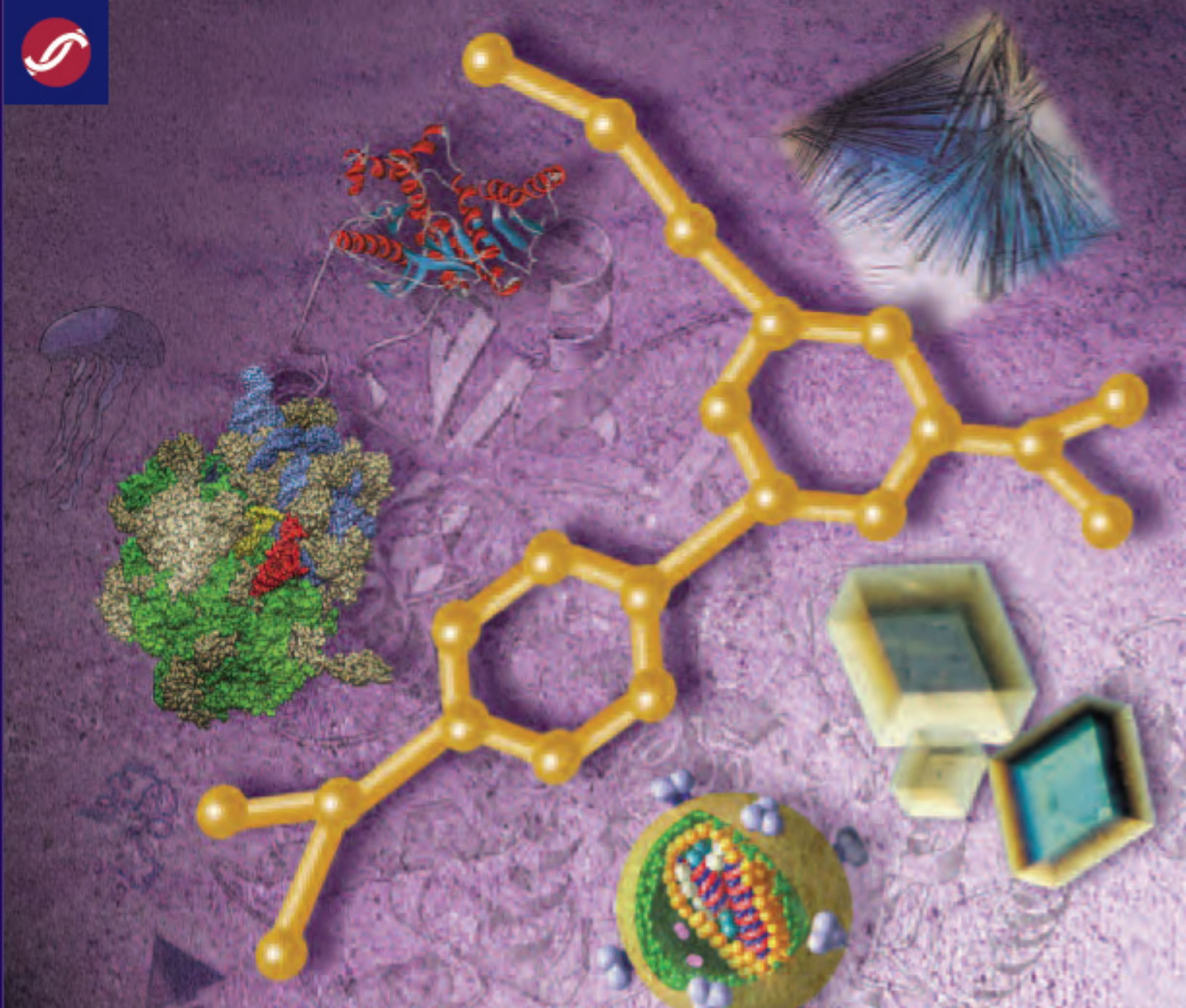




The Structures of Life



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

PREFACE: WHY STRUCTURE?

iv

CHAPTER 1: PROTEINS ARE THE BODY'S WORKER MOLECULES

2

Proteins Are Made From Small Building Blocks	3
<i>Proteins in All Shapes and Sizes</i>	4
<i>Computer Graphics Advance Research</i>	4
<i>Small Errors in Proteins Can Cause Disease</i>	6
Parts of Some Proteins Fold Into Corkscrews	7
<i>Mountain Climbing and Computational Modeling</i>	8
The Problem of Protein Folding	8
<i>Provocative Proteins</i>	9
Structural Genomics: From Gene to Structure, and Perhaps Function	10
<i>The Genetic Code</i>	12

CHAPTER 2: X-RAY CRYSTALLOGRAPHY: ART MARRIES SCIENCE

14

<i>Viral Voyages</i>	15
Crystal Cookery	16
<i>Calling All Crystals</i>	17
<i>Student Snapshot: Science Brought One Student From the Coast of Venezuela to the Heart of Texas</i>	18
Why X-Rays?	20
Synchrotron Radiation—One of the Brightest Lights on Earth	21
<i>Peering Into Protein Factories</i>	23
Scientists Get MAD at the Synchrotron	24

CHAPTER 3: THE WORLD OF NMR: MAGNETS, RADIO WAVES, AND DETECTIVE WORK	26
<i>A Slam Dunk for Enzymes</i>	27
<i>NMR Spectroscopists Use Tailor-Made Proteins</i>	28
<i>NMR Magic Is in the Magnets</i>	29
The Many Dimensions of NMR	30
<i>NMR Tunes in on Radio Waves</i>	31
Spectroscopists Get NOESY for Structures	32
The Wiggling World of Proteins	32
<i>Untangling Protein Folding</i>	33
<i>Student Snapshot: The Sweetest Puzzle</i>	34
 CHAPTER 4: STRUCTURE-BASED DRUG DESIGN: FROM THE COMPUTER TO THE CLINIC	 36
<i>The Life of an AIDS Virus</i>	36
Revealing the Target	38
<i>Structure-Based Drug Design: Blocking the Lock</i>	42
A Hope for the Future	44
<i>How HIV Resistance Arises</i>	44
<i>Homing in on Resistance</i>	45
<i>Student Snapshot: The Fascination of Infection</i>	46
Gripping Arthritis Pain	48
 CHAPTER 5: BEYOND DRUG DESIGN	 52
Muscle Contraction	52
Transcription and Translation	53
Photosynthesis	54
Signal Transduction	54
 GLOSSARY	 56

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

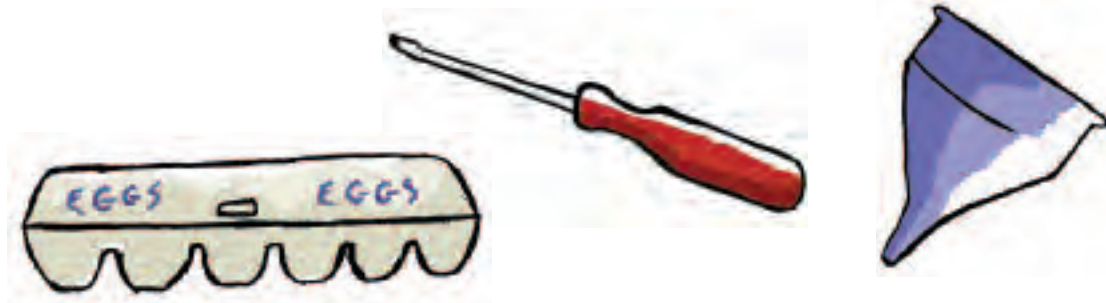
protein offers clues about the role it plays in the body. It may also hold the key to developing new medicines, materials, or diagnostic procedures.

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

In addition to teaching about our bodies, these

“structures of life” may hold the key to developing new medicines, materials, and diagnostic procedures.

crystallography and nuclear magnetic resonance spectroscopy—that structural biologists use to study the detailed shapes of proteins and other biological molecules.



▲ 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

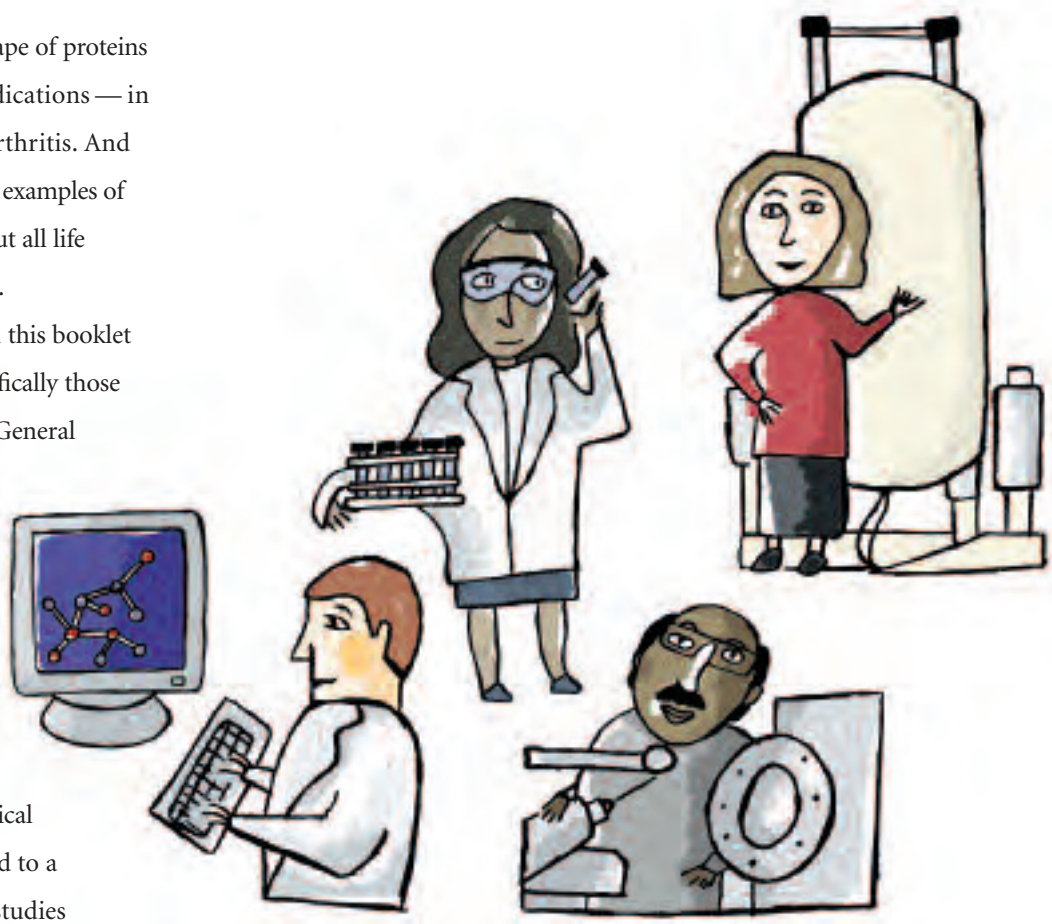
brim and narrow neck enable the transfer of liquids into a container with a small opening. The shape of a protein—although much more complicated than the shape of a common object—teaches us about that protein’s role in the body.

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.

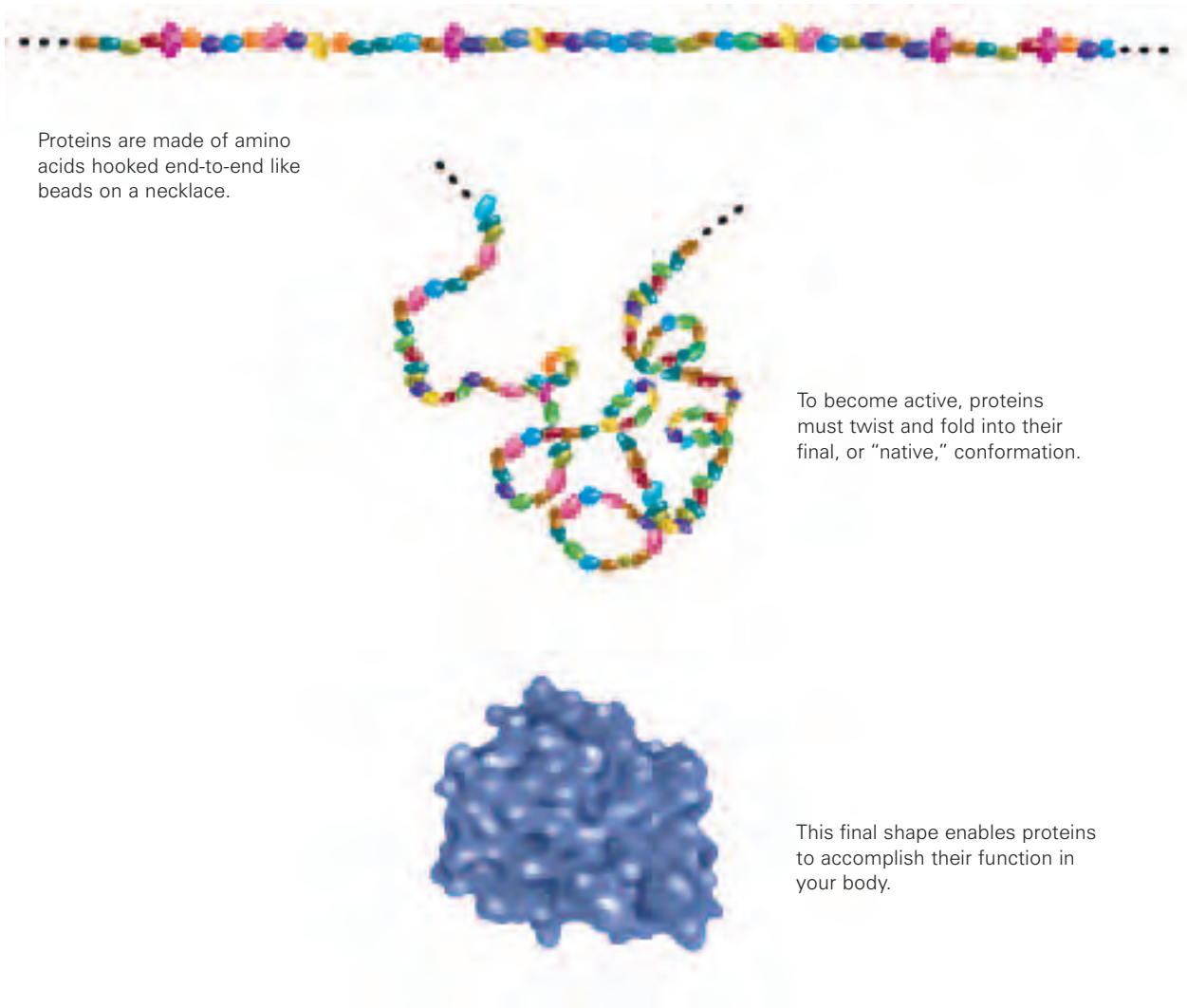
Proteins Are Made From Small Building Blocks

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.

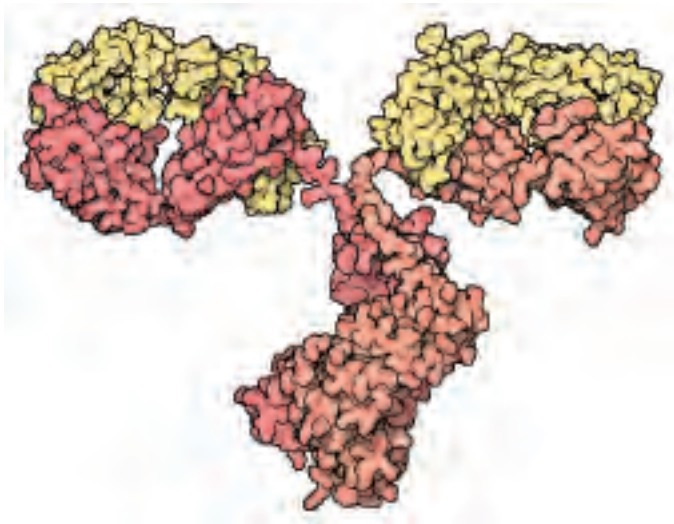


Proteins in All Shapes and Sizes

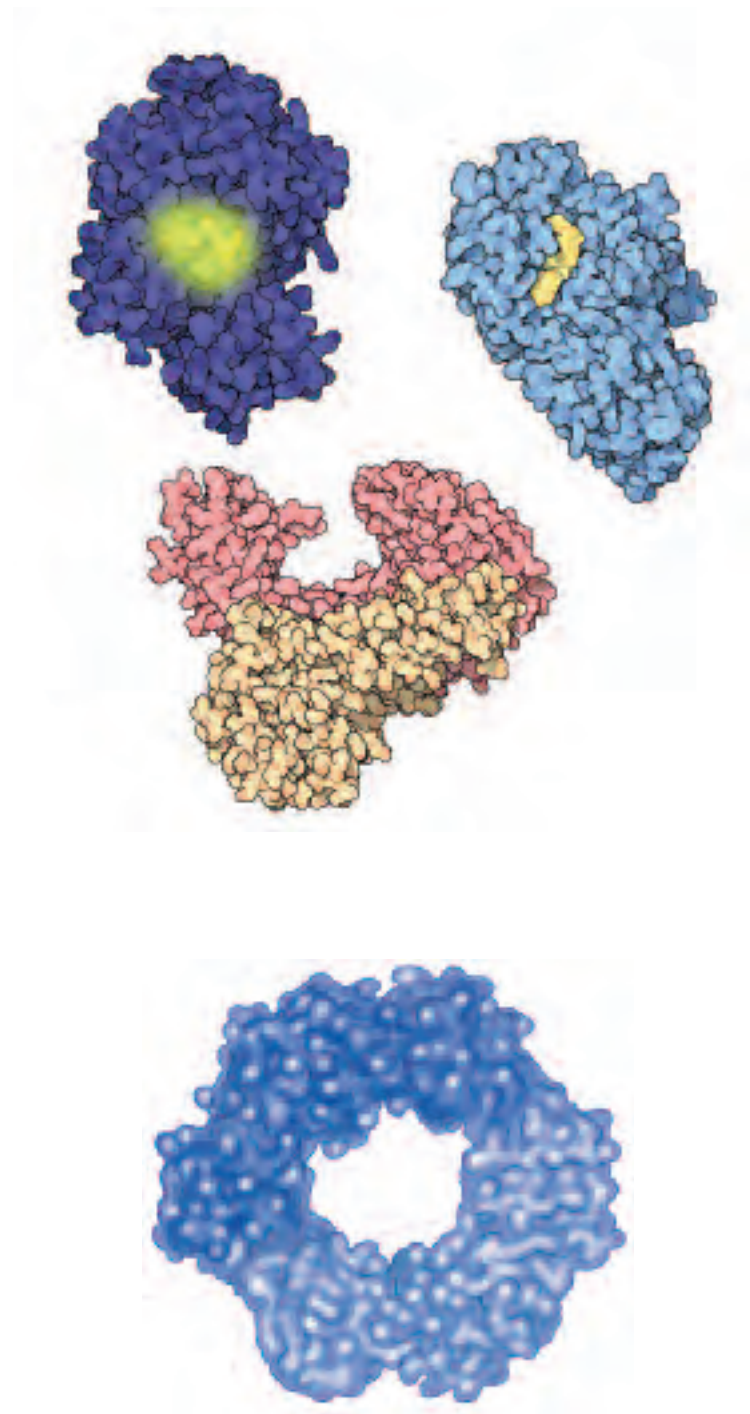
Because proteins have diverse roles in the body, they come in many shapes and sizes. Studies of these shapes teach us how the proteins function in our bodies and help us understand diseases caused by abnormal proteins.

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

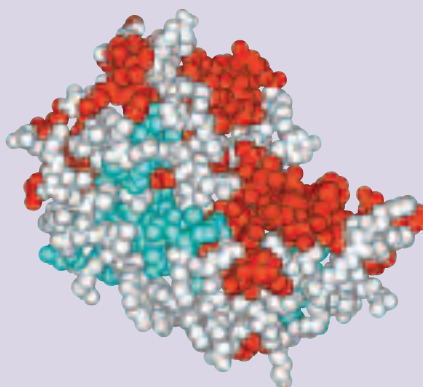
Decades ago, scientists who wanted to study three-dimensional molecular structures spent days, weeks, or longer building models out of rods, balls, and wire scaffolding.

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.

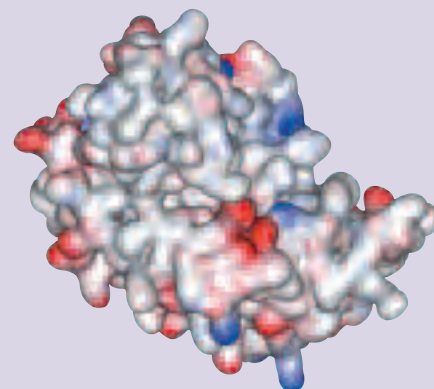
To try one of these computer graphics programs, go to <http://www.proteinexplorer.org> or <http://www.pdb.org>.



- ▲ 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.

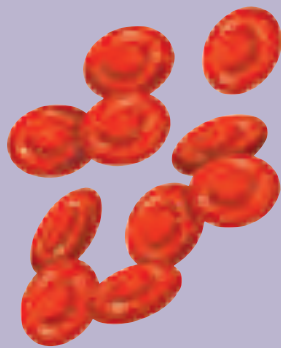


- ▲ A surface rendering of the same protein shows its overall shape and surface properties. The red and blue coloration indicates the electrical charge of atoms on the protein's surface.

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.

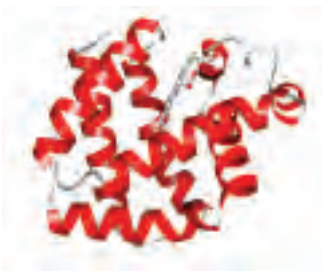
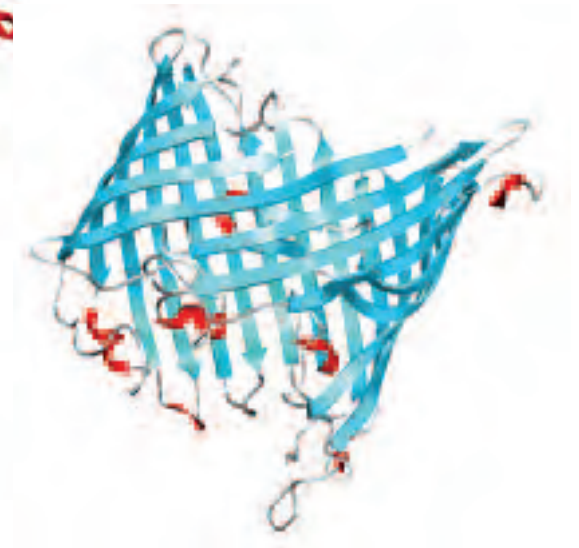
The disease is caused when a protein called CFTR is incorrectly folded. This misfolding is usually caused by the deletion of a single amino acid in CFTR. The function of CFTR, which stands for cystic fibrosis transmembrane conductance regulator, is to allow chloride ions (a component of table salt) to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



- 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.

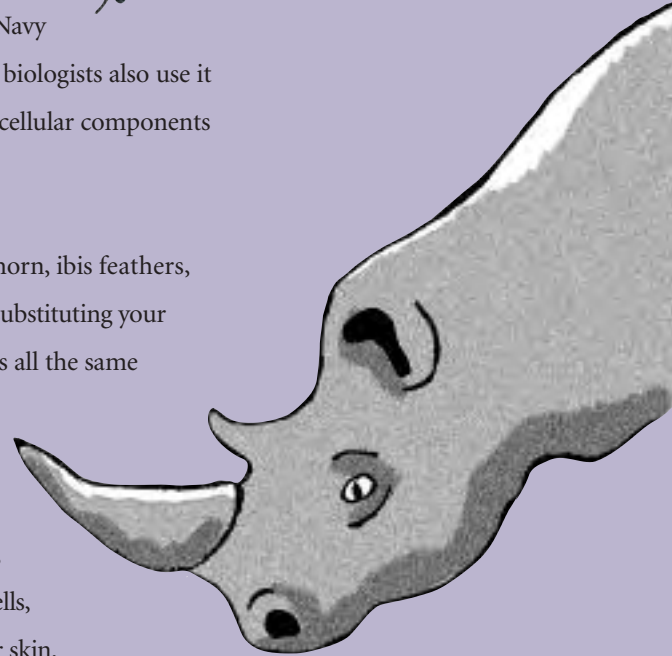


- The light of fireflies (also called lightning bugs) is made possible by a protein called luciferase. Although most predators stay away from the bitter-tasting insects, some frogs eat so many fireflies that they glow!

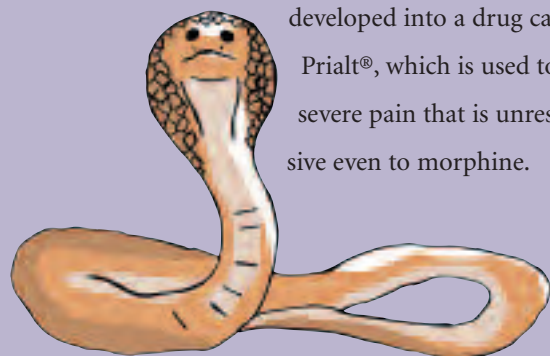


- The deadly venoms of cobras, scorpions, and puffer fish contain small proteins that act as nerve toxins. Some sea snails stun their prey (and occasionally, unlucky humans) with up to 50 such toxins. One of these toxins has been

- 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.



developed into a drug called Prialt®, which is used to treat severe pain that is unresponsive even to morphine.



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.

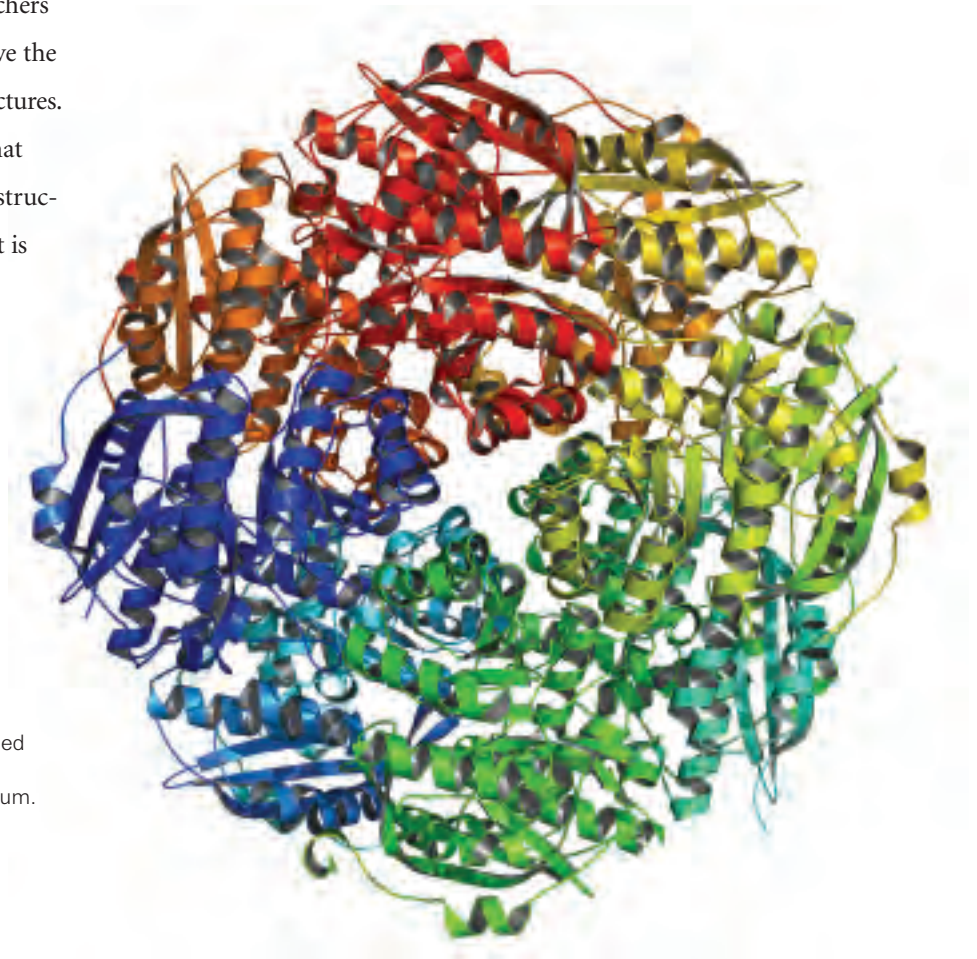
Perhaps even more significant, PSI researchers have developed new technologies that improve the speed and ease of determining molecular structures. Many of these new technologies are robots that automate previously labor-intensive steps in structure determination. Thanks to these robots, it is

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.

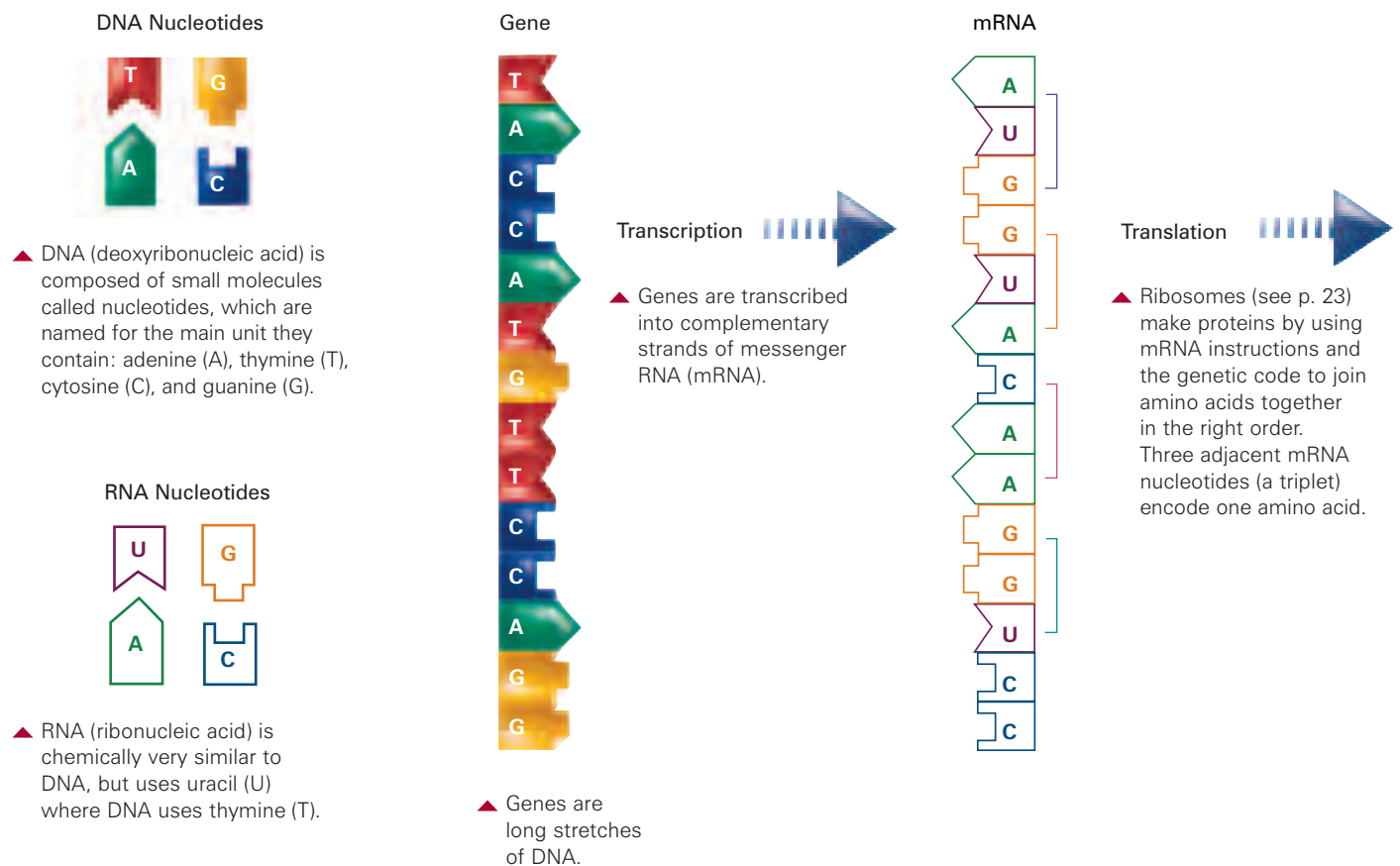
Courtesy of the New York Structural GenomiX Consortium

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
U	UUU	phenylalanine
	UUC	phenylalanine
	UUA	leucine
	UUG	leucine
C	CUU	leucine
	CUC	leucine
	CUA	leucine
	CUG	leucine
A	AUU	isoleucine
	AUC	isoleucine
	AUA	isoleucine
	AUG	methionine
G	GUU	valine
	GUC	valine
	GUA	valine
	GUG	valine



Thank You for previewing this eBook

You can read the full version of this eBook in different formats:

- HTML (Free /Available to everyone)
- PDF / TXT (Available to V.I.P. members. Free Standard members can access up to 5 PDF/TXT eBooks per month each month)
- Epub & Mobipocket (Exclusive to V.I.P. members)

To download this full book, simply select the format you desire below

