Organic Origami

Scientists study the art of protein folding

By RICK WEISS

Imagine a string of large, colored beads. Keep in mind that yellow ones are attracted to each other like magnets, green ones are repelled by yellow ones, and pink ones are drawn weakly to blue ones, unless there's a brown one in the vicinity. Also remember that red beads like to nestle between yellow ones and are moderately attracted to blue ones, unless there are two greens next to each other within four beads of a purple.

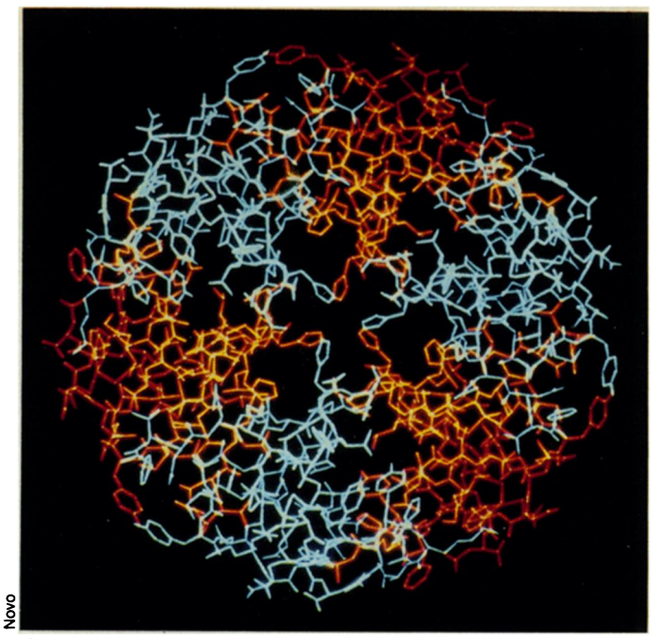
Now make up a random sequence of a few hundred beads and predict the three-dimensional structure that the string will spontaneously fold into.

Need more time? So do the molecular biologists who play this game — and they're playing a much more complicated version than this one. It's known as the protein folding problem, and the colored beads represent amino acids — the variously charged, pearls-on-a-string components of proteins. Scientists hope the game's solution will herald a new era of biological engineering in which highly specific enzymes, antibodies and a variety of drugs will be custom-designed from computer models. They envision the creation of potent chemicals capable of changing our moods, boosting immunity or speeding the healing process.

But to do all this requires insight into protein folding. And those secrets are proving to be extremely well kept.

"The fundamental fact is that you really don't know what a protein is doing until you know its three-dimensional structure," says Robert Langridge, a specialist in protein modeling at the University of California at San Francisco. "Knowing a protein's [amino acid] sequence is fine, and that's terribly important, but in order to really know what it's doing you have to know the three-dimensional structure."

Amino acid sequencing has become almost routine, but scientists are still struggling to learn the basic rules by which sequence affects a protein's three-dimensional form. The aim is not simply to understand the function of existing proteins. Rather it's to predict how a specific amino acid substitution might change the structure — and hence the function — of existing proteins in useful ways. Once the rules become even better understood, scientists foresee designing entirely new proteins to perform specific biological functions.



Novo

A computer-generated model of a human insulin molecule in its hexamer form. While the molecule's six subunits remain linked, the insulin is absorbed into the blood only very slowly. Individual monomers are absorbed rapidly.

In a laboratory, however, it can take years to determine a protein's three-dimensional structure using X-ray crystallography or magnetic resonance imaging—the current "reference" methods for mapping protein topography. Meanwhile, amino acid sequences are being determined every day, creating a growing backlog of sequenced proteins whose structures remain unknown.

"The take home," Langridge says, "is that we've got to figure out how to *predict* from the sequence how these things are going to fold."

What kinds of proteins would scientists like to redesign? The list is growing quickly, as it becomes increasingly clear that proteins are a major class of regulatory chemicals in the body, affecting a host of biological functions ranging from immunity to mood.

Promising work is already underway with engineered versions of insulin, the protein that regulates blood sugar levels and that must be given by injection to diabetics. Scientists have for years sought to design a modified, injectable insulin molecule that would be absorbed either more quickly or more slowly than conventional insulin. Making use of their growing understanding of protein structure and function, researchers at Novo Industri A/S in Denmark have recently designed two new varieties of insulin. One is already in clinical trials, and the other is to be tested next year.

"It was a very rational approach," says Lise G. Heding, of the insulin research team at Novo. The difficulty in getting insulin to absorb more quickly, she explains, is that it tends to form a relatively insoluble six-unit complex after injection. "In order to have a quick-acting insulin we had to prevent the insulin monomers from forming dimers and hexamers, which normally fit together like keys and locks. For example, you can introduce a new amino acid that's slightly bigger, causing a steric hindrance [blockage]. Or you can introduce something with a charge that will keep these molecules from being attracted to each other."

In fact, the puzzle was a bit more difficult than this because the researchers had to restrict their tampering to areas that wouldn't affect the insulin molecule's biological activity. Computer

graphics of insulin's three-dimensional structure, based on previous crystallography work, were very helpful, Heding says.

"But even with that we sometimes made mistakes. There were some cases where we lost biological activity when we didn't expect to. Now we are learning, because we are making all these analogs and each time we get a new one we learn something about what contributes to biological activity."

In order to develop a slow-acting, or protracted, insulin, the researchers took a different approach. Most protracted insulins are suspensions of tiny, insoluble hexamer crystals that need to be resuspended before each injection. Ideally, however, insulin should be soluble rather than crystallized before injection, so that dosage won't vary depending on how well the suspension is mixed.

By changing some of the chemical characteristics of the molecule, the researchers were able to make it soluble while stored in a vial at a low pH. Only after injection and contact with the body's slightly higher pH should the insulin precipitate into millions of small crystals, thus slowing the absorption once it's in the body.

Such predictions are complicated because of the tremendous number of variables involved, and because proteins engage in two magnitudes of folding. The first magnitude, known as the protein's secondary structure (the primary structure is the amino acid sequence itself), refers to the orientation of the amino acids along an imaginary, straight central axis. The most common secondary structures are the alpha helix and the ribbon-like beta-pleated sheet (see top diagram, p.346).

In turn, this configuration affects the protein's tertiary structure, or the way its secondary structure folds upon itself. That final jumble of bends is mediated by a variety of short- and long-range interactions, including positive and negative ionic charges, van der Waals forces, and hydrogen and disulfide bonds—all acting within a matrix of water-soluble and water-repelling molecular subunits.

Most important, a protein's tertiary structure determines its characteristic function. In the case of an antibody, for example, it is the tertiary structure that determines whether a foreign antigen will be bound (SN: 10/10/87, p.228). In the case of an enzyme, tertiary structure defines the type of substrate that can be catalyzed.

For years, researchers quibbled in the scientific literature over approaches for predicting these molecular structures. Finally, in 1974, a sort of scientific contest was held. G.E. Schulz of the Max Planck Institute in West Germany distributed the amino acid sequence of a protein for which he had recently determined the three-dimensional structure using X-ray crystallography. He challenged researchers in the field to use their favorite prediction methods to calculate the protein's secondary structure from its amino acid sequence.

Ten predictions of structure were submitted. Alpha helices were found to be easiest to predict, with various researchers correctly predicting the location of seven, eight or nine of the molecule's 10 helices. One group predicted a helix where there was none. Beta-pleated sheets were correctly predicted approximately half the time, and beta bends were predicted with even slightly better accuracy.

"The interesting thing," Langridge recalls of the study, "was that the average of these looked better than any one of them. So it suggests that there were germs of truth in every one of the prediction methods."

Thirteen years later, however, the determinants of secondary structure are not much better understood. And the much more relevant art of predicting tertiary structure is, in the words of one researcher, little more than a "hodgepodge of inspired guesses."

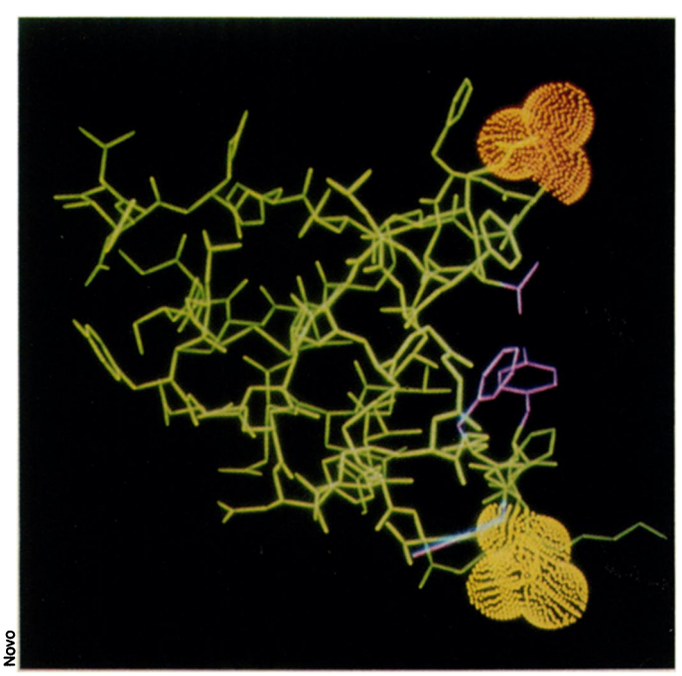
Part of the challenge inherent in the protein folding problem is that a protein in solution does not actually remain locked in a single, stable configuration. Even when at "rest" in solution, a protein will flip-flop among different conformations. And when proteins come into contact with neighboring molecules, any number of changes might be expected to occur.

"The real problem with protein folding is that you're dealing with very transient states," says Peter Wright, chairman of the molecular biology department at Scripps Clinic and Research Foundation in La Jolla, Calif. These transformations can happen very quickly, he notes — sometimes in as little as one-tenth of a second. This dynamic makes it difficult to measure experimentally the "ideal" structure of even simple proteins, from which scientists would like to learn some of the basic rules of protein folding.

"We've been playing games, changing the amino acid sequence, making substitutions and seeing how these changes affect the structure," Wright says. "A single amino acid substitution at a certain place in a beta turn, for example, can either enhance the turn in the original peptide or totally eliminate the turn."

Other researchers with an interest in the folding problem are working with computers rather than with actual proteins. Timothy Havel, also of Scripps, is using distance geometry — a branch of theoretical mathematics — to get a new perspective on protein folding. Havel hopes to use distance geometry to understand how a folded protein can change into another conformation without first completely unfolding.

Those studies may help scientists gain a better understanding of the ways in which proteins interact with each other — an important step toward designing job-specific proteins. For example, in order for an enzyme to catalyze a biochemical reaction it must first undergo a three-dimensional change into its "active" form. That all-important conformational change is the result of the interaction between the enzyme and its specific substrate. If scientists understood how these conformational changes occurred, they'd be a step closer to designing proteins capable of interacting in specific ways with biological targets.



Novo

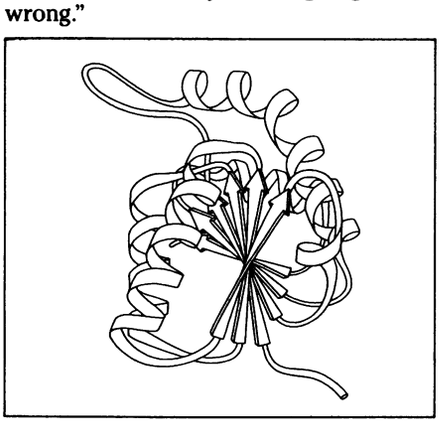
A computerized, skeletal view of an insulin molecule that has been modified by substituting two of its 51 amino acids (shown in orange). The two new amino acids are shown in a way that measures the extent of their weakly attractive van der Waals forces to help scientists see the extent to which the substitution may affect surrounding amino acids. Areas critical to insulin's binding to other molecules are shown in purple.

Other applications for such mutated proteins, or "muteins," abound. Food technologists at Cornell University in Ithaca, N.Y., are replacing the amino acid cysteine in specific locations on the beta lactoglobulin molecule – a common whey protein in bovine milk. They hope to find a version of the lactoglobulin molecule that is stable at warmer temperatures, making it more useful for certain food applications.

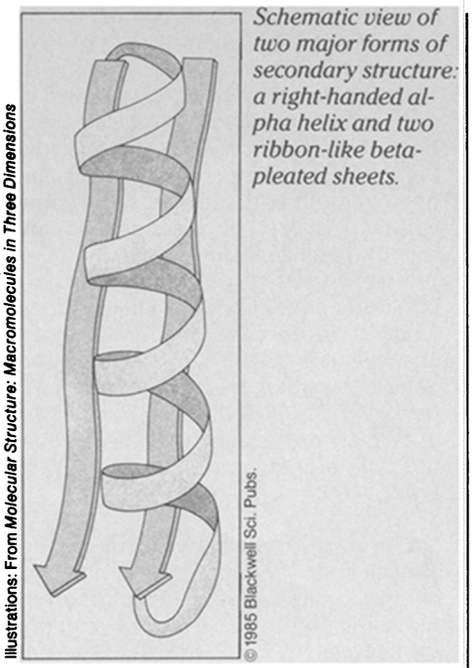
Other researchers are looking into amino acid substitutions that might change the structure and activity of epidermal growth factor (EGF). EGF plays an important role in mammalian cell growth, healing and cancer formation. New research suggests that a small, induced change in the molecule's structure may create a mutant EGF that still binds to its substrate but can't stimulate cancerous growth. In cases where cancer is caused by an excess of natural EGF, this non-cancer-causing EGF might be injected to compete for and block EGF binding sites.

Meanwhile, scientists at Eastman Kodak Research Laboratories in Rochester, N.Y., are experimenting with rationally altered redox enzymes – proteins that catalyze electron transfer reactions. These enzymes are used in medical laboratories for diagnostic assays of blood or tissues. According to Richard Ciccarelli, a member of Kodak's protein engineering group, enzyme structures can be altered to make them more stable at desired temperatures or pH levels by making site-specific amino acid substitutions. Enzymes are extremely specialized molecules, however, and it's difficult to alter them without simultaneously losing biological activity.

"These proteins are stabilized by millions of years of evolution, and you're bound to have problems when you try to speed up evolution by 10 million years," Ciccarelli says. "How do you decide which amino acid to substitute? It's difficult because there is still no real database. There are no ground rules, and a lot of the time you are going to be wrong."



Part of the tertiary structure of the enzyme lactate dehydrogenase, showing the spatial arrangement of its alpha helices and beta-pleated sheets.



Schematic view of two major forms of secondary structure: a right-handed alpha helix and two ribbon-like beta-pleated sheets.

Illustrations: From *Molecular Structure: Macromolecules in Three Dimensions*

In fact, a protein engineering database is gradually accumulating. As researchers keep track of the structural and functional effects of various amino acid substitutions, the rules of protein folding are beginning to emerge. Ultimately, however, everything that is garnered from experimental studies must be written into rules and then incorporated into software – for it is computers that will finally use what is known about protein folding to rationally design new molecules.

In theory, says Langridge, we should be able to collect all this information "and just let a computer chew on it." In practice, however, "this is a problem that could just take all the available computers in the world right now and you still wouldn't make a dent in it. The number of variables, the number of possible conformations a protein can adopt – it's a huge problem. The problem can expand to fill any available computer."

Even supercomputers are not nearly powerful enough, Langridge says, noting that he and his colleagues used up 1,500 hours of supercomputer time in 1985 just to calculate some very small changes in chemical bond energies.

"What you need," he says, "is massively parallel machines," computers capable of processing huge numbers of operations simultaneously rather than serially. Such computers are still in the early stages of development (SN: 8/1/87, p.76). However, he says, "There's no reason why in the long run if we understand this fully we'd not be able to say, 'I want this particular function,' and then just go ahead and write down the appropriate amino acid sequence.

"We're still a long way from that," he adds. "We're really having great difficulty in doing some of the things that you might think were fairly simple." □