrom equals one ten-billionth of a meter, or $10^{-10}$m. That's more than 10 million times smaller than the diameter of the period at the end of this sentence.

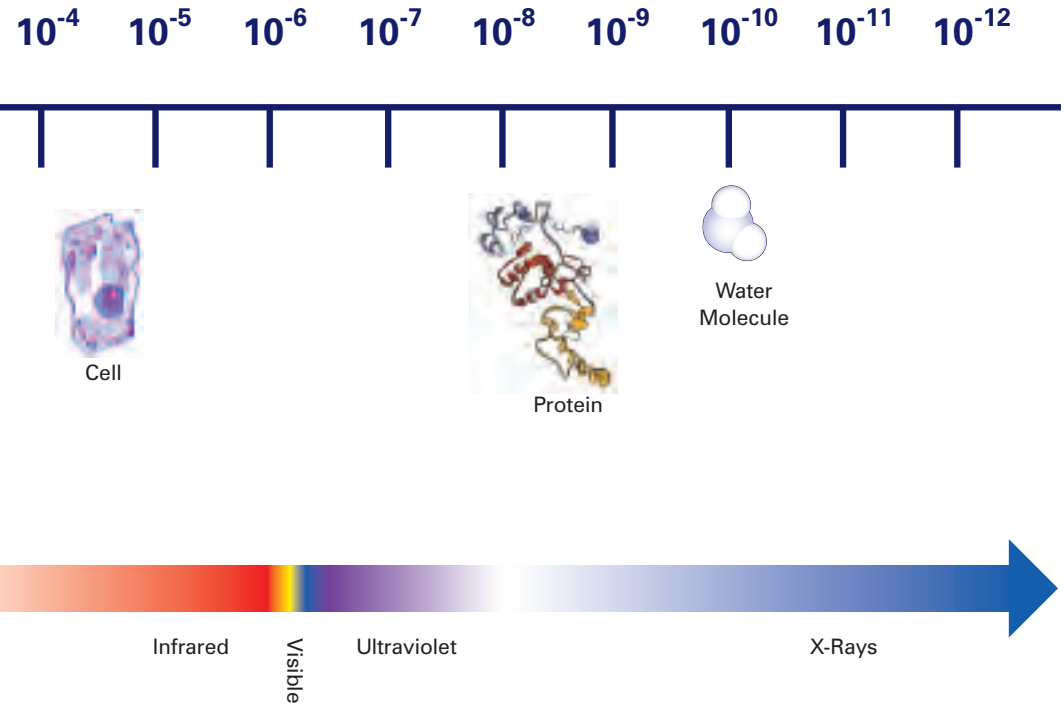
The perfect "rulers" to measure angstrom distances are X-rays. The X-rays used by crystallographers are approximately 0.5 to 1.5 angstroms long — just the right size to measure the distance between atoms in a molecule. There is no better place to generate such X-rays than in a synchrotron.



## Synchrotron Radiation—One of the Brightest Lights on Earth

Imagine a beam of light 30 times more powerful than the Sun, focused on a spot smaller than the head of a pin. It carries the blasting power of a meteor plunging through the atmosphere. And it is the single most powerful tool available to X-ray crystallographers.

This light, one of the brightest lights on earth, is not visible to our eyes. It is made of X-ray beams generated in large machines called synchrotrons. These machines accelerate electrically charged particles, often electrons, to nearly the speed of light, then whip them around a huge, hollow metal ring.




When using light to measure an object, the wavelength of the light needs to be similar to the size of the object. X-rays, with wavelengths of approximately 0.5 to 1.5 angstroms, can measure the distance between atoms. Visible light, with a wavelength of 4,000 to 7,000 angstroms, is used in ordinary light microscopes because it can measure objects the size of cellular components.


Argonne National Laboratory

▲ The Advanced Photon Source (APS) at Argonne National Laboratory near Chicago is a "third-generation" synchrotron radiation facility. Biologists were considered parasitic users on the "first-generation" synchrotrons, which were built for physicists studying subatomic particles. Now, many synchrotrons, such as the APS, are designed specifically to optimize X-ray production and support the research of scientists in a variety of fields, including biology.

Synchrotrons were originally designed for use by high-energy physicists studying subatomic particles and cosmic phenomena. Other scientists soon clustered at the facilities to snatch what the physicists considered an undesirable byproduct — brilliant bursts of X-rays.

The largest component of each synchrotron is its electron storage ring. This ring is actually not a perfect circle, but a many-sided polygon. At each corner of the polygon, precisely aligned magnets bend the electron stream, forcing it to stay in the ring (on their own, the particles would travel straight ahead and smash into the ring's wall). Each time the electrons' path is bent, they emit bursts of energy in the form of electromagnetic radiation.

This phenomenon is not unique to electrons or to synchrotrons. Whenever any charged particle changes speed or direction, it emits energy. The type of energy, or radiation, that particles emit depends on the speed the particles are going and how sharply they are bent. Because particles in a synchrotron are hurtling at nearly the speed of light, they emit intense radiation, including lots of high-energy X-rays.

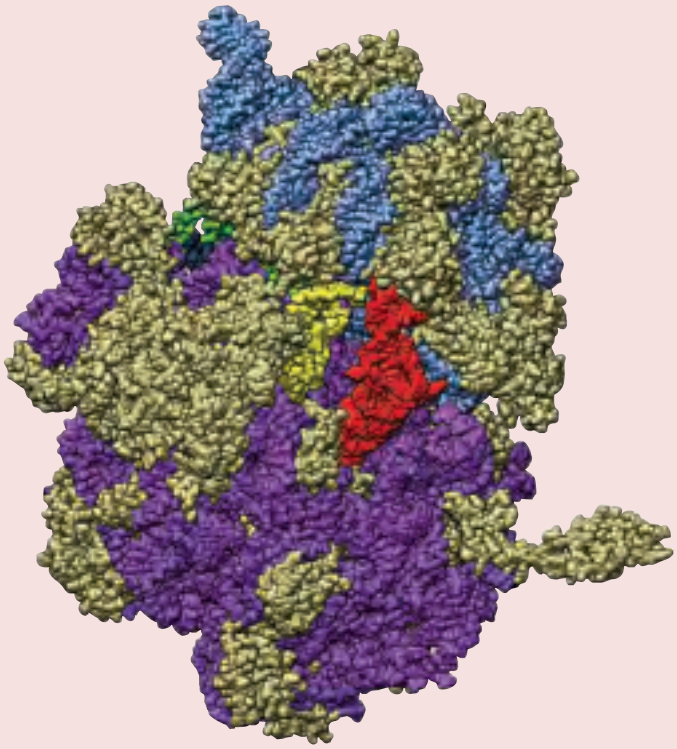
# *Peering Into Protein Factories*

Ribosomes make the stuff of life. They are the protein factories in every living creature, and they churn out all proteins ranging from bacterial toxins to human digestive enzymes.

To most people, ribosomes are extremely small—tens of thousands of ribosomes would fit on the sharpened tip of a pencil. But to a structural biologist, ribosomes are huge. They contain three or four strands of RNA and more than 50 small proteins. These many components work together like moving parts in a complex machine—a machine so large that it has been impossible to study in structural detail until recently.

In 1999, researchers determined the crystal structure of a complete ribosome for the first time. The work was a technical triumph for crystallography. Even today, the ribosome remains the largest complex structure obtained by crystallography. (Some larger virus structures have been determined, but the symmetry of these structures greatly simplified the process.)

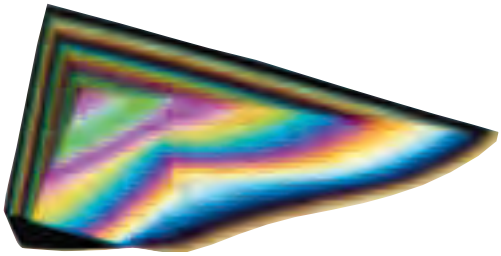
This initial snapshot was like a rough sketch that showed how various parts of the ribosome fit together and where within a ribosome new proteins are made. Today, researchers have extremely detailed images of ribosomes in which they can pinpoint and study every atom.



▲ Examining ribosomal structures in detail will help researchers better understand the fundamental process of protein production. It may also aid efforts to design new antibiotic drugs or optimize existing ones.

Courtesy of Catherine Lawson, Rutgers University and the RCSB Protein Data Bank

In addition to providing valuable insights into a critical cellular component and process, structural studies of ribosomes may lead to clinical applications. Many of today's antibiotics work by interfering with the function of ribosomes in harmful bacteria while leaving human ribosomes alone. A more detailed knowledge of the structural differences between bacterial and human ribosomes may help scientists develop new antibiotic drugs or improve existing ones.

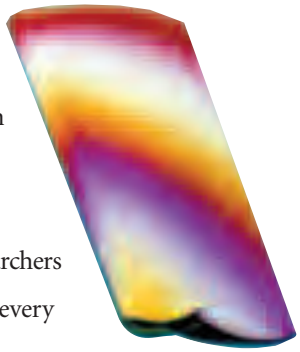


## Scientists Get MAD at the Synchrotron

Synchrotrons are prized not only for their ability to generate brilliant X-rays, but also for the "tunability" of these rays. Scientists can actually select from these rays just the right wavelength for their experiments.

In order to determine the structure of a molecule, crystallographers usually have to compare several versions of a crystal — one pure crystal and several others in which the crystallized molecule is soaked in, or "doped" with, a different heavy metal, like mercury, platinum, or uranium.

Because these heavy metal atoms contain many electrons, they scatter X-rays more than do the smaller, lighter atoms found in biological molecules. By comparing the X-ray scatter patterns of a pure crystal with those of various metal-containing crystals, the researchers can determine the location of the metals in the crystal. These metal atoms serve as landmarks that enable researchers to calculate the position of every other atom in the molecule.




A Berkeley, CA

B Menlo Park, CA

C Baton Rouge, LA

D Argonne, IL

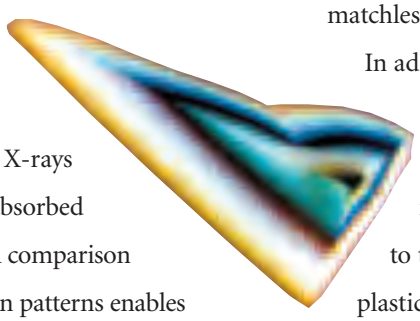
E Upton, NY

F Ithaca, NY

▲ There are half a dozen major synchrotrons used for X-ray crystallography in the United States.

But when using X-ray radiation from the synchrotron, researchers do not have to grow multiple versions of every crystallized molecule — a huge savings in time and money. Instead, they grow only one type of crystal that contains the chemical element selenium instead of sulfur in every methionine amino acid. They then "tune" the wavelength of the synchrotron beam to match certain properties of selenium. That way, a single crystal serves the purpose of several different metal-containing crystals. This technique is called MAD, for Multi-wavelength Anomalous Diffraction.

Using MAD, the researchers bombard the selenium-containing crystals three or four different times, each time with X-ray beams of a different wavelength — including one blast with X-rays of the exact wavelength absorbed by the selenium atoms. A comparison of the resulting diffraction patterns enables researchers to locate the selenium atoms, which again serve as markers, or reference points, around which the rest of the structure is calculated.



The brilliant X-rays from synchrotrons allow researchers to collect their raw data much more quickly than when they use traditional X-ray sources, which are small enough to fit on a long laboratory table and produce much weaker X-rays than do synchrotrons. What used to take weeks or months in the laboratory can be done in minutes at a synchrotron. But then the data still must be analyzed, refined, and corrected before the protein can be visualized in its three-dimensional structural splendor.

The number and quality of molecular structures determined by X-ray diffraction has risen sharply in recent years, as has the percentage of these structures obtained using synchrotrons. This trend promises to continue, due in large part to new techniques like MAD and to the matchless power of synchrotron radiation.

In addition to their role in revealing molecular structures, synchrotrons are used for a variety of applications, including to design computer chips, to test medicines in living cells, to make plastics, to analyze the composition of geological materials, and to study medical imaging and radiation therapy techniques.

Crystal photos courtesy of Alex McPherson, University of California, Irvine

## Got It?

What is meant by the detailed, three-dimensional structure of proteins?

What is X-ray crystallography?

Give two reasons why synchrotrons are so valuable to X-ray crystallographers.

What is a ribosome and why is it important to study?

CHAPTER 3

# *The World of NMR: Magnets, Radio Waves, and Detective Work*

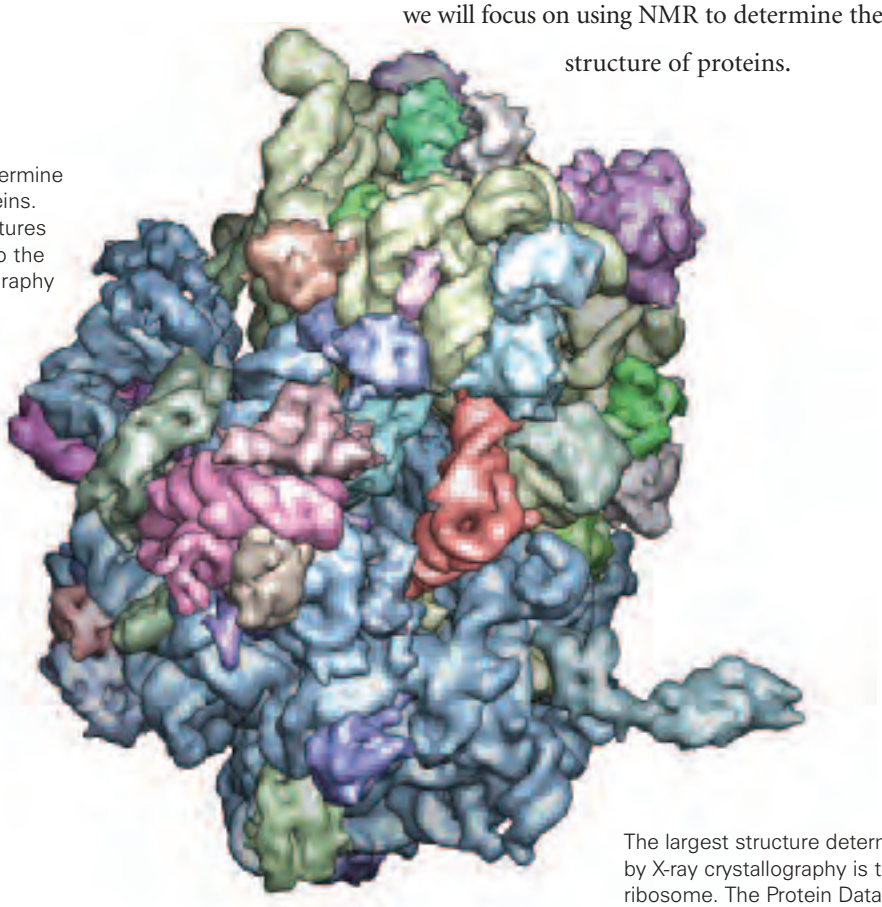
Did you ever play with magnets as a kid? That's a large part of what scientists do when they use a technique called nuclear magnetic resonance (NMR) spectroscopy.

An NMR machine is essentially a huge magnet. Many atoms are essentially little magnets. When placed inside an NMR machine, all the little magnets orient themselves to line up with the big magnet.

By harnessing this law of physics, NMR spectroscopists are able to figure out physical, chemical, electronic, and structural information about molecules.

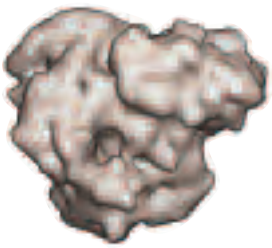
Next to X-ray diffraction, NMR is the most common technique used to determine detailed molecular structures. This technique, which has nothing to do with nuclear reactors or nuclear bombs, is based on the same principle as the magnetic resonance imaging (MRI) machines that allow doctors to see tissues and organs such as the brain, heart, and kidneys.

Although NMR is used for a variety of medical and scientific purposes — including determining the structure of genetic material (DNA and RNA), carbohydrates, and other molecules — in this booklet we will focus on using NMR to determine the structure of proteins.



Currently, NMR spectroscopy is only able to determine the structures of small and medium-sized proteins. Shown here to scale is one of the largest structures determined by NMR spectroscopy compared to the largest structure determined by X-ray crystallography (the ribosome).

Images courtesy of Catherine Lawson, Rutgers University and the RCSB Protein Data Bank



One of the largest structures determined by NMR is malate synthase G, with a mass of 82 kilodaltons.

The largest structure determined by X-ray crystallography is the ribosome. The Protein Data Bank includes many structures of ribosomes, the largest more than 2,000 kilodaltons.

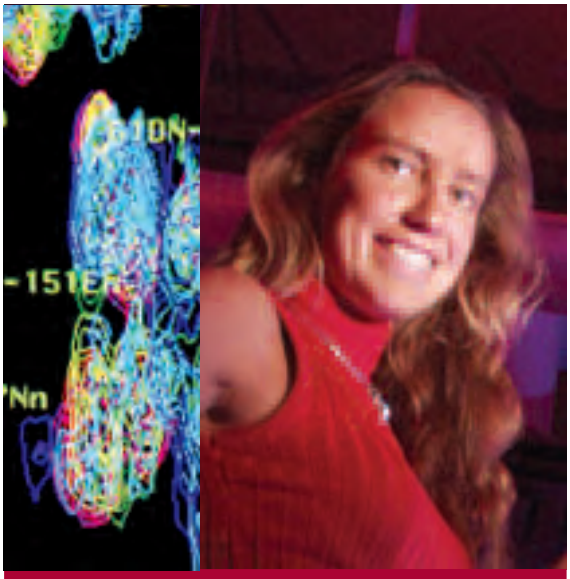
Methods for determining structures by NMR spectroscopy are much younger than those that use X-ray crystallography. As such, they are constantly being refined and improved.

The most obvious area in which NMR lags behind X-ray crystallography is the size of the structures it can handle. Most NMR spectroscopists focus on molecules no larger than 60 kilodaltons (about 180 amino acids). X-ray crystallographers have solved structures up to 2,500 kilodaltons—40 times as large.

But NMR also has advantages over crystallography. For one, it uses molecules in solution, so it is not limited to those that crystallize well. (Remember that crystallization is a very uncertain and time-consuming step in X-ray crystallography.)

NMR also makes it fairly easy to study properties of a molecule besides its structure — such as the flexibility of the molecule and how it interacts with other molecules. With crystallography, it is often either impossible to study these aspects or it requires an entirely new crystal. Using NMR and crystallography together gives researchers a more complete picture of a molecule and its functioning than either tool alone.

NMR relies on the interaction between an applied magnetic field and the natural "little magnets" in certain atomic nuclei. For protein structure determination, spectroscopists concentrate on the atoms that are most common in proteins, namely hydrogen, carbon, and nitrogen.



## A Slam Dunk for Enzymes

*NMR spectroscopy is ideal for studying how enzymes change shape as they do their jobs. Take it from Dorothee Kern, a former professional basketball player who is now an NMR researcher at Brandeis University. Read about her work at http://publications.nigms.*

Before the researchers begin to determine a protein's structure, they already know its amino acid sequence — the names and order of all of its amino acid building blocks. What they seek to learn through NMR is how this chain of amino acids wraps and folds around itself to create the three-dimensional, active protein.

Solving a protein structure using NMR is like a good piece of detective work. The researchers conduct a series of experiments, each of which provides partial clues about the nature of the atoms in the sample molecule — such as how close two atoms are to each other, whether these atoms are physically bonded to each other, or where the atoms lie within the same amino acid. Other experiments show links between adjacent amino acids or reveal flexible regions in the protein.

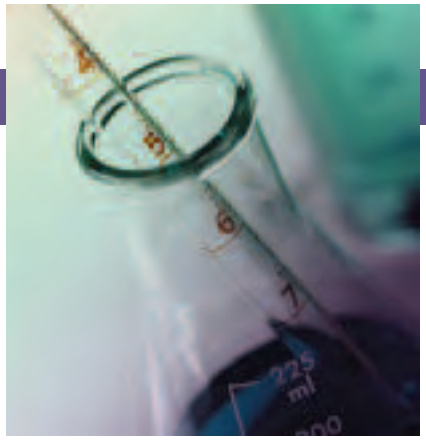
The challenge of NMR is to employ several sets of such experiments to tease out properties unique to each atom in the sample. Using computer programs, NMR spectroscopists can get a rough idea of the protein's overall shape and can see possible arrangements of atoms in its different parts. Each new set of experiments further refines these possible structures. Finally, the scientists carefully select 10 to 20 solutions that best represent their experimental data and present the average of these solutions as their final structure.

## *NMR Spectroscopists Use Tailor-Made Proteins*

Only certain forms, or isotopes, of each chemical element have the correct magnetic properties to be useful for NMR. Perhaps the most familiar isotope is $^{14}C$, which is used for archeological and geological dating.

You may also have heard about isotopes in the context of radioactivity. Neither of the isotopes most commonly used in NMR, namely $^{13}C$ and $^{15}N$, is radioactive.

Like many other biological scientists, NMR spectroscopists (and X-ray crystallographers) use harmless laboratory bacteria to produce proteins for their studies. They insert into these bacteria the gene that codes for the protein under study. This forces the bacteria, which grow and multiply in swirling flasks, to produce large amounts of tailor-made proteins.



To generate proteins that are "labeled" with the correct isotopes, NMR spectroscopists put their bacteria on a special diet. If the researchers want proteins labeled with $^{13}C$, for example, the bacteria are fed food containing $^{13}C$. That way, the isotope is incorporated into all the proteins produced by the bacteria.

# *NMR Magic Is in the Magnets*

The magnets used for NMR are incredibly strong. Those used for high resolution protein structure determination range from 500 megahertz to 900 megahertz and generate magnetic fields thousands of times stronger than the Earth's.

Although the sample is exposed to a strong magnetic field, very little magnetic force gets out of the machine. If you stand next to a very powerful NMR magnet, the most you may feel is a slight tug on hair clips or zippers. But don't get too close if you are wearing an expensive watch or carrying a wallet or purse—NMR magnets are notorious for stopping analog watches and erasing the magnetic strips on credit cards.

NMR magnets are superconductors, so they must be cooled with liquid helium, which is kept at 4 Kelvin (-452 degrees Fahrenheit). Liquid nitrogen, which is kept at 77 Kelvin (-321 degrees Fahrenheit), helps keep the liquid helium cold.


Varian NMR Systems

▲ Most NMR spectroscopists use magnets that are 500 megahertz to 900 megahertz. This magnet is 900 megahertz.

## The Many Dimensions of NMR

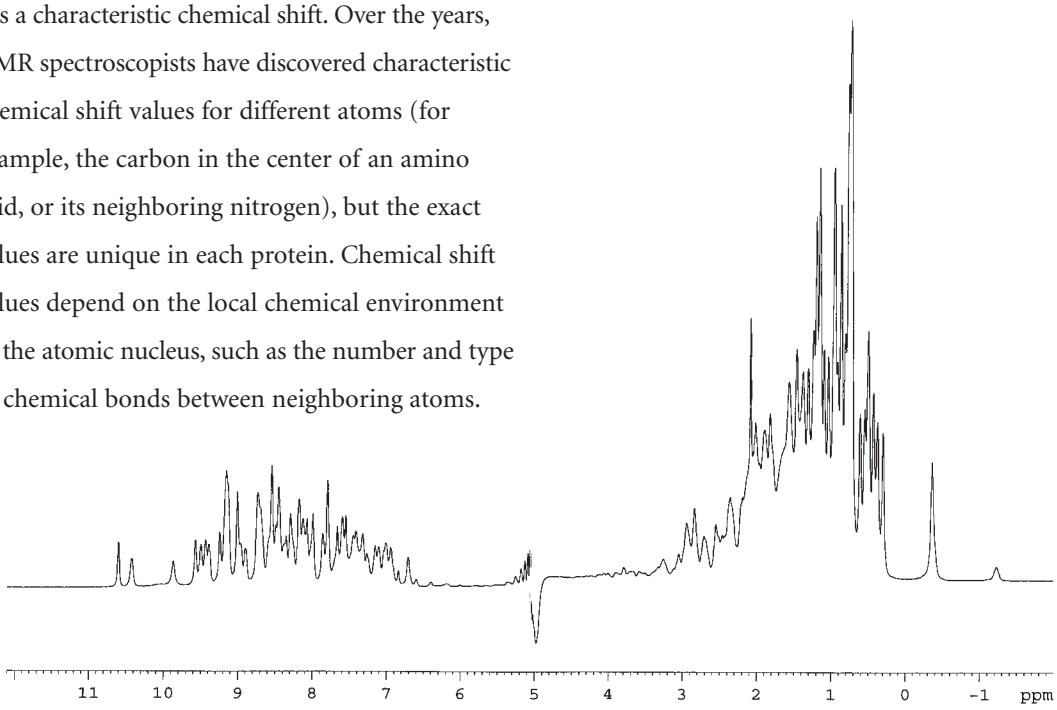
To begin a series of NMR experiments, researchers insert a slender glass tube containing about a half a milliliter of their sample into a powerful, specially designed magnet. The natural magnets in the sample's atoms line up with the NMR magnet just as iron filings line up with a toy magnet. The researchers then blast the sample with a series of split-second radio wave pulses that disrupt this magnetic equilibrium in the nuclei of selected atoms.

By observing how these nuclei react to the radio waves, researchers can assess their chemical nature. Specifically, researchers measure a property of the atoms called chemical shift.

Every type of NMR-active atom in the protein has a characteristic chemical shift. Over the years, NMR spectroscopists have discovered characteristic chemical shift values for different atoms (for example, the carbon in the center of an amino acid, or its neighboring nitrogen), but the exact values are unique in each protein. Chemical shift values depend on the local chemical environment of the atomic nucleus, such as the number and type of chemical bonds between neighboring atoms.

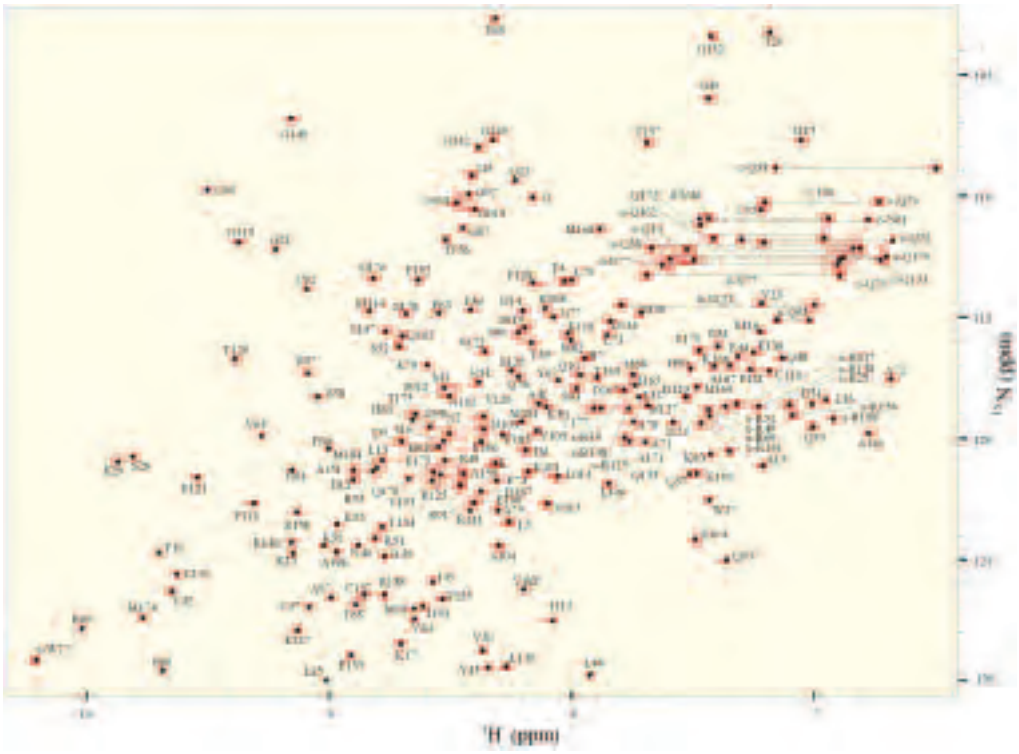
The pattern of these chemical shifts is displayed as a series of peaks in what is called a one-dimensional NMR spectrum. Each peak corresponds to one or more hydrogen atoms in the molecule. The higher the peak, the more hydrogen atoms it represents. The position of the peaks on the horizontal axis indicates their chemical identity.

The overlapping peaks typical of one-dimensional NMR spectra obscure information needed to determine protein structures. To overcome this problem, scientists turn to a technique called multi-dimensional NMR. This technique combines several sets of experiments and spreads out the data into discrete spots. The location of




▲ This one-dimensional NMR spectrum shows the chemical shifts of hydrogen atoms in a protein from streptococcal bacteria.

Spectrum courtesy of Ramon Campos-Olivas, National Institutes of Health



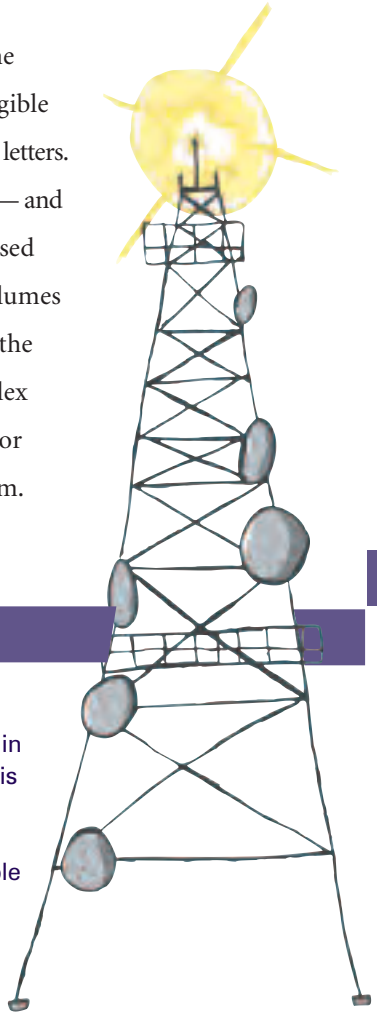
A two-dimensional NMR spectrum of a protein with labeled spots.

The laboratory of Xiaolian Gao, University of Houston

each spot indicates unique properties of one atom in the sample. The researchers must then label each spot with the identity of the atom to which it corresponds.

For a small, simple protein, computational programs require only a few days to accurately assign each spot to a particular atom. For a large, complex protein, it could take months.

To better understand multi-dimensional NMR, we can think of an encyclopedia. If all the words in the encyclopedia were condensed into one dimension, the result would be a single, illegible line of text blackened by countless overlapping letters. Expand this line to two dimensions — a page — and you still have a jumbled mess of superimposed words. Only by expanding into multiple volumes is it possible to read all the information in the encyclopedia. In the same way, more complex NMR studies require experiments in three or four dimensions to clearly solve the problem.



## *NMR Tunes in on Radio Waves*

Each NMR experiment is composed of hundreds of radio wave pulses, each separated by no more than a few milliseconds. Scientists enter the experiment they'd like to run into a computer, which then sends precisely timed pulses to the sample and collects the resulting data.

This data collection process can require as little as 20 minutes for a single, simple experiment. For a complex molecule, it could take weeks or months.

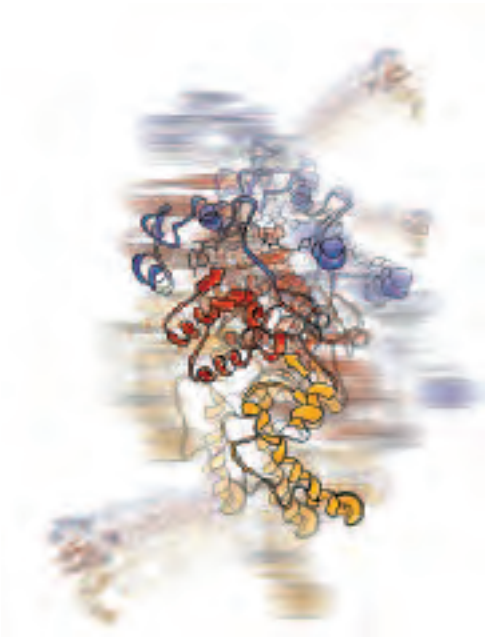
NMR's radio wave pulses are quite tame compared to the high-energy X-rays used in crystallography. In fact, if an NMR sample is prepared well, it should be able to last for many years, allowing the researchers to conduct further studies on the same sample at a later time.

## Spectroscopists Get NOESY for Structures

To determine the arrangement of the atoms in the molecule, scientists use a multi-dimensional NMR technique called NOESY (pronounced "nosy") for Nuclear Overhauser Effect Spectroscopy. This technique works best on hydrogen atoms, which have the strongest NMR signal and are the most abundant atoms in biological systems. They are also the simplest — each hydrogen nucleus contains just a single proton.

The NOESY experiment reveals how close different protons are to each other in space. A pair of protons very close together (typically within 3 angstroms) will give a very strong NOESY signal. More separated pairs of protons will give weaker signals, out to the limit of detection for the technique, which is about 6 angstroms.

From there, the scientists (or, to begin with, their computers) must determine how the atoms are arranged in space. It's like solving a complex, three-dimensional puzzle with thousands of pieces.



## The Wiggling World of Proteins

Although a detailed, three-dimensional structure of a protein is extremely valuable to show scientists what the molecule looks like, it is really only a static "snapshot" of the protein frozen in one position. Proteins themselves are not rigid or static — they are dynamic, rapidly changing molecules that can move, bend, expand, and contract. NMR researchers can explore some of these internal molecular motions by altering the solvent used to dissolve the protein.

A three-dimensional NMR structure often merely provides the framework for more in-depth studies. After you have the structure, you can easily probe features that reveal the molecule's role and behavior in the body, including its flexibility, its interactions with other molecules, and how it reacts to changes in temperature, acidity, and other conditions.

# *Untangling Protein Folding*

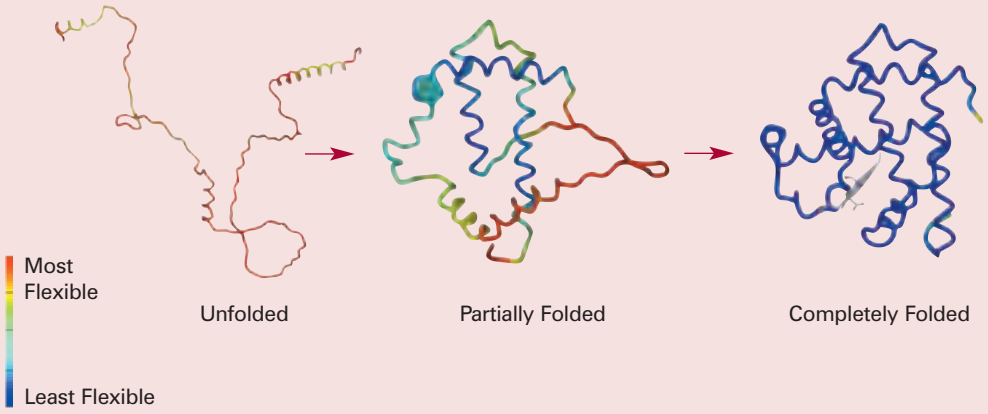
A hundred billion years. That's the time scientists estimate it could take for a small protein to fold randomly into its active shape. But somehow, Nature does it in a tenth of a second.

Most proteins start out like a loose string flopping around in a lake, possibly with short coiled sections. The molecules contort quickly into various partially folded states before congealing into their final form. Because the process is so fast, scientists cannot study it directly. But NMR is well suited to certain studies of protein folding.

By changing the temperature, acidity, or chemical composition of a protein's liquid environment, spectroscopists can reverse and interrupt protein folding. By capturing a protein in different stages of unraveling, researchers hope to understand how proteins fold normally.

H. Jane Dyson and Peter Wright, a husband-and-wife team of NMR spectroscopists at the Scripps Research Institute in La Jolla, California, used this technique to study myoglobin in various folding states.

Myoglobin, a small protein that stores oxygen in muscle tissue, is ideal for studying the structure and dynamics of folding. It quickly folds into a compact, alpha-helical structure. Dyson and Wright used changes in acidity to reveal which regions are most flexible in different folding states. The first two "structures" below each represent one of many possible conformations of a floppy, partially folded molecule.




Adapted with permission from *Nature Structural Biology* 1998, 5:499–503

Understanding how proteins fold so quickly and correctly (most of the time) will shed light on the dozens of diseases that are known or suspected to result from misfolded proteins. In addition, one of the greatest challenges for the biotechnology industry is to coax bacteria into making vast quantities of properly folded human proteins.

**STUDENT SNAPSHOT**

# *The Sweetest Puzzle*

"Getting a protein structure using NMR is a lot of fun," says Chele DeRider, a graduate student at the University of Wisconsin-Madison. "You're given all these pieces to a puzzle and you have to use a set of rules, common sense, and intuitive thinking to put the pieces together. And when you do, you have a protein structure."

Jeff Miller, University of Wisconsin-Madison

DeRider is working at UW-Madison's national NMR facility. She is refining the structure of brazzein, a small, sweet protein. Most sweet-tasting molecules are sugars, not proteins; so brazzein is quite unusual. It also has other remarkable properties that make it attractive as a sugar substitute. It is 2,000 times sweeter than table sugar — with many fewer calories. And, unlike aspartame (NutraSweet®), it stays sweet even after 2 hours at nearly boiling temperatures.

In addition to its potential impact in the multimillion-dollar market of sugar substitutes, brazzein may teach scientists how we perceive some substances as sweet. Researchers know which amino acids in brazzein are responsible for its taste — changing a single one can either enhance or eliminate this flavor — but they are still investigating how these amino acids react with tongue cells to trigger a sensation of sweetness.

*"Getting a protein structure using NMR is a lot of fun . . . .*
*You start out with just dots on a page*
*and you end up with a protein structure."*

Chele DeRider
Graduate Student
University of Wisconsin-Madison

DeRider became interested in NMR as an undergraduate student at Macalester College in St. Paul, Minnesota. She was studying organic chemistry, but found that she spent most of her time running NMR spectra on her compounds. "I realized that's what I liked most about my research," she says.

After she finishes her graduate work, DeRider plans to obtain a postdoctoral fellowship to continue using NMR to study protein structure and then to teach at a small college similar to her *alma mater.*

H.M. Hadik

▲ The plum-sized berries of this African plant contain brazzein, a small, sweet protein.

## Got It?

Give one advantage and one disadvantage of NMR when compared to X-ray crystallography.

What do NMR spectroscopists learn from a NOESY experiment?

Why is it important to study protein folding?

CHAPTER 4

# *Structure-Based Drug Design: From the Computer to the Clinic*

In 1981, doctors recognized a strange new disease in the United States. The first handful of patients suffered from unusual cancers and pneumonias. As the disease spread, scientists discovered its cause—a virus that attacks human immune cells. Now a major killer worldwide, the disease is best known by its acronym, AIDS.

AIDS or acquired immunodeficiency syndrome, is caused by the human immunodeficiency virus, or HIV.

Although researchers have not found a cure for AIDS, structural biology has greatly enhanced their understanding of HIV and has played a key role in the development of drugs to treat this deadly disease.

## *The Life of an AIDS Virus*

HIV was quickly recognized as a retrovirus, a type of virus that carries its genetic material not as DNA, as do most other organisms on the planet, but as RNA. After entering a cell, retroviruses "reverse transcribe" their RNA into DNA.

Long before anyone had heard of HIV, researchers in labs all over the world studied retroviruses, some of which cause cancers in animals. These scientists traced out the life cycle of retroviruses and identified the key proteins the viruses use to infect cells.

When HIV was identified as a retrovirus, these studies gave AIDS researchers an immediate jump-start. The previously identified viral proteins became initial drug targets.

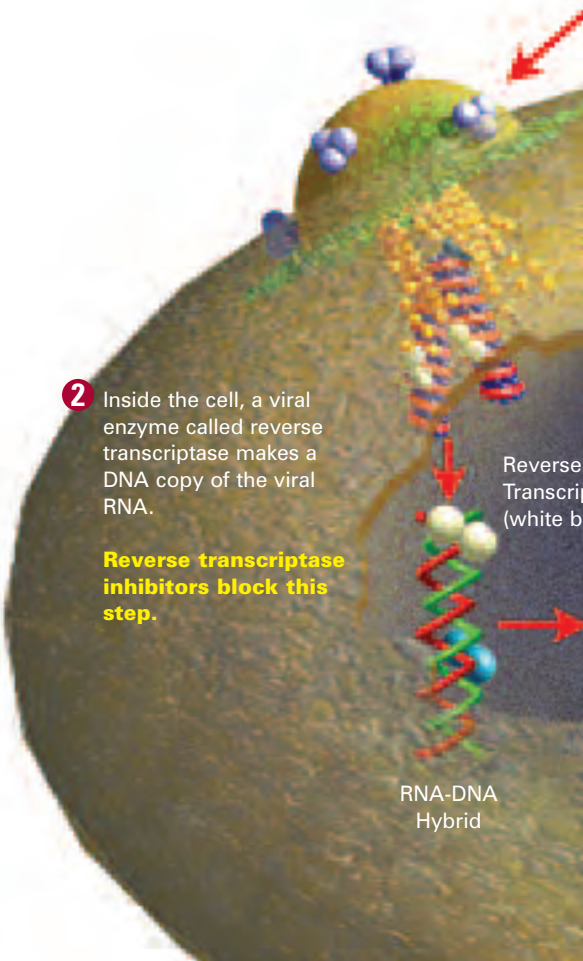
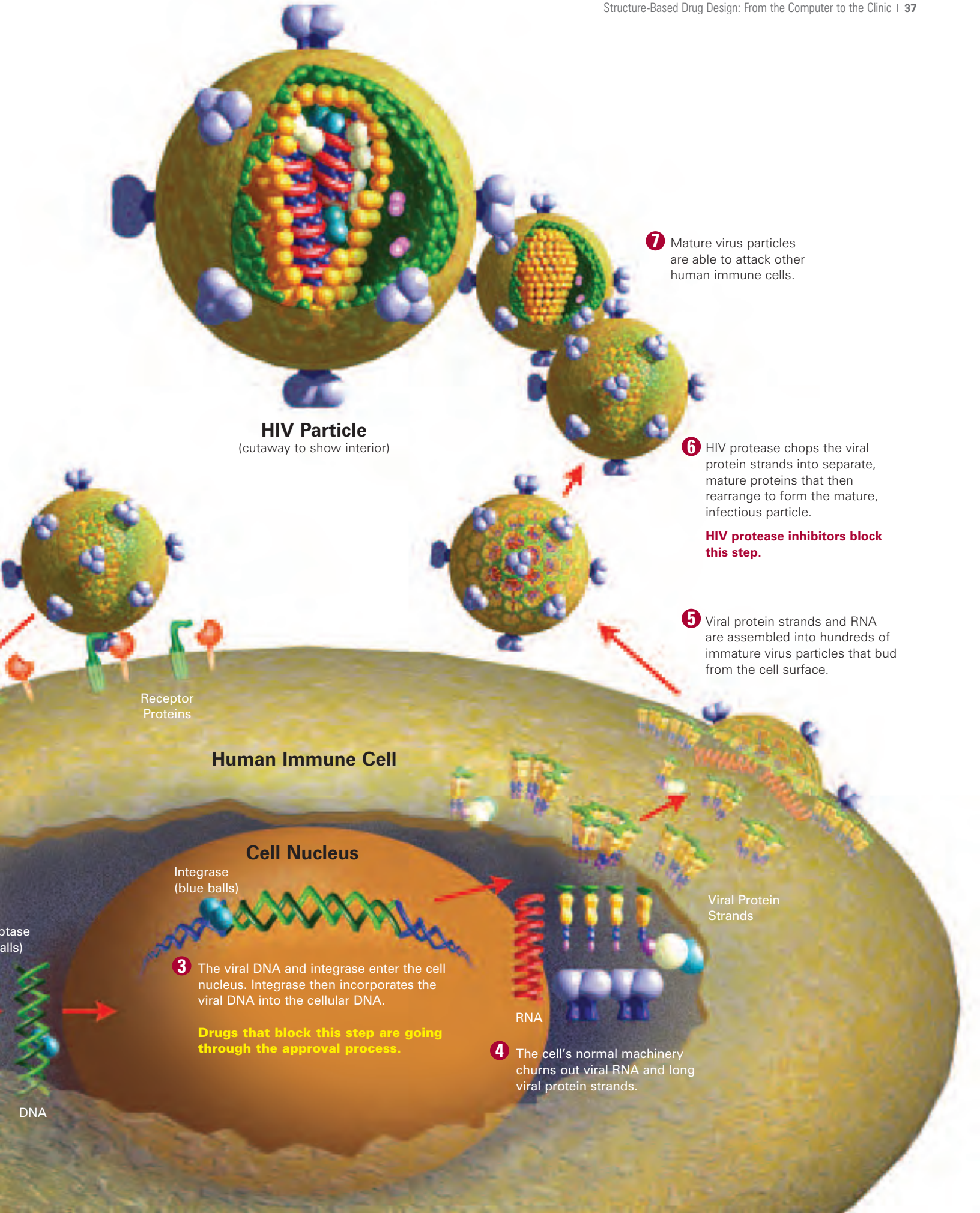
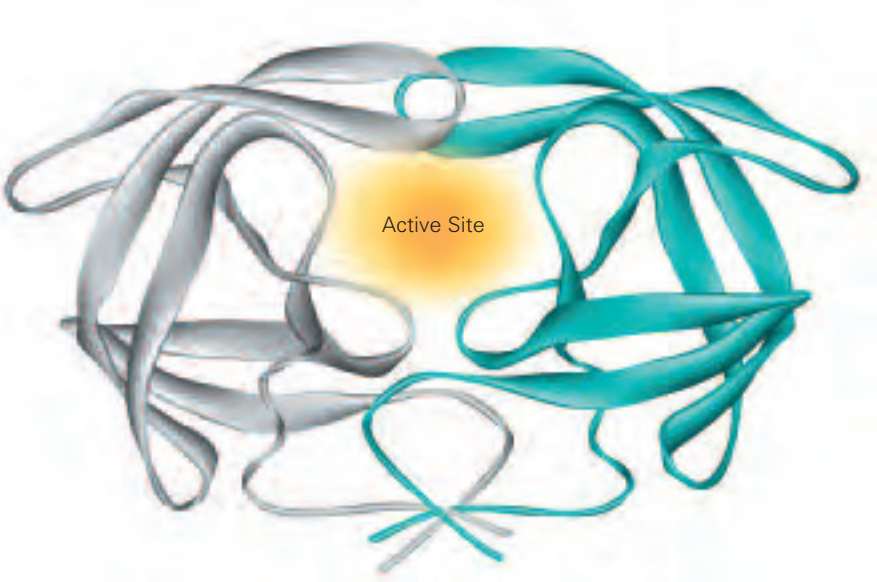



Illustration courtesy of Louis E. Henderson, Senior Scientist (emeritus, retired) AIDS Vaccine Program, National Cancer Institute (Frederick, MD)


HIV Particle
(cutaway to show interior)
7 Mature virus particles are able to attack other human immune cells.
6 HIV protease chops the viral protein strands into separate, mature proteins that then rearrange to form the mature, infectious particle.
HIV protease inhibitors block this step.
5 Viral protein strands and RNA are assembled into hundreds of immature virus particles that bud from the cell surface.
Receptor Proteins
Human Immune Cell
Cell Nucleus
Integrase (blue balls)
Viral Protein Strands
3 The viral DNA and integrase enter the cell nucleus. Integrase then incorporates the viral DNA into the cellular DNA.
Drugs that block this step are going through the approval process.
RNA
4 The cell's normal machinery churns out viral RNA and long viral protein strands.
DNA

## Revealing the Target

Our story begins in 1989, when scientists determined the X-ray crystallographic structure of HIV protease, a viral enzyme critical in HIV's life cycle. Pharmaceutical scientists hoped that by blocking this enzyme, they could prevent the virus from spreading in the body.

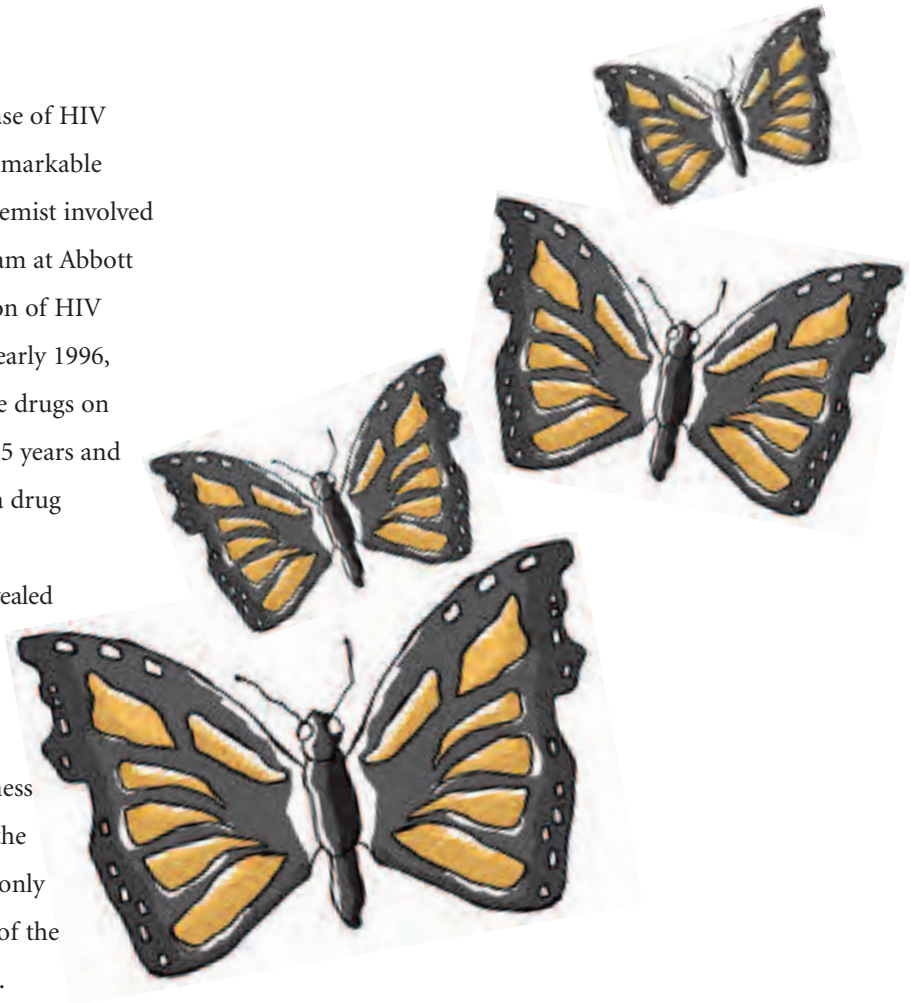



HIV protease is a symmetrical molecule with two equal halves and an active site near its center.

Molecular models of HIV protease in this chapter were generated by Alisa Zapp Machalek

With the structure of HIV protease at their fingertips, researchers were no longer working blindly. They could finally see their target enzyme — in exhilarating, color-coded detail. By feeding the structural information into a computer modeling program, they could spin a model of the enzyme around, zoom in on specific atoms, analyze its chemical properties, and even strip away or alter parts of it.

Most importantly, they could use the computerized structure as a reference to determine the types of molecules that might block the enzyme. These molecules can be retrieved from chemical libraries or can be designed on a computer screen and then synthesized in a laboratory. Such structure-based drug design strategies have the potential to shave off years and millions of dollars from the traditional trial-and-error drug development process.

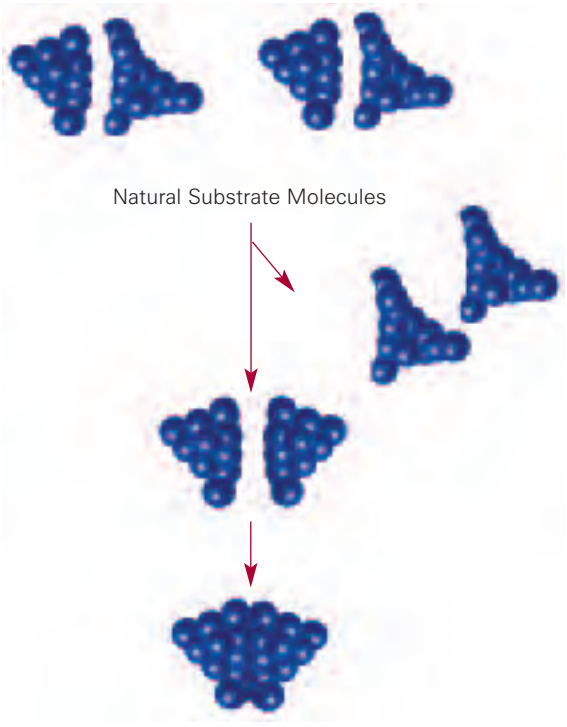


These strategies worked in the case of HIV protease inhibitors. "I think it's a remarkable success story," says Dale Kempf, a chemist involved in the HIV protease inhibitor program at Abbott Laboratories. "From the identification of HIV protease as a drug target in 1988 to early 1996, it took less than 8 years to have three drugs on the market." Typically, it takes 10 to 15 years and more than $800 million to develop a drug from scratch.

The structure of HIV protease revealed a crucial fact — like a butterfly, the enzyme is made up of two equal halves. For most such symmetrical molecules, both halves have a "business area," or active site, that carries out the enzyme's job. But HIV protease has only one such active site — in the center of the molecule where the two halves meet.

Pharmaceutical scientists knew they could take advantage of this feature. If they could plug this single active site with a small molecule, they could shut down the whole enzyme — and theoretically stop the virus' spread in the body.


HIV Protease

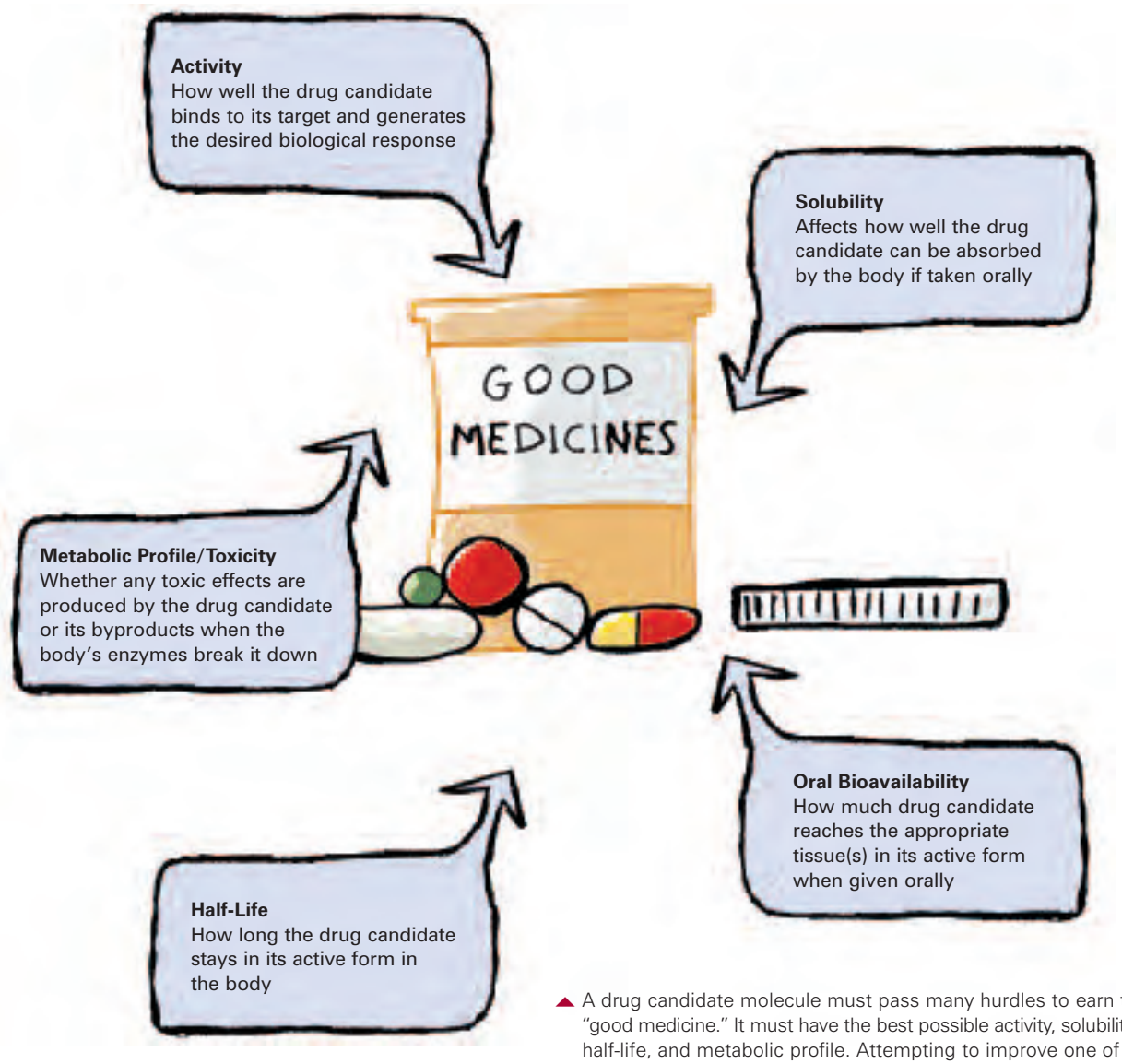



Knowing that HIV protease has two symmetrical halves, pharmaceutical researchers initially attempted to block the enzyme with symmetrical small molecules. They made these by chopping in half molecules of the natural substrate, then making a new molecule by fusing together two identical halves of the natural substrate.

Several pharmaceutical companies started out by using the enzyme's shape as a guide. "We designed drug candidate molecules that had the same two-fold symmetry as HIV protease," says Kempf. "Conceptually, we took some of the enzyme's natural substrate [the molecules it acts upon], chopped these molecules in half, rotated them 180 degrees, and glued two identical halves together."

To the researchers' delight, the first such molecule they synthesized fit perfectly into the active site of the enzyme. It was also an excellent inhibitor — it prevented HIV protease from functioning normally. But it wasn't water-soluble, meaning it couldn't be absorbed by the body and would never be effective as a drug.

Abbott scientists continued to tweak the structure of the molecule to improve its properties. They eventually ended up with a nonsymmetrical molecule they called Norvir® (ritonavir).




▲ A drug candidate molecule must pass many hurdles to earn the description "good medicine." It must have the best possible activity, solubility, bioavailability, half-life, and metabolic profile. Attempting to improve one of these factors often affects other factors. For example, if you structurally alter a lead compound to improve its activity, you may also decrease its solubility or shorten its half-life. The final result must always be the best possible compromise.

# *Structure-Based Drug Design: Blocking the Lock*

Traditionally, scientists identify new drugs either by fiddling with existing drugs or by testing thousands of compounds in a laboratory. If you think of the target molecule — HIV protease in this case — as a lock, this approach is rather like trying to design a key perfectly shaped to the lock if you're given an armload of tiny metal scraps, glue, and wire cutters.

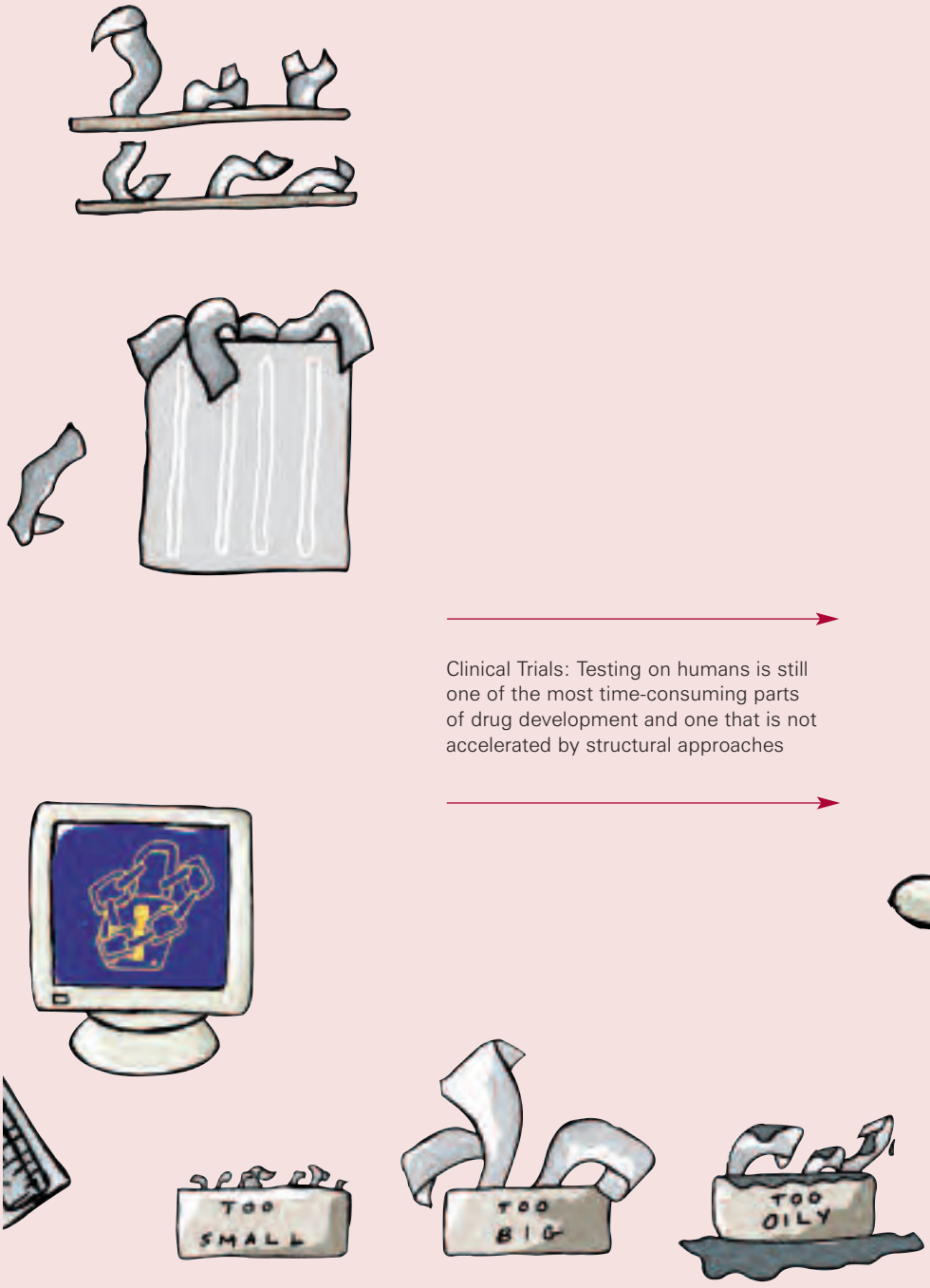
Using a structure-based strategy, researchers have an initial advantage. They start with a computerized model of the detailed, three-dimensional structure of the lock and of its key (the natural molecule, called a substrate, that fits into the lock, triggering viral replication). Then scientists try to design a molecule that will plug up the lock to keep out the substrate key.

Knowing the exact three-dimensional shape of the lock, scientists can discard any of the metal scraps (small molecules) that are not the right size or shape to fit the lock. They might even be able to design a small molecule to fit the lock precisely. Such a molecule may be a starting point for pharmaceutical researchers who are designing a drug to treat HIV infection.

Of course, biological molecules are much more complex than locks and keys, and human bodies can react in unpredictable ways to drug molecules, so the road from the computer screen to pharmacy shelves remains long and bumpy.

Traditional drug design often requires random testing of thousands — if not hundreds of thousands — of compounds (shown here as metal scraps)

By knowing the shape and chemical properties of the target molecule, scientists using structure-based drug design strategies can approach the job more "rationally." They can discard the drug candidate molecules that have the wrong shape or properties.




Clinical Trials: Testing on humans is still one of the most time-consuming parts of drug development and one that is not accelerated by structural approaches

## A Hope for the Future

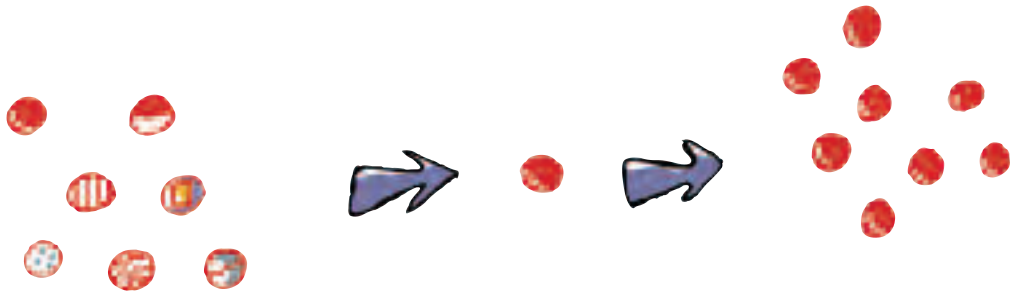
Between December 1995 and March 1996, the Food and Drug Administration approved the first three HIV protease inhibitors — Hoffman-La Roche's Invirase™ (saquinavir), Abbott's Norvir™ (ritonavir), and Merck and Co., Inc.'s Crixivan® (indinavir). Initially, these drugs were hailed as the first real hope in 15 years for people with AIDS. Newspaper headlines predicted that AIDS might even be cured.

Although HIV protease inhibitors did not become the miracle cure many had hoped for, they represent a triumph for antiviral therapy. Antibiotics that treat *bacterial* diseases abound (although they are becoming less effective as bacteria develop resistance), but doctors have very few drugs to treat *viral* infections.

Protease inhibitors are also noteworthy because they are a classic example of how structural biology can enhance traditional drug development. "They show that with some ideas about structure and rational drug design, combined with traditional medicinal chemistry, you can come up with potent drugs that function the way they're predicted to," says Kempf.

"That doesn't mean we have all the problems solved yet," he continues. "But clearly these compounds have made a profound impact on society." The death rate from AIDS went down dramatically after these drugs became available. Now protease inhibitors are often prescribed with other anti-HIV drugs to create a "combination cocktail" that is more effective at squelching the virus than are any of the drugs individually.

## *How HIV Resistance Arises*

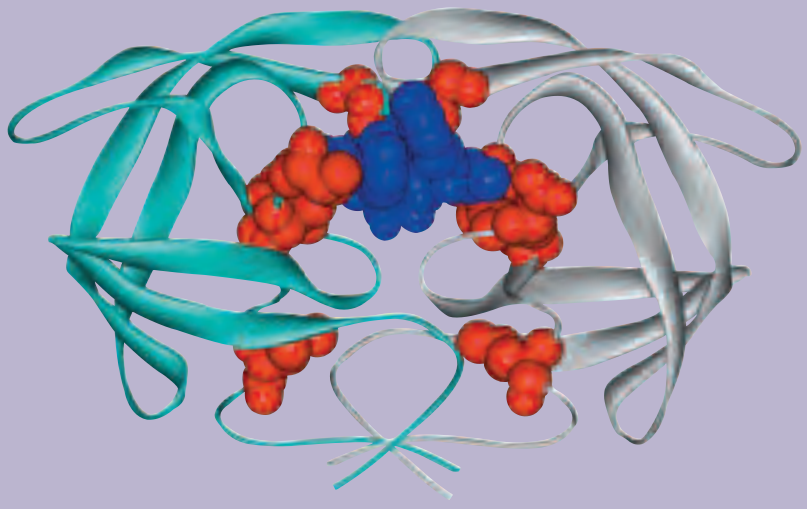


HIV produces many different versions of itself in a patient's body (although the huge majority are the normal form).

Drugs kill all of these virus particles except those that are resistant to the drugs.

The resistant virus particles continue to reproduce. Soon the drug is no longer effective for the patient.

# *Homing in on Resistance*



Scientists have identified dozens of mutations (shown in red) that allow HIV protease to escape the effects of drugs. The protease molecules in some drug-resistant HIV strains have two or three such mutations. To outwit the enzyme's mastery of mutation, researchers are designing drugs that interact specifically with amino acids in the enzyme that are critical for the enzyme's function. This approach cuts off the enzyme's escape routes. As a result, the enzyme — and thus the entire virus — is forced to succumb to the drug.

HIV is a moving target. When it reproduces inside the body, instead of generating exact replicas of itself, it churns out a variety of slightly altered daughter virus particles. Some of these mutants are able to evade, or "resist," the effects of a drug — and can pass that resistance on to their own daughter particles. While most virus particles initially succumb to the drug, these resistant mutants survive and multiply. Eventually, the drug loses its anti-HIV activity, because most of the virus particles in the infected person are resistant to it.

Some researchers now are working on new generations of HIV protease inhibitors that are designed to combat specific drug-resistant viral strains.

Detailed, computer-modeled pictures of HIV protease from these strains reveal how even amino acid substitutions far away from the enzyme's active site can produce drug resistance. Some research groups are trying to beat the enzyme at its own game by designing drugs that bind to these mutant forms of HIV protease. Others are designing molecules that latch onto the enzyme's Achilles' heels — the aspartic acids in the active site and other amino acids that, if altered, would render the enzyme useless. Still others are trying to discover inhibitors that are more potent, more convenient to take, have fewer side effects, or are better able to combat mutant strains of the virus.

**STUDENT SNAPSHOT**

# *The Fascination of Infection*

"I really like to study retroviruses," says Kristi Pullen, who majored in biochemistry at the University of Maryland, Baltimore County (UMBC). "I also like highly infectious agents, like Ebola. The more virulent something is, the less it's worked on, so it opens up all sorts of fascinating questions. I couldn't help but be interested."

In addition to her UMBC classwork, Pullen helped determine the structure of retroviruses in the NMR spectroscopy laboratory of Michael Summers. This research focuses on how retroviruses package "RNA warheads" that enable them to spread in the body. Eventually, the work may reveal a new drug target for retroviral diseases, including AIDS.

Kelly Burns Photography, Columbia, Maryland

*"Working in Dr. Summers' lab and other labs teaches you that research can be fun. It's not just a whole lot of people in white coats. We went biking and skiing together. All the people were great to work with."*

Kristi Pullen
Graduate Student
University of California, Berkeley

Until her senior year in high school, Pullen wanted to be an orthopedic surgeon. But after her first experience working in a lab, she recognized "there's more to science than medicine." Then, after taking some science courses, she realized she had an inner yearning to learn science and to work in a lab.

Pullen is now a graduate student at the University of California, Berkeley in the Department of Molecular and Cell Biology. She plans to continue studying structural biology, to earn a Ph.D., and possibly also to earn an M.D.

She also has some longer-term goals. "Ultimately what I want to do way, way, way down the line is head the NIH [National Institutes of Health] or CDC [Centers for Disease Control and Prevention] and in that way affect the health of a large number of people — the whole country."

## Gripping Arthritis Pain

While the HIV protease inhibitors are classic examples of structure-based drug design, they are also somewhat unusual — at least for now. Although many pharmaceutical companies have entire divisions devoted to structural biology, most use it as a complementary approach, in partnership with other, more traditional, means of drug discovery. In many cases, the structure of a target molecule is determined after traditional screening, or even after a drug is on the market.

This was the case for Celebrex®. Initially designed to treat osteoarthritis and adult rheumatoid arthritis, Celebrex® became the first drug approved to treat a rare condition called FAP, or familial adenomatous polyposis, that leads to colon cancer.

Normally, the pain and swelling of arthritis are treated with drugs like aspirin or Advil® (ibuprofen), the so-called NSAIDs, or non-steroidal anti-inflammatory drugs. But these medications can cause damage to gastrointestinal organs, including bleeding ulcers. In fact, a recent study found that such side effects result in more than 100,000 hospitalizations and 16,500 deaths every year. According to another study, if these side effects were included in tables listing mortality data, they would rank as the 15th most common cause of death in the United States.

National Institutes of Health

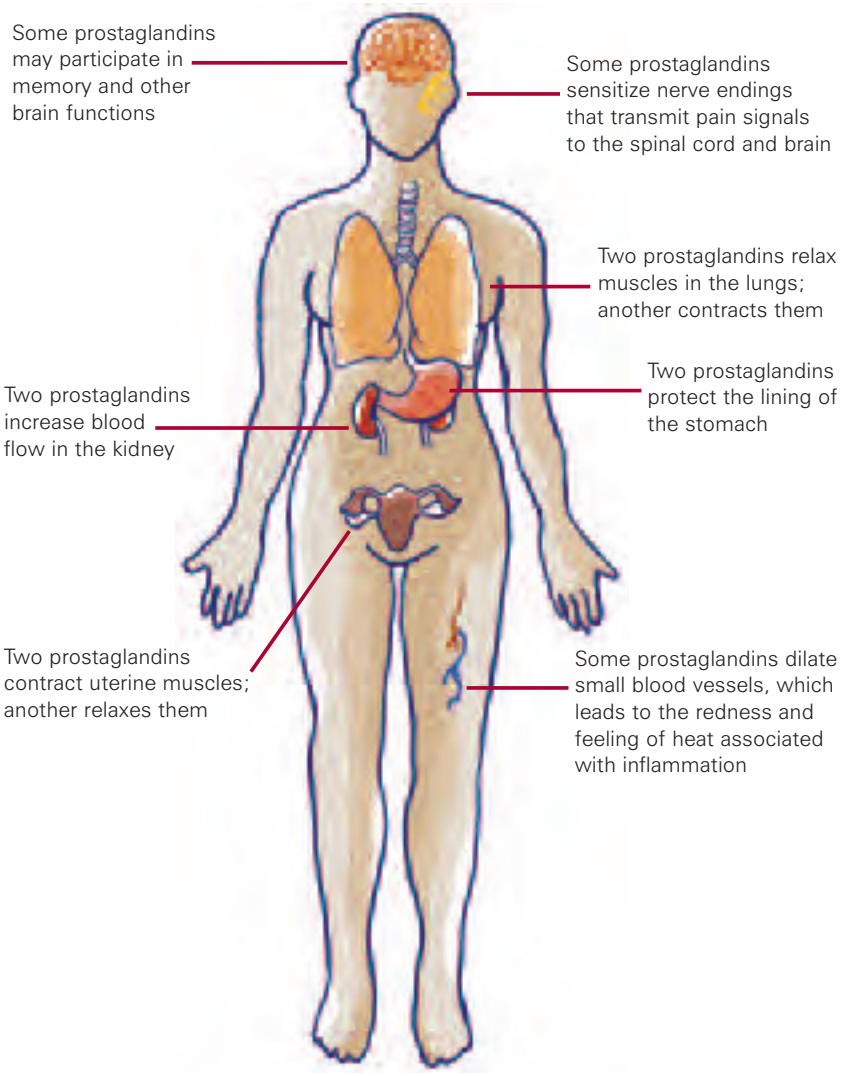
Rheumatoid arthritis is an immune system disorder that affects more than 2 million Americans, causing pain, stiffness, and swelling in the joints. It can cripple hands, wrists, feet, knees, ankles, shoulders, and elbows. It also causes inflammation in internal organs and can lead to permanent disability. Osteoarthritis has some of the same symptoms, but it develops more slowly and only affects certain joints.

A fortunate discovery enabled scientists to design drugs that retain the anti-inflammatory properties of NSAIDs without the ulcer-causing side effects.

By studying the drugs at the molecular level, researchers learned that NSAIDs block the action of two closely related enzymes called cyclooxygenases. These enzymes are abbreviated COX-1 and COX-2.

Although the enzymes share some of the same functions, they also differ in important ways. COX-2 is produced in response to injury or infection and activates molecules that trigger inflammation and an immune response. By blocking COX-2, NSAIDs reduce inflammation and pain caused by arthritis, headaches, and sprains.

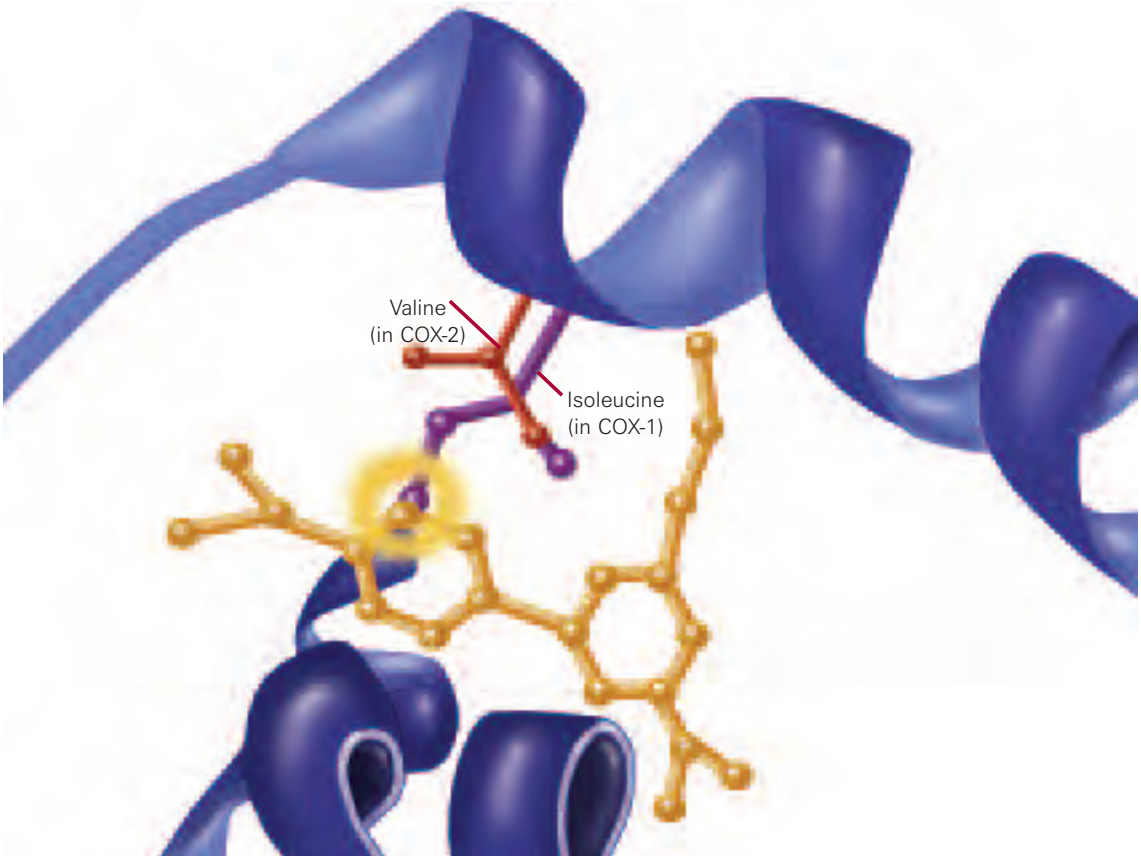
In contrast, COX-1 produces molecules, called prostaglandins, that protect the lining of the stomach from digestive acids. When NSAIDs block this function, they foster ulcers.




Both COX-1 and COX-2 produce prostaglandins, which have a variety of different — and sometimes opposite — roles in the body. Some of these roles are shown here.

To create an effective painkiller that doesn't cause ulcers, scientists realized they needed to develop new medicines that shut down COX-2 but not COX-1. Such a compound was discovered using standard medicinal chemistry and marketed under the name Celebrex®. It quickly became the fastest selling drug in U.S. history, generating more prescriptions in its first year than the next two leading drugs combined.

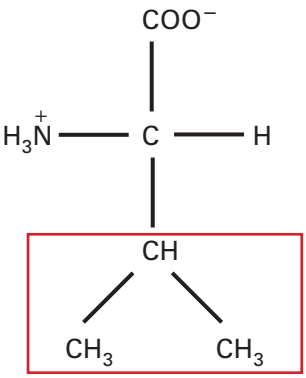
At the same time, scientists were working out the molecular structure of the COX enzymes. Through structural biology, they could see exactly why Celebrex® plugs up COX-2 but not COX-1.



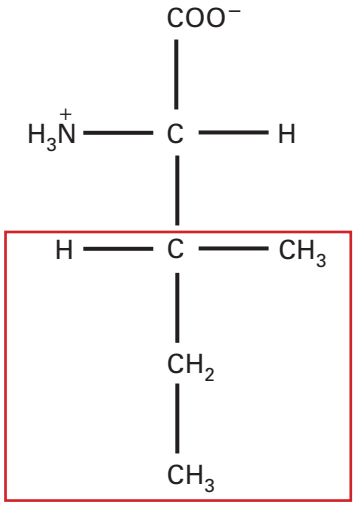

This close-up view of the active sites of COX-1 and COX-2 (ribbons) reveal why Celebrex® can bind to one of the COX enzymes, but not to the other. A single amino acid substitution makes all the difference. In a critical place in the protein, COX-2 contains valine, a small amino acid that creates a pocket into which the drug (in yellow) can bind. In the same position, COX-1 contains isoleucine, which elbows out the drug.

The three-dimensional structures of COX-2 and COX-1 are almost identical. But there is one amino acid change in the active site of COX-2 that creates an extra binding pocket. It is this extra pocket into which Celebrex® binds.

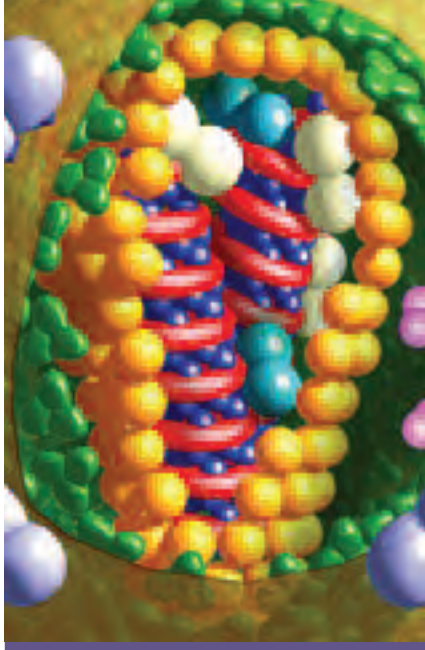
In addition to showing researchers in atom-by-atom detail how the drug binds to its target, the structures of the COX enzymes will continue to provide basic researchers with insight into how these molecules work in the body.




Valine




Isoleucine



## Got It?

What is structure-based drug design?

How was structure-based drug design used to develop an HIV protease inhibitor?

How is the structural difference between COX-1 and COX-2 responsible for the effectiveness of Celebrex®?

How do viruses become resistant to drugs?

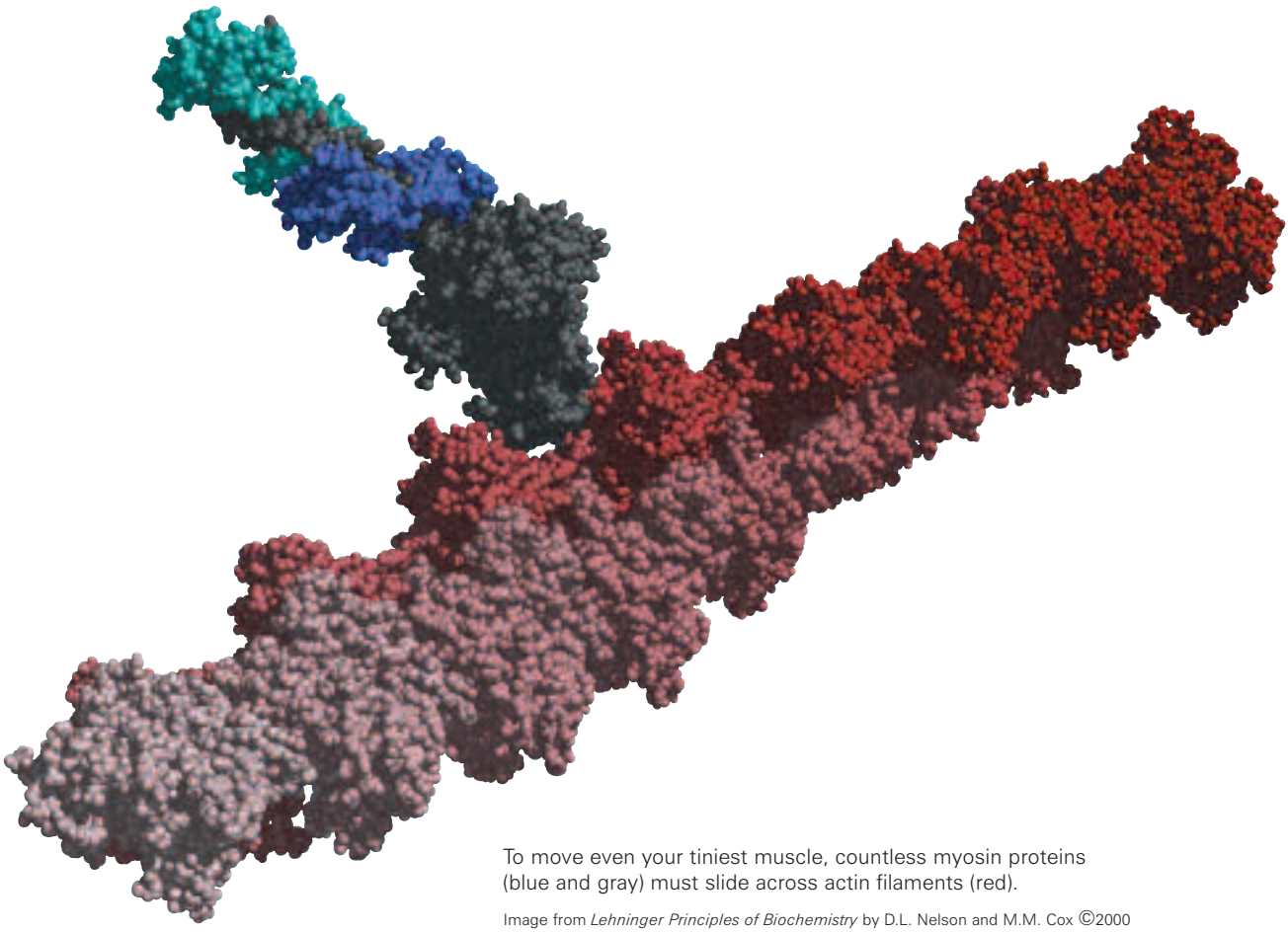
CHAPTER 5

# *Beyond Drug Design*

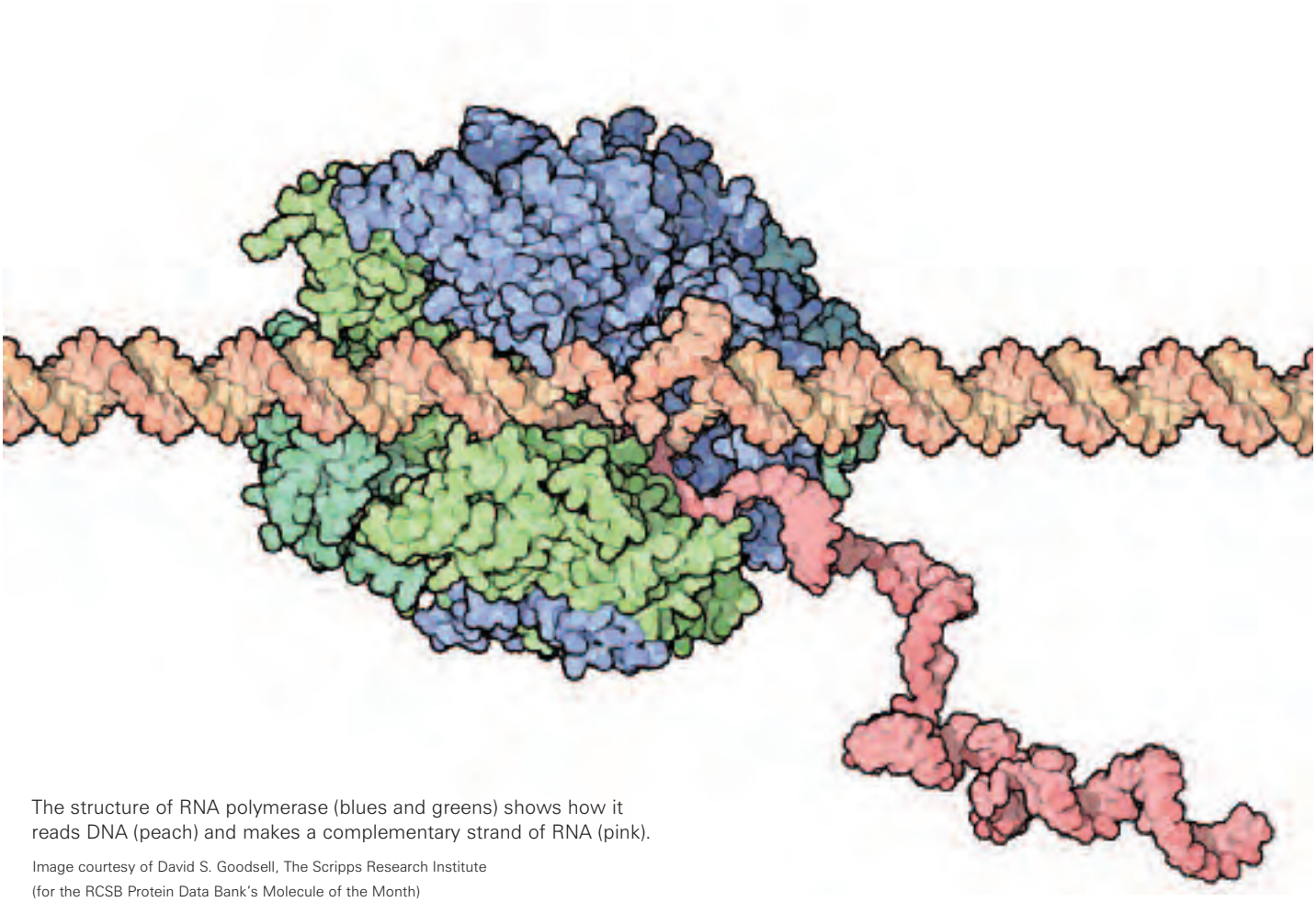
This booklet has focused on drug design as the most immediate medical application of structural biology. But detailed studies of protein structure have value and potential far beyond the confines of the pharmaceutical industry. At its root, such research teaches us about the fundamental nature of biological molecules. The examples below provide a tiny glimpse into areas in which structural biology has, and continues to, shed light.

## Muscle Contraction

With every move you make, from a sigh to a sprint, thick ropes of myosin muscle proteins slide across rods of actin proteins in your cells. These proteins also pinch cells in two during cell division and enable cells to move and change shape — a process critical both to the formation of different tissues during embryonic development and to the spread of cancer. Detailed structures are available for both myosin and actin.



To move even your tiniest muscle, countless myosin proteins (blue and gray) must slide across actin filaments (red).



The structure of RNA polymerase (blues and greens) shows how it reads DNA (peach) and makes a complementary strand of RNA (pink).

Image courtesy of David S. Goodsell, The Scripps Research Institute (for the RCSB Protein Data Bank's Molecule of the Month)

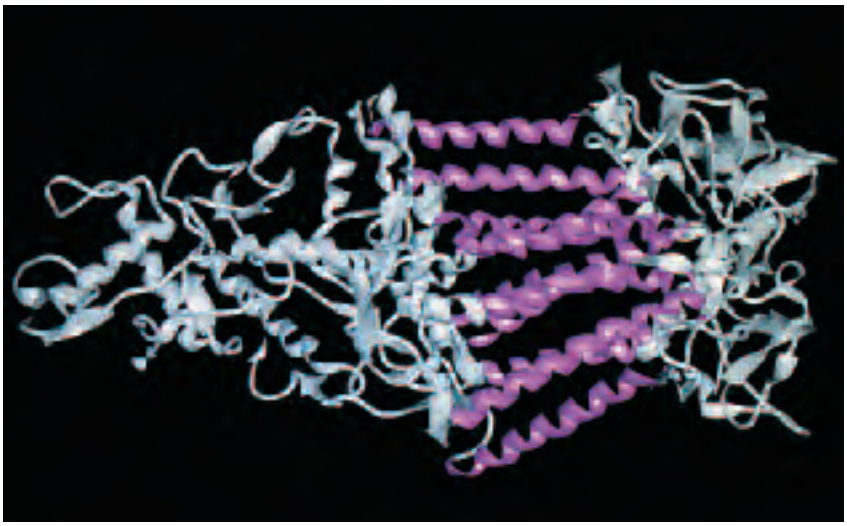
## Transcription and Translation

Cells use DNA instructions to make proteins. Dozens of molecules (mostly proteins) cling together and separate at carefully choreographed times to accomplish this task. The structures of many of these molecules are known and have provided a better understanding of transcription and translation.

A key example is RNA polymerase, an enzyme that reads DNA and synthesizes a complementary strand of RNA. This enzyme is a molecular machine composed of a dozen different small proteins. In 2001, Roger Kornberg, a crystallographer at Stanford University, determined the structure of RNA polymerase in action. This crystal structure suggested a role for each of RNA polymerase's proteins. Kornberg was awarded the 2006 Nobel Prize in Chemistry for this work.

## Photosynthesis

"Photosynthesis is the most important chemical reaction in the biosphere, as it is the prerequisite for all higher life on Earth," according to the Nobel Foundation, which awarded its 1988 Nobel Prize in chemistry to three researchers who determined the structure of a protein central to photosynthesis.



Alisa Zapp Machalek

This bacterial photosynthetic reaction center was the first membrane protein to have its structure determined. The purple spirals (alpha helices) show where the protein crosses the membrane. In the orientation above, the left part of the molecule protrudes from the outside of the bacterial cell, while the right side is inside the cell.

This protein, from a photosynthetic bacterium rather than from a plant, was the first X-ray crystallographic structure of a protein embedded in a membrane. The achievement was remarkable, because it is very difficult to dissolve membrane-bound proteins in water — an essential step in the crystallization process. To borrow further from the Nobel Foundation: "[This] structural determination … has considerable chemical importance far beyond the field of photosynthesis. Many central biological functions in addition to photosynthesis … are associated with membrane-bound proteins. Examples are transport of chemical substances between cells, hormone action, and nerve impulses" — in other words, signal transduction.

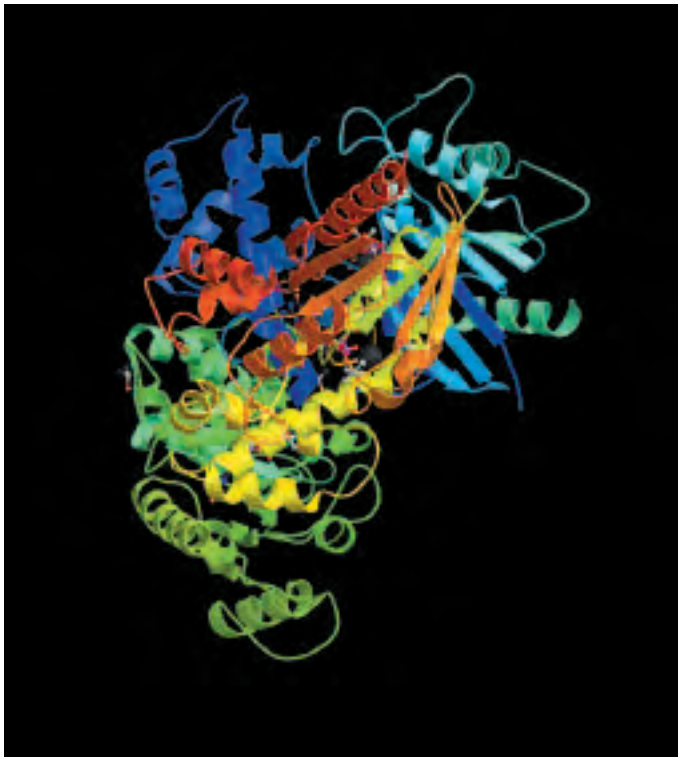
## Signal Transduction

Hundreds, if not thousands, of life processes require a biochemical signal to be transmitted into cells. These signals may be hormones, small molecules, or electrical impulses, and they may reach cells from the bloodstream or other cells. Once signal molecules bind to receptor proteins on the outside surface of a cell, they initiate a cascade of reactions involving several other molecules inside the cell. Depending on the nature of the target cell and of the signaling molecule, this chain of reactions may trigger a nerve impulse,

a change in cell metabolism, or the release of a hormone. Researchers have determined the structure of some molecules involved in common signal transduction pathways.

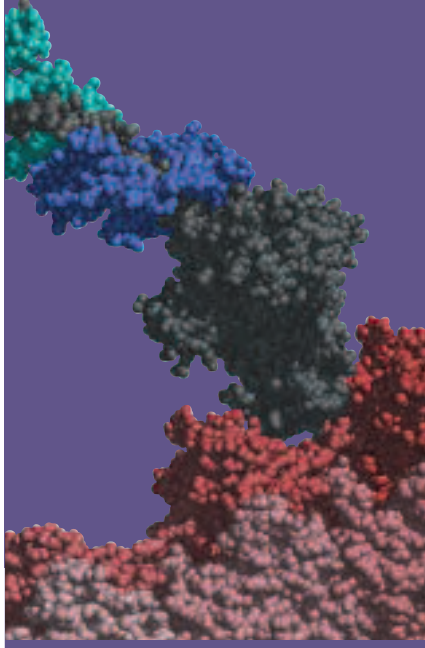
The receptor proteins that bind to the original signal molecule are often embedded in the cell's outer membrane so, like proteins involved in photosynthesis, they are difficult to crystallize. Obtaining structures from receptor proteins not only teaches us more about the basics of signal transduction, it also brings us back to the pharmaceutical industry. At least 50 percent of the drugs on the market target receptor proteins — more than target any other type of molecule.

As this booklet shows, a powerful way to learn more about health, to fight disease, and to deepen our understanding of life processes is to study the details of biological molecules — the remarkable structures of life.



RCSB Protein Data Bank (http://www.pdb.org)

Members of a family of molecules, called G proteins, often act as conduits to pass the molecular message from receptor proteins to molecules in the cell's interior.



## Got It?

Considering this booklet as a whole, how would you define structural biology?

What are the scientific goals of those in the field?

If you were a structural biologist, what proteins or systems would you study? Why?

# *Glossary*

**Acquired immunodeficiency syndrome (AIDS) |** A viral disease caused by the human immunodeficiency virus (HIV).

**Active site |** The region of an enzyme to which a substrate binds and at which a chemical reaction occurs.

**AIDS |** Acquired immunodeficiency syndrome — an infectious disease that is a major killer worldwide.

**Alpha helix |** A short, spiral-shaped section within a protein structure.

**Amino acid |** A chemical building block of proteins. There are 20 standard amino acids. A protein consists of a specific sequence of amino acids.

**Angstrom |** A unit of length used for measuring atomic dimensions. One angstrom equals $10^{-10}$ meters.

**Antibiotic-resistant bacteria |** A strain of bacteria with slight alterations (mutations) in some of their molecules that enable the bacteria to survive drugs designed to kill them.

**Atom |** A fundamental unit of matter. It consists of a nucleus and electrons.

**AZT (azido-deoxythymidine) |** A drug used to treat HIV. It targets the reverse transcriptase enzyme.

**Bacterium (*pl.* bacteria) |** A primitive, one-celled microorganism without a nucleus. Bacteria live almost everywhere in the environment. Some bacteria may infect humans, plants, or animals. They may be harmless or they may cause disease.

**Base |** A chemical component (the fundamental information unit) of DNA or RNA. There are four bases in DNA: adenine (A), thymine (T), cytosine (C), and guanine (G). RNA also contains four bases, but instead of thymine, RNA contains uracil (U).

**Beta sheet |** A pleated section within a protein structure.

**Chaperones |** Proteins that help other proteins fold or escort other proteins throughout the cell.

**Chemical shift |** An atomic property that varies depending on the chemical and magnetic properties of an atom and its arrangement within a molecule. Chemical shifts are measured by NMR spectroscopists to identify the types of atoms in their samples.

**COX-1 (cyclooxygenase-1) |** An enzyme made continually in the stomach, blood vessels, platelet cells, and parts of the kidney. It produces prostaglandins that, among other things, protect the lining of the stomach from digestive acids. Because NSAIDs block COX-1, they foster ulcers.

**COX-2 (cyclooxygenase-2) |** An enzyme found in only a few places, such as the brain and parts of the kidney. It is made only in response to injury or infection. It produces prostaglandins involved in inflammation and the immune response. NSAIDs act by blocking COX-2. Because elevated levels of COX-2 in the body have been linked to cancer, scientists are investigating whether blocking COX-2 may prevent or treat some cancers.

**Cyclooxygenases |** Enzymes that are responsible for producing prostaglandins and other molecules in the body.

**Deoxyribose |** The type of sugar in DNA.

**DNA (deoxyribonucleic acid) |** The substance of heredity. A long, usually double-stranded chain of nucleotides that carries genetic information necessary for all cellular functions, including the building of proteins. DNA is composed of the sugar deoxyribose, phosphate groups, and the bases adenine, thymine, guanine, and cytosine.

**Drug target |** See *target molecule.*

**Electromagnetic radiation |** Energy radiated in the form of a wave. It includes all kinds of radiation, including, in order of increasing energy, radio waves, microwaves, infrared radiation (heat), visible light, ultraviolet radiation, X-rays, and gamma radiation.

**Enzyme |** A substance, usually a protein, that speeds up, or catalyzes, a specific chemical reaction without being permanently altered or consumed. Some RNA molecules can also act as enzymes.

**Gene |** A unit of heredity. A segment of DNA that contains the code for a specific protein or protein subunit.

**Genetic code |** The set of triplet letters in DNA (or mRNA) that code for specific amino acids.

**HIV protease |** An HIV enzyme that is required during the life cycle of the virus. It is required for HIV virus particles to mature into fully infectious particles.

**Human immunodeficiency virus (HIV) |** The virus that causes AIDS.

**Inhibitor |** A molecule that "inhibits," or blocks, the biological action of another molecule.

**Isotope |** A form of a chemical element that contains the same number of protons but a different number of neutrons than other forms of the element. Isotopes are often used to trace atoms or molecules in a metabolic pathway. In NMR, only one isotope of each element contains the correct magnetic properties to be useful.

**Kilodalton |** A unit of mass equal to 1,000 daltons. A dalton is a unit used to measure the mass of atoms and molecules. One dalton equals the atomic weight of a hydrogen atom ($1.66 \times 10^{-24}$ grams).

**MAD |** See *multi-wavelength anomalous diffraction.*

**Megahertz |** A unit of measurement equal to 1,000,000 hertz. A hertz is defined as one event or cycle per second and is used to measure the frequency of radio waves and other forms of electromagnetic radiation. The strength of NMR magnets is often reported in megahertz, with most NMR magnets ranging from 500 to 900 megahertz.

**Messenger RNA (mRNA) |** An RNA molecule that serves as an intermediate in the synthesis of protein. Messenger RNA is complementary to DNA and carries genetic information to the ribosome.

**Molecule |** The smallest unit of matter that retains all of the physical and chemical properties of that substance. It consists of one or more identical atoms or a group of different atoms bonded together.

**mRNA |** Messenger RNA.

**Multi-dimensional NMR |** A technique used to solve complex NMR problems.

**Multi-wavelength anomalous diffraction (MAD) |** A technique used in X-ray crystallography that accelerates the determination of protein structures. It uses X-rays of different wavelengths, relieving crystallographers from having to make several different metal-containing crystals.

**NMR |** Nuclear magnetic resonance.

**NMR-active atom |** An atom that has the correct magnetic properties to be useful for NMR. For some atoms, the NMR-active form is a rare isotope, such as $^{13}C$ or $^{15}N$.

**NOESY |** Nuclear Overhauser effect spectroscopy.

**Non-steroidal anti-inflammatory drugs |** A class of medicines used to treat pain and inflammation. Examples include aspirin and ibuprofen. They work by blocking the action of the COX-2 enzyme. Because they also block the COX-1 enzyme, they can cause side effects such as stomach ulcers.

**NSAIDs |** Non-steroidal anti-inflammatory drugs such as aspirin or ibuprofen.

**Nuclear magnetic resonance (NMR) spectroscopy |** A technique used to determine the detailed, three-dimensional structure of molecules and, more broadly, to study the physical, chemical, and biological properties of matter. It uses a strong magnet that interacts with the natural magnetic properties in atomic nuclei.

**Nuclear Overhauser effect spectroscopy (NOESY) |** An NMR technique used to help determine protein structures. It reveals how close different protons (hydrogen nuclei) are to each other in space.

**Nucleotide |** A subunit of DNA or RNA that includes one base, one phosphate molecule, and one sugar molecule (deoxyribose in DNA, ribose in RNA). Thousands of nucleotides join end-to-end to create a molecule of DNA or RNA. See *base, phosphate group.*

**Nucleus (*pl.* nuclei) |** 1. The membrane-bounded center of a cell, which contains genetic material. 2. The center of an atom, made up of protons and neutrons.

**Phosphate group |** A chemical group found in DNA and RNA, and often attached to proteins and other biological molecules. It is composed of one phosphorous atom bound to four oxygen atoms.

**Photosynthesis |** The chemical process by which green plants, algae, and some bacteria use the Sun's energy to synthesize organic compounds (initially carbohydrates).

**Prostaglandins |** A hormone-like group of molecules involved in a variety of functions in the body, including inflammation, blood flow in the kidney, protection of the stomach lining, blood clotting, and relaxation or contraction of muscles in the lungs, uterus, and blood vessels. The formation of prostaglandins is blocked by NSAIDs.

**Protein |** A large biological molecule composed of amino acids arranged in a specific order determined by the genetic code and folded into a specific three-dimensional shape. Proteins are essential for all life processes.

**Receptor protein |** Specific proteins found on the cell surface to which hormones or other molecules bind, triggering a specific reaction within the cell. Receptor proteins are responsible for initiating reactions as diverse as nerve impulses, changes in cell metabolism, and hormone release.

**Resistance |** See *antibiotic-resistant bacteria.* Viruses can also develop resistance to antiviral drugs.

**Retrovirus |** A type of virus that carries its genetic material as single-stranded RNA, rather than as DNA. Upon infecting a cell, the virus generates a DNA replica of its RNA using the enzyme reverse transcriptase.

**Reverse transcriptase |** An enzyme found in retroviruses that copies the virus' genetic material from single-stranded RNA into double-stranded DNA.

**Ribose |** The type of sugar found in RNA.

**Ribosomal RNA |** RNA found in the ribosome.

**RNA (ribonucleic acid) |** A long, usually single-stranded chain of nucleotides that has structural, genetic, and enzymatic roles. There are three major types of RNA, which are all involved in making proteins: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). RNA is composed of the sugar ribose, phosphate groups, and the bases adenine, uracil, guanine, and cytosine. Certain viruses contain RNA, instead of DNA, as their genetic material.

**Side chain |** The part of an amino acid that confers its identity. Side chains range from a single hydrogen atom (for glycine) to a group of 15 or more atoms.

**Signal transduction |** The process by which chemical, electrical, or biological signals are transmitted into and within a cell.

**Structural biology |** A field of study dedicated to determining the detailed, three-dimensional structures of biological molecules to better understand the function of these molecules.

**Structural genomics |** A field of study that seeks to determine a large inventory of protein structures based on gene sequences. The eventual goal is to be able to produce approximate structural models of any protein based on its gene sequence. From these structures and models, scientists hope to learn more about the biological function of proteins.

**Structure-based drug design |** An approach to developing medicines that takes advantage of the detailed, three-dimensional structure of target molecules.

**Substrate |** A molecule that binds to an enzyme and undergoes a chemical change during the ensuing enzymatic reaction.

**Synchrotron |** A large machine that accelerates electrically charged particles to nearly the speed of light and maintains them in circular orbits. Originally designed for use by high-energy physicists, synchrotrons are now heavily used by structural biologists as a source of very intense X-rays.

**Target molecule (or target protein) |** The molecule on which pharmaceutical researchers focus when designing a drug. Often, the target molecule is from a virus or bacterium, or is an abnormal human protein. In these cases, the researchers usually seek to design a small molecule — a drug — to bind to the target molecule and block its action.

**Transcription |** The first major step in protein synthesis, in which the information coded in DNA is copied (transcribed) into mRNA.

**Translation |** The second major step in protein synthesis, in which the information encoded in mRNA is deciphered (translated) into sequences of amino acids. This process occurs at the ribosome.

**Virus |** An infectious microbe that requires a host cell (plant, animal, human, or bacterial) in which to reproduce. It is composed of proteins and genetic material (either DNA or RNA).

**Virus particle |** A single member of a viral strain, including all requisite proteins and genetic material.

**X-ray crystallography |** A technique used to determine the detailed, three-dimensional structure of molecules. It is based on the scattering of X-rays through a crystal of the molecule under study.

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
National Institutes of Health
National Institute of General Medical Sciences

NIH Publication No. 07-2778
Reprinted July 2007
http://www.nigms.nih.gov