ISSN 1860-6768 · BJIOAM 2 (2) 155–280 (2007) · Vol. 2 · February 2007

Biotechnology journal
v. 2, no. 2 (Feb. 2007)
General Collection
W1 BI918QK
2007-03-29 11:34:45

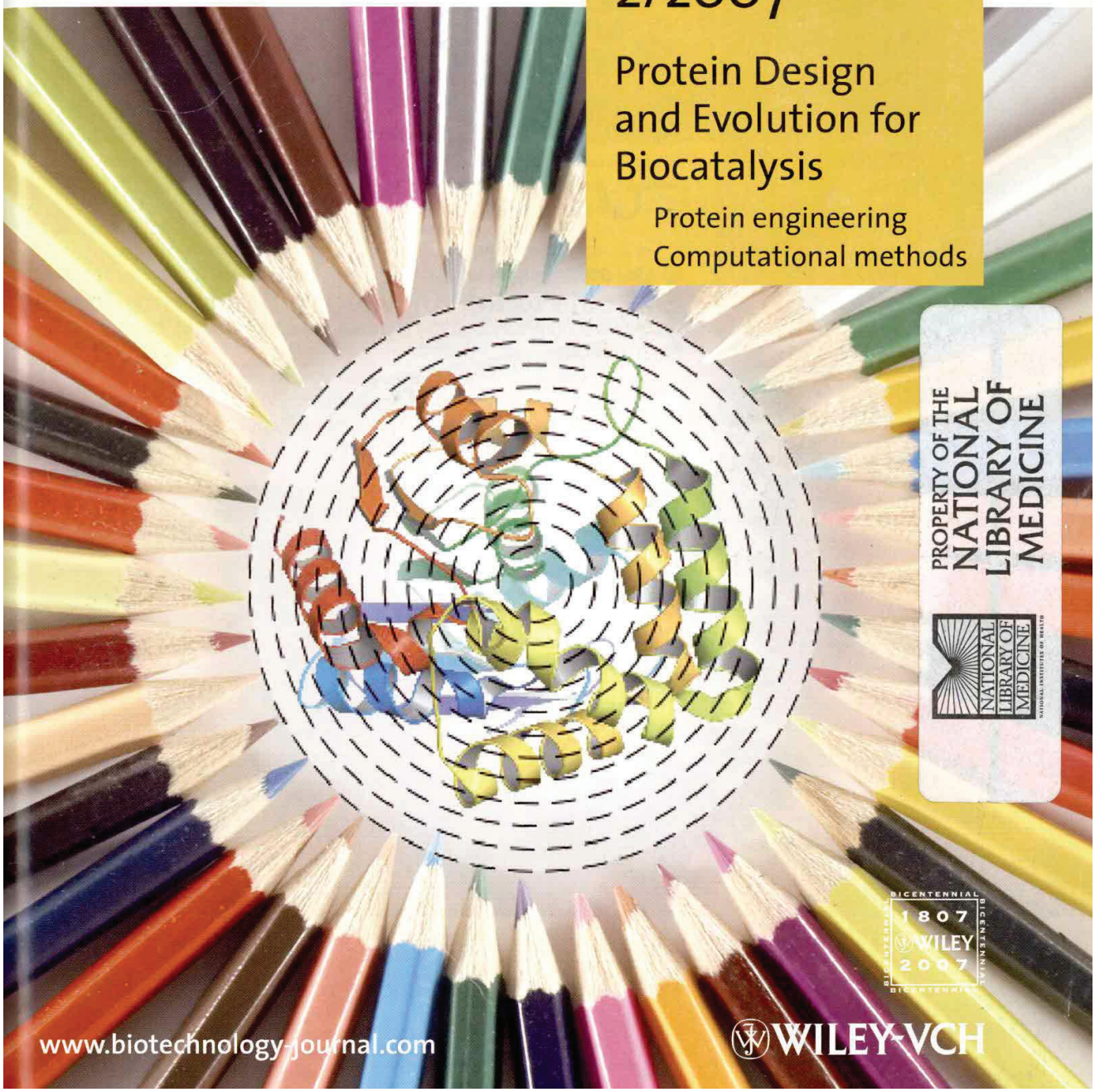
Biotechnology Journal

Healthcare · Nutrition · Technology

2/2007

Protein Design and Evolution for Biocatalysis

Protein engineering
Computational methods



PROPERTY OF THE NATIONAL LIBRARY OF MEDICINE

www.biotechnology-journal.com

WILEY-VCH

Biotechnology Journal

Healthcare · Nutrition · Technology

Editor-in-Chief

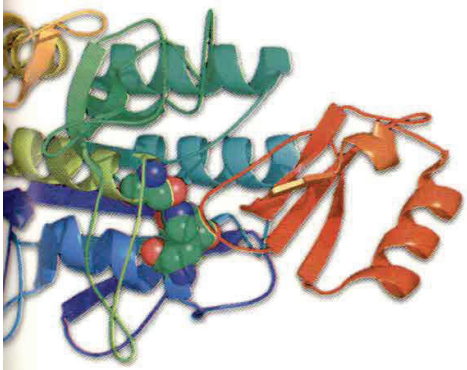
Hans Günter Gassen (*Darmstadt, Germany*)
E-mail: gassen@bc.chemie.tu-darmstadt.de

Senior Editors

Pushpa M. Bhargava (*Hyderabad, India*)
Marc Blondel (*Brest, France*)
Stefan Buchholz (*Hanau, Germany*)
Zhuan Cao (*Beijing, China*)
Giorgio Carta (*Charlottesville, VA, USA*)
Henry Daniell (*Orlando, FL, USA*)
Theo Dingermann (*Frankfurt, Germany*)
Carl J. Douglas (*Vancouver, Canada*)
Gerhard Eisenbrand (*Kaiserslautern, Germany*)
Alois Jungbauer (*Vienna, Austria*)
Sang Yup Lee (*Daejeon, Korea*)

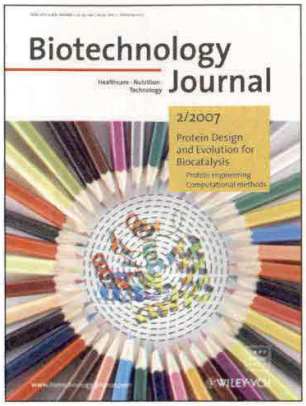
Managing Editor

Barbara Janssens
Biotechnology Journal
Wiley-VCH Verlag GmbH & Co. KGaA
Boschstrasse 12
D-69469 Weinheim, Germany
Tel.: +49-6201-606 495
Fax: +49-6201-606 525
E-mail: biotechnology-journal@wiley-vch.de



Editorial Board

Petr Bartunek (*Prague, Czech Republic*)
Lawrence W. Bergman (*Philadelphia, PA, USA*)
Anne Bertolotti (*Cambridge, UK*)
Miguel A. Blázquez (*Valencia, Spain*)
Andrea Buchacher (*Vienna, Austria*)
Janusz M. Bujnicki (*Warsaw, Poland*)
Donald Cowan, (*Cape Town, South Africa*)
Jiri Damborsky (*Brno, Czech Republic*)
Massimo Delledonne (*Verona, Italy*)
Luc Douay (*Paris, France*)
Ruth Freitag (*Bayreuth, Germany*)
Hikmet Geckil (*Malatya, Turkey*)
Klaus Graumann (*Kundl, Austria*)
Elmar Heinzle (*Saarbrucken, Germany*)
László Homolya (*Budapest, Hungary*)
Suwan N. Jayasinghe (*London, UK*)
Goetz Laible (*Hamilton, New Zealand*)
Ulf Landegren (*Uppsala, Sweden*)
Gen Larsson (*Stockholm, Sweden*)
David Leys (*Leicester, UK*)
Walter Lichtensteiger (*Zurich, Switzerland*)
Christine Machold (*Thousand Oaks, CA, USA*)
Stephan Michnick (*Montreal, Canada*)
Phillipe Mondon (*Labege, France*)
Maria Mota (*Lisboa, Portugal*)
Daniel Otzen (*Aalborg, Denmark*)
Livio Pellizzoni (*Rome, Italy*)
Mikhail L. Rabinovich (*Bayreuth, Germany*)
Barur R. Rajeshkumar (*Worcester, MA, USA*)
John Rossi (*Duarte, CA, USA*)
Sven J. Saupe (*Bordeaux, France*)
Hiroaki Suga (*Tokyo, Japan*)
Kazunari Taira (*Tokyo, Japan*)
Nektarios Tavernarakis (*Heraklion, Greece*)
Didier Trouche (*Toulouse, France*)
Victoriano Valpuesta (*Malaga, Spain*)



Cover illustration

Soon designing proteins on demand? This is the vision expressed by the designer pencils pointing at the logo of the meeting held in September 2006 in Greifswald, Germany. This special issue was edited by Prof. Uwe Bornscheuer from Greifswald, who selected papers from keynote speakers at this meeting.
With special thanks to Prof. Romas Kazlauskas, University of Minnesota, for the design of the conference logo.
Pencils © FOTOLIA.

Publisher

Wiley-VCH GmbH & Co. KGaA
P.O. Box 101161
D-69451 Weinheim, Germany

http://www.biotechnology-journal.com

WILEY-VCH

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table of Contents



Editorial

155 **Protein design and evolution for biocatalysis**



Forum

News and views
159 News & views
161 Meetings
163 Paper watch

News from industry
166 New on the market
167 Biotech startups
167 Biotech industry highlights
169 Facts and figures
170 Industrial prizes
171 Patents

News from academia
172 Public-private partnerships
173 Academic awards
174 Novel technologies
174 Teaching biotechnology
175 Biotechnology in our society



176 Meeting report
Protein design and evolution for biocatalysis
Jiri Damborsky

Scientific

180 Review
Better library design: data-driven protein engineering
Javier F. Chaparro-Riggers, Karen M. Polizzi and Andreas S. Bommarius

192 Review
Recent progress in engineering α/β hydrolase-fold family members
Zhen Qian, Christina J. Fields, Ying Yu and Stefan Lutz

201 Review
Updating the metagenomics toolbox
Esther Gabor, Klaus Liebeton, Frank Niehaus, Juergen Eck and Patrick Lorenz



© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 3 of 16
Bayer v. Corteva - IPR2023-01036

207 Short Communication
Biochemical and structural features of a novel cyclodextrinase from cow rumen metagenome
Manuel Ferrer, Ana Beloqui , Olga V. Golyshina, Francisco J. Plou, Alexander Neef, Tatyana N. Chernikova, Lucia Fernández-Arrojo, Iraj Ghazi, Antonio Ballesteros, Kieran Elborough, Kenneth N. Timmis and Peter N. Golyshin
Supporting information see www.biotechnology-journal.com

214 Research Article
Volsurf computational method applied to the prediction of stability of thermostable enzymes
Paolo Braiuca, Andreas Buthe, Cynthia Ebert, Paolo Linda and Lucia Gardossi

221 Research Article
Fluorogenic substrates for lipases, esterases, and acylases using a TIM-mechanism for signal release
Renaud Sicart, Marie-Pierre Collin and Jean-Louis Reymond
Supporting information see www.biotechnology-journal.com

232 Research Article
An optimized ATP/PP_i-exchange assay in 96-well format for screening of adenylation domains for applications in combinatorial biosynthesis
Linda G. Otten, Michelle L. Schaffer, Benoit R.M. Villiers, Torsten Stachelhaus and Florian Hollfelder
Supporting information see www.biotechnology-journal.com

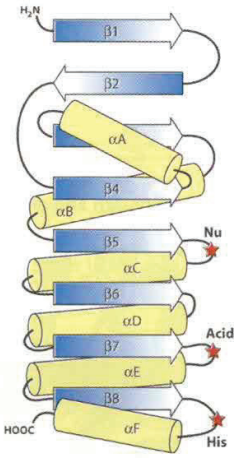
241 Research Article
Directed evolution of glucose oxidase from *Aspergillus niger* for ferrocenemethanol-mediated electron transfer
Ziwei Zhu, Mian Wang, Amit Gautam, Jovana Nazor, Carmen Momeu, Radivoje Prodanovic and Ulrich Schwaneberg
Supporting information see www.biotechnology-journal.com

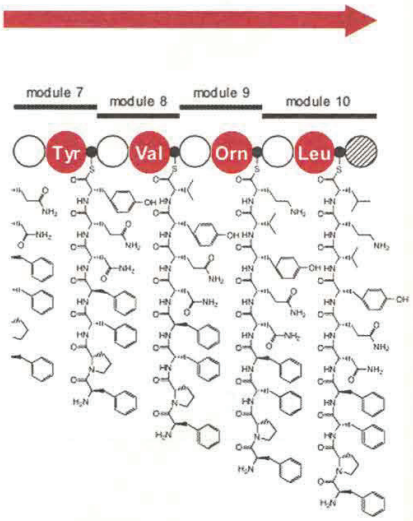
249 Research Article
A versatile esterase from *Bacillus subtilis*: Cloning, expression, characterization, and its application in biocatalysis
Marlen Schmidt, Erik Henke, Birgit Heinze, Robert Kourist, Aurelio Hidalgo and Uwe T. Bornscheuer

254 Research Article
Involvement of alanine 103 residue in kinetic and physicochemical properties of glucose isomerases from *Streptomyces* species
Mohamed Ali Borgi, Moez Rhimi and Samir Bejar

260 Research Article
Microbial synthesis of ethyl (*R*)-4,4,4-trifluoro-3-hydroxybutanoate by asymmetric reduction of ethyl 4,4,4-trifluoroacetoacetate in an aqueous-organic solvent biphasic system
Junyao He, Xudong Mao, Zhihao Sun, Pu Zheng, Ye Ni and Yan Xu

266 Research Article
Expression of an acetyl-CoA synthase and a CoA-transferase in *Escherichia coli* to produce modified taxanes *in vivo*
Catherine Loncaric, Amanda F. Ward and Kevin D. Walker





Resources

275 Meetings and Conferences
277 BTJ future topical issues
278 Books
279 Instructions
280 Impressum

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 4 of 16
Bayer v. Corteva - IPR2023-01036

Biotechnology Journal

DOI 10.1002/biot.200600170

Biotechnol. J. 2007, 2, 180–191

This material may be protected by Copyright law (Title 17 U.S. Code)

Review

Better library design: data-driven protein engineering

Javier F. Chaparro-Riggers[1]*, *Karen M. Polizzi*[1]* and *Andreas S. Bommarius*[1, 2]

[1]School of Chemical and Biomolecular Engineering, Parker H. Petit Institute of Bioengineering and Bioscience, Atlanta, GA, USA
[2]School of Chemistry/Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA

Data-driven protein engineering is increasingly used as an alternative to rational design and combinatorial engineering because it uses available knowledge to limit library size, while still allowing for the identification of unpredictable substitutions that lead to large effects. Recent advances in computational modeling and bioinformatics, as well as an increasing databank of experiments on functional variants, have led to new strategies to choose particular amino acid residues to vary in order to increase the chances of obtaining a variant protein with the desired property. Strategies for limiting diversity at each position, design of small sub-libraries, and the performance of scouting experiments, have also been developed or even automated, further reducing the library size.

Received 29 August 2006
Revised 2 November 2006
Accepted 12 November 2006

Keywords: Biocatalysis · Consensus approach · Directed evolution · Protein engineering

1 Introduction

1.1 The case for data-driven protein engineering

Data-driven protein engineering is increasingly used as an alternative to rational design and combinatorial engineering because it uses available knowledge to limit library size, while still allowing for the identification of unpredictable substitutions that lead to large effects. One illustration of the efficiency of a data-driven approach is the work of Parikh and Matsumura [1], who showed that a single round of structure-guided site-saturation mutagenesis on β-galactosidase using the crystal structure with a bound substrate analog identified a variant that was 180-fold more active and 700 000-fold more specific for fucoside substrates than the wild-type enzyme. Previous mutants identified in seven rounds of directed evolution and screening were only 10-fold more active and 2-fold more specific than wild type, and required considerably more time and effort to screen [1, 2]. Table 1 summarizes recent work in data-driven protein design. For each type of design outlined below, one specific example is discussed in detail. The other examples are provided in the table to give a general summary (rather than the "best-case scenario") of the work in this area and to direct interested readers towards other references worth examining.

In a broad sense, data-driven protein design uses all available knowledge to limit library size. Crystal structures of the protein of interest (with or without the substrate bound) are, of course, useful, but data-driven protein engineering is not limited to such cases. In fact, the view that data-driven protein engineering can be equated with site-saturation mutagenesis is too narrowly focused as data-driven methods can also be applied to methods such as site-directed mutagenesis or recombination. Besides crystal structures, a wide variety of tools can be applied, ranging from sequence alignments or phylogenetic trees to computational models, and small, experimental libraries used to "scout" for interesting residues. Conversely, such tools can be used to identify areas of the protein that must not be manipulated. Additionally, it is important to note that many useful data-driven techniques (*e.g.*, the consensus concept) for the improvement of protein stability were pioneered by protein therapeutic engineers. The concepts can be often directly transferred to enzyme engineering, where they work equally well.

Correspondence: Professor Andreas S. Bommarius, 315 Ferst Dr., Atlanta, GA 30332-0363, USA
E-mail: andreas.bommarius@chbe.gatech.edu
Fax: +1-404-894-2291

Abbreviations: epPCR, error-prone PCR; T_m, melting temperature

*These authors contributed equally to this work.

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

WILEY InterScience® DISCOVER SOMETHING GREAT

Corteva Exhibit 2035 - Page 5 of 16
Bayer v. Corteva - IPR2023-01036

Table 1. Summary of important work in data-driven protein engineering (giving numerous examples for each type of library to give a general overview and indicate other references worth examining). Where applicable, an example from classical combinatorial engineering is provided for comparison[a)]

Enzyme	Method	No. of mutation sites	No. of rounds	Mutants	Improved property		Library size	Refs
					Property	Fold-improvement		
Choosing residues to target								
Structure-based								
Penicillin acylase	SM	3	1	3	Product inhibition	4–5	700	[22]
N-Acetylneuraminic acid lyase	SM	3	1	1	k_{cat}/K_M Specificity	49 690	96*3	[24]
Phosphotriesterase	SM SM	2 1	1 1	2 1	k_{cat}/K_M	100 (wt) 1000 (wt)	1200 100?	[9]
Epoxide hydrolase	SM SM SM	4 2 1		 3	Specific activity	 10	4×300 2×300 1×100	[23]
Lactate dehydrogenase	SDM	Loop length			k_{cat}/K_M Selectivity	$3.9*10_5$ 1700	2	[28]
Cytochrome P450	Rec Comp	3 genes	1		Proper folding Activity	46% >33%	6561	[37]
Toluene 4-monooxygenase	SM	1 2	1	1 2	Specific activity	4 7	300 350	[89]
β-Galactosidase to β-Fucosidase	SM	3	1	2	k_{cat}/K_M Specificity	180 $7*10^5$	$1*10^4$	[1]
	Rec		7	8	k_{cat}/K_M Specificity	10 1000	7× $1*10^4$	[2]
Esterase	SM	4	1	1	Enantio-selectivity	5	4× 500	[4]
	epPCR		1	1		1.5	288	[90]
Cre recombinase	SM	11	1	5	Specificity	New recognition	$3×10^8$	[25]
Homology-based								
Subtilisin	SS	55	1		Heat treatment	4	1500	[39]
Epidermal growth factor	SS	19			Agonist activity	123	$3×10^6$	[40]
Glyphosate N-acetyltransferase	epPCR/Rec/SS		11	30–35	k_{cat}/K_M	9563	$1.7×10^8$	[41]
Phytase	CC SDM	38	3	11	T_M	27.9	44	[44]
Isocitrate dehydrogenase	CC SDM	6	1	2	T_M	17.3	5	[45]
3-Isopropylmalate dehydrogenase	CC	12	2	1	T_M	3.9	12	[46]
β-Lactamase	SDM SDM	29	2	11	T_M	9.1	90 352	[91]

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 6 of 16
Bayer v. Corteva - IPR2023-01036

Table 1. Continued

Enzyme	Method	No. of mutation sites	No. of rounds	Improved property: Mutants	Improved property: Property	Improved property: Fold-improvement	Library size	Refs
Phenylcyclohexanone monooxygenase	SDM	Loop length	0	2	Specificity	E>200	3	[47]
2,4-Dinitrotoluene dioxygenase	SM	1	174	1	Specific activity	174	200	[92]
Phytase	SDM	24	1	24	Specific activity	1.4	1	[29]
Scouting								
Benzoformate decarboxylase	epPCR./ SDM	1	2	1	Specific activity	5	4200	[51]
Glutaryl acylase	SS SM	3	2	1	k_{cat}/K_M	2	$1*10^4$ $1*10^6$	[52]
Lipase	epPCR SM		3	6	Enantio-selectivity	46.4	4×10^4	[54]
β-Glucuronidase	i) Rec ii) SM	3	b) 2	3	k_{cat}/K_M Enantio-selectivity	200 $2.9*10^5$	a) $3.45*10^4$ b) $2*10^4$	[53]
Lipase B	i) comp ii) epPCR		2		Thermostability	7.5	$1*10^4$	[55]
Combinations								
Lipase	epPCR epPCR SDM structure		3	3	T_{50}	300	2000 2000	[7]
D-Hydantoinase	SDM SM	3		1 2	Specificity k_{cat}	200 10	 $1.6*10^4$	[56]
Penicillin G acylase	CC structure SDM	21	1	10	Thermostability	50% up to 2.5-fold	21	[8]
Leucine zipper	i) SM, structure ii) other	28			Affinity	7×10^5	8.3×10^4 6.2×10^4	[93]
Further restrictions to limit library size								
Cytochrome P450 BM	epPCR	1		 1–6 1–5	Proper folding Improved activity Specificity	61% 7 variants 4 variants	522 8160	[86]
Dihydrofolate reductase	SM+RCS	4	1	3	k_{cat}/K_M	6.2	5000	[15]
Nuclear receptor	SM epPCR	14	4 1	7	Ligand specificity	1×10^8	50×190 3.3×10^6	[20]
Nuclear receptor	SM+RCS		1	4	Ligand specificity	5000	$3.8*10^5$	[57]
Antibody	i).SM+RCS ii) Other iii) Other	57	3	21	Ligand affinity	870	a) 1×10^7 b) 6×10^5 2×10^5 c) 1×10^7	[58]

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 7 of 16
Bayer v. Corteva - IPR2023-01036

Table 1. Continued

Enzyme	Method	No. of mutation sites	No. of rounds	Improved property			Library size	Refs
				Mutants	Property	Fold-improvement		
Nuclear receptor	epPCR	4	4	4	Ligand selectivity	7000	1×10^6	[63]
Lipase	i) SM ii) SDM	10 4	1 1	1 3	Thermostability	a) 34 b) 145	a) 5×3000 b) 6	[61, 94]
Lipase	SM	15	5	9	Enantio-selectivity	25	2×10^4	[62]
Xylanase	SM/Rec	189 9	2	9	T_M	35 °C	189 × 384 4000	[95]
Subtilisin	SM	12	12	12	Stability	~ 7500	2400–3600	[5]
Computational approaches								
Modeling of sequence-activity relationship								
β-Lactamase	CPS	5×10^5	1	6	Specificity	1280	$2*10^5$	[14]
Lipase	SM	181	1	1	Enantio-selectivity	>1000	$2.1*10^4$	[68]
	CPS	174	1	1	Enantio-selectivity			
Substance P	Comp	12	3		Affinity	300	a) 36 b) 6 c) 3	
Ecotin	SM Comp	4	1		Inhibition	Similarly potent	7.5×10^6 1.4×10^5	[67]
Lipase	Comp	4	1	4	Enantio-selectivity	1.7	4	[80]
Active site replacement								
Glutathione transferace	i) Comp Rec ii) Rec		2		k_{cat}/K_M	65	1) 96 2) 1031	[81]
Sesquiterpene synthase γ-humulene synthase	Structure SM Comp SDM	19	2	4	Specificity	18	2500	[82]
Acetylcholinesterase	i) optimize length ii) 8synthetic genes iii) Comp	2	2	2	Surface display	3.5	a) 9 b) 8	[73]
Antibacterial peptide	i) Comp ii) Comp iii) 39 synthetic genes	15	3		Antibacterial	2	60	[74]

a) CC: consensus concept; Comp: computational modeling; CPS: computational pre-screening; epPCR: error-prone PCR; RCS: reduced codon set; Rec: recombination; SDM: site-directed mutagenesis; SM: saturation mutagenesis; SS: synthetic shuffling; T_{50}: half live; T_m: melting temperature

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 8 of 16
Bayer v. Corteva - IPR2023-01036

To our knowledge the presented subdivision of data-driven approaches is the first attempt to give a broad overview of such methods. We are conscious that many data-driven methods often fit into different categories, but to prevent redundant citations we divided them into just one specific subdivision.

1.2 Considerations when using data-driven engineering

Targeted protein evolution first must begin with an analysis of which portions of a protein are most likely to impact the property of interest. To evolve some properties, such as substrate specificity or enantioselectivity, mutations in and around the active site, particularly of residues implicated in binding and catalysis, are most likely to be successful [3–6]. Other properties, such as stability and specific activity, can be improved by mutations either near or far from the active site [3]. Techniques such as random mutagenesis are more likely to sample mutations in regions of the protein away from the active site because the active site constitutes only a very small fraction of whole protein. Therefore, focused libraries can be especially effective for certain evolutionary targets. However, it is also possible to use a crystal structure find important residues for properties that are not connected with the active site (e.g. [7, 8]).

Another important consideration when designing libraries is that many mutations are also context dependent, meaning they show synergism with or antagonism to other sites in the protein [9–12]. To detect synergistic effects, both mutations must be present at the same time. From a probabilistic standpoint, this is unlikely when using random mutagenesis, particularly since error-prone PCR (epPCR) can only access three to seven amino acid via a single base exchange, meaning only a small fraction of sequence space can be explored via mutagenesis [13]. Focused libraries can access all diversity at a single position, and often results in variants that could not be obtained easily via point mutations [14, 15]. Overall, simultaneous randomization at targeted positions has a greater chance of detecting synergistic effect that could not have been predicted by mutating the positions individually. Active-site mutations are frequently coupled in this manner [12]. On the other hand, some properties (particularly those such as stability, which have many routes to success) are often independent [16]. In one example of computational engineering, only one interaction term was necessary to describe the fitness landscape of 14 different mutations [17].

Finally, in many cases, second-shell mutations have a large impact on activity [18]. These will likely be missed when using focused libraries; however, a combination of active-site randomization with combinatorial mutagenesis to find second-shell mutations is often advantageous [19, 20].

2 Choosing residues or areas of the protein to target

The most important part of creating a focused library is determining a method with which to select the residues to change. If the incorrect residues are targeted, the library will not contain a protein with the desired effect, and the experimental time in constructing and screening the library will be wasted. There are a variety of methods to choose important residues for a given problem: crystal-structure-based studies such as molecular modeling and molecular dynamics; sequence alignment or homology-based methods; and "scouting" experiments. These may be used alone or in concert to predict which residues hold the most promise to yield a library with the desired properties.

2.1 Structure-based studies

In an analysis of data from published single mutants, Morley and Kazlauskas [3] showed that improvements in enantioselectivity and substrate specificity were more closely linked to mutations close to the active site, while mutations that improved overall activity level and thermal stability were equally likely to be near the active site or in more distant regions. Thus, for cases in which a crystal structure has been obtained and for which a change in substrate specificity is the goal, one logical way to determine which residues to target is to utilize 3-D structures to determine which residues are likely to be in contact with the substrate of interest. Structure-based methods include docking of substrates into a crystal structure or homology model and molecular dynamics simulations of the protein behavior during catalytic cycles.

Many successes with targeted randomization based on crystal structure data have been shown (reviewed in Chica *et al.* [19]) in diverse examples such as oxidations [21], antibiotic syntheses [22], hydrolytic reactions [4, 23], aldol condensations [24], degradation of chemical warfare agents [9], and DNA recombinases [25]. In one example, the activity of *N*-acetylneuraminic acid lyase towards 6-dipropylcarboxamides was increased using structure-guided site-saturation mutagenesis. Here, the crystal structures from the *E. coli* and *H. influenzae* enzymes with multiple substrate analogs were used to identify residues in contact with the polar group of the native substrate that would be more hydrophobic in the desired substrate. Each of the three residues in contact was then subjected to single site-saturation mutagenesis and one 96-well plate of each library was screened to find variants with the desired activity. The best single mutant E192N had an increased k_{cat} and a decreased K_m for the desired substrate as well as a decreased k_{cat} and an increased K_m for the native substrate, a 690-fold shift in substrate specificity overall [24]. When a crystal structure containing a bound substrate is not available, docking the substrate or sub-

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 9 of 16
Bayer v. Corteva - IPR2023-01036

strates of interest into the active-site cavity can also be used to determine which residues are in contact with the substrate upon binding [15].

Another strategy is to exchange structural elements, domains, or residues between one or more proteins. This concept has been applied to both antibodies and enzymes. Due to their inherent flexibility, loops are particularly suited for exchange. The concept of loop grafting is widespread in antibody engineering whereby antibodies can be raised in a non-human species and the hypervariable loops exchanged into a human antibody structure to humanize the therapy [26]. Antigen-recognition domains can also be grafted into non-antibody scaffolds [27]. One early report of loop grafting in enzymes relied on a rational strategy whereby a computational search of the protein data bank was used to identify candidate loops that would fit the peptide environment and bind the desired substrate [28]. In another report, the entire active site of a phytase was exchanged, resulting in more favorable catalytic properties being transferred into a more thermostable scaffold. Once the active site residues of the thermostable scaffold phytase in contact with the substrate were identified, all of those that were divergent, plus all residues immediately adjacent to the active site that were different, were simultaneously replaced with the corresponding residues from the more catalytically active phytase. The resulting variant retained approximately half the thermostability of the scaffold [melting temperature (T_m) 7°C higher than the less thermostable parent, but 8°C lower than the thermostable scaffold] and had a specific activity increase of nearly 20 U/mg over the thermostable (less active) parent [29].

Combinatorial strategies, also known as domain swapping or shuffling [30–33], can also be employed and a surprisingly large fraction of domain-swapped variants form proper secondary structure [34]. In addition, computational techniques have been developed to determine the best placement of crossover points to minimize disruption of the product proteins [35, 36]. Using the SCHEMA algorithm [35] to virtually screen 5000 randomly generated libraries of 6561 chimeras, 14 crossover sites that retained functionality in cytochrome P450 were identified. Further *in silico* screening of all possible seven crossover libraries identified one library with high diversity and evenly distributed crossover points. The library was generated using an experimental technique to create chimeras with crossovers directed to certain areas of the protein [32], yielding a family of cytochrome P450s, approximately half of which were properly folded, 72% of these had activity [37]. While this example did not specifically yield improved proteins, the technique is very valuable since creating libraries that have a higher percentage of folded members will increase the chance of obtaining a mutant with the desired characteristics.

2.2 Homology-based methods

Several strategies have been developed for choosing mutations based on homologous sequences, either from a multiple sequence alignment or from a phylogenetic tree (reviewed in Minshull *et. al.* [38]). In these methods, orthologs (speciated sequences from a common ancestor, usually with the same function) can be used to identify residues that can tolerate mutation without radically altering function. Therefore, positions that are conserved among the orthologous genes provide potential substitutions to improve properties such as thermostabilization. On the other hand, paralogs (proteins diverged from a common ancestor, but evolved to new function) are useful for identifying the residues responsible for functional variation among proteins and can give clues for the evolution of new functions.

Incorporation of substitutions from homologs is the underlying premise of DNA shuffling experiments. In addition, specific substitutions can be incorporated into a gene using synthetic oligonucleotides that code for the changes identified from an alignment [39, 40]. Representing diversity from more parental genes can overcome evolutionary plateaus reached after a few rounds of DNA shuffling [41], and analysis of successful variants will often lead to the discovery that substitutions could have been predicted from sequence analyses [42, 43]. In the evolution of glyphosphate tolerance via the glyphosphate *N*-acetyltransferase enzyme, 11 rounds of DNA shuffling beginning with three parental genes improved the k_{cat}/K_m of the enzyme for glyphosphate more than 9500-fold. However, in intermediate rounds, plateaus in activity improvement were reached. To overcome these plateaus, synthetic oligonucleotides representing diversity from two additional homologs at 27 sites were added to increase diversity in the 5th and 8th rounds. Although a very large number of colonies were screened (5000 per round), the strategy of adding synthetic diversity from homologs resulted in a smaller library size than shuffling of seven parental genes would have created [41].

The strict consensus approach utilizes a multiple sequence alignment of highly homologous enzymes to find and directly synthesize a variant with improved stability, although it is beneficial to analyze the effects of individual mutations so that the destabilizing ones can be removed [44]. If a residue is present significantly more often in a sequence alignment, the consensus method holds that substitution of that residue into a protein of interest should lead to increased stability. In contrast, random mutagenesis predominantly introduces neutral or destabilizing mutations [96]. Variations use phylogenetic distances to retrace the sequence of a thermostable common ancestor [45, 46]. The underlying concept is the same; positions are detected where a substitution to a residue predominating in homologs should confer greater stability. Using a phylogenetic analysis of 18 sequences, Watanabe

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 10 of 16
Bayer v. Corteva - IPR2023-01036

et al. [46] designed 12 variants of 3-isopropylmalate dehydrogenase that had a single amino acid substitution that came from the putative common ancestor of the phylogenetic tree. Of these 12 single mutants, 50% had increased thermostability compared with the already very thermostable parental sequence. The best mutant had a threefold increase in half-life at 86°C and a nearly 4°C increase in T_m compared with wild type.

Structural overlays of homologs can also be informative in choosing mutations. Overlaying two or more related proteins can point out areas of structure that are problematic for the adoption of new function. In one case, truncation of a loop in phenylacetone monooxygenase based on a homology model from a related enzyme was shown to enhance both substrate acceptance (decreased selectivity for natural substrate over others) and enantioselectivity (a more than 100-fold increase in E value), when only three variants were created [47].

Structures of homologs can also be used to analyze substrate binding, even if all proteins of interest were not crystallized with the substrate. Once substrate binding residues are identified by this "evolutionary tracing", these can be chosen for randomization to create libraries towards new functions [48, 49]. Functionally important residues can also be detected statistically, particularly to recognize proteins that are likely to be part of the same subfamily, and therefore have similar functions [50].

2.3 "Scouting" experiments

Clues as to which residues are most important can also be gleaned from prior experimental data, or from small "scouting" experiments. For instance, sites identified by prior combinatorial experiments such as epPCR can be subjected to site-saturation mutagenesis, alone or in concert, to find the optimum combination for the property of interest [19, 51–54]. This strategy was employed to change the substrate specificity of glutaryl acylase toward adipyl compounds. One round of exploratory mutagenesis followed by selection of a library of 10 000 colonies identified three residues that affected the ability of the enzyme to accept adipylated substrates. These three residues were subjected to site-saturation mutagenesis and selection. The best mutant identified by this strategy had a threefold increase in catalytic efficiency [52].

When epPCR, rather than randomization strategies are chosen, the optimal mutation load can be determined from small libraries with varying mutation load. Each library is assayed to determine the number of mutations that lead to the maximum rate of evolution of the property of interest. Using this strategy, the thermal stability of *Candida antarctica* lipase B was increased 7.5-fold by screening a library of only 10 000 colonies [55].

2.4 Combinations

The above strategies have also been used in combination to narrow the residues of interest. The combination of sequence alignment and structural data has been used to evolve a D-hydantinoase with altered substrate specificity [56] and single mutants of penicillin G acylase with improved thermostability [8]. In the latter case, structural data were used to narrow the field of possible mutations suggested by the consensus method.

In a slightly different approach, the three-dimensional structure can be used to exclude potentially disrupting mutations collected by epPCR. In one example, the half-life of denaturation of a lipase at 55°C was increased 300-fold using known rules on compatibility of amino acid residues to eliminate unfavorable mutations [7].

3 Further restrictions to limit library size

After the identification of positions through one of the above methods, the decision of which amino acids to allow at each site and how many substitutions to allow simultaneously must be made. The simplest choice would be to allow each of the 20 amino acids at each site at once. However, in cases where multiple sites are to be changed, the library size quickly becomes unmanageable if all possibilities are allowed. To keep library sizes within a screenable range, only a few sites are changed at once, or a reduced number of amino acids are allowed at each position.

Restricting the diversity allowed at chosen positions can be of interest particularly when the desired property of the amino acid is known (*e.g.*, small amino acids to create space in a binding pocket, large amino acids to narrow the binding pocket, hydrophobic contacts to bind a particular substrate) [57]. Even when the desired property is unpredictable, a small library can be constructed using only representative amino acids of each type to identify which amino acid types are of interest for a given position [15, 58]. Libraries with reduced choice of amino acids at all of the sites of interest have a better probability of detecting interactions between sites than partial libraries at a few sites with exhaustive codon usage. The process of designing oligonucleotides with specific codons at chosen sites has been automated [59]. The affinity of an anti-TNF-α antibody was increased from a starting K_d value of 1 nM to the picomolar range (870-fold) using a procedure called "look-through mutagenesis". Oligonucleotides encoding 9 of the 20 possible amino acid substitutions were used to create libraries in one, two, or three positions. Following selection, positive substitutions were mapped and combined to yield the greatly improved antibody [58].

Partial libraries, where only a few sites are changed at the same time can also reduce the sequence search

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 11 of 16
Bayer v. Corteva - IPR2023-01036

space. Although synergistic effects between sites, leading to path-dependent evolution, cannot be ruled out [60], many successful amino acid substitutions are nonobvious and would defy attempts at prediction. Thus, exhaustively mutating each site [20] or small combinations of sites [61] can be a powerful strategy. Additional improvements can be obtained by iterating this process [62] or by additional rounds of epPCR to include further improvements caused by distant mutations [20]. One such method, iterative combinatorial active site mutagenesis (CASTing), was used to increase the enantioselectivity of the epoxide hydrolase from *A. niger* by screening partial libraries randomized simultaneously in two to three positions near the active site. After screening a small (non-exhaustive) sample of the library, beneficial mutations were locked and other sites randomized to further improve the protein. Following CASTing in nine total sites, the E value of the enzyme was increased from 4.6 to 115 with only 20 000 total clones subjected to screening [62].

The selected positions can also be changed in a stepwise fashion by first evolving the protein of interest towards one or more intermediate substrates with chemical properties that are between the natural substrate and the desired one. Gradually changing the functionality of the molecule is particularly suited to evolution of proteins that bind a certain molecule (*e.g.* receptors). Chen and Zhao [63] evolved a variant of the human estrogen receptor that could bind corticosterone, by first evolving variants that could bind testosterone and progesterone, two hormones with small structural changes between the natural ligand estradiol and the desired ligand. In four rounds of mutagenesis, a total of 10^6 variants were screened to shift the ligand selectivity 7000-fold. A similar approach was also recently reported with extends this concept to enzyme substrate specificity [64].

4 Computational approaches

Computing power can be substituted for experimental effort and is used to design more efficient libraries. Computational tools allow the prescreening of mutations to eliminate unfruitful ones, modeling of structure-function relationships, placement of residues into a protein scaffold, and optimization of crossover points [65, 66].

4.1 *In silico* library design

Various computational methodologies have been developed for the prescreening of potential mutations. Most often, the goal is to eliminate proteins that are unlikely to have desired properties, rather than to choose which variants are likely to be successful. In this way, the immense possibilities are narrowed down to a library size that can be handled experimentally. Algorithms have been designed to evaluate the effect of the substitution on the protein fold [14], to perform *in silico* site-saturation mutagenesis [67] or alanine scanning [68] to identify important residues. Funke *et al.* [68] used computation alanine scanning, followed by experimental corroboration to identify functionally important residues in the lipase from *Bacillus subtilis*. Each of the 175 amino acid side chains, with the exception of the catalytic triad, that were resolved in the crystal structure were computationally mutated to alanine individually, and the effect on the activity was computed. The same residue identified as a hot spot by computational scanning was identified in experimental mutagenesis as having a large effect on the enantioselectivity of the reaction.

Recombination of genes can often result in non-functional proteins due to incompatibility of residues in the 3-D structure of the chimeric protein. However, libraries with more folded members have a better chance of yielding variants with the desired characteristics. Thus, the optimal placement of crossover points to preserve proper folding [36, 69], or algorithms aimed at identifying mutations in the parents that alleviate problems in the recombined library [70], or which identify pairs of interacting residues that must be preserved [71] can be used to design libraries with fewer nonfunctional variants due to misfolding. For instance, the FamClash algorithm uses a combination of crystal data and sequence alignments to predict the changes in volume and hydrophobicity for each of the pairs of amino acids in a theoretical chimera. Number of clashing interactions in that theoretical chimera is noted, and the crossover point moved along the proteins until all possible chimeras have been evaluated. A plot of the number of predicted clashes versus the measured activity of the chimera for a library of dihydrofolate reductases created by nonhomologous recombination shows that when the algorithm predicts many clashes, the specific activity is usually decreased. Algorithms such as FamClash can be used to predict which crossover points to avoid [36].

4.2 Modeling of sequence-activity relationships

Structure-activity relationships, widespread in the field of chemistry, are used to correlate the physicochemical or other properties of molecules to their biological function via mathematical modeling [72]. The analogy for proteins and peptides is the sequence-activity relationship, where the amino acid sequence is related to the desired outcome. Data from sequences with known function or function level can be modeled to determine sequence-activity relationships for a given protein. From such models, tolerated and nontolerated substitutions can be evaluated, residues important for function can be identified, and optimum sequences predicted. In many cases, the number of input sequences required to develop the model is rather small. For example, from a dataset of 45 mutants of prolyl endopeptidase, Aita *et al.* [17] were able to predict the ac-

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 12 of 16
Bayer v. Corteva - IPR2023-01036

tivities of all possible combinations of a 14-mutant protein.

The first models of sequence-activity relationships were restricted to small peptides because of the impossibility of substituting 20 amino acids at every position in a full-size protein (since this exceeds the number of atoms in the universe) [73–75]. Simple statistical models also work for nucleic acid molecules, such as promoters, where only four substitutions at each position are possible [76]. Subsequent work by Fox [77] showed that the concept can be extended to the combinatorial evolution of full-sized proteins, particularly when a limited set of the total protein residues is altered. Machine-learning techniques, trained on small sets of experimental data, can be used to predict important positions in proteins and to design new libraries aimed toward the optimization of protein properties [78–80]. The advantage of machine-learning techniques is their ability to detect hidden patterns relating the input (sequence) and output (function) that can be used for predictive purposes.

Another approach is to use the substrate "fingerprint" of a variant or set of variants to develop rules for substrate specificity. These rules can be used to develop variants with divergent substrate profiles [81, 82]. Using a homology structure of γ-humulene synthase, 16 active site residues were identified, four of which had a large impact on the product distribution produced by the enzyme. Assuming that each residue was independent, an algorithm was developed that mapped product profile to the particular mutation, such that new specificities could be evolved by new combinations of mutations. Fewer than 2500 mutants were screened to develop seven enzymes with seven different selectivities [82].

4.3 Active-site placement

New enzymes can be created by the introduction of the proper active site into an existing scaffold. This can be done by computationally designing a binding shell plus catalytic residues in the correct orientation into a protein [83] or by grafting elements of one enzyme into another [84]. In both cases, activity can be further tuned by the introduction of random point mutations. Although neither case resulted in a new enzyme with activity as high as a wild-type enzyme, both designed enzymes were biologically active with rate accelerations of 10^5–10^6 in the example of computational redesign and 10^2 in the case of loop grafting. In both cases, only a handful of mutants were created.

5 Future trends

Until recently, rational design or directed evolution has been predominantly used alone. Future research should aim to combine these proven techniques to gain even greater benefit. For example, a mutant, computationally stabilized by core redesign was explored via randomization at the designed positions to find even better amino acid exchanges for these positions. Two variants were identified with even further increased stability, demonstrating that the computational and combinatorial methods can be used complementarily to pinpoint residues [85].

Additionally, the evolution of properties is often treated as a 1-D problem, while in reality the optimization of one property often comes at the expense of another unless controls in the screening process are used to lock all other properties above a minimum value. Therefore, the optimal solution is often a higher dimensional one. For example, the link between stability and evolvability, has been established [86, 87]. Since mutations accumulated in the course of evolution of new function are often destabilizing, it is worth the time investment to first create variants of the wild-type protein with enhanced stability and then to evolve for change of function. Additionally, it has been shown that extremely high mutation rates in epPCR experiments can lead to the emergence of new function [55, 88]. These libraries show a higher success rate when the starting protein has sufficient stability.

Finally, one area of future development should be the creation of more user-friendly tools for the mining of information from various databases. Tool development should keep in mind the needs of the wide variety of researchers working in the field from evolutionary biologists and protein engineers, to chemists, and computer scientists. Tools that are understandable and useable by such disparate groups would make an important contribution to future growth in the field.

In summary, data-driven protein engineering limits library size and screening effort, while still creating proteins with improved properties. Methods to detect which residues are important in a given template include structure, homology, experimental evidence, and computational algorithms. After identification, important residues can be mutated according to various schemes. The overall goal is to use available information to perform smarter protein engineering.

6 References

[1] Parikh, M., Matsumura, I., Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of beta-fucosidase from beta-galactosidase. *J. Mol. Biol.* 2005, *352*, 621–628.

[2] Zhang, J., Dawes, G., Stemmer, W., Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci. USA* 1997, *94*, 4504–4509.

[3] Morley, K., Kazlauskas, R., Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* 2005, *23*, 231–237.

[4] Park, S., Morley, K., Horsman, G., Holmquist, M. *et al.*, Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations. *Chem. Biol.* 2005, *12*, 45–54.

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 13 of 16
Bayer v. Corteva - IPR2023-01036

[5] Strausberg, S., Ruan, B., Fisher, K., Alexander, P. *et al.*, Directed coevolution of stability and catalytic acitivity in calcium-free subtilisin. *Biochemistry* 2005, *44*, 3272–3279.
[6] Dalby, P., Optimising enzyme function by directed evolution. *Curr. Opin. Struct. Biol.* 2003, *13*, 500–505.
[7] Acharya, P., Rajakumara, E., Sankaranarayanan, R., Rao, N., Structural basis of selection and thermostability of laboratory evolved *Bacillus subtilis* lipase. *J. Mol. Biol.* 2004, *341*, 1271–1281.
[8] Polizzi, K., Chaparro-Riggers, J., Vazquez-Figueroa, E., Bommarius, A., Structure-guided consensus approach to create a more thermostable penicillin G acylase. *Biotechnol. J.* 2006, *1*, 531–536.
[9] Hill, C., Li, W., Thoden, J., Holden, H. *et al.*, Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active site. *J. Am. Chem. Soc.* 2003, *125*, 8990–8991.
[10] Matsumura, I., Ellington, A., *In vitro* evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* 2001, *305*, 331–339.
[11] Midelfort, K., Wittrup, K., Context-dependent mutations predominate in an engineered high-affinity single chain antibody fragment. *Protein Sci.* 2006, *15*, 324–334.
[12] Mildvan, A., Inverse thinking about double mutants of enzymes. *Biochemistry* 2004, *43*, 14517–14520.
[13] Wong, T., Roccatano, D., Zacharias, M., Schwanenberg, U., A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol.* 2006, *355*, 858–871.
[14] Hayes, R., Bentzien, J., Ary, M., Hwang, M. *et al.*, Combining computational and experimental screening for rapid optimization of protein properties. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 15926–15931.
[15] Schmitzer, A., Lepine, F., Pelletier, J., Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations. *Protein Eng. Des. Sel.* 2004, *17*, 809–819.
[16] Eijsink, V. G., Bjork, A., Gaseidnes, S., Sirevag, R. *et al.*, Rational engineering of enzyme stability. *J. Biotechnol.* 2004, *113*, 105–120.
[17] Aita, T., Hamamatsu, N., Nomiya, Y., Uchiyama, H. *et al.*, Surveying a local fitness landscape of a protein. *Biopolymers* 2002, *64*, 95–105.
[18] Tomatis, P., Rasia, R., Segovia, L., Vila, A., Mimicking natural evolution in metallo-β-lactamases through second-shell ligand mutations. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 13761–13766.
[19] Chica, R., Doucet, N., Pelletier, J., Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* 2005, *16*, 378–384.
[20] Chockalingham, K., Chen, Z., Katzenellenbogen, J., Zhao, H., Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 5691–5696.
[21] El Hawrani, A., Sessions, R., Moreton, K., Holbrook, J., Guided evolution of enzymes with new substrate specificities. *J. Mol. Biol.* 1996, *264*, 97–110.
[22] Gabor, E., Janssen, D., Increasing the synthetic performance of penicillin acylase PAS2 by structure-inspired semi-random mutagenesis. *Protein Eng. Des. Sel.* 2004, *17*, 571–579.
[23] Rui, L., Cao, L., Chen, W., Reardon, K. *et al.*, Active site engineering of the epoxide hydrolase from *Agrobacterium radiobacter* AD1 to enhance aerobic mineralization of cis-1,2-dichloroethylene in cells expressing an evolved toluene *ortho*-monooxygenase. *J. Biol. Chem.* 2004, *45*, 46810–46817.
[24] Williams, G., Woodhall, T., Nelson, A., Berry, A., Structure-guided saturation mutagenesis of *N*-acetylneuraminic acid lyase for the synthesis of sialic acid mimetics. *Protein Eng. Des. Sel.* 2005, *18*, 239–246.
[25] Santoro, S., Schultz, P., Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 4185–4190.
[26] Reichmann, L., Clark, M., Waldmann, H., Winter, G., Reshaping human antibodies for therapy. *Nature* 1988, *332*, 323–327.
[27] Nicaise, M., Valerio-Lepiniec, M., Minard, P., Desmadril, M., Affinity transfer by CDR grafting on a nonimmunoglobulin scaffold. *Protein Sci.* 2004, *13*, 1882–1891.
[28] Wilks, H., Moreton, K., Halsall, D., Hart, K. *et al.*, Design of a specific phenyllactate dehydrogenase by peptide loop exchange on the *Bacillus stearothermophilus* lactate dehydrogenase framework. *Biochemistry* 1992, *31*, 7802–7806.
[29] Lehmann, M., Lopez-Ulibarri, R., Loch, C., Viarouge, C. *et al.*, Exchanging the active site between phytases for altering the functional properties of the enzyme. *Protein Sci.* 2000, *9*, 1866–1872.
[30] Gibbs, M., Nevalainen, K., Bergquist, P., Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene* 2001, *271*, 13–20.
[31] O'Maille, P., Bakhtina, M., Tsai, M., Structure-based combinatorial protein engineering (SCOPE). *J. Mol. Biol.* 2002, *321*, 677–691.
[32] Hiraga, K., Arnold, F., General method for sequence-independent site-directed chimeragenesis. *J. Mol. Biol.* 2003, *330*, 287–296.
[33] Saftalov, L., Smith, P., Friedman, A., Bailey-Kellogg, C., Site-directed combinatorial construction of chimaeric genes: general method for optimizing assembly of gene fragments. *Proteins* 2006, *64*, 629–642.
[34] Tsuji, T., Yoshida, K., Satoh, A., Kohno, T. *et al.*, Foldability of barnase mutants obtained by permutation of modules or secondary structure units. *J. Mol. Biol.* 1999, *268*, 1581–1596.
[35] Voigt, C., Mayo, S., Arnold, F., Wang, Z., Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. USA* 2001, *98*, 3778–3783.
[36] Saraf, M., Horswill, A., Benkovic, S., Maranas, C., FamClash: A method for ranking the activity of engineered enzymes. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 4142–4147.
[37] Otey, C., Landwehr, M., Endelman, J., Hiraga, K. *et al.*, Structure-guided recombination creates an artificial family of cytochromes P450. *PLOS Biol.* 2006, *4*, e112–e121.
[38] Minshull, J., Ness, J., Gustafsson, C., Govinarajan, S., Predicting enzyme function from protein sequence. *Curr. Opin. Chem. Biol.* 2005, *9*, 202–209.
[39] Ness, J., Kim, S., Gottman, A., Pak, R. *et al.*, Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* 2002, *20*, 1251–1255.
[40] Coco, W., Encell, L., Levinson, W., Crist, M. *et al.*, Growth factor engineering by degenerate homoduplex gene family recombination. *Nat. Biotechnol.* 2002, *20*, 1246–1250.
[41] Castle, L., Siehl, D., Gorton, R., Patten, P. *et al.*, Discovery and directed evolution of a glyphosphate tolerance gene. *Science* 2004, *304*, 1151–1154.
[42] Cochran, J., Kim, Y., Lippow, S., Rao, B. *et al.*, Improved mutants from directed evolution are biased to orthologous substitutions. *Protein Eng. Des. Sel.* 2006, *19*, 245–253.
[43] Lassila, J., Keeffe, J., Oelschlaeger, P., Mayo, S., Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity. *Protein Eng. Des. Sel* 2005, *18*, 161–163.
[44] Lehmann, M., Loch, C., Middendorf, A., Studer, D. *et al.*, The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* 2002, *15*, 403–411.
[45] Iwabata, H., Watanabe, K., Ohkuri, T., Yokobori, S. *et al.*, Thermostability of ancestral mutants of *Caldococcus noboribetus* isocitrate dehydrogenase. *FEMS Microbiol. Lett.* 2005, *243*, 393–398.
[46] Watanabe, K., Ohkuri, T., Yokobori, S., Yamagishi, A., Designing thermostable proteins: Ancestral mutants of 3-isopropylmalate dehydrogenase designed using a phylogenetic tree. *J. Mol. Biol.* 2006, *355*, 664–674.
[47] Bocola, M., Schulz, F., Leca, F., Vogel, A. *et al.*, Converting phenylacetone monooxygenase into phenylcylohexanone monooxygenase

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 14 of 16
Bayer v. Corteva - IPR2023-01036

by rational design: Towards practical Baeyer-Villiger monooxygenases. *Adv. Synth. Catal.* 2005, *347*, 979–986.

[48] Lichtarge, O., Bourne, H., Cohen, F., An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 1996, *257*, 342–358.

[49] Lichtarge, O., Yao, H., Kristensen, D., Madabushi, S. *et al.*, Accurate and scalable identification of functional sites by evolutionary tracing. *J. Struct. Funct. Genomics* 2003, *4*, 159–166.

[50] del Sol Mesa, A., Pazos, F., Valencia, A., Automatic methods for predicting functionally important residues. *J. Mol. Biol.* 2003, *326*, 1289–1302.

[51] Lingen, B., Grotzinger, J., Kolter, D., Kula, M. *et al.*, Improving the carboligase activity of benzoylformate decarboxylase from *Pseudomonas putida* by a combination of directed evolution and site-directed mutagenesis. *Protein Eng.* 2002, *15*, 585–593.

[52] Sio, C., Riemens, A., van der Laan, J., Verhaert, R. *et al.*, Directed evolution of a glutaryl acylase into an adipyl acylase. *Eur. J. Biochem.* 2002, *269*, 4495–4504.

[53] Geddie, M., Matsumura, I., Rapid evolution of beta-glucuronidase specificity by saturation mutagenesis of an active site loop. *J. Biol. Chem.* 2004, *279*, 26462–26468.

[54] Reetz, M., Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 5716–5722.

[55] Chordorge, M., Fourage, L., Ullmann, C., Duvivier, V. *et al.*, Rational strategies for directed evolution of biocatalysts- Application to *Candida antarctica* lipase B (CALB). *Adv. Synth. Catal.* 2005, *347*, 1022–1026.

[56] Cheon, Y., Park, H., Kim, J., Kim, Y. *et al.*, Manipulation of the active site loops of D-hydantoinase, a $(\beta/\alpha)_8$-barrel protein, for modulation of the substrate specificity. *Biochemistry* 2004, *43*, 7413–7420.

[57] Schwimmer, L., Rohatgi, P., Azizi, B., Seley, K. *et al.*, Creation and discovery of ligand-receptor pairs for transcriptional control with small molecules. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 14707–14712.

[58] Rajpal, A., Beyaz, N., Haber, L., Cappuccilli, G. *et al.*, A general method for greatly improving the affinity of antibodies by using combinatorial libraries. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 8466–8471.

[59] Mena, M., Daugherty, P., Automated design of degenerate codon libraries. *Protein Eng. Des. Sel.* 2005, *18*, 559–561.

[60] Weinreich, D., Delaney, N., DePristo, M., Hartl, D., Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 2006, *312*, 111–114.

[61] Reetz, M., Bocola, M., Carballeira, J., Zha, D. *et al.*, Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed. Engl.* 2005, *44*, 4192–4196.

[62] Reetz, M., Wang, L., Bocola, M., Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem. Int. Ed. Engl.* 2006, *45*, 1236–1241.

[63] Chen, Z., Zhao, H., Rapid creation of a novel protein function by *in vitro* coevolution. *J. Mol. Biol.* 2005, *348*, 1273–1282.

[64] Vedha-Peters, K., Gunawardana, M., Rozzell, J. D., Novick, S. J., Creation of a broad-range and highly stereoselective D-amino acid dehydrogenase for the one-step synthesis of D-amino acids. *J. Am. Chem. Soc.* 2006, *128*, 10923–10929.

[65] Gustafsson, C., Govinarajan, S., Minshull, J., Putting engineering back into protein engineering: bioinformatic approaches to catalyst design. *Curr. Opin. Biotechnol.* 2003, *14*, 366–370.

[66] Park, S., Kono, H., Want, W., Boder, E. *et al.*, Progress in the development and application of computational methods for probabilistic protein design. *Comp. Chem. Eng.* 2005, *29*, 407–421.

[67] Laboissiere, M., Young, M., Pinho, R., Todd, S. *et al.*, Computer-assisted mutagenesis of ecotin to engineer its secondary binding site for urokinase inhibition. *J. Biol. Chem.* 2002, *277*, 26623–26631.

[68] Funke, S., Otte, N., Eggert, T., Bocola, M. *et al.*, Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution. *Protein Eng. Des. Sel.* 2005, *18*, 509–514.

[69] Voigt, C., Martinez, C., Wang, Z., Mayo, S. *et al.*, Protein building blocks preserved by recombination. *Nat. Struct. Biol.* 2002, *9*, 553–558.

[70] Saraf, M., Moore, G., Goodey, N., Cao, V. *et al.*, IPRO: An iterative computation protein library redesign and optimization procedure. *Biophys. J.* 2006, *90*, 4167–4180.

[71] Dubey, A., Realff, M., Lee, J., Bommarius, A., Identifying the interacting positions of a protein using Boolean learning and support vector machines. *Comput. Biol. Chem.* 2006, *30*, 268–279.

[72] Klebe, G., Abraham, U., Mietzner, T., Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 1994, *37*, 4130–4146.

[73] Bucht, G., Wikstrom, P., Hjalmarsson, K., Optimising the signal peptide for glycosyl phosphatidylinositol modification of human acetylcholinesterase using mutational analysis and peptide-quantitative structure-activity relationships. *Biochim. Biophys. Acta* 1999, *1431*, 471–482.

[74] Mee, R., Auton, T., Morgan, P., Design of active analogues of a 15-residue peptide using D-optimal design, QSAR and a combinatorial search algorithm. *J. Pept. Res.* 1997, *49*, 89–102.

[75] Norvinder, U., Rivera, C., Unden, A., A quantitative structure-activity relationship study of some substance P-related peptides. A multivariate approach using PLS and variable selection. *J. Pept. Res.* 1997, *49*, 155–162.

[76] Jensen, K., Alper, H., Fischer, C., Stephanopoulos, G., Identifying functionally important mutations fro phenotypically diverse sequence data. *Appl. Environ. Microbiol.* 2006, *72*, 3696–3701.

[77] Fox, R., Roy, A., Govinarajan, S., Minshull, J. *et al.*, Optimizing the search algorithm for protein engineering by directed evolution. *Protein Eng.* 2003, *16*, 589–597.

[78] Dubey, A., Realff, M., Lee, J., Bommarius, A., Support vector machines for learning to identify the critical positions of a protein. *J. Theor. Biol.* 2005, *234*, 351–361.

[79] Fox, R., Directed molecular evolution by machine learning and the influence of nonlinear interactions. *J. Theor. Biol.* 2005, *234*, 187–199.

[80] Kato, R., Nakano, H., Konishi, H., Kato, K. *et al.*, Novel strategy for protein exploration: high-throughput screening assisted with fuzzy neural network. *J. Mol. Biol.* 2005, *351*, 683–692.

[81] Larsson, A., Emren, L., Bardsley, W., Mannervik, B., Directed enzyme evolution guided by multidimensional analysis of substrate-activity space. *Protein Eng. Des. Sel.* 2004, *17*, 49–55.

[82] Yoshikuni, Y., Ferrin, T., Keasling, J., Designed divergent evolution of enzyme function. *Nature* 2006, *440*, 1078–1082.

[83] Dwyer, M., Looger, L., Hellinga, H., Computational design of a biologically active enzyme. *Science* 2004, *304*, 1967–1971.

[84] Park, H., Nam, S., Lee, J., Yoon, C. *et al.*, Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 2006, *311*, 535–538.

[85] Wunderlich, M., Martin, A., Staab, C., Schmid, F., Evolutionary protein stabilization in comparison with computational design. *J. Mol. Biol.* 2005, *351*, 1160–1168.

[86] Bloom, J., Labthavikul, S., Otey, C., Arnold, F., Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA* 2006, *103*, 5869–5874.

[87] Bloom, J., Wilke, C., Arnold, F., Adami, C., Stability and the evolvability of function in a model protein. *Biophys. J.* 2004, *86*, 2758–2764.

[88] Drummond, D., Iverson, B., Georgiou, G., Arnold, F., Why high error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J. Mol. Biol.* 2005, *350*, 806–816.

[89] Tao, Y., Fishman, A., Bentley, W., Wood, T., Altering toluene 4-monooxygenase by active-site engineering for the synthesis of 3-

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 15 of 16
Bayer v. Corteva - IPR2023-01036

methoxycatechol, methoxyhydroquinone, and methylhydroquinone. *J. Bacteriol.* 2004, *186*, 4705–4713.
[90] Horsman, G., Liu, A., Henke, E., Bornscheuer, U. *et al.*, Mutations in distant residues moderately increase the enantioselectivity of *Pseudomonas fluorescens* esterase towards methyl-3-bromo-2-methylpropanoate and ethyl-3-phenylbutyrate. *Chemistry* 2003, *9*, 1933–1939.
[91] Amin, N., Liu, A., Ramer, S., Aehle, W. *et al.*, Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.* 2004, *17*, 787–793.
[92] Keenan, B., Leungsakul, T., Smets, B., Wood, T., Saturation mutagenesis of *Burkholderia cepacia* R34 2,4-dinitrotoluene dioxygenase at DntAc valine 350 for synthesizing nitrohydroquinone, methylhydroquinone, and methoxyhydroquinone. *Appl. Environ. Microbiol.* 2004, *70*, 3222–3231.
[93] Mason, J., Schmitz, M., Muller, K., Arndt, K., Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc. Natl. Acad. Sci.USA* 2006, *103*, 8989–8994.
[94] Reetz, M., Carballeira, J., Peyralans, J., Hobenreich, H. *et al.*, Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chemistry* 2006, *12*, 3061–3068.
[95] Palackal, N., Brennan, Y., Callen, W., Dupree, P. *et al.*, An evolutionary route to xylanase process fitness. *Protein Sci.* 2004, *13*, 494–503.
[96] Guo, H., Choe, J., Loeb, L., Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 9205–9210.

Life's Simple Pleasures!



No need to waste precious time looking for the right information – Register now for the free **Wiley-VCH Alerting Service.**

It's simple – and it's fast.

To receive regular news per e-mail tailored precisely to your needs and interests, just fill in the registration form at **www.wiley-vch.de/home/pas/**

WILEY-VCH

© 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Corteva Exhibit 2035 - Page 16 of 16
Bayer v. Corteva - IPR2023-01036