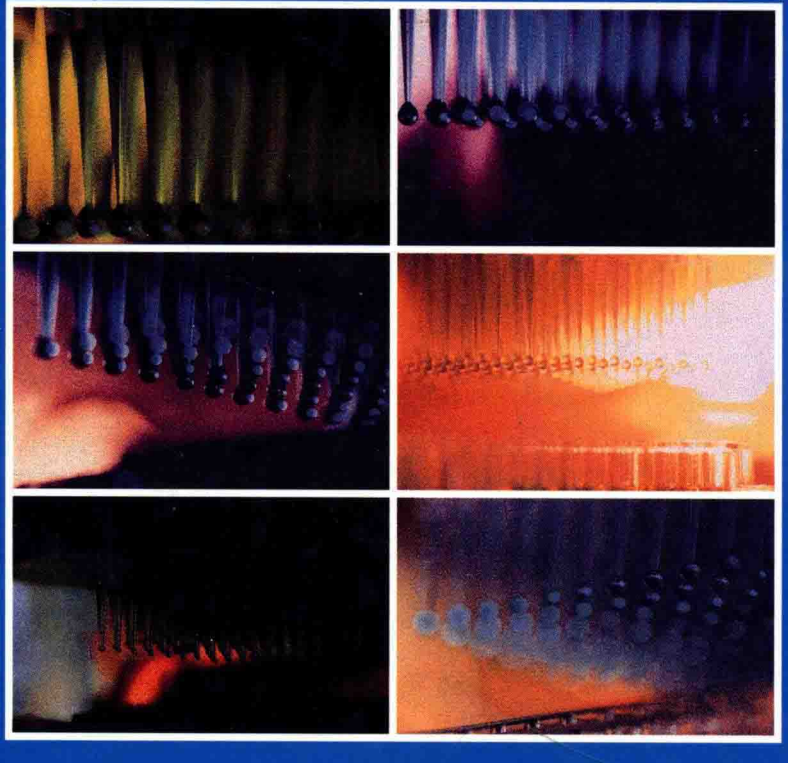
WILEY-VCH

Directed Molecular Evolution of Proteins

or How to Improve Enzymes for Biocatalysis

Edited by
Susanne Brakmann and Kai Johnsson



Directed Molecular Evolution of Proteins

or How to Improve Enzymes for Biocatalysis

Edited by
Susanne Brakmann and Kai Johnsson

WILEY-VCH

The Editor of this volume

Dr. Susanne Brakmann
AG „Angewandte Molekulare Evolution"
Institut für Spezielle Zoologie
Universität Leipzig
Talstraße 33
04103 Leipzig, Germany

Prof. Dr. Kai Johnsson
Institute of Molecular
and Biological Chemistry
Swiss Federal Institute of
Technology Lausanne
CH-1015 Lausanne, Switzerland

Library of Congress Card No.:
applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Die Deutsche Bibliothek – CIP Cataloguing-in-Publication Data
A catalogue record for this publication is available from Die Deutsche Bibliothek.

Printed on acid-free paper.

Printed in the Federal Republic of Germany.

Composition Mitterweger & Partner Kommunikationsgesellschaft mbH, Plankstadt
Printing betz-druck GmbH, Darmstadt
Bookbinding Großbuchbinderei J. Schäffer GmbH & Co. KG, Grünstadt

ISBN 3-527-30423-1

Cover Illustration Recent advances in automation and robotics have greatly facilitated the high-throughput screening for proteins with desired functions. Among other devices liquid handling tools are integral parts of most screening robots. Depicted are 96-channel pipettors for the microliter- and submicroliter range (illustrations kindly provided by Cybio AG, Jena).

Directed Molecular Evolution of Proteins

Edited by S. Brakmann and K. Johnsson

Related Titles from Wiley-VCH

Kellner,R.; Lottspeich, F.; Meyer, H. E.

Microcharacterization of Proteins

1999
ISBN 3-527-30084-8

Bannwarth,W.; Felder, E.; Mannhold, R.; Kubinyi, H.; Timmermann, H.

Combinatorial Chemistry. A Practical Approach

2000
ISBN 3-527-30186-0

Gualtieri,F.; Mannhold, R.; Kubinyi, H.; Timmermann, H.

New Trends in Synthetic Medicinal Chemistry

2001
ISBN 3-527-29799-5

Clark,D. E.; Mannhold, R.; Kubinyi, H.; Timmermann, H.

Evolutionary Algorithms in Molecular Design

2000
ISBN 3-527-30155-0

List of Contributors

Prof. Dr. Frances H. Arnold
Chemical Engineering 210–41
California Institute of Technology
1201 East California Boulevard
Pasadena, California 91125, USA

Prof. Dr. Stephen Benkovic, Dr. Stefan Lutz
Department of Chemistry
The Pennsylvania State University
414 Wartik Laboratory
University Park, Pennsylvania 16802, USA

Prof. Dr. Uwe Bornscheuer
Institut für Chemie und Biochemie
Ernst-Moritz-Arndt-Universität
Soldmannstraße 16
17487 Greifswald

Prof. Dr. Virginia W. Cornish
Columbia University
Department of Chemistry
3000 Broadway, MC 3167
New York, NY 10027-6948, USA

Dr. Rolf Daniel
Institut für Mikrobiologie und Genetik
Georg-August-Universität
Grisebachstraße 8
37077 Göttingen

Prof. Dr. Jacques Fastrez
Laboratoire de Biochimie Physique
et de Biopolymères
Université Catholique de Louvain
Place L. Pasteur, 1.Bte 1B
B-1348 Louvain-la-Neuve, Belgium

Prof. Dr. Donald Hilvert
ETH Hönggerberg
Laboratorium für Organische Chemie
HCI, F339
CH-8093 Zürich, Schweiz

Prof. Dr. Lawrence A. Loeb
Department of Pathology
School of Medicine
University of Washington
Box 357705
Seattle, Washington 98195–7705, USA

Prof. Dr. Alfred Pingoud
Institut für Biochemie
Justus-Liebig-Universität
Heinrich-Buff-Ring 58
35392 Giessen

Prof. Dr. Manfred T. Reetz
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1
45740 Mülheim

Prof. Dr. Peter K. Schuster
Institut für Theoretische Chemie und Strahlenchemie
Universität Wien
Währingerstraße 17
A-1090 Wien, Österreich

Dr. Andreas Schwienhorst
Institut für Mikrobiologie und Genetik
Georg-August-Universität
Grisebachstraße 8
37077 Göttingen

Prof. Dr. K. Dane Wittrup
Dept. Chemical Engineering & Div. Bioengineering
and Environmental Health
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Contents

1
Introduction

Kai Johnsson, and Susanne Brakmann

The application of evolutionary and combinatorial techniques to study and solve complex biological and chemical problems has become one of the most dynamic fields in chemistry and biology. The book presented here is a loose collection of articles aiming to provide an overview of the current state of the art of the directed evolution of proteins as well as highlighting the challenges and possibilities in the field that lie ahead.

Although the first examples of directed molecular evolution date back to the pioneering experiments of S. Spiegelman *et al.* and of M. Eigen and W. Gardiner, who proposed that evolutionary approaches be adapted for the engineering of biomolecules [1, 2], it was the success of methods such as phage display for *in vitro* selection of peptides and proteins as well the selection of functional nucleic acids using the SELEX procedure (Systematic Evolution of Ligands by Exponential enrichment) that brought the power of this concept to the attention of the general scientific community [3, 4]. In the last decade, directed evolution has become a key technology for biomolecule engineering. The success of the evolutionary approach, however, not only depends on the potency of the method itself but is also a result of the limitations of alternative approaches, as our lack of understanding of the structure-function relationship of proteins in general hinders the rational design of biomolecules with new functions. What are the prerequisites for a successful directed evolution experiment? In its broadest sense, (directed) evolution can be considered as repeated cycles of variation followed by selection. In the first chapter of the book, the underlying principles of this concept and their application to the evolutionary design of biomolecules are reviewed by P. Schuster – one of the pioneers in the field of molecular evolution.

Naturally, the first step of each evolutionary project is the creation of diversity. The most straightforward approach to create a library of proteins is to introduce random mutations into the gene of interest by techniques such as error-prone PCR or saturation mutagenesis. The success of random mutagenesis strategies is witnessed by their ample appearances in the different chapters of this book describing case studies of particular classes of proteins and enzymes. In addition, recombination of mutant

genes by DNA shuffling or related techniques can be used to create additional diversity and to accumulate rapidly beneficial and additive point mutations [5]. This is a key technique that also surfaces in the majority of the chapters. The sequence space searched by these approaches is, however, quite limited. DNA shuffling between homologous genes, which has also been called family shuffling, allows yet unexplored regions of sequence space to be accessed [6]. In the chapter by S. Lutz and S. J. Benkovic, an approach to create chimeras even between non-homologous genes and its application in protein engineering is described.

An interesting alternative to the generation of libraries with *in vitro* methods is the generation of so-called environmental libraries, described by R. Daniel. Here, advantage is taken of natural microbial diversity by isolating and cloning environmental DNA and by using the resulting libraries to search for novel biocatalysts.

After the creation of diversity, i.e. the generation of a library of different mutants, the protein(s) with the desired phenotype (function or activity) have to be selected from the library. This can be achieved by either selection or screening procedures. The principal advantage of selection is that much larger libraries can be examined: the number of clones that can be subjected to selection is, in general, five orders of magnitudes above those that can be sorted by advanced screening methods. Impressive examples for the power of true selection, where the survival of the host is directly coupled to the desired phenotype, can be found in the chapters written by D. Hilvert *et al.* and J. F. Davidson *et al.*. The major challenge of most selection approaches is to couple the desired phenotype, such as the catalysis of an industrially important reaction, to the survival of the host. But what can be done if the desired phenotype cannot provide a direct selective advantage to a given host organism? Different approaches appear feasible: if the desired property binds to a given molecule, display systems for the protein of interest such as phage display, ribosomal display or mRNA display, and the subsequent *in vitro* selection of binders by so-called panning procedures are established technologies [3, 7, 8]. A recent publication by the group of J. W. Szostak describes the employment of *in vitro* selection of functional proteins from libraries of completely randomized 80mers (actual library size $\sim 10^{13}$) using mRNA display. This work highlights the power of *in vitro* selection, and is a striking example of an experiment that would simply be impossible to perform using screening procedures [9]. In the chapter written by P. Soumillion and J. Fastrez, an interesting extension of this approach, the *in vitro* selection of novel enzymatic activities using phage display, is reviewed. Here, clever selection schemes link the immobilization of the phage to the desired reactivity.

Another approach to the selection of biomolecules with novel functionalities, i.e. binding, or even enzymatic activity, is based on the yeast two- and three-hybrid system. The potential and limitations of these and related approaches are reviewed in the chapter contributed by the group of V. W. Cornish *et al.*

Despite their inferiority in terms of number of clones examined, screening procedures have become increasingly important over the last years. One important reason for this is the enormous technological progress that has been achieved in automation and miniaturization, allowing up to 10^6 different mutants to be screened in a reasonable timeframe. An overview of advanced screening strategies is given in the article of A. Schwienhost. In the chapter written by K. D. Wittrup a discussion of the prerequisites for a successful screening process is given, analyzing the outcome of the directed evolution of proteins displayed on cell surfaces as a function of the screening conditions. The power of intelligently designed screening processes is demonstrated in the following contributions: M. T. Reetz and K.-E. Jaeger describe screening techniques to engineer the enantioselectivity of enzymes; T. Lanio *et al.* present their approaches for the evolutionary generation of restriction endonucleases, U. T. Bornscheuer reports on the functional optimization of lipases, and last but not least, P. C. Cirino and F. H. Arnold give an overview of directed evolution experiments with heme enzymes.

Clearly, there are various developments and applications in the field of directed evolution that are not covered by any of the articles published in this book. Nevertheless, we hope to provide a snapshot of this rapidly developing field that will inspire and support scientists with different backgrounds and intentions in planning their own experiments.

Finally, we would like to thank all authors for their contributions, and P. Gölitz and K. Kriese of Wiley-VCH for their continuous motivation and help in getting this book published.

References

[1] S. Spiegelman, I. Haruna, I. B. Holland, G. Beaudreau, D. Mills, *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 919–927.
[2] M. Eigen, W. Gardiner, *Pure Appl. Chem.* **1984**, *56*, 967–978.
[3] G. P. Smith, *Science* **1985**, *28*, 1315–1317.
[4] a) C. Tuerk, L. Gold, *Science* **1990**, *249*, 505–510; b) A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.
[5] W. P. Stemmer, *Nature* **1994**, *370*, 389–391.
[6] A. Crameri, S. A. Raillard, E. Bermudez, W. P. Stemmer, *Nature* **1998**, *391*, 288–291.
[7] J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **1997**, *91*, 4937–4942.
[8] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297–12302.
[9] A. D. Keefe, J. W. Szostak, *Nature* **2001**, *410*, 715–718.

2
Evolutionary Biotechnology – From Ideas and Concepts to Experiments and Computer Simulations

Peter Schuster

Research on biological evolution entered the realm of science in the 19th century with the centennial publications by Charles Darwin and Gregor Mendel. Molecular models for evolution under controlled conditions became available only in the second half of the twentieth century after the initiation of molecular biology. This chapter presents an account of the origins of molecular evolution and develops the concepts that have led to successful applications in the evolutionary design of biopolymers with predefined properties and functions.

2.1
Evolution *in vivo* – From Natural Selection to Population Genetics

Nature is the unchallenged master in design by variation and selection and since Charles Darwin's epochal publication of the "Origin of Species" [1, 2] the basic principles of the mechanism behind natural selection have become known. Darwin deduced his principle of evolution from observations "in the field" and compared species adapted to their natural habitats with the results achieved through artificial selection by animal breeders and in nursery gardens. Natural selection introduces changes in populations by differential fitness, which is tantamount to the instantaneous differences in the numbers of decedents between two competing variants. In artificial selection the animal breeder or the gardener interferes with the natural selection process by discarding the part of the progeny with undesired properties. Only shortly after the publication of Darwin's "Book of the Century" the quantitative rules of genetics were discovered by Gregor Mendel [1, 2]. It took, nevertheless, about seventy years before Darwin's theory was united successfully with the consequences of Mendel's results in the development of population genetics [2, 3].

The differential equations of population genetics are commonly derived for sexually replicating species and thus deal primarily with recombination as the dominant source

of variation. Mutation is considered as a rather rare event. In evolutionary design of biopolymers *the opposite is true: Mutation is the common source of variation and recombination occurs only with special experiments, "gene shuffling" [4], for example*. In the formulation of the problem we shall consider here the asexual case exclusively. The mathematical expression dealing with selection through differential fitness is then of the form

$$\frac{dx_k}{dt} = x_k\ (f_k - \Sigma_{j=1}^{n} f_j x_j) = x_k(f_k - \Phi); \quad k = 1, 2, \ldots, n. \tag{1}$$

The fraction of variant I_k is denoted by x_k with $\sigma_k\ x_k = 1$; f_k is its fitness value. Accordingly, we introduced $\phi = \sigma_k f_k\ x_k$ as the mean fitness of the population. The mathematical role of ϕ is to maintain the normalization of variables. The interpretation of Eq. (1) is straightforward: Whenever the differential fitness, f_k-ϕ, of a variant I_k is positive or its fitness is above average, $f_k > \phi$, dx_k/dt is positive and this variant will increase in frequency. The opposite is true if $f_k < \phi$, then the fraction of the corresponding variant will decrease and ultimately approach zero: The variant has died out. Selection thus chooses the variant I_m with the highest fitness value, $f_m = \max\{f_k, k = 1,2,\ldots,n\}$, and after sufficiently long time only this variant will be present in the population, $\lim_{t\to\infty} x_m = 1$. In other words, if we wait long enough, all less fit variants will have died out, and the population becomes homogeneous.

The typical evolutionary scenario considered by population genetics is characterized by low mutation rates. Then the arrival of a new variant by mutation, I_λ in a currently optimized population (containing exclusively I_m) is a rare event and the dynamics of Eq. (1) is visualized in response to such an instant. Apart from a stochastic initial phase, during which the new species is in danger of dying out by accident, the course and the outcome of the selection process is determined exclusively by the difference in fitness values: $s = f_\lambda - f_m$. The value of s is reflected by the number of generations that are required to select the advantageous mutant (see Fig. 2.1). In nature selective advantages of emerging mutants are commonly very small and hence thousands of generations are required before a new variant can take over in the population.

Population genetics saw a major extension by Motoo Kimura [5] who suggested that adaptive mutations were extremely rare, most mutants were selectively neutral, and the predominant role of evolution was the elimination of deleterious variants. Kimura's view was strongly supported by the data obtained from comparative sequence analysis of proteins and nucleic acids [6], which became the basis of current molecular phylogeny. Genotypes are changing steadily and this also during epochs of phenotypic stasis. Despite overwhelming indirect hints for neutral evolution from molecular data, the first direct proof came only recently from experiments on bacterial evolution under controlled conditions: The change in phenotypic properties, like cell size, shows clear