Ernst Schering Research Foundation
Workshop 51

# Biocombinatorial Approaches for Drug Finding

W. Wohlleben
T. Spellig
B. Müller-Tiemann
(Editors)

Springer

Ernst Schering Research Foundation Workshop 51
Biocombinatorial Approaches for Drug Finding

Ernst Schering Research Foundation
Workshop 51

# Biocombinatorial Approaches for Drug Finding

W. Wohlleben, T. Spellig, B. Müller-Tiemann
Editors

With 113 Figures

Springer

Series Editors: G. Stock and M. Lessl

Library of Congress Control Number: 2004113532

ISSN 0947-6075
ISBN 3-540-22092-5 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springeronline.com
© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. Product liability: The publishers cannot guarantee the accuracy of any information about dosage and application contained in this book. In every individual case the user must check such information by consulting the relevant literature.

Editor: Dr. Ute Heilmann, Heidelberg
Desk editor: Wilma McHugh, Heidelberg
Production editor: Andreas Gösling, Heidelberg
Cover design: design & production GmbH, Heidelberg
Typesetting: K+V Fotosatz GmbH, Beerfelden

Printed on acid-free paper – 21/3130/AG-5 4 3 2 1 0

# Preface

Genome- and proteome-based research is generating a significant increase in the number of available drug targets. As a result of such target discovery and validation, the requirements for screening resources, and especially the need for novel compounds, will increase dramatically.

Combinatorial approaches are powerful new tools for the generation of novel chemical structures. Combinatorial chemistry leads to a huge increase in the size and diversity of chemical libraries. In addition, lead structures isolated from such large compound pools by high-throughput screening can be improved by further rounds of combinatorial chemistry.

Nevertheless, although there have been major investments and efforts to establish and integrate these new techniques in drug discovery programs, this has not led to an increase in the number of novel drugs entering the market or at least in advanced development yet. Even worse, random million-compound libraries assembled by combinatorial chemistry in the last decade often contained hardly any hits. Reasons for this include the limited potential of combinatorial chemistry to synthesize complex and highly diverse compound libraries, as well as the improbability that the compounds will interact with biological targets.

Natural compounds overcome these limitations. Created by nature, they have been biologically validated over long periods of evolution. This is strikingly reflected by the number of drugs on the market that are either natural or based on the modification of natural compounds. A recent survey found that more than of half of the

drugs approved between 1981 and 2002 are natural compounds, natural compound-derived, or at least mimic structural elements of natural compounds (J Nat Prod 66:1022–1037). This experience, that a small collection of 'smart' compounds may be more valuable than a much larger random library, is a major stimulus to invest in combinatorial biosynthesis to assemble compound libraries that take 'nature's wisdom' as a guiding principle.

Recently, there has been a dramatic increase in the isolation and characterization of gene clusters encoding polyketide synthases (PKS), nonribosomal peptide synthetases (NRPS) and mixed PKS/NRPS enzymes. The field has been revolutionized by the successful expression of these enzyme complexes – the largest enzymes known so far – in appropriate, well-established heterologous hosts (Science 265:509–512; for review see Microbiol Mol Biol Rev 65:106–118). The establishment of these techniques, together with recently described methods for evolving new enzyme specificities, now enable the creation of many novel compounds by biocombinatorial approaches, using domain and module shuffling, precursor-directed biosynthesis, or molecular evolution technology (Science 277:367–369, Science 279:199–202, Science 291:1790–1792).

It has been calculated that a library of polyketides diversified by such techniques could contain up to $10^7$ different new polyketides (Curr Opinion Biotechnol 9:403–411). Such chemical diversity is further enhanced by the introduction of tailoring enzymes into the expression hosts (Chem Biol 8:547–555). In this way, the number and diversity of novel compounds with biological activity should be much greater than those resulting from purely chemical approaches. Biocombinatorial approaches should therefore make a major contribution to bridging the gap between the increasing number of validated drug targets and the number of appropriate lead candidates to interact with them.

This volume comprises contributions from different areas of academic and industrial research and will discuss biocombinatorial approaches for drug discovery in terms of:

- The significance of natural compounds for state-of-the-art drug discovery
- The underlying basic principle for the biosynthesis of highly complex compounds as well as the tailoring enzymes that theoretically can create an almost infinite chemical diversity
- The scope and limitations of combinatorial biosynthesis regarding formation, identification, isolation and manufacturing of novel biologically active entities, as well as the optimization of lead compounds.

D. A. Hopwood, T. Spellig, W. Wohlleben

# Contents

# List of Editors and Contributors

## Editors

*Wohlleben, W.*
Microbiology/Biotechnology, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 28, 72076 Tübingen, Germany
e-mail: wolfgang.wohlleben@biotech.uni-tuebingen.de

*Spellig, T.*
Corporate Research, Research Center Europe, Proteinchemistry,
Schering AG, Müllerstr. 178, 13342 Berlin, Germany
e-mail: tilman.spellig@schering.de

*Müller-Tiemann, B.*
Corporate Research, Research Center Europe, Proteinchemistry,
Schering AG, Müllerstr. 178, 13342 Berlin, Germany
e-mail: beate.mueller-tiemann@schering.de

## Contributors

*Bamonte, R.*
Vicuron Pharmaceutical, Via Lepetit 34, 21040 Gerenzano, Italy

*Bechthold, A.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany
e-mail: andreas.bechthold@pharmazie.uni-freiburg.de

*Berner, M.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Bihlmeier, C.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Bischoff, D.*
Institute of Chemistry, Eberhard-Karls-Universität, Tübingen,
Auf der Morgenstelle 18, 72076 Tübingen, Germany

*Boll, R.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Busti, E.*
Vicuron Pharmaceutical, Via Lepetit 34, 21040 Gerenzano, Italy

*Cavaletti, L.*
Vicuron Pharmaceutical, Via Lepetit 34, 21040 Gerenzano, Italy

*Chen, M.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, 777 Highland Avenue,
Madison WI 53705, USA

*Cheng, Y.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Donadio, S.*
Vicuron Pharmaceutical, Via Lepetit 34, 21040 Gerenzano, Italy
e-mail: sdonadio@vicuron.it

*Du, L.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Dürfahrt, T.*
Fachbereich Chemie/Biochemie, Philipps-Universität Marburg,
Hans-Meerwein-Strasse, 35032 Marburg, Germany

*Dürr, C.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Edwards, D. J.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Eliasson, A.*
Center for Microbial Biochtechnology, BioCentrum-DTU,
Building 223, 2800 Kgs. Lyngby, Denmark

*Frerich, A.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*George, N. P.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Gongli, T.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Hofmann, C.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Hopwood, D.A.*
John Innes Centre, Department of Molecular Microbiology,
Norwich Research Park, Colney Norwich NR4 7UH, United Kingdom
e-mail: david.hopwood@bbsrc.ac.uk

*Huang, Y.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Kittel, C.*
Microbiology/Biotechnology, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 28, 72076 Tübingen, Germany

*Kling, E.*
Institut für Biochemie, TU Dresden, 01062 Dresden, Germany

*Koch, M.A.*
Max Planck Institute of Molecular Physiology,
Department of Chemical Biology and Fachbereich Organische Chemie,
University of Dortmund, Otto-Hahn-Straße 11, 44227 Dortmund,
Germany

*Luzhetskyy, M.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Marahiel, M.A.*
Fachbereich Chemie/Biochemie, Philipps-Universität Marburg,
Hans-Meerwein-Strasse, 35032 Marburg, Germany
e-mail: marahiel@chemie.uni-marburg.de

*Mayer, A.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Mazza, P.*
Vicuron Pharmaceutical, Via Lepetit 34, 21040 Gerenzano, Italy

*Méndez, C.*
Departamento de Biologia Funcional e Instituto Universitario
de Oncologia (I.U.O.P.A.), Universidad de Oviedo,
33006 Oviedo, Spain

*Monciardini, P.*
Vicuron Pharmaceutical, Via Lepetit 34, 21040 Gerenzano, Italy

*Nielsen, J.*
Center for Microbial Biotechnology, BioCentrum-DTU,
Building 223, 2800 Kgs. Lyngby, Denmark
e-mail: jn@biocentrum.dtu.dk

*Oh, T.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*van Pée, K.-H.*
Institut für Biochemie, TU Dresden, 01062 Dresden, Germany
e-mail: Karl-Heinz.vanPee@chemie.tu-dresden.de

*Pelzer, S.*
Microbiology/Biotechnology, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 28, 72076 Tübingen, Germany
e-mail: Pelzer@combinature.com

*Puk, O.*
Microbiology/Biotechnology, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 28, 72076 Tübingen, Germany

*Recktenwald, J.*
Microbiology/Biotechnology, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 28, 72076 Tübingen, Germany

*Salas, J.A.*
Departamento de Biologia Funcional e Instituto Universitario
de Oncologia (I.U.O.P.A.), Universidad de Oviedo,
33006 Oviedo, Spain
e-mail: jasalas@correo.uniovi.es

*Sanchez, C.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Schmid, C.*
Institut für Biochemie, TU Dresden, 01062 Dresden, Germany

*Shen, B.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin- Madison, Madison WI 53705, USA
e-mail: bshen@pharmacy.wisc.edu

*Sosio, M.*
Vicuron Pharmaceutical, Via Lepetit 34, 21040 Gerenzano, Italy

*Stegmann, E.*
Microbiology/Biotechnology, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 28, 72076 Tübingen, Germany

*Süßmuth, R.*
Chemisches Institut, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 18, 72076 Tübingen, Germany

*Tang, G.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Treede, I.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Unversucht, S.*
Institut für Biochemie, TU Dresden, 01062 Dresden, Germany

*Vente, A.*
Combinature Biopharm AG, Robert-Rössle-Strasse 10,
13125 Berlin, Germany

*Vosburg, D.A.*
Department of Biological Chemistry and Molecular Pharmacology,
Harvard Medical School, 240 Longwood Avenue, Boston,
Massachusetts 02115, USA

*Wage, T.*
Institut für Biochemie, TU Dresden, 01062 Dresden, Germany

*Waldmann, H.*
Max Planck Institute of Molecular Physiology, Department
of Chemical Biology and Fachbereich Organische Chemie,
University of Dortmund, Otto-Hahn-Straße 11, 44227 Dortmund,
Germany
e-mail: Herbert.waldmann@mpi-dortmund.mpg.de

*Walsh, C.T.*
Department of Biological Chemistry and Molecular Pharmacology,
Harvard Medical School, 240 Longwood Avenue, Boston,
Massachusetts 02115, USA
e-mail: Christopher_walsh@hms.harvard.edu

*Weissman, K.J.*
Department of Biochemistry, University of Cambridge,
80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom
e-mail: kjw21@cus.cam.ac.uk

*Weist, S.*
Institute of Chemistry, Eberhard-Karls-Universität Tübingen,
Auf der Morgenstelle 18, 72076 Tübingen, Germany

*Weitnauer, G.*
Institut für Pharmazeutische Wissenschaften,
Albert-Ludwigs-Universität Freiburg, Stefan-Meier-Strasse 19,
79104 Freiburg, Germany

*Wendt-Pienkowski, E.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Wohlert, S.-E.*
Combinature Biopharm AG, Robert-Rössle-Straße 10, 13125 Berlin,
Germany

*Yi, F.*
Division of Pharmaceutical Sciences and Department of Chemistry,
University of Wisconsin-Madison, Madison WI 53705, USA

*Zehner, S.*
Institut für Biochemie, TU Dresden, 01062 Dresden, Germany

# 1 Protein Domain Fold Similarity and Natural Product Structure as Guiding Principles for Compound Library Design

M. A. Koch, H. Waldmann



## 1.1 Introduction

The last decade brought a tremendous gain in biological information through large-scale and global approaches addressing the aspects of DNA sequence (genomics), protein structure (structural genomics), and protein expression and interactions (proteomics). Bioinformatics tools help to convert this vast amount of basic data into actual knowledge exploitable for the benefit of mankind – in particular the development of new therapies of diseases. Of particular interest is the relationship between protein structure and function as its under-

standing will help to find small molecules which alter protein function by either selective inhibition or activation (Stockwell 2000; Alaimo et al. 2001). Compound development via combinatorial chemistry techniques will become the method of choice to undertake this Herculean task. But as the universe of thinkable chemical compounds is almost infinite (Bohacek et al. 1996), one important question arises: Where in the universe of chemical structures are compounds with the desired biological properties to be found?

The original expectation that the synthesis of million compound libraries will produce as many or even more drug candidates as historical libraries of pharmaceutical companies and that it will thereby overcome the problem of efficient hit and lead finding was not fulfilled. It was soon recognized that not sheer numbers will determine the quality of a library, but rather its 'diversity' (Golebiowski et al. 2001; Mason and Hermsmeier 1999; Schreiber 2000), its 'drug-likeness' (Walters et al. 1999; Ajay et al. 1998; Sadowski and Kubinyi 1998; Ghose et al. 1999; Lee and Schneider 2001) and its 'biological relevance' (Breinbauer et al. 2002; Koch et al. 2003; Koch and Waldmann 2004). A central and crucial task is the identification of compound classes that represent already biologically validated starting points in structural space, to find a synthetic access to them that is amenable to combinatorial variation and to design and synthesize combinatorial libraries centered on the identified underlying structural frameworks of these compound classes. Biologically active natural products, usually low molecular weight chemical compounds which are synthesized by biological organisms and influence biological processes, are viable, biologically validated starting points for library design. They permit finding of hit or lead compounds with enhanced probability and quality if these libraries are included in high throughput screening (Brohm et al. 2002; Breinbauer et al. 2002; Koch et al. 2003). Scaffolds of certain natural products and non-natural compounds embody so-called 'privileged structures'. This term was originally coined by the group of B.E. Evans and co-workers at Merck who recognized in their pharmacological studies of benzodiazepines (Evans et al. 1988) that derivatives within this compound class bind not only to benzodiazepine receptors of the central nervous system, but also to cholecystokinin receptors and to the unrelated class of peripheral benzodiazepine receptors. As peptidomimetics benzodiazepines can be as-

sumed to have an intrinsic good binding affinity to several proteins, which bind similar regions of peptides or other proteins. According to the definition of Evans, privileged structures constitute a compound class which can bind to various proteinaceous receptor surfaces (Horton et al. 2003). The 'biological relevance' of natural products and privileged structures can be understood in the light of structural and/or functional relationship of proteins.

## 1.2 Protein Folds and Protein Function

Proteins can be regarded as modularly built biomolecules assembled from individual building blocks. These building blocks are called 'domains', parts of the proteins that fold independently from the rest of the structure to a compact arrangement of secondary structures interconnected via more or less complex linker peptides. The term 'domain family' as it is used in this review refers to a family of related sequences which have a common ancestor, i.e., which have developed via divergent evolution. Different sequence families (i.e., domains) can adopt the same fold. This could be regarded either as convergence due to functional and physical constraints because of the limited number of acceptable spatial arrangements of secondary structural elements or as a result of divergent evolution to an extent that the sequence relationship is not recognizable anymore (Lupas et al. 2001; Govindarajan et al. 1999). Protein domains can be regarded as structurally conserved yet genetically mobile units (Ponting et al. 2000).

Although the estimate for the number of different proteins in humans ranges from 100,000 to 450,000, there is a common agreement that the number of domain families and – even more – of topologically distinct folds will be much smaller. At present approximately 600 folds are known, derived by classifying all structurally characterized proteins according to their three-dimensional structure (Murzin et al. 1995; Thornton et al. 1995; Wolf et al. 2000; Govindarajan et al. 1999). Data from the ongoing genome sequencing projects allow estimation of the number of existing folds and families in Nature. Current estimates vary between 600 and 8,000 distinct folds, and 4,000–60,000 sequence families (Koonin et al. 2003; Wolf et al.