200 Topics in Current Chemistry

Biocatalysis – From Discovery to Application

Volume Editor: W.-D. Fessner

With contributions by
A. Archelas, D. C. Demirjian, R. Furstoss,
H. Griengl, K.-E. Jaeger, F. Morís-Varas,
R. Öhrlein, M. T. Reetz, J.-L. Reymond, M. Schmidt,
S. Servi, P. C. Shah, W. Tischer, F. Wedekind

Springer

This series presents critical reviews of the present position and future trends in modern chemical research. It is addressed to all research and industrial chemists who wish to keep abreast of advances in the topics covered.

As a rule, contributions are specially commissioned. The editors and publishers will, however, always be pleased to receive suggestions and supplementary information. Papers are accepted for "Topics in Current Chemistry" in English.

In references Topics in Current Chemistry is abbreviated Top. Curr. Chem. and is cited as a journal.

Springer WWW home page: http://www.springer.de
Visit the TCC home page at http://www.springer.de/

ISSN 0340-1022
ISBN 3-540-64942-5
Springer-Verlag Berlin Heidelberg New York

Library of Congress Catalog Card Number 74-644622

Printed in Germany

Cover design: Friedhelm Steinen-Broo, Barcelona; MEDIO, Berlin
Typesetting: Fotosatz-Service Köhler GmbH, 97084 Würzburg

SPIN: 10649288 66/3020 - 5 4 3 2 1 0 - Printed on acid-free paper

Volume Editor

Prof. Dr. Wolf-Dieter Fessner
Institut für Organische Chemie
Technische Universität Darmstadt
Petersenstraße 22
D-64287 Darmstadt, Germany
E-mail: fessner@tu-darmstadt.de

Editorial Board

Preface

Worldwide, industrial chemists are endeavoring to meet the criteria of a sustainable development. Ideally, an environmentally benign reaction utilizes nontoxic reagents and solvents for a quantitative conversion at the optimum level of product selectivity. Catalytic procedures are clearly the most economical means of effecting selective processes in organic synthesis. Therefore, Biocatalysis continues to attract considerable attention of the synthetic chemist's community due to its competitive "natural" advantages, particulary in its intrinsic capacity for the asymmetric synthesis of enantiomerically pure compounds. Despite the initial high promise, however, industrialization of biocatalytic processes so far has only been realized for a few large scale operations. Limited substrate tolerance of available enzymes, tedious and costly process development, and the need for an extended knowledge base across many scientific disciplines have often hampered a straightforward replacemant of traditional chemical operations by the utilization of enzyme catalysis.

This volume in the series of Topics in Current Chemistry attempts to familiarize the synthetic organic community with a number of important new developments in the Biocatalysis arena in consideration of both enabling and able technologies. A number of leading contributors from the forefront of this exciting technology address the state of the art of biocatalysis in eight authoritative and timely reviews, from discovery through development to application. Recent landmark advances in molecular biology have the potential to profoundly alter the shape of the field of applied biocatalysis and the pace of its future progress at the beginning of the next millenium. The latest screening and selection technologies allow the rapid identification of enzyme activities that offer properties suitable for organic synthetic applications. Thus, unique enzymes with improved or with novel properties are becoming available from diverse, previously inaccessible sources, by ingenious DNA recombination techniques along an evolutionary approach, or by eliciting monoclonal antibodies with directable binding specificities. Besides appropriate techniques of protein handling and reaction engineering, for future successful biocatalytic processes it will be mandatory to have ready access to an extended database that precisely defines scope and limitations for each synthetically useful enzyme. Thus, half of the chapters illustrate the synthetic potential of recently emerging biocatalysts that have a capacity for the synthesis or modification of the most important classes of pharmaceutically interesting compounds, particulary in the phospholipid, epoxide, cyanohydrin, and oligosaccharide fields.

Unique new enzymes are now readily accessible in quantity with properties that are amenable to modification on demand. It is my firm belief that such fascinating possibilities not only open new playgrounds for creative minds but will also assist each practising scientist with effective tools for tackling the future challenges in Organic Synthesis, and it is my hope that this volume will actively support this goal.

Darmstadt, September 1998 Wolf-Dieter Fessner

Contents of Volume 184

Bioorganic Chemisty
Models and Application

Volume Editor: F. P. Schmidtchen

ISBN 3-540-61388-9

Contents

Screening for Novel Enzymes

David C. Demirjian[1] · Pratik C. Shah[2] · Francisco Morís-Varas[1]

[1] ThermoGen, Inc., 2225 W. Harrison, Chicago, IL 60612, USA. *E-mail: ddem@thermogen.com*
[2] NephRx Corporation, 2201 W. Campbell Park, Chicago, IL 60612, USA

The development of new biocatalysts as synthetic tools for chemists has been expanding rapidly over the last several years. It is now possible either to discover or engineer enzymes with unique substrate specificities and selectivities that are stable and robust for organic synthesis applications. This has been made possible by the application of the newest screening and selection technologies that allow rapid identification of enzyme activities from diverse sources. We focus on how to recognize and tackle important issues during the strategic design and implementation of screening for novel enzymes. We also review the approaches available for biocatalyst discovery and relate them to the isolation of thermostable enantioselective esterases and alcohol dehydrogenases for the purpose of illustration and discussion.

Keywords: Enzymes, Biocatalysts, Screening, Enzyme discovery, Biocatalysis, Selections, Biotransformation, Industrial enzymes.

Topics in Current Chemistry, Vol. 200

1 Overview

1.1 A Bit of Biocatalyst History

The first enzymes were discovered in the 1830s: diastase by Payen and Persoz and pepsin by Schwann. The first asymmetric synthesis using an isolated enzyme was carried out by Emil Fischer in 1894 when he applied the cyanohydrin reaction to L-arabinose, establishing that enzymes have an extremely high degree of substrate specificity [1]. By the 1920s several different enzymes were known to exist. While the idea that enzymes could be used for a variety of commercial applications was always in the realm of possibility, it was only in the 1960s and early 1970s that commercial processes using enzymes were widely used. For example, carbohydrate-processing enzymes have been widely used in the food industry for the processing of corn, potato and other starches [2]. Also, the addition of proteases to detergents became an important industrial use of enzymes. While these applications have accounted for a majority of the bulk enzyme sales, an increasingly important application of enzymes has been as a catalytic tool in the synthesis of specialty organic chemicals [3–7].

For example, an early application of biocatalysis was a commercial process to make ascorbic acid starting with glucose, developed by Reichstein in 1934 (reviewed in [8]). The adoption of biocatalysis processes in the production of specialty chemicals has been slower than anticipated. Most synthetic chemistry biocatalysis applications currently in place have been developed using a limited set of commercially available enzyme tools such as lipases, esterases, or proteases. These enzymes were developed for other applications, such as those in

the food and textile industries [9]. They were explored as biocatalytic solutions because they were readily available inexpensively and in large quantities. It is only within the last few years that there has been a concerted effort to discover, engineer, and develop enzymes specifically for chemical synthesis applications [10]. The development of new molecular screening techniques, recombinant DNA technology, and a new focus on the importance of biocatalysis has significantly advanced the field.

1.2 *Tailoring Screening to Meet Biocatalysis Challenges*

There are a number of challenges to the successful development of a commercial biocatalysis process. Most of these challenges can be addressed at the outset by properly setting up a screen in order to identify a novel catalyst which has the properties necessary to carry out a reaction that can be scaled-up. The most critical challenges include the following.

Every reaction may need a different enzyme. Even highly related molecules may need a different catalyst to optimally carry out a reaction. While the same enzyme can often recognize similar substrates, the differences in substrate specificity can be great enough to prevent a process from being economical. There are several approaches to expand the diversity of enzymes available to carry out biotransformations and develop new enzymes specifically tailored or discovered for chemical synthesis applications. Recently several groups, including our own, have begun to overcome this by developing families of enzymes such as hydrolases, oxidoreductases and others through enzyme discovery approaches. We have now developed specific families of enzymes capable of carrying out certain classes of reactions, giving the synthetic chemist a wider range of enzymes off-the-shelf. Another approach to diversify an enzyme collection is to use directed evolution to fine-tune an enzyme's activity [11]. This is even more effective when taken together with the enzymes available from biocatalyst discovery searches that now allow a greater range of starting templates for the reaction.

Enzymes are notoriously unstable and require special handling conditions. It is likely that no other factor has been more detrimental in keeping biocatalyst technologies from practical implementation on large scale than the inherent instability found in most enzymes. If enzymes are to play a significant role in large scale processing of chemicals, they must be able to endure the often harsh conditions associated with industry. Alpha-amylase and subtilisin are the most successful industrial enzymes primarily because of their thermostability and hardiness. In addition, while most enzymes lose a significant portion of their activity in organic solvents, thermostable enzymes are typically more tolerant to the denaturing conditions of many organic solvents [12, 13]. Other technologies such as immobilization [14, 15], cross-linked-enzyme-crystals (CLEC [16]), and directed evolution [11, 17] can also help to further stabilize enzymes, but by setting up an initial screen for stable enzymes, one can ensure an inherent stability in the enzyme that is developed.

Multidisciplinary science. Biocatalysis is truly a multidisciplinary science, incorporating microbiology, molecular biology, enzymology and biochemistry, synthetic chemistry, analytical chemistry, and chemical engineering. Effective screening for new enzymes for biocatalytic applications requires an understanding and integration of these sciences.

It takes a long time to develop a biotransformation process. Finally, the time it takes to develop a biotransformation can prevent the process from reaching a commercial scale. In the pharmaceutical industry time is money, and a delay in the development of a useful process will lead to the adoption of an alternative process – even if it is more costly. Biocatalyst discovery and screening represents one of the longest time factors involved. By pre-establishing a diverse library of enzyme activities, organisms and gene banks, biocatalyst discovery and engineering can be shortened considerably.

We will explore here some examples and how screens were used for developing these enzyme families, and discuss the importance of choosing the right conditions to allow for easier identification of an active enzyme that can readily be produced and scaled-up when needed. By implementing screening solutions that take these factors into account, one can exploit the vast diversity of nature's catalytic repertoire.

1.3 A Working Plan

Figure 1 outlines a general working plan for exploring a biocatalysis solution. The simplest solutions are tested first at the top of the flow-chart. The easiest solution is to find a commercially available enzyme for a given reaction. If one is

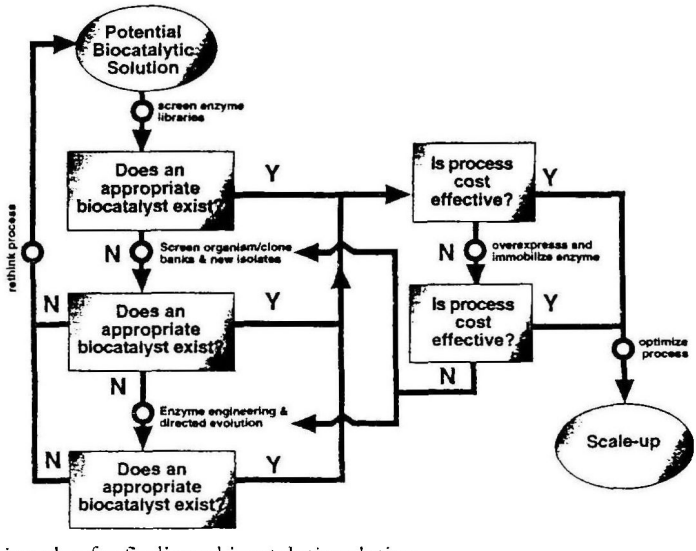

Fig. 1. Working plan for finding a biocatalytic solution

not found, then a more involved screening project needs to be undertaken. By setting up this screening project correctly and methodically, one can save time and effort in the discovery of a new catalyst. We usually screen organism banks, clone banks, and new isolates in parallel because it increases the probability of finding an appropriate enzyme of interest. It is difficult to determine ahead of time that a particular method is likely to lead to an enzyme with the desired activity, and all methods usually yield different results. Finally, if an ideal catalyst cannot be found, newer technologies such as directed evolution and other methods of protein engineering could be employed to customize an appropriate biocatalyst.

One of our activities is to develop new libraries of enzymes for certain classes of chemical reactions so that the speed of developing a new biocatalysis process is increased, thus expanding the options at the top of the diagram.

There are a number considerations which need to be taken into account when screening for new enzymes, and most of these can be generalized to a variety of screening projects.

2
Enzyme Sources

The first question that should be addressed is on the source of the enzymes that are to be screened. There are advantages and disadvantages to each source, and these have been compiled in Table 1.

2.1
Screening from Commercially Available Enzyme Libraries

By far the fastest and easiest route to finding a new enzyme is to find one that exists in a commercial library. This allows for instant access to the catalyst in sufficient quantities to take the project to the next step. For the most part there are still only a few sources. Traditional sources such as Sigma, Amano, Roche Molecular Biochemicals (formerly Boehringer-Mannheim), and Toyobo carry many of the classic enzymes that have been used in the past.

The biggest change over the last several years has been the development of larger commercial enzyme libraries to enhance and simplify biocatalyst discovery. We have been developing enzymes using a methodical plan to discover and develop useful new enzymes for synthetic chemists. As a result, the first new sets of thermostable enzyme families were discovered. These included esterases and lipases, and more recently alcohol dehydrogenases. Other groups have also taken similar approaches [19]. In addition, companies like Altus Biologics have assembled commonly used industrial catalysts for researchers and modified them with technology to stabilize the proteins [20].

Although larger commercial enzyme libraries are becoming available to the researcher, it is still not always possible to find an appropriate enzyme from a commercial source. In this case, a custom screening needs to be performed. This screening can be from a collection of microorganisms or from clone banks that have been generated from these organisms or isolated DNA.

Table 1. Enzyme sources

Source		Advantages	Disadvantages
Commercial enzymes[a]	ThermoGen, Amano, Roche Molecular Biochemicals, Sigma-Aldrich, Fluka, Toyobo, Diversa, Altus Biologics, Biocatalysts	• Available off the shelf and generally in large quantities • Fastest solution • Generally tested by multiple users	• Until recently only limited enzymes available • Generally not available for genetic engineering
Culture sources	International Culture Banks Proprietary banks	• Source of new, undiscovered catalysts	• Need to clone for optimal enzyme expression • Findings limited to those which are expressed in the organism • Expensive and difficult to maintain collections
Clone banks	Proprietary sources	• Source of new, undiscovered catalysts • Enzymes which are not expressed in original host can often be found • Generally easier to move into high volume production • Allows easier genetic engineering of enzyme	• Large libraries need to be maintained • Many proteins will not express in the cloning host
Clone banks from uncultured organisms	Proprietary sources	• Same advantages as clone banks plus source of new, undiscovered catalysts which have not been cultivated by other means	• All the disadvantages of clone banks • Possible redundancies in library • Possibly difficult to express evolutionarily distant genes

[a] For additional suppliers list see [18].

2.2 Culture Sources

Most enzymes of industrial importance developed in the past have been derived from species that are GRAS (Generally Regarded as Safe). These include bacterial species for *Bacillus* and *Lactobacillus*, and *Pseudomonas* and fungi from the *Ascomycota* and *Zygomycota* classes [21].

A partial listing including web addresses of some of the largest culture sources are given in Table 2, and many of these have links to most of the other available strain collections. Some of these include the American Type Culture Collection (ATCC) and the DSMZ German Collection of Microorganisms and Cell Cultures.

Access to a good proprietary collection of enzymes is also extremely helpful for finding an enzyme, especially if those collections have been logically assembled for applications of interest. If one knows which type of enzyme one is screening for, cultures can often be enriched for particular enzyme activities by standard methods [22].

Screening from culture sources has been successful in many cases [23]. There are, however, several challenges. First, since the media for different organisms will be diverse, the systematic screening of organism banks becomes more difficult. A general solution is to group like strains together for screening on different types of media and to find a media recipe that can support growth of the organism. Since the media composition ultimately effects which proteins are produced in the cell, the media choice for a particular enzyme screen can be critical to having the protein of interest expressed [24]. The cost of establishing and maintaining a proprietary strain collection can also be quite high. Several companies provide cost-effective access to their proprietary culture collections.

Strain redundancy is also another concern. When identifying new isolates, verification that a particular strain is unique can be accomplished by one of several methods. The first is a phenotypic characterization of traits such as colony morphology, classical strain typing such as the Bergey system [25], ribosome relationship data [26], or PCR-based comparison of strains [27]. One thing to note, however, is that very small differences in amino acid makeup of related enzymes can lead to significant differences in enzyme activity. This can often be overlooked if using a broader genomic comparison tool to eliminate duplicates.

Table 2. Partial list of online strain collections

Collection	www site
ATCC American Type Culture Collection	www.atcc.org
DSMZ German Collection of Microorg. and Cell Cultures	www.gbf.de/DSMZ
Microbial Information Network of China	sun.im.ac.cn
MSDN Microbial Strain Data Network	www.bdt.org.br/bdt/msdn
World Data Center for Microorganisms (Japan)	wdcm.nig.ac.jp
CGSC *E. coli* Genetic Stock Center	cgsc.biology.yale.edu/top.html

2.3 Screening from Clone Banks

Screening for new enzymes from clone-banks can be extremely rewarding. By setting up the clone banks in a unified or small set of host organisms (like *E. coli*, *Bacillus*, or yeast) only a limited number of different propagation methods need to be implemented, thus allowing systematic screening methods to be carried out more easily. If a cloned enzyme is discovered, it is generally easier to scale-up and produce in larger quantities. In addition, cloning is a prerequisite to most types of genetic modification of the gene including directed evolution. Additionally, in a clone bank the gene of interest is often removed from its regulatory elements that can repress expression and helps purify the gene away from isozymes and other competing activities.

There are a few disadvantages to screening clone banks. Removing a gene from its regulatory elements can turn the gene off instead of on, thus masking the activity. The gene may not express well in the host cloning strain (which typically includes *E. coli*, *Bacillus* and yeast) because of codon usage, nucleic acid structure, or lethality. The activity from enzymes that are post-translationally modified may be altered or destroyed [28]. In addition, the actual restriction sites used to construct the gene bank can affect whether any activity is observed or not, since the distance from the expression signals are altered. Finally, for each organism a clone bank is developed from, one needs to screen thousands to tens of thousands of clones for each organism to cover the entire genome of that organism. Several solutions have been devised, including automated screening and hierarchical screening methods, that will be discussed in more detail below.

The DNA used for cloning can originate from DNA prepared from cultured organisms or DNA prepared from uncultured organisms. It has been estimated that less than 1% of the world's organisms have been cultured [29]. By directly amplifying DNA from soil samples using PCR the DNA from these uncultured organisms can be isolated and used as a cloning source, thus allowing access to a greater diversity of enzymes [19]. There are a few concerns with DNA from uncultured organisms that need to be taken into account. First, it is difficult sometimes to clone DNA between closely related species. Generating DNA fragments from organisms that are less closely related presents even more problems. Second, by nature of the PCR reaction, certain DNA species will be amplified preferentially; thus there is no guarantee that if 100 esterases are isolated from one PCR reaction, they will be different.

One way of alleviating a number of these problems is actually to make several parallel clone banks using different expression vectors and strains, regulatable expression systems, and low copy-number vectors. Vectors choices [28] such as plasmids vs phagemids, high vs low copy number vectors, all can affect what is cloned, and there is not one solution which will satisfy every project. Furthermore, the cloning host also makes a difference. *E. coli* is usually the host of choice, but *Bacillus* can be preferred for secretable enzymes, and yeast for eukaryotic or post-transitionally modified enzymes.

2.4
Organizing the Enzyme Sources

If a commercial or pre-developed enzyme is not available from our library, we typically screen all the other sources simultaneously because one can never tell where a unique or useful activity is likely to be developed from first. In general, the use of formats such as microtiter plates helps to speed screening and creates an array that can be systematically screened, although one needs to take special care to avoid contamination problems. Clone banks can be stored in one of several ways. DNA libraries can be stored untransformed in tubes, or as transformed cultures. These transformed cultures can be stored as a pooled mixture or as an array of individual colonies. Depending on the insert size, a bank of thousands to tens of thousands of individual colonies needs to be stored to cover the complete genome of an organism. This should be done for every organism in the collection if one wants ultimate versatility in screening. The library from a single organism can then be arrayed and multiplexed into a single screening plate to save screening time. Once a positive candidate is identified, the sub-plates can then be tested quickly to determine which clone contains the appropriate activity.

There are advantages and disadvantages for each strategy. First, saving the clone banks as either DNA libraries or transformed pools provides the most flexibility and saves time and effort. The DNA can be transformed into any host that may be desired. Different hosts may be desired for different activities since

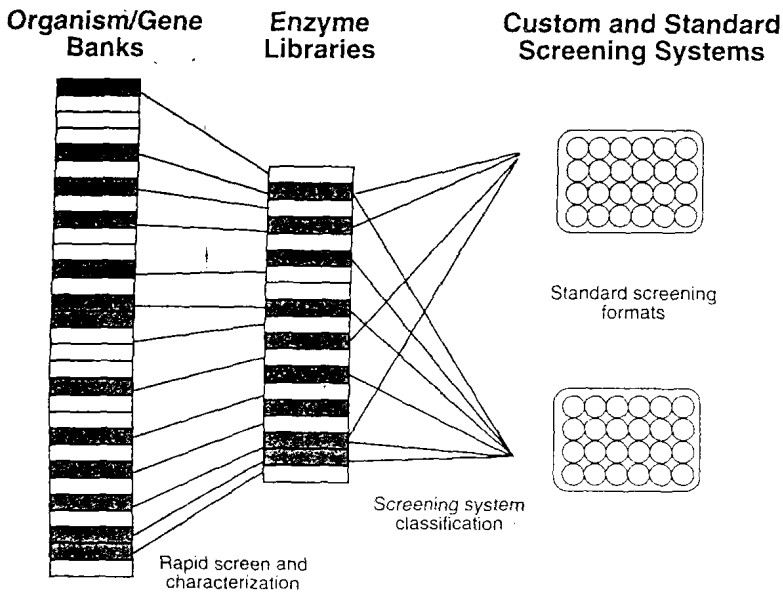

Fig. 2. Enzyme library organization. Enzymes discovered in organisms and gene banks are organized by activity type into enzyme libraries. These are characterized and can be subdivided into screening kits based on their substrate preferences and synthetic utility