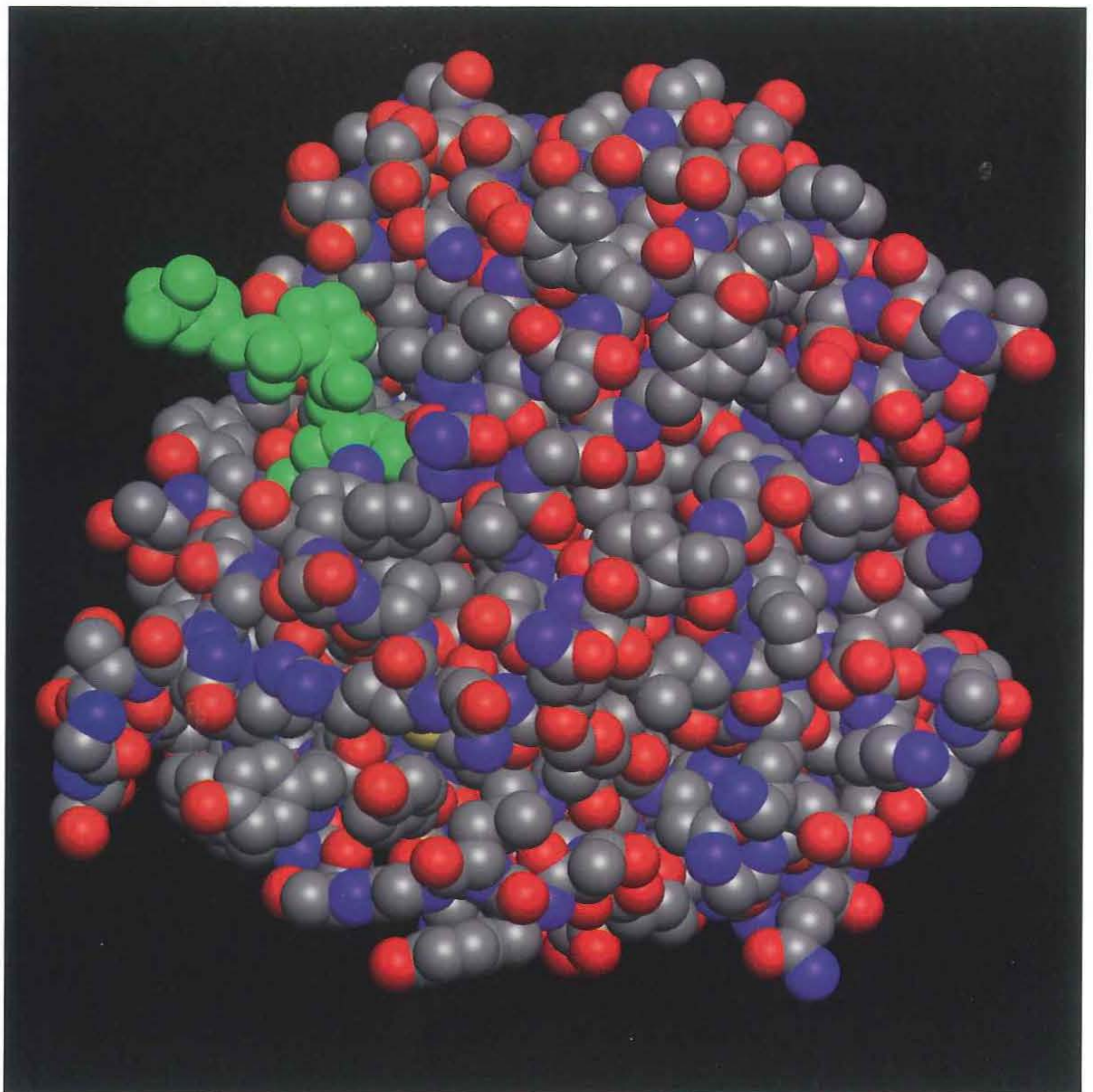
A protein automatically twists its backbone around until it curls itself up into its proper shape—it's as if you could thread all the parts of a turbocharged big-block Chevy V-8 one by one onto a piece of twine, throw the twine into a tub of water, and pull out the fully assembled engine.



No Assembly Required

by Douglas L. Smith

This big bag of marbles is really subtilisin, a bacterial enzyme that chews up proteins and is widely used as a stain-removing agent in laundry detergents. The protein to be cut up, a part of which is shown here in green, fits into a pocket on the enzyme's surface. Only about five percent of subtilisin's 275 amino acids go into forming the pocket; the rest are there primarily to hold that five percent in place. Carbon atoms are shown in gray, oxygens in red, nitrogens in blue; for clarity, hydrogens aren't shown at all.

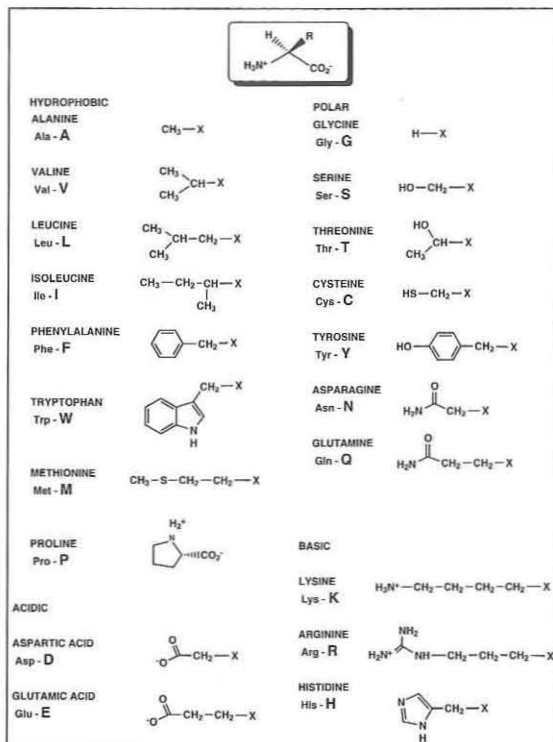
Proteins are the machinery of life, and they work over an astonishing range of conditions—from subzero Antarctic waters, where the fish have protein antifreeze in their blood, to geothermal vents where steam-scalded bacteria live in ambient temperatures above 150° C. Increasingly, proteins are the machines of industry as well. And not just designer drugs (or even generic drugs, such as insulin), although that's the high-profile end of the business, but such humble products as the stain-removing enzyme in your laundry detergent. (An enzyme is a protein designed by nature to make a specific chemical reaction occur rapidly and selectively.) In fact, the worldwide market for nonbiological enzymes was $1.3 billion in 1996, according to World Wide Web pages maintained by the Danish firm Novo Nordisk.

A protein is the architect's maxim of "form follows function" taken to its logical conclusion—the only thing that makes a protein work is the shape into which it folds. An enzyme has a pocket tailored to fit the reacting molecules, and grappling hooks in the right locations to entice the molecules into the pocket and hold them fast while the reaction happens. Chemical, and especially pharmaceutical, companies are getting very good at finding an enzyme that sort of does what they want, and they're starting to learn how to tinker with the structure of the enzyme until it does the right thing. But they'd really like to be able to say, "We want these two molecules to react in this manner, and we need a protein to hold them in this configuration to encourage them to do so. Computer, design me that protein!" In fact, this is one of the central challenges of modern biology. Stephen Mayo, PhD '87, assistant professor of biology at Caltech and assistant investigator with the Howard Hughes Medical Institute, has solved a simpler problem en route to that goal: designing a protein from scratch that will fold up into a predetermined shape.

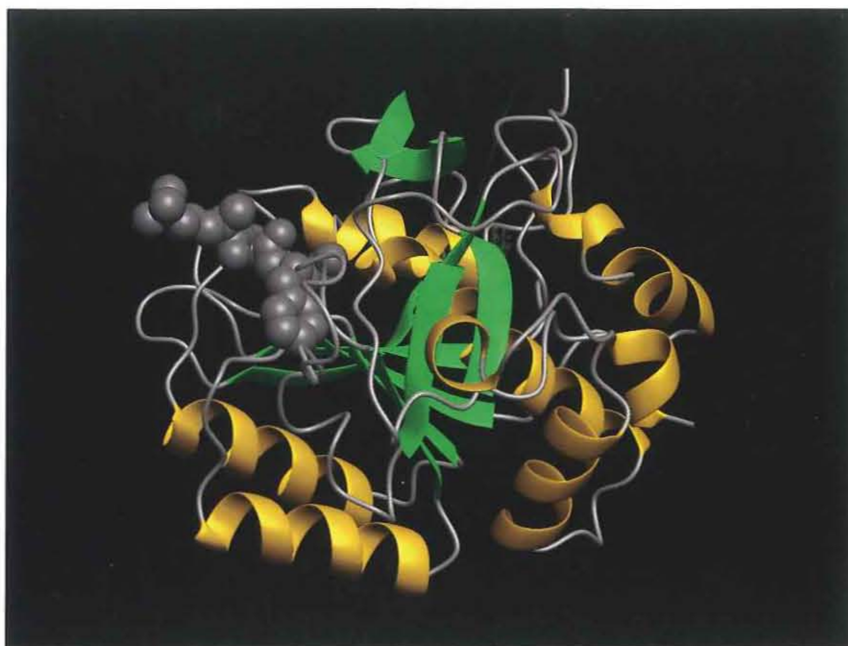
Just as machines have a basic vocabulary of parts—ball bearings, springs, cotter pins, axles, washers, and what have you—proteins are made up of components called amino acids, or residues, of which 20 varieties are commonly found in nature. These amino acids are strung together in a linear sequence, like a train of railroad cars, in what is called the protein's primary structure. And just as parts become subassemblies—carburetors, for example, or distributors—sequences of amino acids can naturally assume certain shapes. These shapes, which include helices, hairpin turns, and wavy sheets, are the protein's secondary structure. (The late Linus Pauling, PhD '25, deduced these structures from crystallographic and bond-angle data at Caltech in the 1950s.) These secondary structures, in turn, come together in specific orientations, called the tertiary structure, to form the biologically active protein. But context matters, too—the tertiary structure surrounding a string of amino acids can influence the secondary structure it chooses to assume. A tertiary structure that occurs over and over again in different proteins is called a motif.

Proteins and machines differ in one crucial respect. Machines have to be built and, as any backyard mechanic knows, seemingly identical parts often aren't interchangeable, and some parts only fit one way. Steps have to be followed in order, or you'll soon find yourself removing things in order to install other things that should have gone on first. And woe betide you if you find left-over parts at the end! But a protein automatically twists its backbone around until it curls itself up into its proper shape—it's as if you could thread all the parts of a turbocharged big-block Chevy V-8 one by one onto a piece of twine, throw the twine into a tub of water, and pull out the fully assembled engine. This shape is entirely determined by the protein's primary structure: a given string of amino acids will scrunch up exactly the same way every time—it doesn't matter if the protein is being made in a cell in your spleen, or a vat

Right: The 20 natural amino acids. The backbone unit is shown in the small box; the "R" stands for any of the side chains beneath. (In the side chains, the backbone is abbreviated as "X.") The amino acids designated "acidic" or "basic" are also polar. Each amino acid has a three-letter and a one-letter code, used for simplicity's sake when writing out sequences. Below: Subtilisin's primary structure (top); the colors correlate to its secondary structure (bottom)—helices are yellow and sheets are green. The protein fragment being chewed is shown in gray.



AQTVPYGIPLIKADKVQAQGFKGANVKVAVLDTGIQASHPDLNVVGGAS
FVAGEAYNTDGNGHGTHVAGTVAALDNTTGVLGVAPSVSLYAVKVLNSS
GSGSYSGIVSGIEWATTNGMDVINMSLGGASGSTAMKQAVDNAYARGV
VVVAAAGNSGNSGSTNTIGYPAKYDSVIAVGAVDSNSNRASFSSVGAELE
VMAPGAGVYSTYPTNTYATLNGTSMASPHVAGAAALILSKHPNLS
ASQVRNRLSSTATYLGSSFYYGKGLINVEAAAQ

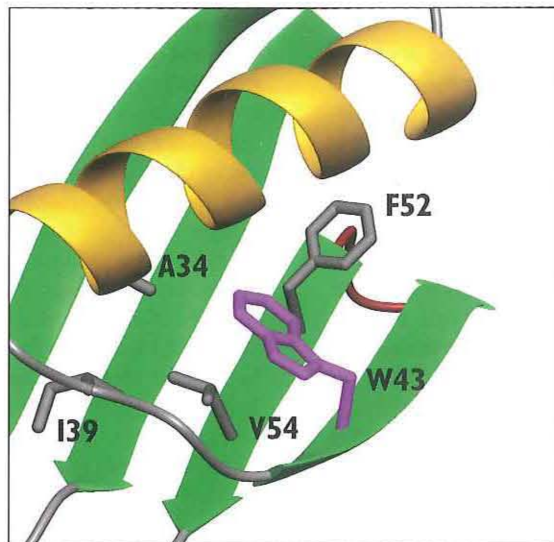


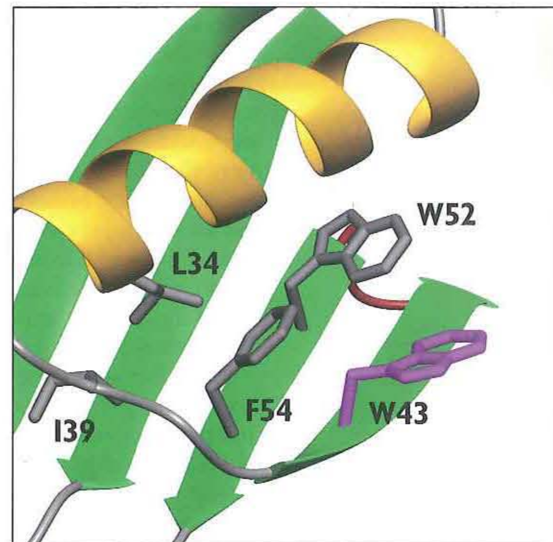
in a factory. (Assuming the protein folds at all, that is—for reasons not well understood, even a nice, stable natural protein, when synthesized outside its normal cellular environment, may lie limp or form a hopeless tangle that refuses to cooperate.)

All amino acids have the same backbone unit, enabling them to be coupled together in any order, but each amino acid has a different side chain dangling off that backbone. These side chains determine the protein's shape and all its other properties. Hydrocarbons, for example, are oily, so side chains made exclusively from carbon and hydrogen atoms mix with water like, ummm… well, oil and water. These hydrophobic side chains flee toward the protein's core, hiding as far from the surrounding water molecules as they can get. But side chains containing nitrogen or oxygen atoms are "polar"—electrically neutral overall, but with just a whiff of negative charge on the nitrogen or oxygen, and a corresponding soupçon of positive charge on an adjoining hydrogen. Water molecules are also polar, so polar side chains like to be on the protein's surface. This compulsion to embrace or avoid water is the hand that wads the protein up. But as the protein curls, the side chains have to accommodate one another. Some side chains are big and bulky and push their brethren aside to make room for themselves; others are quite compact. Some are long and floppy, like overcooked spaghetti; others are flat and stiff, like playing cards. And some form bonds of various kinds with one another. The sum of these manifold attractions and repulsions give the protein its ultimate shape.

So how do you pick the primary structure that will fold itself into the shape you want? Most people have taken one of two approaches. One way is to string amino acids together by eye, as it were—using the biochemical intuition gained through years of working with a particular motif. This has had its successes, but the knowledge gleaned about the subtleties of one motif rarely applies to another. The other method involves synthesizing as many random sequences as you have time and money to make and basically throwing them at the problem—using some sort of screening method (such as reaction-rate enhancement or binding affinity) to see if any of your sequences have the desired result. This approach rapidly gets out of hand—with 20 different amino acids to choose from at each position in the sequence, a string of 10 amino acids, which is about the upper limit for this method, gives you 20^{10} (about 10 trillion) possible primary structures. At this rate, you tend to run out of patience and raw material pretty fast. Things get worse exponentially as the string's length increases—one more position creates 20^{11}, or 205 trillion, possibilities.

Wading through these endless possibilities is clearly a job for computers rather than humans.





This is a close-up of the β1 domain of streptococcal protein G, a protein that resides on the surface of the streptococcus bacterium and is part of the molecular camouflage that allows it to sneak past the immune system. The tryptophan residue in purple (labeled W43 because it's the 43rd residue from the protein's N-terminus) is a classic transition position. In the structure at left, more than 90 percent of the surface area of the tryptophan's hydrophobic side chain is buried in the core. In the structure at right, the side chain is only 46 percent buried, leaving more than half of its surface exposed to water. The bulkier amino acids at positions 34, 52, and 54 have forced it to rotate outward.

After all, amino acids are simple molecules whose structures and properties have been studied in exhaustive detail. In recent years, people have begun to write programs—based on educated guesses as to which chemical properties to include—to generate primary structures that, when made in the lab, often fold up in manners that approximate natural motifs. But most of these programs deal exclusively with helices, sharply limiting the range of motifs that can be made. Furthermore, biological motifs are rigid, like bent coat hangers, while the man-made ones are limp, like tangled yarn. From the molecular machine standpoint, this doesn't cut it. But how to find out what interplay of properties, and in what proportions, would stiffen the tertiary structure into the one true shape?

Mayo realized that he needed a feedback loop in order to close in on the right mix of properties. To do this, he needed a target to shoot for. He decided to take the backbone of a protein whose three-dimensional structure was very precisely known, and attempt to generate a sequence from scratch whose backbone would fold up into an exact match. If the locations of all the backbone atoms matched those of the target, he figured, then the side chains would take care of themselves. (Of course, this left open the question of whether the computer would independently arrive at the original primary structure—just because every natural protein has a unique shape doesn't mean that other sequences might not also assume that same shape.) So the group would choose a set of properties, let the computer generate its best sequence, make that sequence in the lab, determine the three-dimensional structure of its backbone, and see how closely it matched the target backbone. Based on these results, they'd twiddle with the parameters, perhaps pick new ones, and try again. "This design cycle is the key," says Mayo. "Synthesizing trial sequences is a vital reality check, because the data you get from simulations are always refracted through the prism of your expectations. And you have to try sequences from all over the map, because even if you think you know where the answer is going to lie, you may have overlooked something. But this way, the experiments themselves tell us what is important."

Mayo also realized that one set of properties might not be enough—for example, the forces that make the protein's interior a safe refuge for the oily, water-hating amino acids may not hold sway at the surface. So the group broke the problem down into three parts: the core, the surface, and a transitional zone in between, where both the core and the surface properties struggle for supremacy. Finally, both to simplify the lives of the grad students who would actually have to make the stuff, and because the number of possibilities the computer has to look at gets so big so rapidly, they picked their targets from among the shortest naturally occurring primary structures that assumed reproducible shapes.

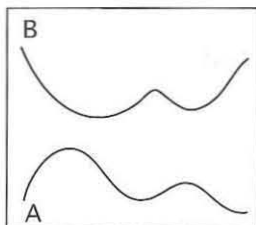
In fact, the computational swamp is deeper than you think. Side chains, as the name implies, are generally floppy, with each link in the chain free to rotate around the chemical bonds that hold them together. Even the flat, stiff side chains can rotate, like solar panels tracking the sun. So packing side chains together is not unlike working a jigsaw puzzle whose pieces are changing shape right before your eyes. (It's quite amazing, really, that such pliable stuff can hold any solid form at all.) Each amino acid has a smooth continuum of rotational shapes available to it, making the swamp, in effect, bottomless. A hint of firmer footing appeared in 1987, when J. W. Ponder and F. M. Richards of Yale University prepared tables of discrete poses, called rotamers, that each amino acid prefers to assume. Depending on the side chain's length and flexibility, the number of rotamers varies from roughly a dozen to perhaps 70, but a typical amino acid has 20 to 30 of them.

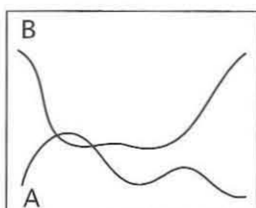
Thus a protein that's 20 residues long, with any one of the 20 naturally occurring amino acids in each position, and with each amino acid having 25 rotamers apiece, has 9.5×10^{53} possible rotamer sequences—a number that, written out, would be more than half again as wide as this column. If you examined a billion sequences per second (a feat far beyond the capacity of even Caltech's best supercomputers), it would take 10^{19}, or 10 quintillion, years to look at them all. This is a stroke of bad luck, as the universe is only about 15 billion years old—job security for the professor, perhaps, but an unwise choice for a grad student. And things get worse—nature's smallest fully functional motifs are some 30 to 40 residues long. The number of possible rotamer sequences for a typical small protein—one that's 100 residues long, say, has 7.9×10^{269} possibilities—is so staggeringly huge that your average supercomputer would gladly gnaw through its own Internet connection in order to escape having anything that big stuffed into it. According to Professor of Astronomy George Djorgovski, the best estimate of the number of protons in the entire universe is a mere 10^{80} or so.

There is a way out of this impasse. Think of the set of all possible rotamer sequences for a given primary structure as an overgrown tree of sufficient impenetrability to guard Sleeping Beauty. One string of rotamers, picked at random, is the tree's trunk. Moving any one rotamer one click to its next pose is a branch off the trunk; moving another rotamer one click as well is a branch off of that branch, and so on. In 1992, a group of Belgians (who were working on the simpler, but related, problem of trying to predict the specific rotamers that a given sequence of amino acids strung on a fixed backbone would assume) developed a procedure, called Dead-End Elimination, that prunes the rotamer tree back severely. The program lets two rotamers compete head-to-head for a single spot in the sequence, as shown in the illustrations below. If there's a clear loser, all branches in which it appears get lopped off. Extending the program to compare rotamers of different amino acids took considerable work, but the resulting computation is still relatively tame. It also takes a certain amount of experience to design the best pruning strategy—you want to get as close to the trunk as possible without wasting lots of time clipping the branches' tips.

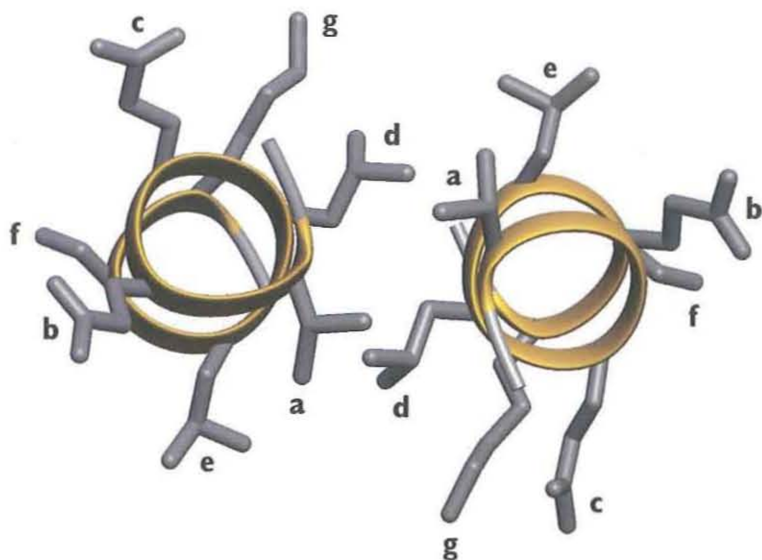
So with all the tools in hand, it was time to roll up the sleeves and go to work. The group's first target backbone was the "coiled coil" motif, which consists of two identical helices wrapped around each other like strands in a rope. The fact that the helices are identical cuts the synthetic work in half—a big plus when you want to test lots of variants quickly. The coiled coil is a critical piece of a larger protein that controls DNA transcription, says grad student Bassil Dahiyat—the cell makes each helix separately, and as the two helices coil up around each other they help zip up the protein. Coiled coils are found in all cells that have nuclei, including ours—this particular motif, which rejoices in the euphonious name of GCN4-p1, comes from yeast. (The high-precision structure of GCN4-p1, which the group used as its starting point, was determined by X-ray crystallography in 1991 by T. Alber et al. at UC Berkeley.)

The group began with the core, whose organization largely determines how the rest of the protein arranges itself, and where the most research had already been done. The coiled-coil helix repeats itself every seven amino acids, as shown on the opposite page. If you label these amino acids **a**, **b**, **c**, **d**, **e**, **f**, and **g**, then residues **a** and **d** will always be buried in the seam where the two strands adhere to each other—in the core, in other words. Because the group was experimenting with the core, the other residues were left identical to the natural protein. All the atoms in these invariant amino acids were nailed down in their natural locations. The group studied strands 33 residues long (four repeating units plus a little extra for stability), giving eight variable amino acids per string, or 16 in all—even though the two strings' primary structures are identical, the rotamers may

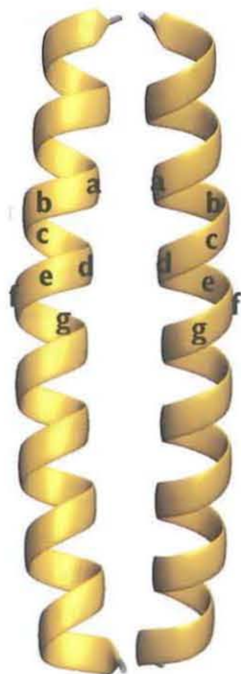




Dead-End Elimination works by computing the attractive and repulsive forces between a given rotamer (here called Rotamer A) and every other atom in the protein—what's called its potential energy. The calculation is then repeated for a different rotamer (Rotamer B) in the same position in the sequence. The lower the potential energy, the more stable the structure. If one rotamer always has a higher potential energy regardless of the rest of the protein's behavior, as does Rotamer B in the upper example, then that rotamer and all its branches are pruned from the tree. But if the curves cross, then either rotamer might be favored and both must be retained.



A cross section through repeat unit of a coiled coil (above). Pairs of a's and d's alternate to form the core, as shown in the 33-residue segment (below).



still be different. The group then limited the computer's choices to the eight (reduced to seven in later work) oily, hydrophobic amino acids that one would expect to find in the core, and let the machine go to town. This much-simplified problem of 238 rotamers filling 16 positions in the sequence still gave 238^{16}, or 10^{38}, possibilities. Ouch! Fortunately, the program is very parallelizable, meaning it can be farmed out to many computers at once. A machine containing eight parallel processors took only three minutes per run to find the best sequence.

Dahiyat and Mayo eventually discovered that a combination of three parameters gives the best results. The first one, the van der Waals potential, measures how hard the rotamers' atoms are being shoved together. The side chains in the core are packed shoulder-to-shoulder, like subway commuters in a Tokyo rush hour, but you can only squeeze them so much without getting the atomic equivalent of an umbrella in the eye. The second parameter measures the amount of hydrophobic surface area that's safely buried and protected from the surrounding water molecules by other residues, and awards a stability bonus as the buried area increases. The third parameter measures the amount of polar, water-loving surface area that is similarly buried, and exacts a stability penalty as the buried area increases. This combination has correctly predicted the relative stability of all the sequences the group made and tested—an encouraging sign that they did, in fact, find the critical forces.

The stabilities were tested by comparing the temperature at which the proteins unfolded in solution—the "melting temperature." A protein that assumes a single, stable shape in solution will have a higher melting temperature than a sloppy, loosely folded protein. The floppier it is, the less energy it takes to finish unfolding it, and the lower the melting temperature.

To nobody's great surprise, the natural sequence proved to be the most stable sequence tested. However, the researchers did discover a family of only slightly less stable sequences containing different amino acids, indicating that there's room for variation, even in a structure this small. On the other hand, some of the least stable sequences were identical to the natural sequence except for a single wrong amino acid at a critical location.

Emboldened by this success, the group moved on to the surface. There are three exposed surface positions per seven-unit coil: **b**, **c**, and **f** in the lettering scheme. (Positions **e** and **g**, which snuggle up against **a** and **d** on the opposite strand, are only partially exposed and make up the transition zone between core and surface.) The group used the same general approach, but with a list of 10 polar, water-loving amino acids. This time, the dominant forces were the propensity of certain amino acids to form helices—a parameter that had been quantified in 1994 by R. L. Baldwin at Stanford, and others—counterbalanced against the number of hydrogen bonds that the overall structure could form. A hydrogen bond is a weak bond formed between a polar atom, such as oxygen, and a nearby hydrogen atom, usually from a different residue. These bonds help brace the structure.

Because the side chains in the surface positions have lots more room to flop around, it had been assumed that they didn't have much influence on the folding process. In fact, Dahiyat, fellow grad student Benjamin Gordon, and Mayo found that altering the surface amino acids caused dramatic changes in stability, as reflected in the melting temperatures. A sequence in which the surface positions were randomly filled from the list refused to curl up, and had by far the lowest melting temperature of any sequence tested. And the best computer-designed sequences were significantly more stable than the original coiled coil, having melting temperatures some 10°–12° C higher. Perhaps nature wasn't particularly interested in optimizing the coiled coil's surface for stability; in any case, it means that there's hope of improving on nature's designs if extra stability is required in the face of harsher-than-natural conditions—for example, when the protein is immersed in some oily organic solvent that's inimical to protein folding (because it won't drive oily residues to the core), but which is necessary to dissolve the chemicals you want the protein to act on.

At about the same time, the group had a go at the transition-zone residues, which they christened the boundary positions. These residues can go either way, and are likely to be swayed by subtle influences. To eliminate the possibility that the coiled coil's repeating, two-stranded structure might somehow skew the outcome, the group turned to the streptococcal G protein—a single strand that doesn't repeat. The key determinant proved to be the amount of hydrophobic surface

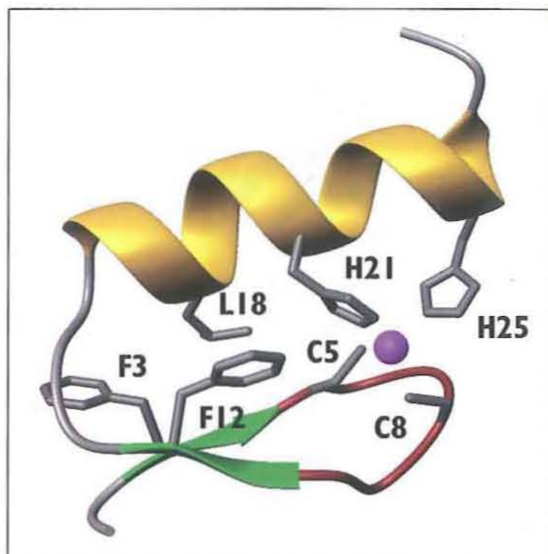
Below: The primary structure of Zif268 (left) and FSD-1 (right). The asterisks mark the zinc-binding amino acids. The orange background shows the core position, red marks the boundary positions, and blue is the surface positions. The six conserved amino acids are the white letters.

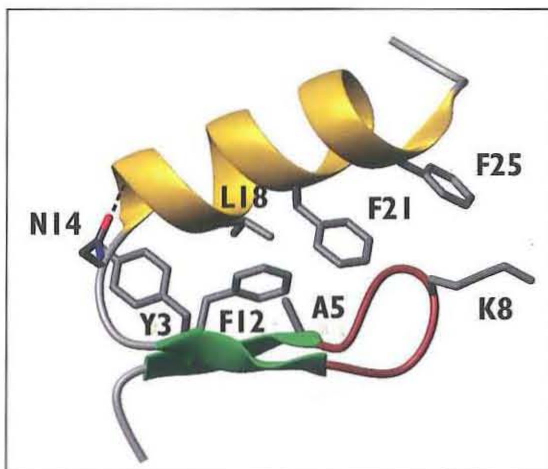
Zif268	FSD-1
K	Q
P	Q
F	Y
Q	T
C*	A
R	K
I	I
C*	K
M	G
R	R
N	T
F	F
S	R
R	N
S	E
D	K
H	E
L	L
T	R
T	D
H*	F
I	I
R	E
T	K
H*	F
T	K
G	G
E	R

area that remained exposed to the water molecules. This exposure penalty is subtly different from the burial bonus that drove the core packing. Let's say that a small and a large hydrophobic residue are competing for the same boundary position, and that both residues have 100 square angstroms of surface area buried. This would give them identical burial bonuses. But the small residue might be almost entirely buried, while the larger one could still be sticking its head and shoulders into the water. So the exposure penalty deters large hydrophobic residues from occupying boundary positions.

Now it was time to try the core, boundary, and surface programs together, which, says Dahiyat, was "going to be a very stringent test. Basically, if you're a little bit off on any piece, it's highly unlikely that they're going to compensate and help each other." The group chose a structure called a "zinc finger," a common feature in proteins that bind to DNA and control the copying of genetic information. The particular one they used—Zif268—is a human variant, from which they selected a 28-residue motif containing a sheet, a helix, and a turn, making it a thorough test of their system. A zinc finger, as the name implies, incorporates a zinc ion to stabilize itself. (In fact, "zinc glove" might be a better name, because the zinc ion sits inside the motif like a finger in a glove.) Professor of Chemistry Barbara Imperiali's lab had demonstrated that some variants can fold up without zinc, but the catch was that these variants included two amino acids not found in nature that were custom-built to stabilize the fold.

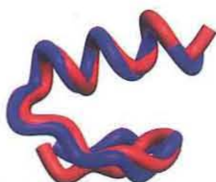
Again, the backbone was held fixed in the natural shape, but this time the identity of every amino acid in every position was left in the computer's hands. The computer decided which positions belonged to the core (only one, because of the motif's small size), the boundary (seven), and the surface (20), based on an analysis of the natural protein. The computer chose the core and surface residues from the previously developed lists of allowable amino acids; for the boundary positions, the computer was allowed to pick any amino acid from either list. All in all, the computer had 1.9×10^{27} possible primary structures to choose from. This is a quintillionfold beyond the reach of physical screening methods—had the group actually synthesized one molecule of each primary structure, the aggregate would have weighed 11.6 metric tons. Factor in the rotamer problem, and the number of possible sequences skyrocketed to 1.1×10^{62}. (This is equivalent to the number of protons in 100,000 suns, says astronomer Djorgovski—comfortably less than the number of protons in the universe, but still a hefty number.) Even so, it only took 90 hours of processing time for a 10-processor system. A nine-hour day for the computer, in other words—pretty good hours for a grad student.





The natural zinc finger (upper) and FSD-1 (lower). The zinc ion is shown in purple. With no zinc ion to cling to, FSD-1 is stabilized by the burial of hydrophobic residues and by hydrogen bonds, especially the one shown as a dashed line between N14's oxygen atom (red) and the helix's backbone. Note how boundary residues 12, 18, 21, and 25 all cluster around the core (residue 5)—even boundary residue 3 is on the interior face of its sheet. (Boundary residues 7 and 22 aren't shown for clarity, but also point toward the core.) The brown stretch of backbone is a turn.

The backbones of Zif268 (red) and FSD-1 (blue), superimposed on each other. Each tube's diameter is 0.8 angstroms, or roughly one-quarter the van der Waals diameter of a carbon atom.



The computer's winning sequence, when synthesized and its three-dimensional structure analyzed, matched the backbone atoms of the target sequence to within, in general, about half a carbon atom's diameter. (The ends were considerably floppier, because there wasn't much to hold them in place.) The computer chose nonpolar residues for all seven boundary positions, placing them in a nice, solid little packet around the one core residue. The computer also came up with well over 1,000 other primary structures that should be only slightly less stable, conserving a few critical amino acids while being tolerant of variation elsewhere. As Mayo remarked in the *Science* paper in which their results appeared, "Even if billions of sequences would successfully achieve the target fold, they would represent only a very small proportion of the 10^{27} possible [primary structures]."

Although the computer-designed backbone was right, its primary structure was completely different. The amino acids that bind the zinc ion, which are essentially invariant in all natural zinc fingers, were completely different in the computer's version, although their side chains remained pointed in the general direction of where a zinc ion would have been, had there been one. Only six of the 28 positions (21 percent, or no better than random chance) contained the same amino acid as did the original protein, and only 11 positions (39 percent) were even close. Furthermore, a search of the sequence database maintained by the National Institutes of Health's National Library of Medicine revealed that the computer's sequence (christened FSD-1, for Full Sequence Design #1) didn't look like any known zinc-finger sequence. In fact, it didn't look like any known protein, period, underlining the fact that the design program relies on chemical first principles and not some hidden biological biases.

"What's important here isn't what FSD-1's primary structure resembles, but that its tertiary structure is correct *and* stable," says Mayo. "A biological zinc finger is so tightly bound to its zinc that it has no melting temperature, but the zinc-finger sequence is so short that most people believed that it would be hopelessly unstable without the zinc to hold it together. To my knowledge, this is the shortest sequence that consists entirely of natural amino acids and assumes a stable fold with no help from metal binding, disulfide bond formation, or other assistance." As was exemplified by the subtilisin protein at the beginning of this article, the bulk of a natural protein is scaffolding: a Dr. Seussian array of props that brace other supports that hold in place the motifs that do the work. But industry can't afford to be as profligate as nature, so getting motifs to hold their shape with the absolute minimum of scaffolding is critical if a protein is to be manufactured from scratch economically.

A 30-amino-acid string is at the bottom end of the realm of functioning motifs; however, current supercomputers can easily handle sequences of 50–60 amino acids, which really gets up into the realm where useful things can be made. And if you're willing to sacrifice academic rigor and run quick-and-dirty approximations, says Dahiyat, you can manage about 100 amino acids.

Furthermore, work by grad student Alyce Su has shown that you can have a fair amount of flexibility in the target backbone, yet still come up with a sequence that will assume the correct fold. This might be the first step to having a computer wrap a backbone of its own design around the shape you want to encase—if the computer could be given some leeway with the backbone, it might simplify the task of incorporating the amino acids that actually grab the reacting molecules and hold them in place. These amino acids need to be in certain spots on the backbone in order to do their jobs, but putting them in those spots could distort the backbone. Figuring out how to put the right grappling hooks in the right spots without messing up the backbone's folding is now high on the group's agenda.

"To my knowledge, this is the shortest sequence that consists entirely of natural amino acids and assumes a stable fold with no help from metal binding, disulfide bond formation, or other assistance."

The group is now trying the system on other motifs that were not used in the development work. If the system successfully re-creates these unfamiliar backbones, it will be a good proof of its generality and a big step toward designing proteins from scratch. Mayo waxes lyrical about the potential for drug design, while Dahiyat likens the possibilities to the burgeoning of consumer plastics. "About 40 years ago, new catalysts were developed for polymer production, and now plastics are everywhere. Today, people are spending hundreds of millions of dollars a year trying to make enzymes more thermally and chemically stable. We can do that with the push of a button in a lot of cases. And by stabilizing the enzymes, you open up the applications you can use them in, and hopefully make the whole thing take off. People will be using proteins everywhere." □

Mayo's work is supported by the Rita Allen Foundation, the Chandler Family Trust, the Booth Ferris Foundation, the Howard Hughes Medical Institute, the David and Lucile Packard Foundation, the Searle Scholars Program, the Chicago Community Trust, and the National Institutes of Health.