BIOCHEMISTRY

Molecule Makers Learn the Rules of a Crooked Game

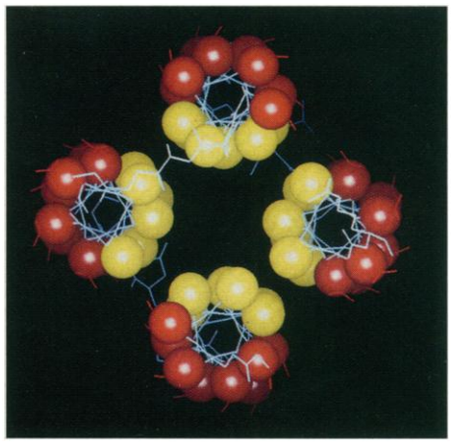
LOGAN, UTAH—It takes more than one language to read the book of life. Biologists are fluent in the four-letter code of the genes, which specifies the sequence of amino acid building blocks in the proteins that carry out the business of life. But knowing a protein's amino acid sequence is not enough to understand its function. For that biologists need to know how the linear chain of amino acids folds up in solution into a specific three-dimensional structure. Today, a protein's structure can only be discovered by painstaking analysis. To ease the job, researchers need another language, one that will enable them to read a linear sequence of amino acids and predict the finished protein's complex architecture of coils, sheets, and folds.

That level of fluency in the language of protein folding is still a long way off, but the chemists who gathered here for a conference on molecular design* say they are starting to grasp some basic elements of its grammar. It's no accident that these researchers were discussing their results at a meeting devoted to tailor-making molecules, because they are learning the rules of protein folding by doing just that: They test possible rules by designing and making synthetic proteins not found in nature, then analyzing them to see whether the proposed rule has predicted the actual three-dimensional form of the protein.

The study of protein folding has traditionally been an esoteric pursuit, in which researchers have sought ways to calculate protein structure from first principles (see box). This newer empirical approach, says Princeton chemist Michael Hecht, is more like engineering. In learning the rules of bridge design, he says, "the ultimate test of understanding is whether you can design and build a bridge from scratch." This strategy has already enabled Hecht and others to pin down some of the basic rules governing which amino acid sequences will coil up into compact bundles of helixes and what's required for the branching side chains in a protein's interior to interlock neatly—two typical structural motifs of functional proteins.

If Hecht and his colleagues could expand these results into a complete grammar of protein folding, they would be handing new powers to biomedical researchers. For genome researchers grinding through billions of characters of the human genetic code, the rules of protein folding would provide what Brandeis protein chemist Gregory Petsko calls a "Rosetta Stone" for interpreting protein function. For geneticists who have implicated a particular gene in a disease but don't know how it works, the structure of the gene's protein could hold clues—and help in devising therapies. And for researchers trying to develop synthetic proteins for medicine and technology, the ability to predict form and function from sequence would be invaluable.

A few years ago, those hopes would have seemed like a pipe dream. In the 1970s, when chemists were following the time-honored tradition of "solving" protein structures by crystallizing each protein and probing it with x-rays, each new folding pattern seemed as distinctive as a fingerprint. But now, regularities are beginning to emerge. About 2 years ago, says Petsko, chemists noticed that these myriad structures form families—and that there are probably fewer than 1000 of these closely related groups. As a result, Petsko concludes, "the whole nature of the game has changed."



S. KAMTEKAR, J. LOEB, M. H. HECHT

Designer protein. An alternating water-repelling *(yellow)* and water-attracting *(red)* amino acids curls into four helixes.

Architectural elements. Those regularities hinted at the existence of a grammar underlying protein folding. Meanwhile, the emerging technology of genetic engineering gave chemists an easy way to test candidate rules: Synthesize a stretch of DNA encoding a sequence of amino acids, slip the DNA into a cell, allow the cell to "express" the artificial protein—and see whether the rule's prediction is borne out in three dimensions.

Early efforts at such protein building, by Jane Richardson of Duke University Medical Center and others, pinned down specific sequences that can mark the beginning and

*International Congress for Design & Modification of Biomolecular Function, 2-4 March.

The Calculated Approach to Protein Folding

Today, when chemists try to understand how a linear sequence of amino acids will fold up inside a cell, many of them take an empirical approach. They formulate rules, then test the rules one at a time by making artificial proteins (see main story). But that's not how researchers have traditionally approached this problem; instead, they have tried to find ways to predict the exact form of any protein from its amino acid sequence alone.

At the Utah meeting, that approach was represented by one of its foremost practitioners: Cornell chemist Harold Scheraga. For the last 35 years, Scheraga has been working on computer algorithms that could take any sequence of amino acids and calculate the most stable structure, based only on the chemical characteristics of its constituent atoms. Scheraga's algorithms consider the possible relative positions of pairs of atoms in a protein to identify the overall structure that has the lowest potential energy.

The difficulty is that, even with massive computing power, such an algorithm can't explore every possible conformation of a protein, and it thus risks missing the most stable one. One way to understand the difficulty, Scheraga explains, is to picture a hypothetical landscape of potential energy. When a protein folds, it should adopt the conformation that corresponds to the deepest valley; the algorithm seeks the lowest-energy conformation by moving downhill from its starting point. But if the landscape has many valleys separated by mountains, the algorithm can easily get "trapped" in a less-deep valley.

Now, Scheraga says his group is close to a solution. The general idea, he says, is to modify the computation to wipe out the mountains, making the deepest valley easier to find. Scheraga says his group has tested this strategy with a 5 amino acid molecule and is now applying it to a 36 amino acid protein. He believes his first-principles method is the "most exciting approach to solving the folding problem." But some other chemists at the meeting place more stock in their experimental approach.

–F.F.

end of the chief "secondary structures" within the larger architecture of proteins: alpha helixes and the folds of the so-called beta pleated sheets. By equipping artificial proteins with these sequences, Richardson and her colleagues were able to mimic the overall shape of many natural proteins: a compact globule consisting of a bundle of helixes.

This approach hasn't stopped being fruitful. Du Pont chemist William DeGrado says he's taken it a step further by identifying short sequences—even single amino acids—that can help constrain a protein's configuration by ruling out alternatives. Since an amino acid sequence can usually fold into several different structures, both natural and artificial sequences must include safeguards to prevent then from snapping into undesirable configurations. "It's as if you are giving someone instructions to go from L.A. to New York and you have to tell them not to get off at every exit before the one you want." For example, when DeGrado tried to design a double-helical structure, he found he had to alter the sequence to prevent one helix from doubling again, forming a triple helix.

Princeton's Hecht, meanwhile, suspected that there might be more general rules for protein structure that don't depend on the presence or absence of specific amino acids. He and others had noted an underlying pattern, for example, in the sequence of natural proteins that have a quadruple coil structure known as a four-helix bundle: Sets of hydrophilic (water attracting) and hydrophobic (water repelling) amino acids alternate at intervals along the sequence. The intervals are just the right length to place the hydrophobic amino acids on the inside of each coil and the hydrophilic ones on the outside.

Hecht and his colleagues mimicked this pattern in a series of synthetic proteins, choosing amino acids that showed the alternating hydrophilic/hydrophobic pattern, but were otherwise chosen at random. The researchers then analyzed the test proteins by measuring their optical properties—a rough indicator of structure. To Hecht's surprise, 60% of the several hundred proteins had folded into four-helix bundles (*Science*, 10 December 1993, p. 1680). "No one thought such semi-random sequences would fold into stable structures," he says. He believes that result implies that overall protein structure doesn't always depend on the presence or absence of specific amino acids. Instead, much more general chemical properties may be capable of determining form.

Interior design. Still, the rules deciphered so far aren't nearly enough to specify the full architecture of a protein, as Hecht, DeGrado, and Richardson have found. The interior of a protein, crowded with branching offshoots known as side chains, is still largely terra incognita. In natural proteins, these chains, whose position is critical to functions such as catalyzing chemical reactions, are locked, jigsaw-puzzle fashion, in just the right pattern. But in many synthetic proteins, Hecht says, the side chains are "flopping in the breeze."

DeGrado is now concentrating on learning how natural proteins pin down their side chains—and how he can do the same, whether by nature's means or his own. In 1992 he showed that added metal atoms, each of which can bind to two or three side chains, can stabilize the interior of a synthetic protein. But while that finding could aid researchers trying to design artificial molecules, it doesn't add much to the understanding of natural proteins, most of which don't rely on metals for stability. More recently DeGrado has experimented with artificial sequences in which the amino acids alternate in size, which should allow them to pack in a stable, orderly array. First published in 1991 in the *Journal of the American Chemical Society*, the approach has since yielded structures with four helixes that, says DeGrado, have the interior order of true proteins.

Researchers eager to design functional proteins aren't waiting for Hecht, DeGrado, and their colleagues to decipher all the rules of protein folding. Leslie Dutton of the University of Pennsylvania, for example, recently applied some of the known rules to devise the first artificial protein that has a stable interior and is also designed to carry out a specified function. Dutton simplifies the problem of controlling a synthetic protein's structure by streamlining it, reducing the molecule to a kind of scale model of the real thing. "We identify what we think are the key elements in a natural protein, and then synthesize these in a test tube without the biological baggage," he says.

In Dutton's simplified protein, the problem of the floppy interior solved itself. As a scaffolding for his protein, he used the multiple helix structure chemists are now adept at designing, then attached four "functional" groups called hemes. The bulky heme groups, he found, propped up the artificial protein's floppy interior, yielding a molecule that had the rigidity of a natural protein, along with a potentially useful property: the ability to transfer electrons, a function crucial in photosynthesis and cellular respiration.

Petsko, too, isn't waiting for the protein-folding rulebook to be completed. To speed up the search for "rational" drugs targeted to disease proteins, he is pushing ahead with traditional x-ray crystallography, cataloguing the structures of natural proteins. But once biologists can reliably predict protein structure from amino acid sequences, his difficult and time-consuming work will be a thing of the past. And Petsko says he's looking forward to that day. "I would love to be put out of business," he says.

–Faye Flam

MOLECULAR BIOLOGY

'Long PCR' Leaps Into Larger DNA Sequences

On Christmas Eve 1992, plant molecular biologist Wayne Barnes showed his son one of Barnes' own new toys: a video machine that displays sharp images of lab gels, which contain molecules separated by size. Barnes, of Washington University School of Medicine in St. Louis, slid in a gel from a recent experiment using the polymerase chain reaction (PCR) technique. PCR amplifies genetic material, making it possible to study specific genetic sequences, and Barnes was pushing the technique to copy longer pieces of DNA than it had ever copied before. When he saw the gel, and the size of the DNA molecule on it, Barnes realized he had succeeded. "I said, 'Boy, this is a fairly big result,'" recalls Barnes.

And, he adds, "big had two meanings." The first meaning of big pertains to the size of the DNA molecules that Barnes can now copy. PCR, invented in 1983, has had a revolutionary impact on molecular biology because it offers an easy way to copy large quantities of DNA, something that was previously a laborious task. Yet until now the technique couldn't reliably copy DNA molecules longer than about 5000 nucleotide bases, putting many molecules of great biological interest out of reach. In the 15 March issue of the *Proceedings of the National Academy of Sciences* (*PNAS*), however, Barnes reports that by optimizing the chemical conditions of the reaction he is able to amplify stretches of DNA seven times as large: 35 kilobases (kb).

The second meaning of big is the impact that large segments could have. "If you had a project that required 10kb before, it was undoable," says Barnes. "Now it's dead easy." In addition to making work easier for scientists studying the human genome, by giving them longer contiguous DNA sequences to analyze, researchers say this "long PCR," as Barnes dubs it, may help them diagnose genetic diseases or copy and study HIV.

Barnes' big leap is only one of several innovations that have begun to make work on PCR seem like molecular biology's version of the broad jump. Scientists at Roche Molecular Systems in Alameda, California, who were trying to make longer PCR products themselves, began collaborating with Barnes last year and have submitted a separate paper to *PNAS* about their progress.