

# MOLECULAR LEGO

A modest collection of small building blocks enables the design and manufacture of nanometer-scale structures programmed to have virtually any shape desired

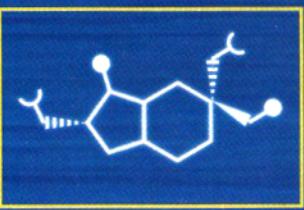
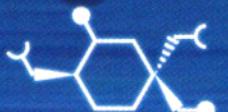
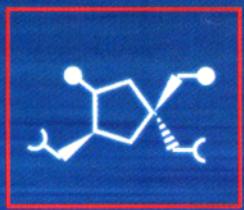
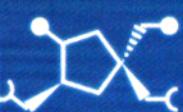
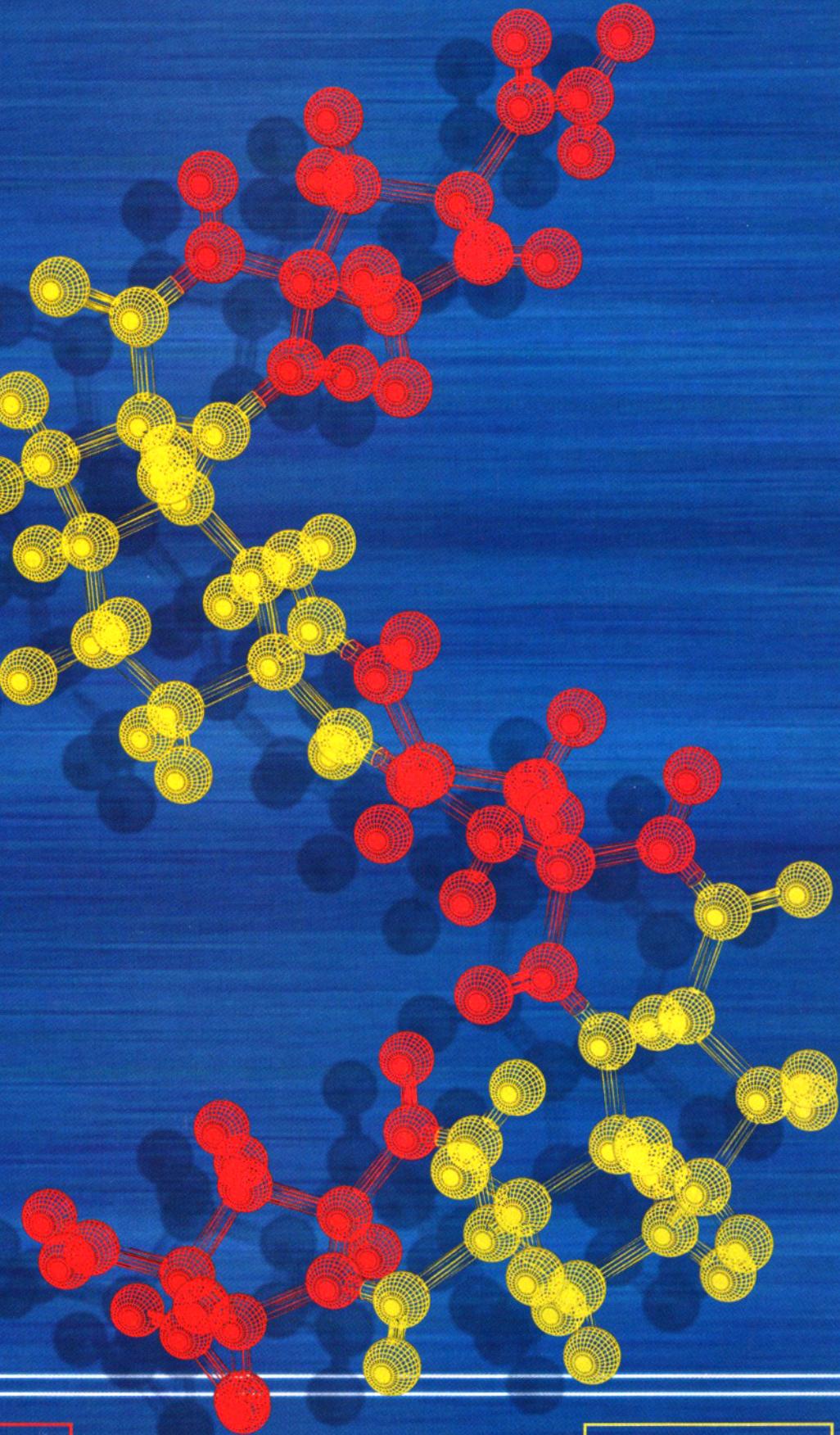
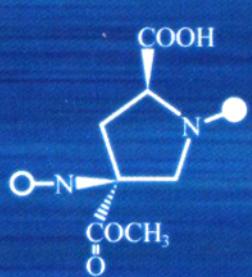
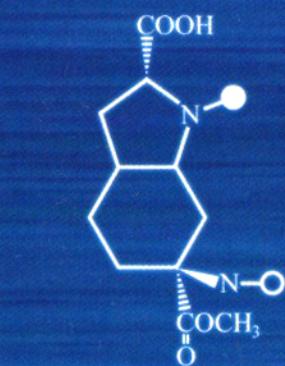
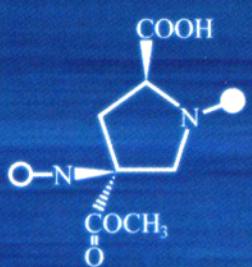
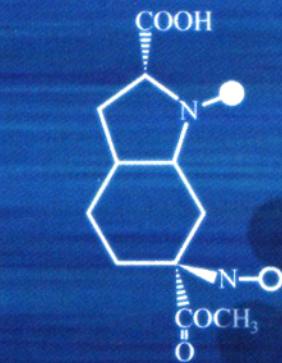
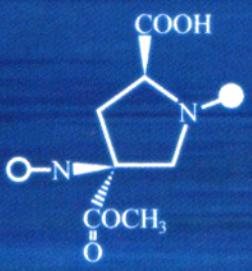
By Christian E. Schafmeister

**P**roteins, the fundamental nanomachines of life, have provided scientists like me with many lessons in our own efforts to create nanomachinery. Proteins are large molecules containing hundreds to thousands of atoms and are typically a few nanometers (billions of a meter) to tens of nanometers across. Our bodies contain at least 20,000 different proteins that, among other things, cause our muscles to contract, digest our food, build our bones, sense our environment and tirelessly recycle hundreds of small molecules within our cells.

As a chemistry undergraduate in 1986, I dreamed of the possibility of designing and synthesizing macromolecules (molecules containing more than 100 atoms) that could do the amazing things that proteins do and more. I have programmed computers since the first TRS-80s came out in the late 1970s, and I thought it would be wonderful if I could build complex molecular machines as easily as I could write software. I wanted to create a “programming language for matter”—a combination of software and chemistry that would enable people to describe a nanomachine’s shape and would then determine the series of chemical processes that a chemist or a robot should carry out to build the nanodevice.

Unfortunately, the idea of inventing nanomachines by designing new proteins runs into a severe obstacle. Every protein generally starts as a simple, linear chain assembled from a specific sequence of amino acids drawn from a repertoire of just 20 amino acids. So far, so good, but the properties of a protein and what functions it can carry out depend on its shape. Shortly after the chain of amino acids is put together in the cell, it collapses into an intricate tangle of helices and other structures through a complex process called protein folding. The sequence of amino acids determines the final shape, but predicting what shape a particular sequence will take on is one of the most significant unsolved challenges of science and engineering (the “protein folding problem”).

CREATING NEW NANOSTRUCTURES similar to proteins is made practical with a collection of building-block molecules (bottom) developed to join together and form rigid structures whose overall shapes are completely preplanned by the designer, like a model made of tiny, oddly shaped Lego bricks.



Some 20 years after I first entertained my vision of the future, my laboratory has at last developed a way to produce large molecules with programmable shapes and the computer software required to design them. Our approach is inspired by the modularity of natural proteins, but it does not rely on amino acid chains to collapse spontaneously into a shape—so it avoids contending with the unsolved folding problem.

We are developing this technology to create molecules that can carry out specific functions. One of our initial goals is to create sensors: large molecules that change shape and color when they bind to particular target molecules, such as glucose, toxins or chemical warfare agents. The binding event triggers the sensor molecule to swing two fluorescent groups together that alter its color, thereby signaling that the target is present in the sample. We are also using our technique to create long, hinged molecules that open and close in response to an external signal—a step toward the creation of molecular actuators, molecular valves and computer memories.

We envisage that our technique will ultimately lead to an even more advanced method of constructing nanomachines: we would use it to fashion complex nanotools such as an assembler that, like the ribosome responsible for constructing proteins inside cells, would assemble other nanomachines under external programmer control. For now, this second dream lies in the future.

## Lessons from Nature

WHEN I FINISHED my undergraduate studies in 1990, I thought that the path to developing nanomachinery lay in deducing the rules of protein folding and using them to develop new proteins. I joined Robert M. Stroud and his protein crystallography group at the University of California, San Francisco. Protein crystallographers grow crystals of proteins and use x-rays to determine the exact three-dimensional arrangement of the proteins' atoms. Using this tool, I developed a deep appreciation of the complexity and beauty of protein structure. I spent four years creating 4HB1, an artificial protein of my own design. I first assembled an artificial gene and then inserted it into bacteria, which "expressed" it—that is, made the protein encoded by the gene's DNA. Next I crystallized the resultant protein and determined its x-ray crystal structure. It was thrilling to discover that 4HB1 had the conformation I had designed it to have!

Yet after all this work, 4HB1 was a molecular doorstop. It did not do anything other than exist as a well-folded artificial protein. Most disturbing was that the experience did not reveal the simple rules I needed to create other proteins of a desired shape. On the contrary, the complexity of protein folding suggested that such simple rules might not exist. While finishing my Ph.D. in 1997, I concluded that a better way to create custom-designed nanomachinery would be to construct them from a limited set of modular building blocks that did not attain

their shape via the folding process of proteins.

This was not a new idea. In 1995 Brent Iverson of the University of Texas at Austin had developed building blocks that could be chained together into short polymers called oligomers. These oligomers then self-assembled into pleated structures as electron-rich donor groups pulled on electron-deficient acceptor groups in the structure.

At about the same time, Sam Gellman of the University of Wisconsin-Madison and Dieter Seebach of the Swiss Federal Institute of Technology in Zurich were developing synthetic molecules called beta-peptides, which are flexible chains of beta-amino acids—molecules that are mostly not naturally occurring and whose general structure is slightly different from that of regular amino acids (alpha-amino acids). Gellman and Seebach's short beta-peptides fold into twisted helices.

These new approaches to constructing macromolecules that held a specific shape were inspiring, but they seemed to trade one folding problem for another. The difficulty is that natural proteins and these new molecules involve chains of molecules connected by single bonds that leave the structure with a lot of freedom to bend at locations all along its length. Which way one of these molecules bends in acquiring its final shape depends on the complex interplay of attractive and repulsive forces arising when different building blocks all along the chain are brought closer together.

I had a more radical approach in mind. I wanted to eliminate the usual folding process altogether and thus gain more control over the shape of the final product. To achieve this goal, I set out to invent rigid building blocks that could be attached to one another through pairs of bonds to create rigid, ladderlike macromolecules. This idea had been tried before: in 1987 J. Fraser Stoddart, then at the University of Sheffield in England, introduced the concept of a "molecular Lego set" by creating molecular belts and collars from building blocks.

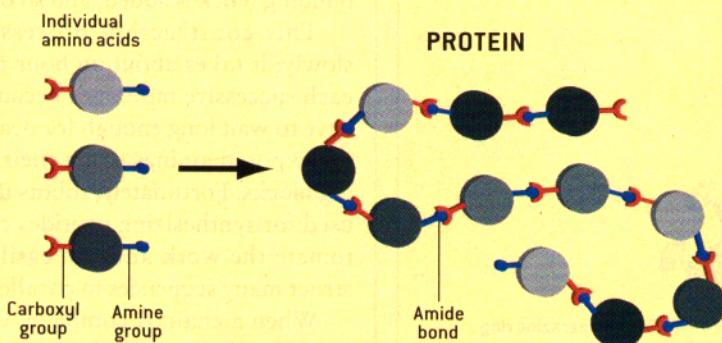
## Overview/Nano Lego

- Proteins are nature's nanomachines, tirelessly carrying out myriad biological tasks. Because proteins are made of flexible chains of amino acids that fold up in a very complicated way, scientists cannot easily predict a new protein's shape (and hence its function).
- Now chemists have developed a library of novel molecular building blocks, called bis-amino acids, that can be strung together to form proteinlike structures that have rigid, readily predictable and designable shapes.
- Potential applications for these "bis-peptides" include medicines, enzymes for catalyzing useful reactions, chemical sensors, nanoscale valves and computer storage devices.

## HOW BIS-PEPTIDES DIFFER FROM PROTEINS

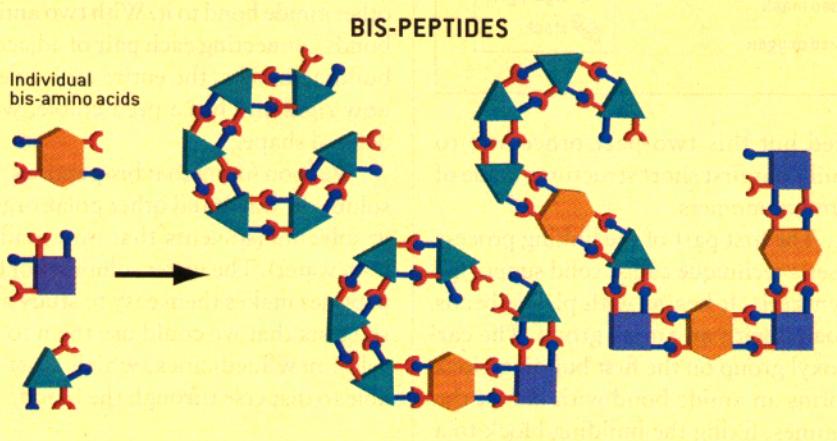
### NATURAL PROTEINS

Organisms make 20 different amino acids that they string together into flexible chains that are generally called peptides when they are short and proteins when they are long. The amino acids are joined by amide bonds, which form when carboxyl and amine groups react together. The final shape of a protein depends on the complex interplay of interactions occurring among amino acids all along its length. This complexity makes it extremely hard to predict what shape a new amino acid sequence will take on. [All figures are highly schematic for clarity.]



### PREDICTABLE BIS-PEPTIDES

Chemists have produced a library of building blocks called bis-amino acids that sport two pairs of carboxyls and amines. When linked together, these building blocks, or monomers, form a rigid chain called a bis-peptide that has a predictable shape directly determined by the sequence of bis-amino acids selected. Therefore, chemists can design and build precise nanostructures simply by combining bis-amino acids in a specific order.



I joined the laboratory of Gregory Verdine at Harvard University to learn synthetic organic chemistry. During two years of synthesizing unnatural amino acids and searching for a route to my larger vision, I came across a paper that described a chemical structure called a diketopiperazine. In this structure, six atoms join into a ring containing two amide bonds [see box on next page]. Amide bonds are the ones that link a protein's constituent amino acids together in a chain, like a line of people holding hands. A diketopiperazine arises

when two amino acids come together like two people facing each other and holding both hands, their arms forming a closed ring. Chemists who synthesize proteins have developed many excellent reactions for forming amide bonds between amino acids, and they are all too familiar with the diketopiperazine structure, because it can form when it is not wanted and interfere with their efforts to synthesize proteins. I figured, though, that I could make use of diketopiperazine formation to link my building blocks.

The rest of the idea soon fell into place. In the “people” analogy, the two “arms” of an amino acid are groups of just a few atoms called the amine group and the carboxyl group. (Unlike arms, however, these groups do not actually stick out very far.) Think of one as the left arm and the other as the right, with an amide bond being a left hand holding a right hand. Each of my new building blocks, or monomers, would be like two people tied rigidly together (for example, back to back) with their arms in front of them. One monomer would connect with the next in the sequence by a person on one holding both hands of a person on the other—forming a diketopiperazine ring.

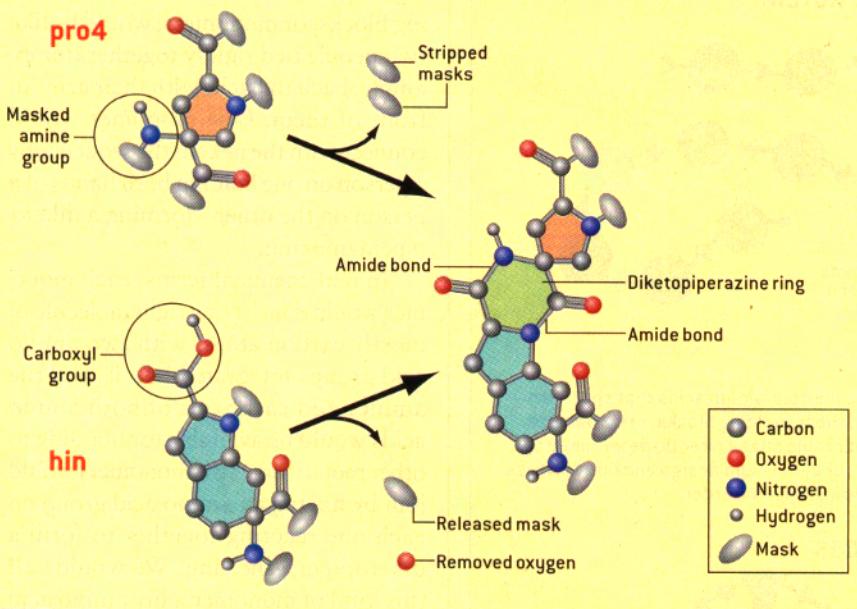
In real chemical terms, each monomer would consist of a rigid molecule of mostly carbon atoms with two amino acid groups integrated into it, and the amines and carboxyls of both amino acids would be available for bonding to other monomers. Two monomers would join by having an amino acid group on each one reacting together to form a diketopiperazine ring. We would call this kind of monomer a bis-amino acid (“bis” meaning “twice”) because each one contains two amino acids. And just as chains of amino acids are called peptides, we would call our chains of bis-amino acids “bis-peptides.”

### Starting from Scratch

WITH BLUEPRINTS for a collection of building blocks in hand, I launched a new lab at the University of Pittsburgh, where my students and I could develop the synthetic chemistry to make this idea work. Within two years Christopher Levins, one of my first graduate students, had synthesized our first bis-amino acids. He started with hydroxyproline, a commercially available component of collagen (the protein that makes cartilage, ligaments and tendons strong) that another group had previously used in making molecules very like our monomer design. Using a nine-step recipe that we worked out together, Levins converted hydroxyproline into four kinds of building blocks, which we named pro4(2S4S), pro4(2S4R),

## THE CHEMISTRY

In practice, chemists synthesize bis-amino acids with protective groups, or masks, to prevent bonds from forming among them indiscriminately. Using a series of steps [not depicted here], the chemists link two monomers—such as pro4 and hin, whose chemical structures are shown at the left—by inducing what is called a diketopiperazine ring (green) to form between them. The rigidity of this ring and of the other carbon rings within the bis-amino acids ensures the stiffness and predictable shape of the resulting chains. (Some hydrogen atoms and details of the protective groups have been omitted for clarity.)



pro4(2R4S) and pro4(2R4R). We call them “pro4” because they all resemble the amino acid proline with an additional amino acid mounted on carbon 4 (chemists identify the carbon atoms in an organic molecule by labeling them with numbers in a systematic fashion). The labels “S” and “R” indicate the orientation of the groups attached to carbon 2 and carbon 4. The completed building blocks are dry powders that are stable for months of storage at room temperature.

We construct our monomer building blocks with protective groups attached to the amines (to prevent amide bonds from forming until we want them to) and with one of the carboxyls in a modified, less reactive form called an ester. To synthesize a bis-peptide, we assemble the building blocks in the desired sequence with single bonds and then join up all the second bonds to rigidify the molecule into its final shape [see box on opposite page]. Levins car-

ried out this two-part procedure to build our first short structures made of pro4 monomers.

The first part of the linking process uses a technique called solid supported synthesis. It begins with plastic beads coated with an amine group. The carboxyl group on the first building block forms an amide bond with one of the amines, fixing the building block to a bead. Using an excess of building blocks ensures that virtually all the amines on the beads have a building block attached. A quick wash with a solvent removes by-products and leftover building blocks. Then a wash with a base removes the protective group from one

of the two amines on the newly added building block (the two amines have different protective groups, so only one of them is stripped). A second building block is added and attaches to the first through its carboxyl and the exposed amine group. The protection is then removed from one of its amines, a third building block is added, and so on.

This construction process goes slowly: it takes about an hour to add each successive monomer because we have to wait long enough for nearly all the exposed amines to get their building blocks. Fortunately, robots usually used for synthesizing peptides can automate the work and can easily construct many sequences in parallel.

When a chain is complete, we use strong acid to remove the beads, then strip the second amine protective group from every building block within the chain. Adding a base solution causes the newly revealed amine on every building block to attack the ester on the preceding building block and form another amide bond to it. With two amide bonds connecting each pair of adjacent building blocks, the entire molecule is now rigid and has a predictable, well-defined shape.

We soon found that bis-peptides are soluble in water and other polar organic solvents (solvents that mix readily with water). The water solubility of bis-peptides makes them easy to study and suggests that we could use them to develop new medicines, which must be able to disperse through the blood.

### Programming Shapes

THE BIS-AMINO ACIDS that make up our bis-peptides join together like strangely shaped Lego bricks. In particular, each bis-amino acid behaves like a brick whose top surface of studs is tilted and twisted relative to its bot-

#### THE AUTHOR

CHRISTIAN E. SCHAFMEISTER is an associate professor of chemistry at Temple University, where he is developing shape-programmable molecules. He received his Ph.D. in biophysics at the University of California, San Francisco, in 1997. As a postdoctoral fellow at Harvard University, he developed a new way of making peptides more resistant to proteases, rendering them more appropriate as potential drugs. He is a member of the working group preparing the Technology Roadmap for Productive Nanosystems for the Foresight Nanotech Institute in Palo Alto, Calif.

tom surface of holes. Repeatedly stacking one type of brick on top of itself allows you to make one curved shape, with the specific shape of the curve depending on which bis-amino acid is chosen. Using just two different kinds of bricks stacked in different sequences, you can make  $2^N$  different shapes ( $N$  is the number of bricks in the stack). A bis-peptide 10 blocks long made out of our four pro4 bis-amino acids could have any one of about a million ( $4^{10}$ ) shapes. The more shapes of building blocks we have, the better we will be able to control the final shape of the macromolecule. The challenge then is to design and synthesize those sequences that have useful functions.

The key to designing bis-peptides with specific shapes is knowing the precise shapes that our individual bis-amino acids take on when they are joined to one another. This information, analogous to knowing the size of each brick and the tilt and twist of its studs, would become the basis for our “programming language for matter.” Having synthesized our first bis-peptides, we could

then carry out measurements to determine how their pieces fit together.

We performed nuclear magnetic resonance experiments to find out which hydrogen atoms on a bis-peptide are close to one another and applied other techniques to measure the orientations of carbon-hydrogen bonds. From the results of these measurements we inferred the shape information that we needed, and we used it to create a computer-aided design program for building bis-peptides called CANDO (for computer-aided *n*anostructure design and optimization).

Gregory Bird, another graduate student in my lab, used CANDO to design molecular rods and curved structures. Recently he assembled these structures, attaching a chemical group called a spin probe to each end of every sequence to verify that the results in the reaction vessel matched the design in the computer. Indeed, sequences of pro4(2S4S) and pro4(2R4R) building blocks had C and S shapes just as CANDO predicted they would.

The pro4 group of bis-amino acids

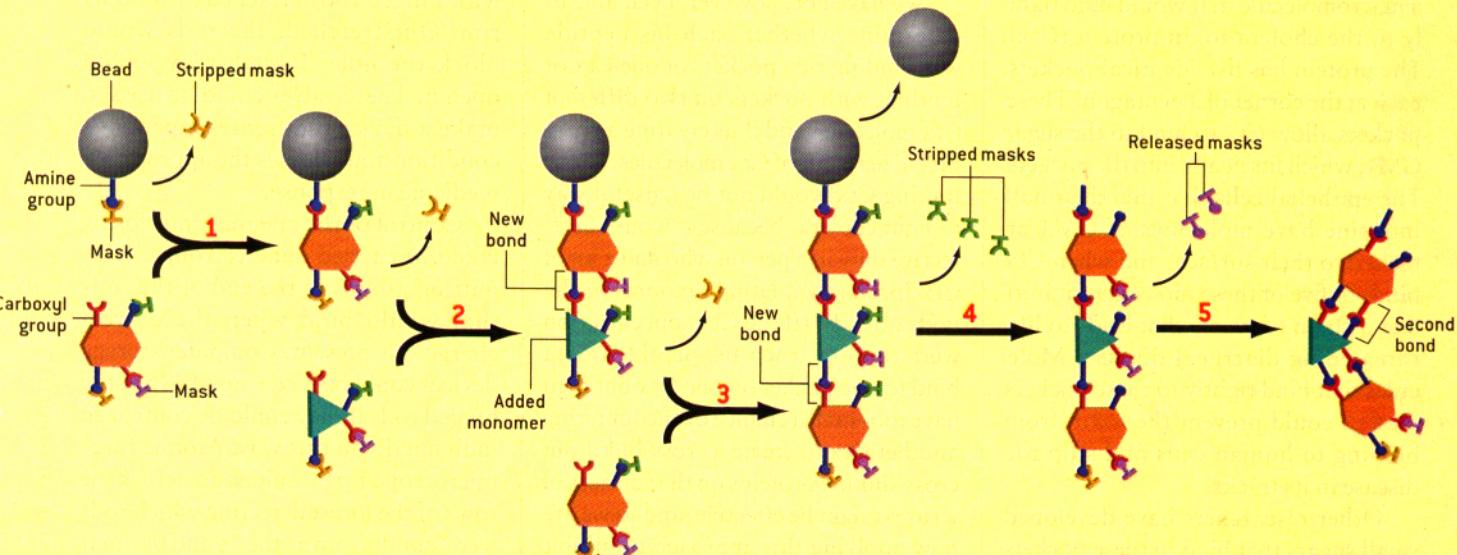
are like Lego bricks that have relatively small tilts, so we can use them to make rodlike and gently curving shapes, which could function like struts to hold chemical groups apart at specific distances. Many useful functions of proteins, however, come about because of cavities that can serve to bind the protein to a specific target or to hold molecules and catalyze reactions. To create compact bis-peptides that have suitable cavities, we needed to expand our repertoire of building blocks. My student Stephen Habay took the first step toward this goal by developing a bis-amino acid we call “hin” that creates a sharp turn in a bis-peptide.

Year by year our collection of monomers continues to grow, and CANDO analyses suggest that our present repertoire of 14 monomers is sufficient to create compact bis-peptides containing cavities. But as we developed new building blocks and incorporated them into bis-peptides, we ran into a problem. The reaction that forms the rigidifying second amide bond was very rapid between pro4 monomers but was

## HOW BIS-PEPTIDES ARE MADE

The synthesis of a bis-peptide proceeds by first assembling selected bis-amino acids in the correct sequence and then rigidifying the structure: A bead is coated with a protected amine group. The protective mask [yellow] is stripped away, and the first bis-amino acid latches onto it via the bis-amino acid's free carboxyl group [1]. The process is repeated [2 and 3] with

more bis-amino acids, producing a chain linked by single bonds. Then the bead is stripped away, as are the protective groups [green] on the unbonded amines [4]. The freed amines react with the nearby masked carboxyls, releasing the masks and forming a second bond between each adjacent pair of monomers [5].



sluggish for all our new building blocks. Raising the reaction temperature sped things up but scrambled the resulting shapes. This problem was a huge obstacle to creating larger and more complex bis-peptides.

My student Sharad Gupta partially overcame this challenge by developing a new approach to closing the second amide bond. On each monomer he changed the ester to one that is more susceptible to the amine's attack, and, inspired by a 1970s paper, he used acetic acid as a catalyst instead of a base. The combination of heat and acid accelerated the ring-closing reaction without scrambling our bis-peptides' shapes in the way that heat and base did.

We took six months to find the combination of ester, protective group, solvent and temperature that we have settled on for now, but we will return to this problem in the future because our solution does not work well for sequences longer than about five monomers. In the meantime, we are focusing on developing some applications with the bis-peptides that we can produce efficiently—those of any length that involve only the pro4 monomers, and sequences of up to five monomers that include the others.

## Developing Applications

ONE OF THE FIRST applications that we have pursued for our bis-peptides is a macromolecule that would bind tightly to the cholera toxin protein (Ctx). The protein has five identical pockets, each at the corner of a pentagon. These pockets allow Ctx to bind to the sugar GM1, which fits neatly into the pockets. The epithelial cells that line the small intestine have molecules of GM1 attached to their surface, and when Ctx binds to five of these molecules, it initiates a chain of events that leads to life-threatening diarrheal disease. Molecules that bind tightly to these pockets on Ctx could prevent the toxin from binding to human cells and stop the disease in its tracks.

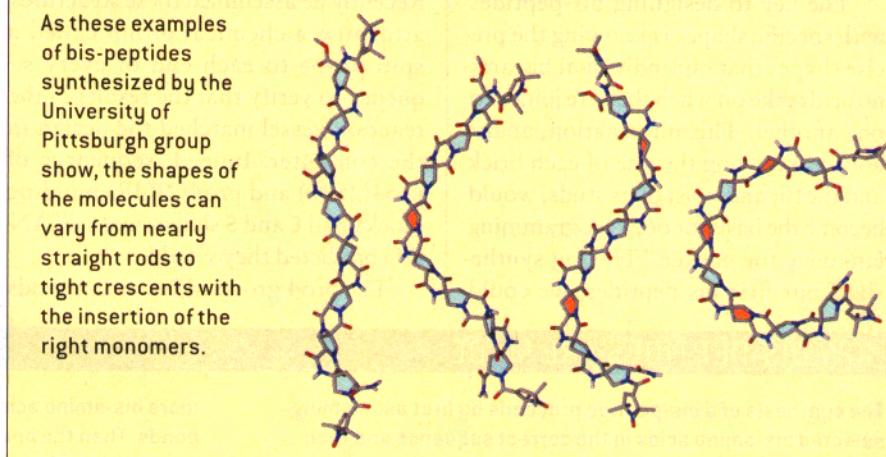
Other researchers have developed small sugars that bind to these pockets individually. But those drugs do not

work well, because they do not bind very tightly to Ctx and cannot compete with the five simultaneous interactions that Ctx makes with GM1 on human cells. We wondered whether we could synthesize a bis-peptide that could plug sugars into two pockets at the same time. We can attach almost anything we want at the ends of a bis-peptide, so for this application we put a small sugar on each end of rod-shaped bis-peptides that just span the distance between adjacent pockets in the Ctx protein. The experiment worked in that bis-peptides with two sugars bound to Ctx more tightly than the individual small sugars, and

As well as attaching groups to the ends of a rigid rod, we have developed molecular actuators in which two rods are joined by a hinge. An actuator is a device that responds to a signal by producing motion. Our rod-hinge-rod actuators are designed to be open normally and to fold over, or close, when groups on the outer ends of the rods bind a metal or a small molecule. My student Laura Belasco made our first version of these, in which the rods are four building blocks long, the hinge is an ordinary amino acid, and a metal triggers the opening and closing. One application would be molecular valves

## DESIGNER SHAPES

As these examples of bis-peptides synthesized by the University of Pittsburgh group show, the shapes of the molecules can vary from nearly straight rods to tight crescents with the insertion of the right monomers.



they bind at least as well as the natural GM1 target does.

We have not, however, been able to determine whether each bis-peptide was binding two pockets of one Ctx or binding with pockets on two different Ctx molecules and thus creating a cross-linked network of Ctx molecules. Cross-linking Ctx would not be a useful way to fight cholera, because it would be effective only in a person who had a lot of Ctx (probably a lethal amount) in the body already. (If the Ctx concentration were too low, each bis-peptide might bind to one pocket on one Ctx but then have too small a chance of encountering another Ctx to create a cross-link.) But cross-linking proteins on the surfaces of viruses might be effective, and so we are now applying this approach to inhibiting viruses, including HIV and Ebola.

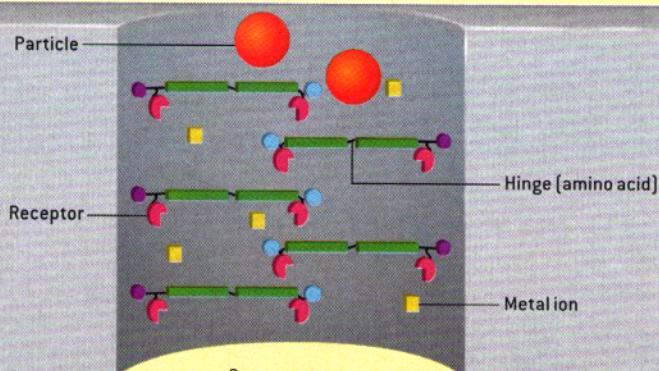
[see box on opposite page]. The valve would consist of a nanoscopic hole with hinged rods attached around its rim. Outstretched, the rods would block the hole; folded, they would open it. These valves could be used to make a device that senses a patient's condition and releases the appropriate medicine in response.

Control of the opening and closing could be carried out electronically by putting groups at the end of the rods that would bind when the correct charge was present. Computer storage devices could be made out of a forest of hinged rods if they could be controlled individually in this way. Atomic force microscope tips would scan across the rows of the forest detecting which rods were standing up as the 1s and 0s, analogous to detecting the pits or no pits of

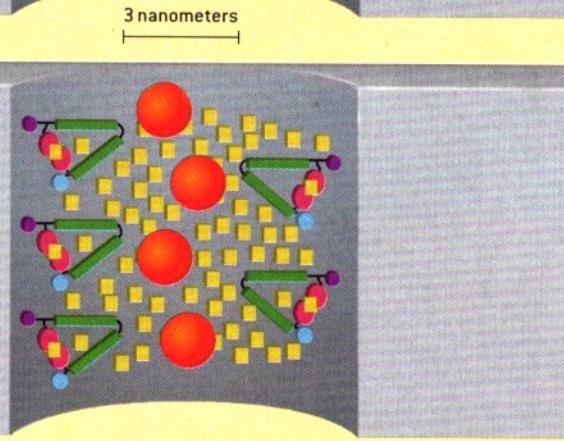
## NANOSCALE VALVES

Valves that open to a mere three-nanometer diameter using bis-peptide actuators are on the drawing board. The actuators, which have been synthesized and demonstrated, consist of two short bis-peptide rods [green] joined by an amino acid that serves as a hinge. In a low concentration of the triggering metal ion [yellow], the actuators extend from the rims of tiny holes etched through aluminum films [gray], blocking larger particles or molecules [orange] from passing (top). At high concentrations the ions bind to receptors [pink], causing the actuators to fold over, opening the channel (bottom).

### CLOSED CHANNEL



### OPEN CHANNEL



IBM's "Millipede" drive. Erasing a pit, which is difficult for the Millipede system, would be as simple as reversing the state of the hinged rod.

The side chains of the 20 amino acids that organisms use to build their proteins are decorated with a variety of chemical groups. Proteins position these chemical groups in configurations whose shape and other properties serve to catalyze reactions, bind small molecules and carry out their many functions. Similarly, in our lab we are developing building blocks that carry an additional chemical group, which will let us create bis-peptides that display chemical groups along their ladderlike back-

bones. So far we have made the first such building block with a side group. If we can make macromolecules with constellations of chemical groups that mimic the active sites of enzymes—the areas where catalysis takes place—we

### MORE TO EXPLORE

**Constructing a Molecular LEGO Set.** John P. Mathias and J. Fraser Stoddart in *Chemical Society Reviews*, Vol. 21, No. 4, pages 215–225; 1992.

**Foldamers: A Manifesto.** S. H. Gellman in *Accounts of Chemical Research*, Vol. 31, No. 4, pages 173–180; April 1998.

**A Field Guide to Foldamers.** D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes and J. S. Moore in *Chemical Reviews*, Vol. 101, No. 12, pages 3893–4011; December 2001.

**Flexibility and Lengths of Bis-Peptide Nanostructures by Electron Spin Resonance.** S. Pornsawan, G. Bird, C. E. Schafmeister and S. Saxena in *Journal of the American Chemical Society*, Vol. 128, No. 12, pages 3876–3877; March 29, 2006.

Christian E. Schafmeister's laboratory site: [www.candolab.org](http://www.candolab.org)

could use them to learn how to create designer enzymes.

Twenty years from now I envision an active community of developers: dozens of groups inventing designer bis-peptide-based macromolecules and learning how to produce artificial enzymes and other useful molecular devices. Some promising anticancer drugs such as halichondrin-B and bryostatin are currently very expensive to synthesize. The rare sponges and sea creatures that produce these compounds cannot provide the quantities needed for widespread use. In 20 years we might be able to create artificial enzymes that efficiently synthesize these and other valuable compounds in an environmentally benign way. Imagine adding a drop of artificial enzymes to a barrelful of high-fructose corn syrup and a few days later harvesting gallons of bryostatin.

If we could develop artificial enzymes that break down plant cellulose into ethanol or that use light energy to combine water and carbon dioxide to create ethanol, such an undertaking would have massive benefits for society. We could even design artificial enzymes to synthesize our bis-amino acid building blocks and join them together, making it much easier to make bis-peptides.

We have developed a combination of chemistry and software for creating macromolecules with programmable shapes. Because it takes only a few days to produce bis-peptides, we can design and assemble them, test their properties and fashion the next generation on a timescale of weeks. The fascinating challenge in coming years will be to learn how to begin with a function and to design the best bis-peptide sequence for carrying it out.