Protein Engineering

Approaches to the Manipulation of Protein Folding

Saran A. Narang

Butterworths

Protein Engineering: Approaches to the Manipulation of Protein Folding

Edited by

Saran A. Narang
Division of Biological Sciences
National Research Council of Canada
Ottawa, Ontario
Canada

Butterworths
Boston London Singapore Sydney Toronto Wellington

Editorial and production supervision by Science Tech Publishers, Madison, WI 53705.

Library of Congress Cataloging-in-Publication Data
Protein engineering: approaches to the manipulation of protein folding/edited by Saran A. Narang.
p. cm.—(Biotechnology ; 14)
Includes bibliographical references.
ISBN 0-409-90116-4
1. Proteins-Biotechnology. I. Narang, Saran A. II. Series.
TP248.65.P76P74 1990
660'.63—dc20 89-22099
CIP

British Library Cataloguing in Publication Data
Protein engineering.
1. Protein engineering
I. Narang, Saran A. II. Series
547.7'5

ISBN 0-409-90116-4

Butterworth Publishers
80 Montvale Avenue
Stoneham, MA 02180

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

Protein Engineering

BIOTECHNOLOGY

JULIAN E. DAVIES, *Editor*
Pasteur Institute
Paris, France

BIOTECHNOLOGY SERIES

1. R. Saliwanchik — *Legal Protection for Microbiological and Genetic Engineering Inventions*
2. L. Vining (editor) — *Biochemistry and Genetic Regulation of Commercially Important Antibiotics*
3. K. Herrmann and R. Somerville (editors) — *Amino Acids: Biosynthesis and Genetic Regulation*
4. D. Wise (editor) — *Organic Chemicals from Biomass*
5. A. Laskin (editor) — *Enzymes and Immobilized Cells in Biotechnology*
6. A. Demain and N. Solomon (editors) — *Biology of Industrial Microorganisms*
7. Z. Vaněk and Z. Hošťálek (editors) — *Overproduction of Microbial Metabolites: Strain Improvement and Process Control Strategies*
8. W. Reznikoff and L. Gold (editors) — *Maximizing Gene Expression*
9. W. Thilly (editor) — *Mammalian Cell Technology*
10. R. Rodriguez and D. Denhardt (editors) — *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*
11. S.-D. Kung and C. Arntzen (editors) — *Plant Biotechnology*
12. D. Wise (editor) — *Applied Biosensors*
13. P. Barr, A. Brake, and P. Valenzuela (editors) — *Yeast Genetic Engineering*
14. S. Narang (editor) — *Protein Engineering: Approaches to the Manipulation of Protein Folding*

CONTRIBUTORS

Barbara A. Blackwell
Plant Research Center
Agriculture Canada, Research Branch
Ottawa, Ontario, Canada

Dani Bolognesi
Department of Surgery
Duke University Medical School
Durham, North Carolina

Edward P. Garvey
Department of Chemistry
The Pennsylvania State University
University Park, Pennsylvania

Kelvin Hill
Department of Biochemistry
Loma Linda University
Loma Linda, California

Kashi Javaherian
Protein Chemistry
Repligen Corporation
Cambridge, Massachusetts

Henry H. Mantsch
Division of Chemistry
National Research Council of Canada
Ottawa, Ontario, Canada

C. Robert Matthews
Department of Chemistry
The Pennsylvania State University
University Park, Pennsylvania

Thomas Matthews
Department of Surgery
Duke University Medical School
Durham, North Carolina

Saran A. Narang
Division of Biological Sciences
National Research Council of Canada
Ottawa, Ontario, Canada

Enrico O. Purisima
Biotechnology Research Institute
National Research Council of Canada
Montreal, Quebec, Canada

Scott D. Putney
Molecular Biology
Repligen Corporation
Cambridge, Massachusetts

David R. Rose
Division of Biological Sciences
National Research Council of Canada
Ottawa, Ontario, Canada

James Rusche
Virology
Repligen Corporation
Cambridge Massachusetts

Paul Schimmel
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts

R.L. Somorjai
Division of Biological Sciences
National Research Council of Canada
Ottawa, Ontario, Canada

Witold K. Surewicz
Division of Chemistry
National Research Council of Canada
Ottawa, Ontario, Canada

Arthur G. Szabo
Center for Protein Structure and Design
National Research Council of Canada
Ottawa, Ontario, Canada

Donald B. Wetlaufer
Department of Chemistry and Biochemistry
University of Delaware
Newark, Delaware

PREFACE

"Vidyavinayasampanne"
(It is when we light the candle that we see how dark it is)

Bhagavadgita

In the protein molecules, nature has used twenty amino acids to manifest its complexity, subtlety, and versatility in terms of their three-dimensional structures. Science seeks the ability to predict the conformation of natural proteins from the sequence of their amino acids. Despite all the progress in the twentieth century, this has not been achieved. It is little wonder that science cannot yet predict the conformation of the natural proteins; perhaps they are not designed to fold predictably. But understanding the rules of protein folding is feasible. The search for the universal truth is more important than finding it because it motivates the creative power of the human mind. About 50 years ago, a group of eminent cosmologists were asked what single question they would ask of an infallible oracle who would answer them only "yes" or "no". George Lemaitre made the wisest choice. He said, "I would ask the oracle not to answer in order that a subsequent generation would not be deprived of the pleasure of searching for and finding the solution."

In this book, I have outlined the complexity of the protein-folding problem and the potential of using genetic tools which, in combination with physical techniques, are expected to shed new light. Many who are actively involved in this field have contributed to this book and I am personally grateful for their enthusiastic support. If this book inspires the imagination of the younger generation, I will be personally satisfied. I could not have undertaken this endeavor without the driving force of my wife Sandhya and the joy of having such a daughter, Ajoo.

Saran A. Narang

INTRODUCTION

The linear array of amino acids in a polypeptide chain does not produce enzymatic activity as such—suitable folding is necessary for the activity to arise. The initial relationship of the particular set of amino acids established in one context has been transformed into a new context. Our real difficulty is the lack of understanding at the interface between the molecular details of the structure and the abstraction of function. In this golden era of molecular biology, we have been generally trained to rationalize biological process based upon the complementarity principle. The central dogma is based on the transfer of information from DNA → RNA → protein, guided by hydrogen-bonding forces. However, when the sequence of the linear chain enters the domain of the tertiary structure of a protein molecule, our concept of information transfer breaks down. It is much like railroad tracks that suddenly end in a sandy wasteland where lies the point of destination. It can get there only if, at the end of the preset linear guidance along tracks, free navigation takes over, steered by a dynamic system. We need to change our outlook toward biology, because we are still looking at the molecular level to understand the structure-function relationship. This situation is a reminder of the story of the drunk who, one dark night, lost his keys. He is seen looking for them under a street light. When asked where he has lost them, he points across the street, where all is dark. "Why, then, are you looking for them here?" He replied, "Because there is more light!" In fact,

the emergence of function can arise through some loss of molecular details of structure. This fact can be recognized, appreciated, and precisely described once we raise our sight from the elemental to the collective level, and this means passing to a higher level of conceptualization such as the structural level of biology.

To understand the process of protein folding, it is important to realize that a complex structure cannot originate instantaneously but is formed in time: it must have evolutionary history. A complex structure is both a partial record of past history and a framework within which future changes occur by the operation of physical laws. During evolution, protein molecules learned not only to survive but also to adapt by random mutagenesis and natural selection, exploiting any chance event which offered the opportunity to improve the structure and its function. It is thus clear that the function must be acquired by a step-by-step process in which new functions do not upset those that have already been established. These arguments suggest that the folding pathway of proteins evolved slowly, step-by-step, in a conservative way at the same time as the structure or function itself. However, the selection pressure in evolution will be mainly on the functions of proteins as manifested by the requirement of the whole organism. These functions are dependent on certain critical amino acids such as those that constitute catalytic or substrate-binding sites and on the maintenance of the three-dimensional structure. It appears that nature conserved these important parts of proteins by maintaining the arrangement of certain amino acid residues in the active site. Conservation of function can be achieved by maintaining the hydrophobic character of the core structure element. For the survival of structure and function in the protein molecule, nature imposed constraints on certain critically placed amino acids in order to have the freedom to choose equivalent amino acids at other places. This paradox, that constraints lead to freedom, is a universal phenomenon observed in living systems. It is thus conceivable that in a protein molecule there is a ZIP code containing the information of the location in space of the essential amino acids and some type of substructure. The code has the information for navigating the pathway of protein folding. This code has to be highly degenerate. The real challenge is to crack this code.

Although the problem of protein folding has been studied for decades, the use of modern genetic tools offers a new hope with great promise. With the recent advances in gene synthesis and recombinant DNA technology, it is now possible to mimic evolution. Any tailor-made DNA can be synthesized and can mutate systematically to introduce substitution, deletion, addition, inversion, frameshift mutation, and transposition, etc.; but the real difficulty is to make predictive changes, which is not possible at present. To put it more bluntly, it is highly unlikely that our current efforts to redesign protein structure rationally will be improved under the present situation. It seems that we may have to resort to an age-old method of learning by making sophisticated use of systematized accidental error as a working hy-

pothesis. A little breakdown of molecular structure is indispensable to opening the door by which new and better knowledge may find entrance. Piet Hein expresses this ubiquitous method in his wonderful verse:

> The way to wisdom? Why, it's plain,
> and easy to express:
> To err, and err, and err again,
> but less, and less, and less.

Saran A. Narang
Ottawa, Ontario

Protein Engineering

CONTENTS