Preface: Applied Molecular Evolution on Rugged Fitness Landscapes

ABOUT THE WORKSHOP

This book grows out of a conference held at the Santa Fe Institute March 27 to 31, 1989. The conference was organized to discuss a central emerging area of biological science: the study of adaptive processes that optimize on rugged fitness landscapes. Since Sewall Wright introduced the concept of fitness landscapes,[9] where each genotype can be thought of as a point in a discrete genotype space and has a fitness, it has become almost second nature for biologists to conceive of adaptive evolution as "hill climbing" towards fitness peaks. Much of the basic research in population genetics, whether concerned with haploid or diploid systems, has involved analysis of the behavior of adapting populations, due to mutation, recombination, and selection, as they flow over fitness landscapes.[4] Classical population genetics has also emphasized frequency and density-dependent selection, coevolution, the evolution of the genetic mechanisms of recombination and sex, and other issues. What then are the new strands which we sought to explore at the Santa Fe Institute meeting?

The dominant answers are three. First, new mathematical tools now permit us to study the *structures* of complex, rugged, multipeaked fitness landscapes. Second, new *experimental techniques* are available to analyze molecular fitness landscapes. Third, current recombinant cloning techniques now allow for the capacity to carry

Molecular Evolution on Rugged Landscapes, SFI Studies in the Sciences of Complexity, vol. IX, Eds. A. Perelson and S. Kauffman, Addison-Wesley, 1991

out *applied molecular evolution* to evolve biomolecules of medical and industrial use.

Fitness landscapes underlie both molecular and morphological evolution. Mathematical description of such landscapes can be expected to lead to new experimental studies that actually test and establish the structure of fitness landscapes. Sidestep for the moment the normal biological definition of "fitness." In this book we shall typically be concerned with quite concrete issues where the "fitness" of an enzyme is its capacity to catalyze a given reaction in standard conditions, or the "fitness" of an antibody is its affinity for a defined epitope. Presumably, the fitness or "affinity landscape" of the diverse set of antibodies generated in response to a specific antigen underlies the maturation of the immune response. Thus, much of the book focuses on the actual facts and processes of maturation of the immune response, and their relation to the new theory that is brewing.

Given concrete cases of molecular fitness landscapes, the new theoretical strands include the following questions:

1. What do fitness landscapes underlying molecular or morphological evolution "look like"? How rugged and multipeaked are such landscapes? Are there families of landscapes, where each family ranges from smooth to rugged? How many such families? What are the salient statistical properties of such landscapes? These must include such features as the number of fitness peaks, the lengths of adaptive walks via fitter mutant neighbors to such peaks, the number of mutations accumulated along such adaptive walks, the rate at which the number of directions "uphill" dwindle as successively fitter variants are found, whether or not the global optimum is attainable, how many local optima are accessible from any starting point, whether the local optima are similar to one another or scattered randomly across the space. Other issues include the distribution of fitnesses of optima, and techniques to measure experimentally the correlation structure of a fitness landscape.
2. A second class of questions concerns the *actual flow of an adapting population* over a fitness landscape of a given structure, as a function of the size of the population, the mutation rate, and, if sexual, the mating structure and effects of recombination.
3. A third class of questions about rugged landscapes concerns what kinds of objects, antibodies, enzymes, DNA regulatory sites, or more macroscopic systems such as entire genomic regulatory networks, have what kinds of fitness landscapes and why.

ABOUT THIS BOOK

The first section of this book lays out a number of the general issues concerning the structure of rugged fitness landscapes, with contributions from Palmer; Amitrano,

Peliti, and Saber; Stein; and Schuster. It is not an accident that many of these workers are solid-state physicists. The expected structures of rugged fitness landscapes is a close cousin to the structure of rugged potential surfaces which arise in spin glasses and other disordered materials. One purpose of this volume is to acquaint biological readers with concepts applicable to biological evolution which have close parallels in statistical physics. These include issues such as the behavior of an electron in a complex potential surface at a finite temperature, as a function of the barrier heights between "valleys." Such processes are related to slow relaxation times and "freezing in" in regions of a rugged fitness landscape as a function of population size and mutation rate.

Rugged, multipeaked landscapes also arise in the protein and nucleic acid folding problems. Here the landscape is a free energy surface and describing movement on the surface and the question of being trapped at local minima versus the global minimum arises. In fact, issues of structure underlie the association of a fitness to a molecular sequence. Thus embedded in the statistical structure of a molecular fitness landscape is information about the relationship between changes in the underlying sequence and the configuration of a molecule. The fourth chapter in this volume by Schuster illustrates this point in the context of RNA evolution.

If the structure of fitness landscapes underlies molecular and morphological evolution, then a central task is to measure and characterize such landscapes. Maturation of the immune response is one favorable system for such an analysis. As described in detail in many chapters in this volume, the immune response to an antigen is initiated by stimulation of mitosis in B cells bearing immunoglobulin receptors of at least modest affinity for the antigen. Activation of B cells is followed by the induction of a hypermutation mechanism that introduces point mutations in the binding region of the antibody molecule, the V region, such that some B-cell variants have reduced affinity for the antigen, others have increased affinity for the antigen. Presumably the latter are stimulated to divide more, leading to selection of B cells synthesizing antibodies of successively higher affinity for the antigen. Thus, maturation of the immune response can be viewed as an adaptive walk in antibody space from initial B cells harboring receptors with "roughed in" molecular matches to the incoming fitter mutant neighbors of the antibody molecules in "antibody space," to or towards high-affinity "peaks" in the fitness landscape.

The second section of the volume, with contributions by experimental immunologists Eisen; Berek and Apel; Manser; Williams, Kieber-Emmons, Weiner, and Greene; and Siskind, and theoretical biologists Macken and Perelson; Kauffman and Weinberger; Weisbuch and Perelson; and Weinand, examines both the history and the current status of experimental work on somatic mutation and the maturation of the immune response, and discusses current attempts to build detailed mathematical models of the affinity landscape. The chapters by Macken and Perelson, and Kauffman and Weinberger show how the ideas of rugged landscaped can be used to understand a number of characteristic features of maturation of the immune response. Weisbuch and Perelson examine simple models for affinity maturation in the context of idiotype networks, while Weinand takes a more global view and

presents an intricate computer model of clonal selection and B cell growth based on a three-dimensional representation of *shape-space*.[8]

Study of the structure of fitness landscapes underlying molecular evolution is not merely of scientific interest.[2,3,6,7] It is likely to become of very great practical interest. It is now possible utilizing cloning techniques to generate very large libraries of random, or partially stochastic DNA sequences, hence their corresponding RNA and protein or peptide products. A variety of selection and screening assays to seek products which might have a specific function are under development. Such products may range from new enzymes to new drugs and vaccines. For example, procedures to obtain peptides which mimic an antigen such as insulin might make use of polyclonal or monoclonal antibodies against insulin to screen for partially stochastic peptides bound by those antibodies. Any such peptide is a candidate to harbor an epitope which is similar to the insulin antigen, and hence may mimic it. Such peptides might potentiate, inhibit, or modulate the activity of insulin. If the incoming antigen is the protective antigen on a pathogen, then the mimetic peptide is a candidate vaccine.[2,3,6,7]

Initial work carrying out applied molecular evolution is underway in the laboratories of Greene, Joyce, Kauffman, Loeb, and Mandecki. Patents held by these workers covering many of the initial techniques are pending or have been issued. Ideas and result discussed in the third section by Robertson and Joyce; Mandecki; and Horwitz, Dube, and Loeb.

The final section discusses one of the current serious models of the origin of life, the *hypercycle* model of Eigen and Schuster. This model is of interest in the current setting for at least three reasons. First, Joyce's experiments evolving DNAase activity in a ribozyme is part of the evidence showing that RNA molecules harbor and evolve a variety of catalytic functions. In the hypercycle model, some kind of catalytic coupling of replicating RNA strand pairs around a cycle of such pairs is required. Second, the hypercycle model is an explicit example of *coevolution* in a system of coupled replicating polymers. It therefore bears a relation to idiotype network models where B cells coevolve with one another as a function of affinities of each antibody for its anti-idiotype. Finally, the Eigen and Schuster model is one of several alternatives being explored by members of the Santa Fe Institute community; the others are by Rothkar, Stein and Anderson, Rasmussen and coworkers, Kauffman[7], Farmer et al.[5], and Bagley et al.[1] The latter models were discussed at the meeting, but are not republished here.

ACKNOWLEDGMENTS

This volume, like others in the Santa Fe Institute Series, is devoted to exploring new issues in the sciences of complexity. We hope the volume, like the series, proves useful to a wide readership. The editors are grateful to the Santa Fe Institute for its continuing hospitality, intellectual vigor, and enthusiasm. We are equally

grateful to and thank the participants for their efforts, intelligence, and capacity to join minds in a rather new effort. Funding for the workshop came from SFI core funding, including major grants from the MacArthur Foundation, the National Science Foundation (PHY8714918), and the U.S. Department of Energy (ER-FG05-88ER25054).

Stuart A. Kauffman
University of Pennsylvania
& Santa Fe Institute

Alan S. Perelson
Theoretical Division
Los Alamos National Laboratory

July 2, 1990

REFERENCES

1. Bagley, R. J., J. D. Farmer, N. H. Packard, A. S. Perelson, and I. M. Stadnyk. "Modeling Adaptive Biological Systems." *Biosystems* **23** (1989):113–138.
2. The Department of Trade and Industry. "Method for Obtaining DNA, RNA, Peptides, Polypeptides, or Proteins by Means of a DNA Recombination Technique." English patent number 2183661, issued to Marc Ballivet and Stuart Allen Kauffman and dated 6/28/89.
3. Deutsches Patentant. "Verfahren zur herstellung von peptiden, polypeptiden, oder proteinen." German patent number 3,590,766.5-41, issued to Marc Ballivet and Stuart Allen Kauffman and dated 8/13/90.
4. Ewens, W. *Mathematical Population Genetics.* New York, NY: Springer-Verlag, 1979.
5. Farmer, J. D., S. A. Kauffman, and N. H. Packard. "Autocatalytic Replication of Polymers." *Physica* **22D** (1986):50–67.
6. French Patent Office. "Procédé d'obtention d'ADN, ARN, peptides, polypeptides ou proteinés par une technique de recombinaison d'ADN." French patent number 863683, issued to Marc Ballivet and Stuart Allen Kauffman, dated 12/24/87 and registered as 2,579,518.
7. Kauffman, S. A. "Autocatalytic Sets of Proteins." *J. Theoret. Biol.* **119** (1986):1–24.
8. Perelson, A. S. and G. F. Oster. "Theoretical Studies of Clonal Selection: Minimal Antibody Repertoire Size and Reliability of Self–Nonself Discrimination." *J. Theoret. Biol.* **81** (1979):645-670.
9. Wright, S. "The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution." In *Proceedings of the Sixth International Congress on Genetics* **1** (1932):356–366.