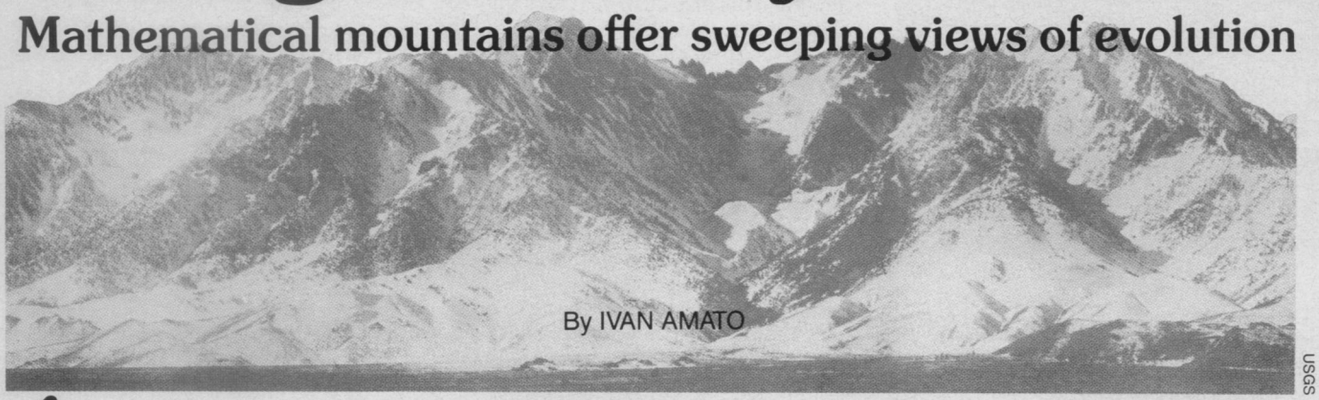
Taking Proteins for a Walk

Mathematical mountains offer sweeping views of evolution



USGS

By IVAN AMATO

A walk in the Alps would hardly seem relevant to protein evolution. But by thinking like mountain trekkers, researchers are learning to assess the merits of various evolutionary pathways, in the ultimate hope of directing protein evolution along selected routes. At the final pinnacle may lie a new set of molecular tools for specific jobs such as shuttling a drug to troubled cells, latching on to bacteria or cleaving chemical bonds in a toxin.

Different terrains demand different pathways. To scale a smooth slope such as Fujiyama, for instance, you'd probably huff straight to the top. In the more rugged Alps, you might have to take an indirect approach, descending from an intermediate peak before finding a tenable route up to your target peak. If the territory is unfamiliar, you might ascend the target peak only to discover that a deep valley separates you from a still higher but prohibitively distant pinnacle. The evolutionary path of biological traits leads through similar landscapes in a mathematical Alps.

This idea dates to the 1930s, when evolutionary theorist Sewall Wright suggested evolution could be viewed as occurring on what he called an "adaptive landscape." Specific collections of genes, or genotypes, endow organisms with traits that help determine their adaptive fitness — how well they can withstand the selection pressures in their environments. Wright reasoned that mapping out the range of possible genotypes (where neighboring genotypes differ by a single mutation) and their associated evolutionary fitnesses produces a landscape that might depict the topography on which the organism's evolution travels.

Starting at a low point on the landscape and "walking" to a higher one, for instance, is analogous to the evolution of traits that improve an organism's ability to survive the selection pressures of its particular environment. Such a trek might depict what happened when, say, the ancestors of today's giraffes evolved progressively longer necks in response to competition from other species for ground-level food.

In 1970, British evolution scientist John Maynard Smith extended the idea of adaptive landscapes to what he called "protein space" — a multidimensional mathematical construct whose points correspond to each of the protein sequences of designated lengths that can be built using the 20 amino acids that serve as the proteins' molecular building blocks. Adjacent points in the space correspond to proteins whose sequences differ by a single amino acid. Assigning each point a fitness value with respect to a relevant protein function, such as the ability to bind to a substrate, produces a space that reflects the relationship between a protein's structure and its function. In this framework, an "adaptive walk" would consist of a series of steps leading from points with lower fitness values to ones with higher values as the protein evolved into more capable versions.

Other researchers have further developed the notions of adaptive landscapes and protein space, and some are even beginning to apply them to a special, accelerated example of protein evolution — namely, the fine-tuning of antibodies during an immune response. These scientists envision using the concept of fitness landscapes to help launch a technology they call "applied molecular evolution." In doing so, they hope to expand evolution from a merely theoretical framework for explaining long-term biological trends into a practical strategy for systematically inventing proteins with predetermined functions.

"The idea is, honest to God, to learn how to evolve proteins to do whatever you want," says biochemist Stuart A. Kauffman of the University of Pennsylvania School of Medicine in Philadelphia. Without a visual display such as a rugged fitness landscape, it's difficult to intuit why some mutations are more adaptive than others, adds applied mathematician Alan S. Perelson of Los Alamos (N.M.) National Laboratory.

"It gives us another view of experimental data," says Perelson, who collaborates with Kauffman. With such a view, protein engineers presumably could examine a protein's fitness landscape (presently a *terra incognita*), plot an evolutionary course that would produce a better version of the protein, and then use known techniques to carry out the necessary mutations in the protein's amino acid sequence.

Right now, protein inventors can do little more than shoot in the dark, triggering specific mutations in a protein and watching to see what happens. A more systematic way of creating proteins for specific jobs, Kauffman suggests, would be to make a huge repertoire of partially random proteins, isolate those that do at least some of the desired job, and then use the guidance of a fitness landscape to evolve the candidates into practical molecular tools.

"Applied molecular evolution is not balderdash," Kauffman stresses. He holds patents in Europe and has others pending there and in the United States that involve molecular repertoires of DNA sequences, which he hopes to translate into a large pool of useful biomolecules such as enzymes.

Provocative as the idea of fitness landscapes is, no one knows what these topologies look like — how rugged they are, how many peaks they have. Several years ago, Perelson met with Kauffman — who with others had already begun developing a general theory of adaptive walks on rugged landscapes — and suggested that the maturation of the immune response might be a phenomenon perfectly suited to honing the theory and getting clues to a fitness landscape's specific ups and downs. Soon

after this conversation, Kauffman began working out some of the general outlines for a mathematical model of protein evolution as an adaptive walk on a rugged landscape. In 1987, he and ecologist Simon A. Levin of Cornell University published the first draft of the model in the JOURNAL OF THEORETICAL BIOLOGY.

Since then, Kauffmann, Perelson, Edward D. Weinberger of the University of Pennsylvania and several co-workers have published refinements and extensions of the theory. In the August 1989 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES (Vol.86, No.16), Perelson and mathematician Catherine A. Macken of the University of Auckland, New Zealand, reported the most sophisticated model of a fitness landscape to date. As before, the researchers described their general sequence space as a multidimensional zone in which each point represents a specific protein structure involving a fixed number of amino acids. In addition, Perelson and Macken randomly assigned to each sequence a fitness value ranging from a low of zero to a maximum of 1. Fitness refers to any of a variety of a protein's functions, such as its ability to bind to another molecule or the rate at which it catalyzes a reaction.

Stepping from one point in the space to an adjacent one represents changing one amino acid of the model protein. The relationship between the fitness values of the two sequences – i.e., whether it's an upward, downward or level step – indicates whether the step is adaptive, maladaptive or neither. Since adjacent sequences in the space are assigned fitnesses at random in these preliminary models, the resulting landscapes have many multidimensional hills and valleys.

In their model, Perelson and Macken assume protein evolution occurs as a direct uphill walk on an adaptive landscape – or, in a biological interpretation, as a series of adaptive point mutations in the DNA codes that translate into a series of progressively fitter proteins. To test how well their model reflects real molecular evolution, the researchers looked to the maturation of the immune response. In this case, the fitness parameter of the model corresponds to the readiness with which an antibody binds to a specific antigen.

In a real body, each time the immune system recognizes a molecular invader, or antigen, and launches an antibody-mediated defense, a highly accelerated and testable version of protein evolution occurs. Antibodies appear a few days after the body's initial exposure to the antigen, and they can last several weeks. During that period, the immune system produces a series of additional antibodies that bind to the antigen with increasing readiness, or affinity. Experiments show that a starting antibody undergoes anywhere from two to 20 mutations in the form of substitutions among its constituent amino acids, and that the antibodies' affinities for the antigen increase by a factor of 10 or so before the affinity peaks.

In agreement with experimental values, the new model predicts that somewhere between two and 15 mutations occur before the antibody response matures, or evolves, to produce antibodies with an affinity that corresponds to an intermediate peak on the affinity landscape. Since getting to a higher peak would require intermediate steps that are maladaptive, the modeled antibody evolution stops at the nearest intermediate peak even though other points in the space represent higher affinities. "It doesn't find the global, or best possible, maximum, but it finds the best it can do by making a small number of changes," Perelson says.

The model also predicts that as the response nears maturity and antibodies emerge with relatively high affinity, more mutation trials are required to produce a new generation of antibodies with still higher affinity – a trend Kauffman likens to a law of diminishing returns. Both Perelson and Kauffman note that this prediction corresponds with the experimentally observed slowdown of affinity improvements in real immune responses.

In a related model of fitness landscapes, Kauffman and Weinberger include the effect other amino acids have on the adaptive value of each particular site in an antibody's amino acid sequence. By assuming, in accordance with experimental results, that about eight mutations must occur before antibodies reach a local affinity peak, the researchers use the model to predict that the binding region of a model antibody is influenced on average by about 40 nearby amino acids. The number of influential amino acids determines the ruggedness of an "antibody space" with topological details that he says correspond surprisingly well with known features of immune-response maturation. For example, the model indicates that roughly 1 percent of an antibody's variants have higher affinities for the antigen. Experiments with actual antibodies confirm the 1 percent prediction.

The same model also uncovers possible behaviors of the immune response that biologists might want to investigate, Kauffman says. For instance, it predicts that a particular antibody can take a variety of divergent mutation paths and reach different local affinity peaks, or different antibodies can mutate convergently into the same structure corresponding to the same peak. "The amount of divergence or convergence that you get somehow is telling you about the landscape," Kauffman says. Lots of divergence, for instance, suggests an Alpine landscape, while total convergence points to a single, Fujiyama-like peak.

Although Perelson and Kauffman are excited about using fitness landscapes for studying protein evolution, the models they have constructed so far ignore important aspects of real molecular systems. For one, the models use randomly assigned fitnesses, whereas some locations in real proteins are more fitness-correlated than others. Weighting fitness values to reflect these differences would make the models more realistic.

Furthermore, says Kauffman, "there's no natural way of scaling the fitnesses of these models to actual chemical affinities," as scientists would do in correlating thermometer readings to physical events like freezing and boiling of water. This uncertainty makes the heights of the peaks rather arbitrary, though it's still possible to glean the evolutionary value of particular steps on the landscapes just by knowing whether the path leading from one point goes up or down.

Perhaps most troublesome is the inescapable fact that fitness landscapes will change, often drastically, depending on the protein property under consideration or perhaps even on the specific antigen eliciting an antibody response. Modeling such transformations would be like viewing different sections of the Alps.

Challenging as these problems are, the researchers are optimistic. Kauffman predicts that models he, Perelson and others are developing will become part of the biotechnological toolshop within the next 20 years or so. Perelson and Macken suggest further refinement and extension of the theory might help researchers understand mutations of the gp120 protein in patients harboring the AIDS virus or chart the evolutionary course hemoglobin has taken in various animal lineages.

"The idea of these rugged landscapes is a very appealing one," comments physicist Daniel L. Stein of the University of Arizona in Tucson. Stein studies complex materials known as spin glasses, which have discordant, disorderly magnetic interactions amongst their constituent atoms, and he sees a connection between mathematical models of spin glasses and the concept of adaptive landscapes. Both spin glasses and proteins contain strongly interacting components (atoms or amino acids) operating under conflicting constraints (aligned vs. nonaligned atomic spins, or adaptive vs. maladaptive mutations).

"Adaptive landscapes form an important bridge between systems in biology and systems in physics," Stein asserts. In particular, he says, they offer a means of communication between physicists and biologists – two groups that often see the world as though they were situated on widely separated Alpine peaks. □