Ernst Schering Research Foundation
Workshop 58

Chemical Genomics

Small Molecule Probes to Study Cellular Function

S. Jaroch
H. Weinmann
(Editors)

Springer

Ernst Schering Research Foundation Workshop 58
Chemical Genomics

Ernst Schering Research Foundation
Workshop 58

Chemical Genomics

Small Molecule Probes to Study Cellular Function

S. Jaroch, H. Weinmann
Editors

With 87 Figures

Springer

Series Editors: G. Stock and M. Lessl

Library of Congress Control Number: 2005931334

ISSN 0947-6075

ISBN-10 3-540-27865-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-27865-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Editor: Dr. Ute Heilmann, Heidelberg
Desk Editor: Wilma McHugh, Heidelberg
Production Editor: Monika Riepl, Leipzig
Cover design: design & production, Heidelberg
Typesetting and production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
21/3100/YL – 5 4 3 2 1 0 Printed on acid-free paper

Preface

In a time when the output of fewer NCEs affords even more significant investments into R&D, the paradigm of conducting drug discovery through a linear process entailing target identification/validation, lead discovery, lead optimization, and finally candidate selection is challenged by an ap-

proach called chemical genomics. Starting with small-molecule probes, synthesized by high-throughput chemistry, to conditionally modulate gene or protein function and to eventually identify therapeutically relevant targets, it places chemistry at the very beginning of the drug discovery process. Depending on a smart library design, the perturbing agent may ideally represent a lead structure, thus reducing development times by running lead identification and target identification/validation processes in parallel.

We initiated an Ernst Schering Research Foundation Workshop to bring together leading scientists from academia and pharmaceutical companies to learn more about progress in chemogenomics, chemical genomics, and chemical biology, and the proceedings of this symposium are detailed in this book.

Though the terms "chemical genomics" and "chemogenomics" are sometimes confused in the literature, the first three chapters point out that chemogenomics is an effort within pharmaceutical companies to integrate data on target protein sequences with molecular structures and selectivity data of small molecules. In contrast to the traditional focus on distinct targets, the chemogenomics approach considers target families and activity profiles. After a general overview by H. Kubinyi, W. Guba, T. Klabunde and R. Jäger present Roche's and Sanofi-Aventis's chemogenomics approach to identifying novel lead structures for drug discovery programs aiming at G-protein coupled receptor (GPCR) modulators.

Chemical genomics, entailing the synthesis of small-molecule probes and their use to study cellular function, depends on diversity-oriented synthesis (DOS) to set up large libraries as a toolbox in order to populate chemical space broadly. As D.R. Spring and colleagues demonstrate, these DOS libraries differ from the old combinatorial chemistry-derived libraries of the 1990s in that they exploit novel and richer chemistries leading to more elaborate architectures in terms of skeletal, stereochemical, and building block diversity. This is in contrast to the sole emphasis on building block diversity in "traditional" combinatorial or parallel high-throughput chemistry approaches.

A chemical genomics application is presented by J.D. Gough and C.M. Crews; they elegantly use the natural product fumagillin (TNP-470) as a chemical probe for studying endothelial biology. Further, they

describe the *Pro*teolysis *Ta*rgeting *C*himerics (PROTAC) approach as a "chemical knockout" tool to study protein function.

R.V. Weatherman and colleagues discuss research in molecular endocrinology, particularly chemical approaches to selectively dissect complex biological processes related to the estrogen receptor.

Natural products already excel through constitutional diversity and are therefore ideal templates for chemical genomics efforts. M.A. Koch and H. Waldmann report on the clustering of natural product frameworks (SCNOP, structural classification of natural products) and their value as biologically validated starting points in structural space for library design.

Before small molecule-protein interactions are studied, a robust screening technology has to be put in place. Opportunities arising through improved capabilities are highlighted by L. Mayr discussing the highly minitiaturized high-throughput nanoscreen and the affinity-based assay technology behind speedscreen at Novartis.

The editors would like to acknowledge the generous support of the Ernst Schering Foundation, which allowed us to set up this exciting workshop. We trust that the reader will share the enthusiasm and excitement in the highly interdisciplinary field of chemical genomics and chemogenomics.

Stefan Jaroch
Hilmar Weinmann

Contents

List of Editors and Contributors

Editors

Jaroch, S.
Corporate Research, Medicinal Chemistry, Schering AG,
Müllerstr. 178, 13342 Berlin, Germany
(e-mail: stefan.jaroch@schering.de)

Weinmann, H.
Corporate Research, Medicinal Chemistry, Schering AG,
Müllerstr. 178, 13342 Berlin, Germany
(e-mail: hilmar.weinmann@schering.de)

Contributors

Bender, A.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: ab454@cam.ac.uk)

Crews, C.M.
Yale University Department of MCDB, Chemistry and Pharmacology,
P.O. Box 208 103, New Haven, CT 06520-8103, USA
(e-mail: craig.crews@yale.edu)

Fergus, S.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: sf341@cam.ac.uk)

Galloway, W.R.J.D.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: wrjdg2@cam.ac.uk)

Glansdorp, F.G.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: fgg20@cam.ac.uk)

Glen, R.C.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: rcg28@cam.ac.uk)

Gough, J.D.
Department of Molecular, Cell and Development Biology,
P.O. Box 208 103, New Haven, CT 06520–8103, USA

Guba, W.
F. Hoffmann-La Roche Ltd., PRBD-CM, Bldg. 92/3.10C,
4070 Basel, Switzerland
(e-mail: wolfgang.guba@roche.com)

Jäger, R.
Sanofi-Aventis Deutschland GmbH, Drug Design, Science & Medical Affairs,
Industriepark Höchst G838, 65926 Frankfurt am Main, Germany
(e-mail: RobertM.Jaeger@sanofi-aventis.com)

Klabunde, T.
Sanofi-Aventis Deutschland GmbH, Drug Design, Science & Medical Affairs,
Industriepark Höchst G878, 65926 Frankfurt am Main, Germany
(e-mail: Thomas.Klabunde@sanofi-aventis.com)

Koch, M.A.
Max Planck Institute of Molecular Physiology, Department of Chemical Biology,
Otto-Hahn-Str. 11, 44227 Dortmund, Germany

Kubinyi, H.
University of Heidelberg,
Donnersbergstr. 9, 67256 Weisenheim am Sand, Germany
(e-mail: Kubinyi@t-online.de)

Marsden, D.M.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: dmm40@cam.ac.uk)

Mayr, L.M.
Novartis Institutes of BioMedical Research (NIBR),
Discovery Technologies/Lead Discovery Center (DT/LDC),
4002 Basel, Switzerland
(e-mail: Lorenz.Mayr@novartis.com)

Nicholson, R.L.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: rln27@cam.ac.uk)

Reyes, P.T.
Department of Medicinal Chemistry and Molecular Pharmacology
and Purdue Cancer Center, Purdue University,
575 Stadium Mall Drive, West Lafayette, IN 47907, USA

Rickert, E.L.
Department of Medicinal Chemistry and Molecular Pharmacology
and Purdue Cancer Center, Purdue University,
575 Stadium Mall Drive, West Lafayette, IN 47907, USA

Spandl, R.J.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: rjs204@cam.ac.uk)

Spring, D.R.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: drspring@ch.cam.ac.uk)

Trebley, J.P.
Department of Medicinal Chemistry and Molecular Pharmacology
and Purdue Cancer Center, Purdue University,
575 Stadium Mall Drive, West Lafayette, IN 47907, USA

Thomas, G.L.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: gt248@cam.ac.uk)

Waldmann, H.
Max Planck Institute of Molecular Physiology, Department of Chemical Biology,
Otto-Hahn-Straße 11, 44221 Dortmund, Germany
and
University of Dortmund, Fachbereich 3, Organische Chemie,
44221, Dortmund, Germany
(e-mail: Herbert.waldmann@mpi-dortmund.mpg.de)

Weatherman, R.V.
Department of Medicinal Chemistry and Molecular Pharmacology
and Purdue Cancer Center, Purdue University,
575 Stadium Mall Drive, West Lafayette, IN 47907, USA
(e-mail: rossw@pharmacy.purdue.edu)

Wyatt, E.E.
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, UK
(e-mail: eew25@cam.ac.uk)

1 Chemogenomics in Drug Discovery

H. Kubinyi

Abstract. Chemogenomics is a new strategy in drug discovery which, in principle, searches for all molecules that are capable of interacting with any biological target. Because of the almost infinite number of drug-like organic molecules, this is an impossible task. Therefore chemogenomics has been defined as the investigation of classes of compounds (libraries) against families of functionally related proteins. In this definition, chemogenomics deals with the systematic analysis of chemical–biological interactions. Congeneric series of chemical analogs are probes to investigate their action on specific target classes, e.g., GPCRs, kinases, phosphodiesterases, ion channels, serine proteases, and others. Whereas such a strategy developed in pharmaceutical industry almost 20 years ago, it is now more systematically applied in the search for target- and subtype-specific ligands. The term "privileged structures" has been defined for scaffolds, such as the

benzodiazepines, which very often produce biologically active analogs in a target family, in this case in the class of G-protein-coupled receptors. The SOSA approach is a strategy to modify the selectivity of biologically active compounds, generating new drug candidates from the side activities of therapeutically used drugs.

1.1 Introduction

Chemical biology, chemical genetics, and chemogenomics are recent strategies in drug discovery. Although definitions in the literature are somehow diffuse and inconsistent, a differentiation of the terms will be attempted here:

Chemical biology may be defined as the study of biological systems, e.g., whole cells, under the influence of chemical libraries. If a new phenotype is discovered by the action of a certain substance, the next step is the identification of the responsible target.

Chemical genetics is the dedicated study of protein function, e.g., signaling chains, under the influence of ligands which bind to certain proteins or interfere with protein–protein interaction; sometimes orthogonal ligand–protein pairs are generated to achieve selectivity for a certain protein.

Chemogenomics defines, in principle, the screening of the chemical universe, i.e., all possible chemical compounds, against the target universe, i.e., all proteins and other potential drug targets. Whereas this task can never be achieved, due to the almost infinite size of the chemical universe, the systematic screening of libraries of congeneric compounds against members of a target family offers unprecedented chances in the search for compounds with significant target or subtype specificity.

1.2 Chemical Biology

In classical drug discovery, research was often based on vague hypotheses on structure–activity relationships. Compounds were synthesized and tested in whole animals. If a biological effect was observed, a medicinal chemistry project started to optimize chemical structures

with respect to activity, pharmacokinetic properties, and lack of toxic side effects. Later on, this approach was replaced by in vitro screening on defined targets, most often human proteins. Only in recent years have we experienced a more systematic investigation of drug-like compounds in biological systems, called chemical biology.

One illustrative example of the chemical biology approach is the discovery of monastrol, a molecule that prevents spindle formation in mitotic cells by inhibiting the kinesin Eg5, a motor protein required for spindle bipolarity (Mayer et al. 1999). In this manner, monastrol stops cell division by mitotic arrest.

Another example of the concept of chemical biology is the discovery of synthetic small molecules that influence embryonic stem (ES) cell fate (Ding et al. 2003). A high-throughput phenotypic cell-based screen identified a 4,6-disubstitued pyrrolo-pyrimidine, which induces the differentiation of ES cells to neurons. Glycogen synthase kinase-3β (GSK-3β) has been identified as the target of this compound.

On the other hand, screening of any compounds may not result in the desired output of results. The production of a 2.18 million-compound natural product library by diversity-oriented synthesis (Tan et al. 1998; Schreiber 2000) generated much hype but, so far, not the anticipated results with respect to biological activities. In a later comment, the author Stu Schreiber had to admit that the chemical diversity of his library was seemingly too narrow – "disappointingly similar" by molecular descriptors; the compounds "tend to cluster in discrete regions of multidimensional descriptor space" (Schreiber 2003). This goes hand in hand with another problem: biologically active compounds seem to be distributed only in certain areas of chemical space, by their physicochemical properties and their structural features (Lipinski and Hopkins 2004). If we consider the chemical universe as a huge ocean, with small islands or groups of islands of biologically active compounds (e.g., the so-called privileged compounds, cf. Sect. 1.4.1), we have to understand and accept that most chemistry-driven approaches will end up in water, instead of discovering new islands. For the broad exploration of biology with small organic molecules (Stockwell 2004), the National Institutes of Health (NIH) has started an initiative to provide a repository of chemically diverse molecules for the public and private sector (Austin et al. 2004).

1.3 Chemical Genetics

Classical genetics sets a (random) mutation, e.g., by irradiation, and tries to conclude from a new phenotype to the genotype. "Chemical genetics" is another new term for a strategy that has also been used since long ago, in a less systematic manner; it describes the purposeful investigation of proteins by small molecules or libraries, for target identification (forward chemical genetics) or target validation (reverse chemical genetics) (Russell and Michne 2004). Sometimes, orthogonal ligand-receptor pairs are constructed if selective ligands are not available. Selective kinase inhibition has been achieved by specifically converting nonspecific, low-affinity inhibitors into larger analogs and to construct certain kinase mutants (e.g., v-Src I338G or Cdk II F80G) that specifically accommodate these originally less well-fitting ligands by their larger binding pocket (Bishop et al. 2000). In this manner, the specific inhibition of a certain kinase can be studied without having developed an inhibitor of comparable specificity against the wild-type kinase.

1.4 Chemogenomics

As well as in the other two cases, chemogenomics defines an approach that has also been used earlier, but less systematically. Since a screening of the chemical universe against the target universe is practically impossible, due to the almost infinite number of potential drug-like compounds, the method defines the screening of congeneric chemical libraries against certain target families, e.g., the G protein-coupled receptors, nuclear receptors, different protease families, kinases, phosphodiesterases, ion channels, transporters, etc. (Caron et al. 2001; Bleicher 2002; Jacoby et al. 2003; Miller 2003; Kubinyi and Müller 2004); this systematic strategy aims to discover highly potent, selective ligands against functionally and evolutionarily related targets, with the least effort.

1.4.1 Privileged Structures

Many drugs have been derived from certain chemotypes, e.g., phenethylamines, tricyclics, steroids, or benzodiazepines, whereas others have

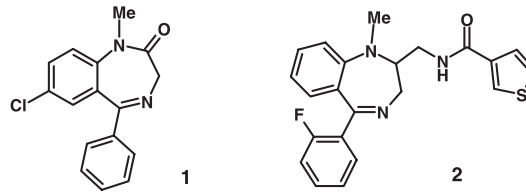


Fig. 1. Diazepam **1** (Valium) was one of the first tranquilizers and the prototype of a series of other GABA receptor agonists, antagonists, and inverse agonists. The chemically closely related benzodiazepine Tifluadom **2** is a κ-opiate receptor agonist and a nanomolar cholecystokinin receptor antagonist

certain structural features in common, e.g., diphenylmethane, diphenylamine, or arylpiperazine groups. The systematic chemical variation of benzodiazepines, e.g., the GABA-agonist diazepam **1** produced not only tranquilizers but also GABA antagonists, inverse agonists, and the strong κ-opiate receptor agonist tifluadom **2** (Fig. 1) (Römer et al. 2002).

When Evans discovered that tifluadom is also a nanomolar cholecystokinin receptor antagonist, he concluded that "these structures appear to contain common features which facilitate binding to various . . . receptor surfaces, perhaps through binding elements different from those employed for binding of the natural ligands . . . " and formulated ". . . what is clear is that certain 'privileged structures' are capable of providing useful ligands for more than one receptor and that judicious modification of such structures could be a viable alternative in the search for new receptor agonists and antagonists" (Evans et al. 1988). Minor chemical modifications of such privileged structures (Fig. 2) (Patchett and Nargund 2000) may result in highly selective ligands or drugs, e.g., the estrogenic, gestagenic, androgenic, glucocorticoid, and mineralocorticoid steroids, or the α-adrenergic, β-adrenergic, and β-antiadrenergic phenethylamines. Others lack such target selectivity: the atypical neuroleptic olanzapine is a highly promiscuous tricyclic ligand, with nanomolar affinities at various GPCRs, including 5-HT_{2A}, 5-HT_{2B}, 5-HT_{2C}, dopaminergic D_1, D_2, D_4, muscarinic M_1, M_2, M_3, M_4, M_5, adrenergic α_1, and histaminic H_1 receptors, as well as the 5-HT_3 ion channel (Bymaster et al. 1996, 1999).

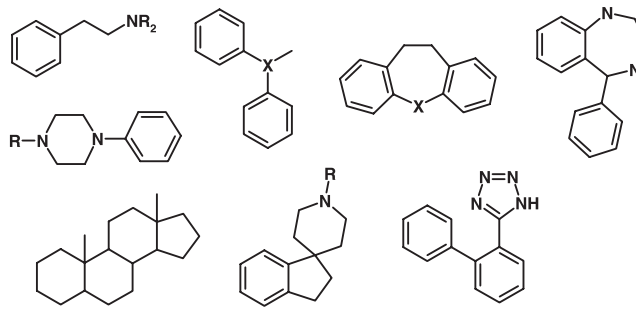


Fig. 2. Privileged structures are scaffolds or substituents that often produce biologically active compounds, e.g., phenethylamines, diphenylmethyl and diphenylamine compounds (X = C or N, respectively), tricyclic compounds (X = C or N), benzodiazepines, arylpiperidines, steroids, spiropiperidines, and tetrazolobiphenyls (from the *upper left* to the *lower right*)

Privileged structures, even if they are promiscuous ligands, should not be confused with some structural classes, which seemingly bind with micromolar affinity to various enzymes. This unspecific binding behavior is caused by an aggregation of the ligands and clumping of these aggregates to the protein (McGovern et al. 2002, 2003; McGovern and Shoichet 2003; Seidler et al. 2003).

1.4.2 Drugs from Side Effects – The SOSA Approach

Many drugs of the past resulted from the experimental or clinical observation of side effects. Diuretic, antihypertonic, antiglaucoma, and antidiabetic drugs were derived from the bacteriostatic sulfonamides; the mood-improving effect of iproniazid was discovered when it was tested as an antituberculous drug; antidepressant inhibitors of neurotransmitter re-uptake, like imipramine and desipramine, stem from the antipsychotic dopamine antagonist chlorpromazine, which itself was derived from H_1 antihistaminics; there are many other stories of this kind (Sneader 1996; Kubinyi 2004). Only recently, Camille Wermuth proposed to investigate the side effects of drugs more systematically, by his “selective optimization of side activities” (SOSA) approach (Wer-

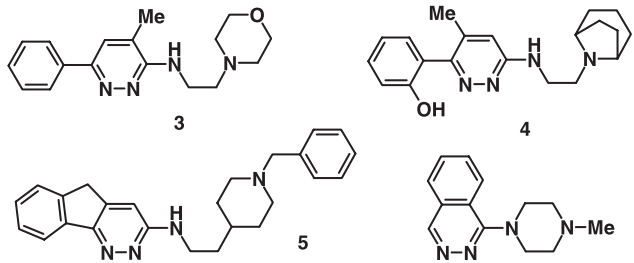


Fig. 3. The antidepressant minaprine **3** is also a weak muscarinic M_1 receptor antagonist ($K_i = 17\ \mu M$) and an acetylcholinesterase inhibitor ($K_i = 600\ \mu M$). By systematic structural variation, these activities could be enhanced to the nanomolar M_1 receptor antagonist **4** ($K_i = 3$ nM) and the acetylcholinesterase inhibitor **5** ($K_i = 10$ nM). A closely related analog of minaprine was optimized to the nanomolar 5-HT_3 receptor antagonist **6** ($IC_{50} = 10$ nM)

muth 2001, 2004). Whenever a side effect of a drug is observed, it might be possible to optimize the candidate to a selective drug with this other biological activity, following a statement by Sir James Black that "the most fruitful basis for discovery of a new drug is to start with an old drug" (Wermuth 2004). Among several other examples, Wermuth demonstrated by his own research the optimization of different weak side effects of the antidepressant minaprine **3** to the nanomolar muscarinic M_1 receptor ligand **4** and the reversible acetylcholinesterase inhibitor **5** (Wermuth 2001, 2004); a closely related analog of minaprine was optimized to the nanomolar 5-HT_3 antagonist **6** (Fig. 3) (Rival et al. 1998). More examples are discussed in refs. (Kubinyi 2004; Wermuth 2001, 2004).

1.4.3 From Target Family-Directed Masterkeys to Selective Drugs

Chemogenomics is mainly based on the masterkey concept of tailor-made privileged structures (Müller 2003, 2004). Starting from such masterkeys, selective ligands can be derived, either by classical medicinal chemistry or by systematic structural variation in combinatorial libraries. The masterkey concept will be illustrated by just one example: selective β_1 and β_2 agonists, as well as β antagonists (β-blockers)

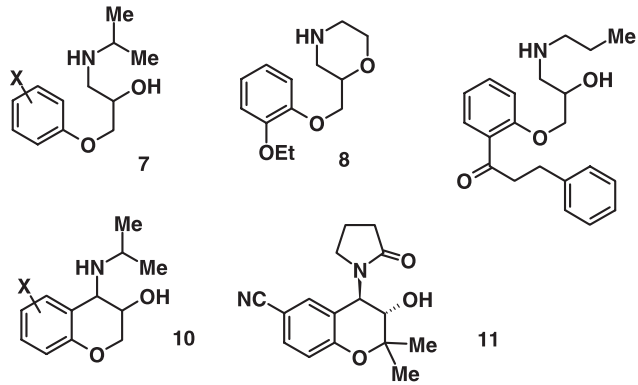


Fig. 4. The β-blocker prototype structure 7, Phenyl-O-CH_2-CH(OR^1)-CH_2NHR^2 is also the key structural element of the antidepressant viloxazine **8** and the class Ic antiarrhythmic propafenone **9**. Structural variation of a cyclic β-blocker analog **10** yielded the potassium channel opener levcromakalim **11**

were derived from the mixed α/β agonist epinephrine. Further chemical variation of the typical β-blocker phenoxypropanolamine structure **7** yielded the antidepressant viloxazine **8** and the class Ic antiarrhythmic propafenone **9**. The optimization of a cyclic β-blocker prototype **10** indeed produced an antihypertensive drug; however, levcromakalim **11** is no longer a β-blocker, it is a vasodilatory potassium channel opener (Fig. 4) (Wermuth 2001, 2004). More examples are discussed in the following sections and in references (Caron et al. 2001; Bleicher 2002; Jacoby et al. 2003; Kubinyi and Müller 2004; Kubinyi 2004; Wermuth 2001, 2004; Müller 2003, 2004).

Enzyme Inhibitors

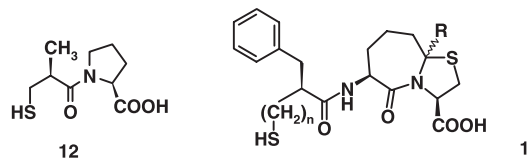
Protease inhibitors are most often derived from the sequence of the amino acids in the positions next to the bond that is cleaved by the enzyme. A simple strategy for a first inhibitor is a conversion of the amide bond of the cleavage site into a noncleavable analog or a group that reacts or coordinates with the catalytic center of the enzyme; the P1, P2, . . . and/or P1′, P2′, . . . amino acids are kept constant.

The structural requirements of the individual protease classes are different:

– For aspartyl protease inhibitors, it is necessary to attach some amino- and carboxy-terminal amino acid side chains to a group that mimics the transition state of the enzymatic cleavage.
– For metalloprotease inhibitors, a metal-coordinating group is introduced at the amino-terminal side of the peptide.
– For serine and cysteine protease inhibitors, the groups that interact with the catalytic center are not necessarily but most often at the carboxy-terminal end of the peptide.

The chemogenomics strategy in the design of protease inhibitors will be illustrated by four examples: the design of HIV protease inhibitors, thrombin and factor Xa inhibitors, selective ACE and dual zinc protease inhibitors, and "dual warhead" MMP/cathepsin inhibitors. Renin is an aspartyl protease, which is involved in blood pressure regulation by converting angiotensinogen into angiotensin I, the substrate of angiotensin-converting enzyme (ACE). Hundreds of person years of research were invested to arrive at orally active peptidomimetics, without much success. When it became known that HIV protease is also an aspartyl protease, the accumulated experience on the design of transition state inhibitors could be transferred to this new project.

The same situation applies to inhibitors of the serine protease thrombin; here also all efforts to arrive at orally active analogs had only limited success. However, structural elements from inhibitors of another serine



12 13

Fig. 5. Captopril **12** was the very first marketed angiotensin-converting enzyme (ACE) inhibitor. The specific ACE inhibitor **13a** (n = 0, R = β-H; K_i ACE = 11.5 nM, K_i NEP24.11 = 2,820 nM) resulted from structural variation, as well as the dual zinc protease inhibitors **13b** (n = 0, R = α-H; K_i ACE = 16 nM, K_i NEP24.11 = 11.5 nM) and **13c** (n = 1, R = α-H; K_i ACE = 5.5 nM, K_i NEP24.11 = 1.1 nM)

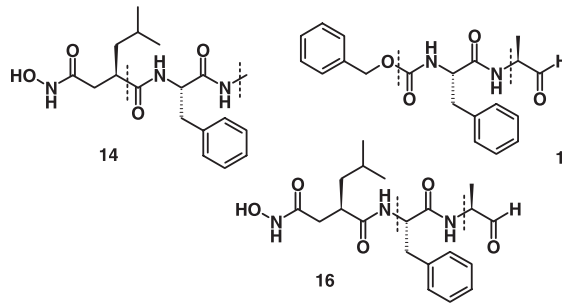


Fig. 6. Compound **14** is a nanomolar metalloprotease inhibitor (IC_{50} MMP-1 = 3 nM; IC_{50} Cat L > 1,000 nM), whereas compound **15** is a nanomolar cysteine protease inhibitor (IC_{50} MMP-1 > 1,000 nM; IC_{50} Cat L = 3 nM). Crossover of the two structures produces the dual inhibitor **16** (IC_{50} MMP-1 = 25 nM; IC_{50} Cat L = 15 nM); the *dashed lines* indicate the common center part of all three molecules

protease, elastase, e.g., the pyrimidone ring system as a substitute for a flexible amino acid, could also be applied to thrombin inhibitors. Later on, the search for inhibitors shifted from thrombin to factor Xa, a serine protease with similar specificity as thrombin.

Captopril **12** was the very first ACE inhibitor that was introduced into human therapy. A multitude of ACE-inhibiting analogs resulted from this drug, e.g., the ACE-specific inhibitor **13a** and the dual ACE/NEP24.11 inhibitors **13b** and **13c** (Fig. 5) (Slusarchyk et al. 1995).

A dual warhead inhibitor resulted from a merger of the structures of a selective matrix metalloprotease (MMP) inhibitor **14** with a cathepsin L inhibitor **15**. Although MMP-1 is a zinc protease and cathepsin L is a cysteine protease, the resulting inhibitor **16**, which bears both "warheads," inhibits both enzymes with nanomolar activity (Fig. 6) (Yamamoto et al. 2002).

Kinases play a most important role in cell signaling. More than 500 different kinases are coded by the human genome; after activation, they phosphorylate either a tyrosine hydroxyl group (tyrosine kinases) or a serine or threonine hydroxyl group (serine/threonine kinases). Some kinase mutants are constitutionally active: they activate a signaling cascade

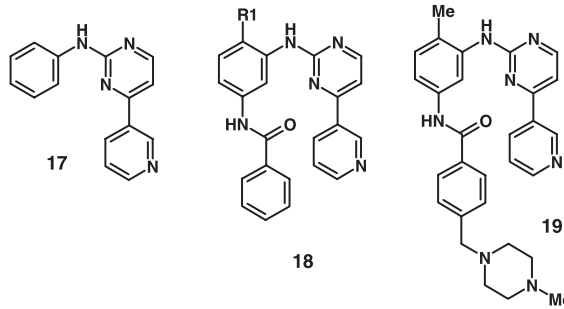


Fig. 7. Structural variation of the protein kinase C (PKC) inhibitor **17** produced the dual PKC/bcr-abl inhibitor **18a** (R = H). A minor structural modification to **18b** (R = CH_3) abolished the undesired PKC activity. After introduction of a methylpiperazine residue, to enhance the aqueous solubility, the bcr-abl inhibitor imatinib **19** (Glivec, Gleevec) resulted

without any external stimulus. Chronic myelogenous leukemia is caused by such a constitutionally active kinase. The coding regions of an abl tyrosine kinase at chromosome 9 and a bcr serine/threonine kinase at chromosome 22 form after reciprocal translocation a bcr-abl coding region at the new, shorter version of the chromosome 9, the so-called Philadelphia chromosome. The resulting bcr-abl tyrosine kinase is constitutionally active. At Novartis, a class of protein kinase C (PKC) inhibitors were optimized to the PKC inhibitor **17**. Amide analogs **18a** of this compound showed activity against PKC and bcr-abl kinase; surprisingly, the methyl analog **18b** inhibited only bcr-abl kinase; finally, an N-methyl-piperazine residue was added to increase solubility (Fig. 7). Imatinib (Gleevec, Glivec), **19**, was clinically developed and is successfully used for the treatment of chronic myelogenous leukemia (Capdeville et al. 2002).

Receptor Ligands

G protein-coupled receptors (GPCRs) are a large group of evolutionarily related seven-transmembrane proteins. They are activated by such different agents as light, ions, odorants, neurotransmitters, peptides, and proteins and transfer the stimulus by the G protein complex. Sero-

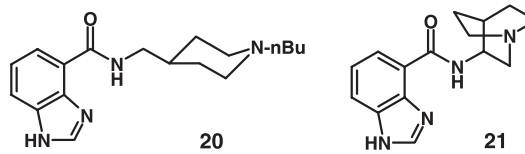


Fig. 8. Compound **20** is a highly selective 5-HT_3 antagonist (K_i 5-HT_3 = 3.7 nM, K_i 5-HT_4 > 1,000 nM), whereas the chemically closely related compound **21** is a selective 5-HT_4 antagonist (K_i 5-HT_3 > 10,000 nM, K_i 5-HT_4 = 13.7 nM)

tonin receptors are made up of 14 subtypes, 13 of which are GPCRs, whereas the 5-HT_3 subtype is a ligand-controlled ion channel. From pharmacophore models, Lopez-Rodriguez et al. designed the structure of a highly selective 5-HT_4 receptor ligand **20**, which shows a selectivity difference of more than five orders of magnitude to its closely related, 5-HT_3-selective analog **21** (Fig. 8) (Lopez-Rodriguez et al. 1997).

Somatostatin receptors are made up of five subtypes: sst1–sst5. In their attempt to obtain selective, peptidomimetic ligands for each subtype, Rohrer et al. synthesized four β-turn-mimicking combinatorial libraries, with up to 350,000 compounds per library. Highly specific ligands resulted for all five subtypes (Rohrer et al. 1998).

Nuclear receptors are another important receptor family. They are made up of a ligand-binding domain and a DNA-binding domain. After activation by their specific ligands, e.g., the steroid hormones, the thyroid hormone or retinoic acid, receptor dimers bind to DNA and activate the expression of certain proteins.

Estrogen receptors exist as two distinct subtypes, ERα and ERβ, which are relatively abundant in several tissues. As their function in all those organs and potential interaction, forming ERα/ERβ heterodimers, has not been completely elucidated so far, it is most important to find selective ligands for both receptors. By homology modeling of the ligand-binding domain of the ERβ receptor, based on the corresponding 3D structure of the ERα receptor, Hillisch et al. inspected the minor differences in the estradiol binding site: in human ERβ, the leucine of ERα at the "top" of the binding site ("top" refers to the β side of the steroid ring) is replaced by a flexible, sterically less demanding methionine, whereas

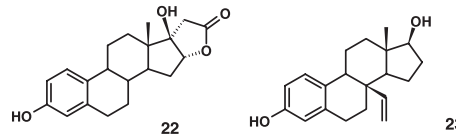


Fig. 9. The estradiol analogs **22** (40% of estradiol activity, ERα-selective) and **23** (50% of estradiol activity, ERβ-selective) have been designed as selective ERα and ERβ receptor ligands. Even though they are less active than estradiol, they show 300-fold and 190-fold selectivity for the different receptor subtypes

at the "bottom" of the binding site, close to ring D, a methionine in ERα is replaced by an isoleucine in ERβ. Using this information on the narrower binding pocket above and below the estradiol binding sites of ERα and ERβ, respectively, the selective ligands **22** and **23** could be designed (Fig. 9) (Hillisch et al. 2004a–c). Whereas **22** has only about 40% of the activity of estradiol at ERα, it shows a 300-fold selectivity against ERβ; on the other hand, compound **23** has only 50% of the activity of estradiol at ERβ but a 190-fold selectivity against ERα.

The thyroid hormone T3 and its less active storage form T4 are iodinated phenoxy-phenylalanines, which bind to two nuclear receptor subtypes TRα and TRβ. Unfortunately, the affinity of T3 to TRα is higher than to TRβ, which causes cardiac side effects, if hypothyroid patients are treated with T3. The alkyl analogs **24** and **25** are less active at TRα than at TRβ (Fig. 10) (Scanlan et al. 2001). Compound **26** binds to both receptor subtypes but has no agonistic activity at TRα and is only a weak partial agonist at TRβ; correspondingly, this compound might be used to treat hyperthyroid patients (Baxter et al. 2002). Other patients suffer from a R320C mutant of TRβ; due to the exchange of the strongly basic arginine side chain against the neutral cysteine, T3 binds with much lower affinity to this receptor, causing a hypothyroid condition. Treatment with T3 or compound **25** is impossible, due to the high affinity of these compounds to the TRα receptor. Conversion of the acid **25** into the neutral analog **27** solved the problem: **27** has a higher affinity to the TRβ mutant than to TRα (Fig. 10) (Ye et al. 2001).

Integrins are another group of receptors. They are expressed at cell surfaces and their endogenous ligands, e.g., fibrinogen at the GP IIb/IIIa integrin (also called fibrinogen receptor) or vitronectin at the $\alpha_v\beta_3$ in-

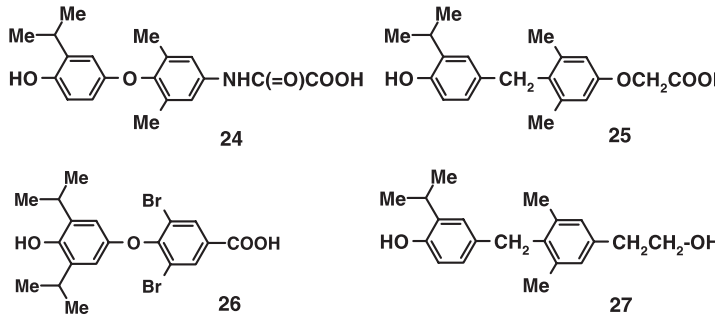


Fig. 10. Compounds **24** (CGS-23425) and **25** (GC-1, UCSF) are alkyl analogs of the thyroid hormone T3; in contrast to T3, which has a higher activity at TRα, these analogs have a higher activity at the TRβ. Compound **26** is a thyroid hormone antagonist at TRα and a weak partial agonist at TRβ. Neither T3 (EC_{50} hTRα = 0.14 nM, EC_{50} hTRβ = 0.66 nM, EC_{50} hTRβ R320C mutant = 4.3 nM) nor compound **25** (EC_{50} hTRα = 6.6 nM, EC_{50} hTRβ = 3.7 nM, EC_{50} hTRβ R320C mutant = 38 nM) have sufficient activity at a hTRβ R320C mutant. Compound **27** is a neutral, weakly active but TRβ R320C mutant-selective thyromimetic (EC_{50} hTRα = 38 nM, EC_{50} hTRβ = 32 nM, EC_{50} hTRβ R320C mutant = 7.0 nM)

tegrin (also called vitronectin receptor), mediate cell–cell contacts. The recognition motif of these two receptors is the Arg-Gly-Asp (RGD) sequence of the ligands, obviously in different conformations. Research at SmithKline Beecham led to the discovery of ligands that showed, after minor chemical modification of a basic side chain, some selectivity

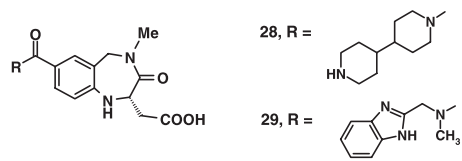


Fig. 11. Compound **28** (lotrafiban, K_i GP IIb/IIIa = 2.5 nM, K_i $\alpha_v\beta_3$ = 10,340 nM; failed in phase III clinical trials) is a specific fibrinogen receptor antagonist, whereas compound **29** (K_i GP IIb/IIIa = 30,000 nM, K_i $\alpha_v\beta_3$ = 2 nM) is a specific vitronectin receptor antagonist

for each of these two receptors (Samanen et al. 1996). After extensive structural modification, the highly selective ligands **28** (SB 214 857, lotrafiban) and **29** (SB 223 245) resulted (Fig. 11) (Samanen et al. 1996; Keenan et al. 1997; Miller et al. 2000). They differ in their selectivity by more than seven orders of magnitude.

1.5 Summary and Conclusions

Chemical similarity principles and bioisosterism are the guidelines of structural modification in classical medicinal chemistry. However, sometimes chemically similar compounds show very different biological activities and/or selectivities (Kubinyi 1998). In the early years of combinatorial chemistry, its potential output was significantly overestimated. An unprecedented number of new drugs was expected from chemistry-driven combinatorial syntheses. However, the output was just zero; sheer numbers did not contribute to drug discovery. Using the comparison of the "drug islands in an ocean," combinatorial chemistry was far away from those islands. The technology was able to deliver active analogs and to speed up drug discovery only after significant evolution. Instead of a combinatorial production of thousands of meaningless compounds, often in undefined mixtures, parallel syntheses of smaller libraries of single, purified compounds are now performed, driven by medicinal chemistry. In this manner, combinatorial chemistry is especially valuable in the very first steps of screening hit exploitation and lead structure optimization, in order to derive first structure–activity relationships (SARs) and to improve affinity, selectivity, and ADME properties to a certain point.

Chemogenomics is a complementary strategy for the investigation of chemically related compounds and libraries against various members of a target family. It is largely based on the proper application of automated parallel synthesis. The advantages of such a systematic approach are manifold:

- Specific analogs within a target family are discovered more easily.
- Results from one target may be used to explore a related target.
- Different subtype selectivities may be observed.
- Structure–activity relationships (SARs) result earlier.

– Coverage of chemical space and therefore patent coverage is more complete.

Of course, other rational approaches, such as molecular modeling, pharmacophore searches, virtual screening, and structure-based ligand design support this new strategy. The final steps of drug optimization will always need dedicated structural modifications, following the accumulated know-how of classical medicinal chemistry.

Only a few examples of chemogenomic applications could be discussed in this review. More illustrative applications are presented in a recent monograph on chemogenomics in drug discovery (Kubinyi and Müller 2004).

References

Austin CP, Brady LS, Insel TR, Collins FS (2004) NIH molecular libraries initiative. Science 306:1138–1139

Baxter JD, Goede P, Apriletti JW, West BL, Feng W, Mellstrom K, Fletterick RJ, Wagner RL, Kushner PJ, Ribeiro RC, Webb P, Scanlan TS, Nilsson S (2002) Structure-based design and synthesis of a thyroid hormone receptor (TR) antagonist. Endocrinology 143:517–524

Bishop AC, Ubersax JA, Petsch DT, Matheos DP, Gray NS, Blethrow J, Shimizu E, Tsien JZ, Schultz PG, Rose MD, Wood JL, Morgan DO, Shokat KM (2000) A chemical switch for inhibitor-sensitive alleles of any protein kinase. Nature 407:395–401

Bleicher KH (2002) Chemogenomics: bridging a drug discovery gap. Curr Med Chem 9:2077–2084

Bymaster FP, Calligaro DO, Falcone JF, Marsh RD, Moore NA, Tye NC, Seeman P, Wong DT (1996) Radioreceptor binding profile of the atypical antipsychotic olanzapine. Neuropsychopharmacology 14:87–96

Bymaster FP, Nelson DL, DeLapp NW, Falcone JF, Eckols K, Truex LL, Foreman MM, Lucaites VL, Calligaro DO (1999) Antagonism by olanzapine of dopamine D_1, serotonin$_2$, muscarinic, histamine H_1, and α_1-adrenergic receptors in vitro. Schizophr Res 37:107–22

Capdeville R, Buchdunger E, Zimmermann J, Matter A (2002) Glivec (ST571, Imatinib), a rationally developed, targeted anticancer drug. Nature Rev Drug Discov 1:493–502

Caron PR, Mullican MD, Mashal RD, Wilson KP, Su MS, Murcko MA (2001) Chemogenomic approaches to drug discovery. Curr Opin Chem Biol 5:464–470

Ding S, Wu TY, Brinker A, Peters EC, Hur W, Gray NS, Schultz PG (2003) Synthetic small molecules that control stem cell fate. Proc Natl Acad Sci U S A 100:7632–7637

Evans BE, Rittle KE, Bock MG, DiPardo RM, Freidinger RM, Whitter WL, Lundell GF, Veber DF, Anderson PS, Chang RSL, Lotti VJ, Cerino DJ, Chen TB, Kling PJ, Kunkel KA, Springer JP, Hirshfield J (1988) Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. J Med Chem 31:2235–2246

Hillisch A, Peters O, Kosemund D, Müller G, Walter A, Elger W, Fritzemeier KH (2004a) Protein structure-based design, synthesis strategy and in vitro pharmacological characterization of estrogen receptor α and β selective compounds. Ernst Schering Res Found Workshop 46:47–62

Hillisch A, Peters O, Kosemund D, Müller G, Walter A, Schneider B, Reddersen G, Elger W, Fritzemeier KH (2004b) Dissecting physiological roles of estrogen receptor α and β with potent selective ligands from structure-based design. Mol Endocrinol 18:1599–1609

Hillisch A, Pineda LF, Hilgenfeld R (2004c) Utility of homology models in the drug discovery process. Drug Discov Today 9:659–669

Jacoby E, Schuffenhauer A, Floersheim P (2003) Chemogenomics knowledge-based strategies in drug discovery. Drug News Perspect 16:93–102

Keenan RM, Miller WH, Kwon C, Ali FE, Callahan JF, Calvo RR, Hwang SM, Kopple KD, Peishoff CE, Samanen JM, Wong AS, Yuan CK, Huffman WF (1997) Potent, selective, orally active 3-oxo-1,4-benzodiazepine GPIIb/IIIa integrin antagonists. J Med Chem 40:2289–2292

Kubinyi H (1998) Similarity and dissimilarity – a medicinal chemist's view. In: Kubinyi H, Folkers G, Martin YC (eds) 3D QSAR in drug design, vol. II. Ligand–protein interactions and molecular similarity. Kluwer/ESCOM, Dordrecht, 1998, pp 225–252; also published in Persp Drug Design Discov 1998, 9–11:225–252

Kubinyi H (2004) Drug discovery from side effects. In: Kubinyi H, Müller G, (eds) Chemogenomics in drug discovery – a medicinal chemistry perspective, vol. 22 of Methods and principles in medicinal chemistry. Mannhold R, Kubinyi H, Folkers G (eds) Wiley-VCH, Weinheim, pp 43–67

Kubinyi H, Müller G (eds) (2004) Chemogenomics in drug discovery – a medicinal chemistry perspective, vol. 22, Methods and principles in medicinal chemistry. Mannhold R, Kubinyi H, Folkers G (eds) Wiley-VCH, Weinheim

Lipinski C, Hopkins A (2004) Navigating chemical space for biology and medicine. Nature 432:855–861

Lopez-Rodriguez ML, Morcillo MJ, Benhamu B, Rosado ML (1997) Comparative receptor mapping of serotoninergic 5-HT_3 and 5-HT_4 binding sites. J Comput-Aided Mol Design 11:589–599

Mayer TU, Kapoor TM, Haggarty SJ, King RW, Schreiber SL, Mitchison TJ (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. Science 286:971–974

McGovern SL, Shoichet BK (2003) Kinase inhibitors: not just for kinases anymore. J Med Chem 46:1478–1483

McGovern SL, Caselli E, Grigorieff N, Shoichet BK (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem 45:1712–1722

McGovern SL, Helfand BT, Feng B, Shoichet BK (2003) A specific mechanism of nonspecific inhibition. J Med Chem 46:4265–4272

Miller WH, Keenan RM, Willette RN, Lark MW (2000) Identification and in vivo efficacy of small-molecule antagonists of integrin $\alpha_v\beta_3$ (the vitronectin receptor). Drug Discov Today 5:397–408

Müller G (2003) Medicinal chemistry of target family-directed masterkeys. Drug Discov Today 8:681–691

Müller G (2004) Target family-directed masterkeys in chemogenomics. In: Kubinyi H, Müller G (eds) Chemogenomics in drug discovery – a medicinal chemistry perspective, vol. 22 of Methods and principles in medicinal chemistry. Mannhold R, Kubinyi H, Folkers G (eds) Wiley-VCH, Weinheim, pp 7–41

Patchett AA, Nargund RP (2000) Privileged structures – an update. Annu Rep Med Chem 35:289–298

Rival Y, Hoffmann R, Didier B, Rybaltchenko V, Bourguignon J-J, Wermuth CG (1998) 5-HT_3 Antagonists derived from aminopyridazine-type muscarinic M_1 agonists. J Med Chem 41:311–317

Rohrer SP, Birzin ET, Mosley RT, Berk SC, Hutchins SM, Shen DM, Xiong Y, Hayes EC, Parmar RM, Foor F, Mitra SW, Degrado SJ, Shu M, Klopp JM, Cai SJ, Blake A, Chan WW, Pasternak A, Yang L, Patchett AA, Smith RG, Chapman KT, Schaeffer JM (1998) Rapid identification of subtype-selective agonists of the somatostatin receptor through combinatorial chemistry. Science 282:737–740; erratum Science 282:1646

Römer D, Büscher HH**,** Hill RC, Maurer R, Petcher TJ, Zeugner H, Benson W, Finner E, Milkowski W, Thies PW (1982) An opioid benzodiazepine. Nature 298:759–760

Russell K, Michne WF (2004) The value of chemical genetics in drug discovery. In: Kubinyi H, Müller G (eds) Chemogenomics in drug discovery. Wiley-VCH, Weinheim, pp 69–96

Samanen JM, Ali FE, Barton LS, Bondinell WE, Burgess JL, Callahan JF, Calvo RR, Chen W, Chen L, Erhard K, Feuerstein G, Heys R, Hwang S-M, Jakas DR, Keenan RM, Ku TW, Kwon C, Lee C-P, Miller WH, Newlander KA, Nichols A, Parker M, Peishoff CE, Rhodes G, Ross S,

Shu A, Simpson R, Takata D, Yellin TO, Uzsinskas I, Venslavsky JW, Yuan CK, Huffman WF (1996) Discovery of potent nonpeptide vitronectin receptor ($\alpha_v\beta_3$) antagonists. J Med Chem 39:4867–4870

Scanlan TS, Yoshihara HA, Nguyen N-H, Chiellini G (2001) Selective thyromimetics: tissue-selective thyroid hormone analogs. Curr Opin Drug Disc Dev 4:614–622

Schreiber SL (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. Science 287:1964–1969

Schreiber SL (2003) The small-molecular approach to biology. C&EN 81:51–61; cf. Rouhi AM (2003) Moving beyond natural products. C&EN 81:104–107

Seidler J, McGovern SL, Doman TN, Shoichet BK (2003) Identification and prediction of promiscuous aggregating inhibitors among known drugs. J Med Chem 46:4477–4486

Slusarchyk WA, Robl JA, Taunk PC, Asaad MM, Bird JE, DiMarco J, Pan Y (1995) Dual metalloprotease inhibitors. V. Utilization of bicyclic azepinonethiazolidines and azepinonetetrahydrothiazines in constrained peptidomimetics of mercaptoacyl dipeptides. Bioorg Med Chem Lett 5:753–758

Sneader W (1996) Drug prototypes and their exploitation. John Wiley, Chichester

Stockwell BR (2004) Exploring biology with small organic molecules. Nature 432:846–854

Tan DS, Foley MA, Shair MD, Schreiber SL (1998) Stereoselective synthesis of over two million compounds having structural features both reminiscent of natural products and compatible with miniaturized cell-based assays. J Am Chem Soc 120:8565–8566

Wermuth CG (2001) The "SOSA" approach: an alternative to high-throughput screening. Med Chem Res 10:431–439

Wermuth CG (2004) Selective optimization of side activities: another way for drug discovery. J Med Chem 47:1303–1314

Yamamoto M, Ikeda S, Kondo H, Inoue S (2002) Design and synthesis of dual inhibitors for matrix metalloproteinase and cathepsin. Bioorg Med Chem Lett 12:375–378

Ye HF, O'Reilly KE, Koh JT (2001) A subtype-selective thyromimetic designed to bind a mutant thyroid hormone receptor implicated in resistance to thyroid hormone. J Am Chem Soc 123:1521–1522

2 Chemogenomics Strategies for G-Protein Coupled Receptor Hit Finding

W. Guba

Abstract. Targeting protein superfamilies via chemogenomics is based on a similarity clustering of gene sequences and molecular structures of ligands. Both target and ligand clusters are linked by generating binding affinity profiles of chemotypes vs a target panel. The application of this multidimensional similarity paradigm will be described in the context of Lead Generation to identify novel chemical hit classes for G-protein coupled receptors.

2.1 Introduction

The term "chemogenomics" appeared first in 2000 in a Vertex press release (Vertex Pharmaceuticals Inc.), where it was described as an approach to "rapidly and simultaneously design multiple lead classes

of drugs directed at protein targets in gene families." In a follow-up landmark publication in 2001 (Caron et al. 2001), chemogenomics was defined as "the discovery and description of all possible drugs to all possible drug targets." In 2001, there was only one further publication from E. Jacoby (Jacoby 2001), which mentioned chemogenomics in the title. Since then the number of chemogenomics-related publications has risen steadily and this concept has become so popular in the pharmaceutical industry that now a book titled *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective* (Kubinyi and Müller 2004) has been devoted to this topic.

However, there is neither a precise definition of chemogenomics nor a clear differentiation from chemical genetics and chemical genomics. Nevertheless, chemogenomics is claimed to be an integral component in industrial drug discovery with sometimes different wordings (e.g., Sanofi-Aventis' "chemical biology"). The common theme of chemogenomics-labeled approaches is the focus on protein superfamilies (GPCRs, kinases, proteases, ligand-gated ion channels, nuclear hormone receptors, etc.) to exploit cross-target synergies by capitalizing on family-specific recognition motifs for ligand and library design (Frye 1999; Savchuk et al. 2004; Müller 2003; Bleicher 2002; Jacoby et al. 1999, 2003; Bredel and Jacoby 2004; Schuffenhaurer et al. 2002; Crossley 2004). Chemogenomics at Roche is not considered as a technology platform but it is an integrative process to accelerate and systematize targeted drug discovery by cross-linking classification and annotation schemes of chemical, biological, and genetic data.

2.2 G-Protein Coupled Receptors as Drug Targets

The following discussion will focus on G-protein coupled receptors (GPCRs) as a key target family of special interest to Roche. GPCRs represent the most attractive gene family for drug discovery covering all major therapeutic indications with a proportion of roughly 25% among the top 100 selling drugs. Approximately 350 GPCRs are considered as possible drug targets, 150 of which are still orphan GPCRs. Currently, marketed drugs act on less than 10% of all known nonsensory GPCRs; therefore, this target family offers a huge potential for new and improved

therapies to treat human disease. The interest in GPCRs as drug targets is reflected by the fact that they account for approximately 30% of all drug discovery projects across the pharmaceutical industry (Wise et al. 2002; Porter and Alanine 2005).

Pharmaceutically relevant GPCRs are divided into three main families, which mainly differ in their extracellular N-terminal domains, but they all share a common motif, the seven transmembrane helical domains. Family A (rhodopsin-like) contains roughly 90% of all GPCRs and includes many validated drug targets for therapeutic intervention such as biogenic amine, opioid and purinergic receptors. GPCRs are activated by a broad range of extracellular effector molecules and trigger a variety of signaling cascades to modulate cell function. So far, only the X-ray structure of bovine rhodopsin (Palezewski et al. 2000) is available, which serves as a template for defining a consensus drug-binding site within the seven helix bundle. The amino acids surrounding this common binding site are of pivotal importance for classifying GPCRs and defining structural motifs for ligand–target recognition (Jacoby 2001; Jacoby et al. 1999; Crossley 2004).

2.3 The Similarity Principle

A common strategy in focused library design is the utilization of the well-known similarity principle (Maggiora and Johnson 1999), i.e., structurally similar molecules are likely to have similar biological properties. However, experience has shown (Martin et al. 2002) that this empirical guideline often fails since minor structural modifications may abolish the pharmacological activity of a ligand. Furthermore, virtual screening campaigns that are based on topological similarity to known active seed compounds do not tend to retrieve novel molecular architectures unless the molecular similarity descriptors allow for a sufficient degree of fuzziness. Less stringent similarity requirements imply a higher number of compounds to be submitted to biological testing and, consequently, a lower hit rate due to an increasing number of inactive molecules. The combination of 3D pharmacophore information and topological constraints was demonstrated to successfully identify novel chemical hit classes from a small compound collection without the need for HTS

resources (Guba et al. 2005). However, this approach cannot be transformed into a general GPCR hit finding strategy.

2.4 Chemogenomics: A Multidimensional Similarity Paradigm

Chemogenomics moves beyond the traditional similarity principle by cross-linking chemotype (molecular structures) with biotype (biological activity patterns) and target families (sequence space). Bioinformatics is providing tools for pattern recognition to classify protein superfamilies by phylogenetic relationships. The clustering of chemical structures is the realm of cheminformatics where subsets of similar molecules are formed by describing molecular similarity in terms of topological features, substructures, pharmacophoric patterns, and 3D molecular fields. Sequence space and chemotype are linked by the biotype, i.e.,

Molecule	Biogenic Amine Receptors									Peptide GPCRs				others	
1	6.1	7.1	4.7	8.8	7.2	6.9	7.7	5.4	5.6	4.0	4.2	4.5	5.5	4.2	4.6
2	7.0	4.4	4.2	4.3	4.9	5.8	5.8	6.1	4.4	4.0	4.8	7.7	4.8	4.2	4.1
3	5.6	4.4	4.2	4.8	4.4	6.5	6.0	5.7	5.1	4.0	5.3	7.7	5.8	4.2	4.1
4	6.3	5.0	4.2	4.3	4.4	6.1	4.8	5.5	4.4	4.7	5.0	5.7	5.4	4.6	4.1

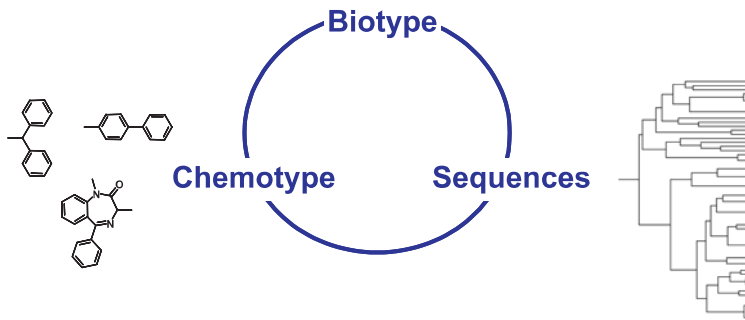


Fig. 1. Chemogenomics is based on a multidimensional similarity paradigm. Protein sequences and molecular structures are grouped with appropriate similarity metrics and both target and ligand clusters are linked by ligand–target annotations such as binding affinity profiles (Biotype)

the bioactivity profile of compounds vs members of target families. As depicted in Fig. 1, chemogenomics is providing the framework to link these three components by classification and annotation schemes. The one-dimensional molecular similarity principle is replaced by a multi-dimensional similarity paradigm enabling a novel strategy in GPCR hit finding as outlined below.

Weinstein et al. (1997) and Kauvar et al. (1995) pioneered affinity fingerprints, the measurement of binding affinities toward a reference panel of proteins. The biological profiles are used as similarity descriptors in lead generation (Dixon and Villa 1998; Beroza et al. 2002) and

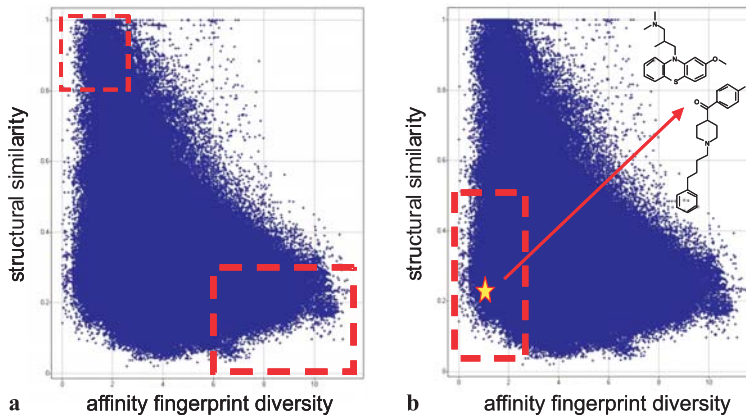


Fig. 2. a The relationship between structural similarity and affinity fingerprint diversity is visualized by plotting the Tanimoto coefficient (Daylight fingerprints [Daylight Chemical Information Systems, Inc. 2002]) of a compound pair against the Euclidean distance of the corresponding affinity profiles. The indicated region in the *upper left* indicates the similarity principle, i.e., similar structures have similar bioactivity profiles, whereas the region in the *lower right* highlights high structural dissimilarity and affinity fingerprint diversity. **b** The *dashed rectangle* highlights a large number of chemically diverse compound pairs with a similar bioactivity profile, as illustrated by two diverse structures with similar affinity fingerprints. Thus, the reverse of the similarity principle, i.e., similar biological properties determine structural similarity, is not valid. In conclusion, similarity searches based on affinity fingerprints are able to retrieve chemically diverse ligands

for establishing quantitative relationships between chemical structures and biological activity spectra (Krejsa et al. 2003; Fliri et al. 2005). Within the context of the GPCR chemogenomics initiative at Roche, approximately 5,000 compounds were submitted to Cerep (2005) for measuring binding affinities against a panel of 15 GPCRs. Further details concerning the design of the GPCR panel library, selection of receptors and binding data will be given in a forthcoming publication. The relationship between structural similarity and affinity fingerprint diversity was systematically analyzed by neighborhood plots (Patterson et al. 1996), where the Tanimoto coefficient (Daylight fingerprints [Daylight Chemical Information Systems, Inc. 2002]) of a compound pair was plotted against the Euclidean distance of the corresponding affinity profiles (Fig. 2a). The triangular shape of the plot is indicative of the similarity principle, i.e., similar structures imply similar bioactivity profiles and affinity fingerprint diversity increases with structural dissimilarity. However, the plot in Fig. 2b also illustrates that there is a considerable proportion of chemically diverse structures with a similar bioactivity profile. Thus, the reverse of the similarity principle, i.e., similar biological properties determine structural similarity, is not valid. As a conclusion, similarity searches based on affinity fingerprints are able to retrieve chemically diverse ligands, and in the following section a chemogenomics strategy for GPCR hit finding will be outlined.

2.5 Chemogenomics Strategy for G-Protein Coupled Receptor Hit Finding

The most challenging scenario in a Lead Generation project is to target a receptor without any information about small molecule ligands. In the case of GPCRs, the first step is to identify phylogenetically related receptors for which lead-like compounds with pharmacological activity are known. These molecules will be called seeds. Following the traditional medicinal chemistry approach, various similarity metrics (e.g., topological or pharmacophore descriptors) would be applied to identify a focused screening subset from the corporate compound depository containing compounds that are similar to the seeds. In order to enhance the probability of identifying structurally novel hits, the follow-

ing scheme has been devised as a consequence of the conclusion drawn from the above analysis of the neighborhood plots. The seeds are first profiled against the reference panel of GPCRs. Then, the affinity profiles of the seeds are compared (e.g., by Euclidean distance) to the profiles of the GPCR panel screen library, and the most closely related molecules are selected for screening. In a recent Roche Lead Generation project, less than 5% of the library was selected, with a hit rate of 24% covering several chemical classes.

2.6 Conclusion

Targeting protein superfamilies via chemogenomics is based on a multidimensional similarity paradigm. Protein sequences and molecular structures are grouped with appropriate similarity metrics and both target and ligand clusters are linked by ligand–target annotations. The comparison of affinity fingerprints of seed structures with a reference library of GPCR chemotypes, which have been profiled against the same target panel, is an efficient and cost-saving route to novel hit structures compared to high-throughput screening. Chemogenomics is not a technology but a process to gather biological data, to enrich the corporate compound depository with target-specific chemotypes, and to provide for a data warehouse to link ligand with target information. Chemogenomics will have a sustained impact on systematizing drug discovery and will play a pivotal role in the ongoing quest for new and better medicines.

Acknowledgements. It is with real pleasure that I wish to thank all my colleagues whose contributions and support made the described work possible and enjoyable, especially Drs. O. Roche, N. Kratochwil, C. Bissantz, H. Mauser, K. Bleicher, L. Green, R. Martin, M. Rogers-Evans, A. Alanine, and M. Stahl. For the same reasons I would also like to express my gratitude to Drs. C. Krejsa and M.-C. Bodinier from Cerep.

References

Beroza P, Villar HO, Wick MM, Martin GR (2002) Chemoproteomics as a basis for post-genomic drug discovery. Drug Discov Today 7:807–814

Bleicher KH (2002) Chemogenomics: bridging a drug discovery gap. Curr Med Chem 9:2077–2084
Bredel M, Jacoby E (2004) Chemogenomics: an emerging strategy for rapid target and drug discovery. Nat Rev Genet 5:262–275
Caron PR, Mullican MD, Mashal RD, Wilson KP, Su MS, Murcko MA (2001) Chemogenomic approaches to drug discovery. Curr Opin Chem Biol 5:464–470
Cerep (2005) http://www.cerep.fr/Cerep/Users/index.asp
Crossley R (2004) The design of screening libraries targeted at G-protein coupled receptors. Curr TopMed Chem 4:581–588
Daylight Chemical Information Systems, Inc., http://www.daylight.com (2002)
Dixon SL, Villar HO (1998) Bioactive diversity and screening library selection via affinity fingerprinting. J Chem Inf Comput Sci 38:1192–1203
Fliri AF, Loging WT, Thadeio PF, Volkmann RA (2005) Biological spectra analysis: linking biological activity profiles to molecular structure. Proc Natl Acad Sci U S A 102:261–266
Frye SV (1999) Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. Chem Biol 6:R3–R7
Guba W, Neidhart W, Nettekoven M (2005) Novel and potent NPY5 receptor antagonists derived from virtual screening and iterative parallel chemistry design. Bioorg Med Chem Lett 15:1599–1603
Jacoby E (2001) Quant Struc-Act Relat 20:115–123
Jacoby E, Fauchere J-L, Raimbaud E, Ollivier S, Michel A, Spedding M (1999) A three binding site hypothesis for the interaction of ligands with monoamine G protein-coupled receptors: implications for combinatorial ligand design. Quant Struct-Act Relat 18:561–571
Jacoby E, Schuffenhauer A, Floersheim P (2003) Chemogenomics knowledge-based strategies in drug discovery. Drug News Perspect 16:93–102
Kauvar LM, Higgins DL, Villar HO, Sportsman JR, Engqvist-Goldstein A, Bukar R, Bauer KE, Dilley H, Rocke DM (1995) Predicting ligand binding to proteins by affinity fingerprinting. Chem Biol 2:107–118
Krejsa CM, Horvath D, Rogalski SL, Penzotti JE, Mao B, Barbosa F, Migeon JC (2003) Predicting ADME properties and side effects: the BioPrint approach. Curr Opin Drug Discov Devel 6:470–480
Kubinyi H, Müller G (2004) (eds) Chemogenomics in drug discovery. Wiley-VCH, Weinheim
Maggiora GM, Johnson MA (1990) (eds) Concepts and applications of molecular similarity. Wiley-Interscience, New York
Martin YC, Kofron JL, Traphagen LM (2002) Do structurally similar molecules have similar biological activity? J Med Chem 45:4350–4358

Müller G (2003) Medicinal chemistry of target family-directed masterkeys. Drug Discov Today 8:681–691

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Trong IL, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. Science 289:739–745

Patterson DE, Cramer RD, Ferguson AM, Clark RD, Weinberger LE (1996) Neighborhood behavior: a useful concept for validation of "molecular diversity" descriptors. J Med Chem 39:3049–3059

Porter R, Alanine A (2005) Screening 1:21–23

Savchuk NP, Balakin KV, Tkachenko SE (2004) Exploring the chemogenomic knowledge space with annotated chemical libraries. Curr Opin Chem Biol 8:412–417

Schuffenhauer A, Zimmermann J, Stoop R, Van der Vyver J-J, Lecchini S, Jacoby E (2002) An ontology for pharmaceutical ligands and its application for in silico screening and library design. J Chem Inf Comput Sci 42:947–955

Vertex Pharmaceuticals Incorporated. http://www.vpharm.com

Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ Jr, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, van Osdol WW, Monks AP, Scudiero DA, Sausville EA, Zaharevitz DW, Bunow B, Viswanadhan VN, Johnson GS, Wittes RE, Paull KD (1997) An information-intensive approach to the molecular pharmacology of cancer. Science 275:343–349

Wise A, Gearing K, Rees S (2002) Target validation of G-protein coupled receptors. Drug Discov Today 7:235–246

3 Chemogenomics Approaches to G-Protein Coupled Receptor Lead Finding

T. Klabunde, R. Jäger

Abstract. G-protein coupled receptors (GPCRs) are promising targets for the discovery of novel drugs. In order to identify novel chemical series, high-throughput screening (HTS) is often complemented by rational chemogenomics lead finding approaches. We have compiled a GPCR directed screening set by ligand-based virtual screening of our corporate compound database. This set of compounds is supplemented with novel libraries synthesized around proprietary scaffolds. These target-directed libraries are designed using the knowledge of privileged fragments and pharmacophores to address specific GPCR subfamilies (e.g., purinergic or chemokine-binding GPCRs). Experimental testing of the GPCR collection has provided novel chemical series for several GPCR targets including the adenosine A1, the $P2Y_{12}$, and the chemokine CCR1 receptor. In addition, GPCR sequence motifs linked to the recognition of GPCR ligands

(termed chemoprints) are identified using homology modeling, molecular docking, and experimental profiling. These chemoprints can support the design and synthesis of compound libraries tailor-made for a novel GPCR target.

3.1 Introduction

GPCRs are transmembrane (TM) receptors comprised of seven TM helices. Activated by binding of their physiological ligand an extracellular signal is transformed by conformational changes of the receptor into an intracellular response. GPCR ligands are of diverse nature: these can be small molecules such as biogenic amines, amino acids, lipids, nucleosides, and nucleotides, small peptides, but also proteins such as chemokines. Excluding olfactory receptors, the current estimate of GPCRs in the human genome is about 400 (Venter et al. 2001; Wise et al. 2002). Among the 100 top-selling drugs, 25 act on GPCRs (Klabunde and Hessler 2002). Currently marketed drugs only address 30 GPCRs, only a small fraction of the GPCR target space. Thus there are 370 further GPCRs representing promising drug targets and providing excellent opportunities for successful drug discovery programs. Therefore, most pharmaceutical companies invest heavily in this target family following different approaches to identify novel GPCR lead series. Often high-throughput screening (HTS) is complemented with rational chemogenomics lead-finding approaches.

Several recent review articles have been published on this novel drug discovery approach, aiming for a clear definition and trying to set this concept into perspective (Bredel and Jacoby 2004; Jimonet and Jäger 2004; Caron et al. 2001). A shift of pharmaceutical research from traditional target-specific case-by-case studies to a cross-target view might be considered as the common underlying idea. Following the chemogenomics concept, targets are no longer viewed as individual and single entities but grouped into sets of related proteins or target families (e.g., kinases, GPCRs) that are systematically explored. Insights into family-wide commonalities of ligand recognition (e.g., privileged chemical structural motifs correlated to conserved binding sites) are translated into the design and synthesis of chemical libraries to accelerate lead finding and target validation.

Often receptors with structurally related physiological ligands can be grouped into subfamilies. Directed libraries designed based on common pharmacophores and privileged building blocks can then be exploited for all members of the respective gene family. This library design concept is applicable in those cases where the recognition site for the natural ligand is structurally conserved among the members of the respective subfamily. This is, however, often not the case, not even for receptors with the same physiological ligand. There is more and more evidence that the activity of most GPCRs can be modulated by ligands binding into a site of the receptor, similar to the one used by retinol in rhodopsin (Klabunde and Hessler 2002). The site is located within the seven-TM domain and is used by many natural ligands, like the biogenic amines for receptor binding and activation. In addition, it has been shown that many small molecule surrogate ligands of peptide and chemokine binding GPCRs appear to use this binding site to modulate the activity of their particular receptors (Klabunde and Hessler 2002). In the context of library design and screening set compilation, it thus appears appropriate to compare and classify GPCRs based on the similarity of their putative ligand binding site or subpockets thereof, as this is the best indication that two receptors would recognize structurally related ligands or similar ligand fragments. In order to support this classification method, thorough analysis and recognition of the ligand–receptor interactions is thus of utmost importance. Although sufficient and reliable 3D structural information on GPCRs and their ligand complexes is still lacking, homology modeling and molecular recognition studies (mainly site-directed mutagenesis and affinity labeling) have provided valuable insights into GPCR structure and ligand–receptor interactions in the recent years (Shi and Javitch 2002). Typical receptor subsites for ligand binding or fingerprints on the primary sequence level for binding of the natural and surrogate ligands have been identified (Bondensgaard et al. 2004; Jacoby et al. 1999).

In this work we describe the design of the companies GPCR targeted compound collection and disclose the first screening results. The collection was compiled by virtual screening of the corporate compound collection and supplemented with libraries around proprietary scaffolds, termed scaffold libraries. These are designed and synthesized to address specific GPCR subfamilies (e.g., chemokine receptors, purinergic

GPCRs). We illustrate how these chemogenomics approaches provided novel lead series for several GPCR projects. Furthermore, our approach to identifying ligand recognition motifs within GPCR sequences (termed chemoprints) by homology modeling, docking, and experimental profiling is described. These chemoprints can guide the selection of building blocks and scaffolds for the design of compound libraries tailor-made for a novel GPCR target.

3.2 G-Protein Coupled Receptor Compound Collection

The companies GPCR screening collection consists of compounds with structural similarity to known GPCR ligands supplemented with scaffold libraries addressing specific GPCR subfamilies (see Fig. 1). As a reference set of GPCR ligands, we have used the MDDR (http://www.mdli.com), the Aureus GPCR ligand database (http://www.aureus-pharma.com), and structural data from internal GPCR hits and leads (GPCR ligand database, Fig. 1). The MDDR covers patent literature, journals, meetings, and congresses and currently contains over 141,000 biologically relevant compounds and well-defined derivatives such as drugs

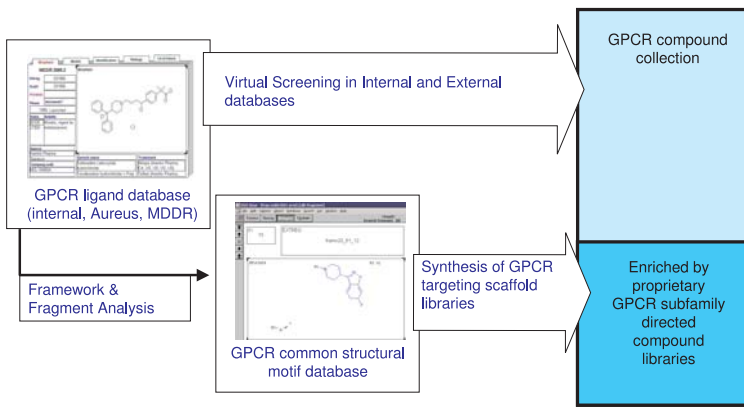


Fig. 1. Design process of the companies compound collection targeting the GPCR gene family. Details are given in the text

launched or in the development phase. The database compiled by Aureus Pharma focuses on GPCR ligands but also covers all biological data including detailed information on experimental conditions. Entries from this GPCR ligand database have been taken as probes to perform similarity searches in internal and external vendor compound collections using molecular 2D descriptors as similarity indices. Compounds with structural similarity to known GPCR ligands were compiled as a base set of the Aventis GPCR compound collection.

3.3 Scaffold Libraries Targeting Purinergic Receptors

The base set of the companies compound collection targeting the entire GPCR gene family is supplemented with proprietary scaffold libraries (as shown in Fig. 1, lower part), which are designed and synthesized to address specific GPCR subfamilies (e.g., chemokine receptors, purinergic GPCRs). The family of purinergic GPCRs represents a group of validated drug targets within a highly competitive field of GPCR research. It can be divided into P1 (binding nucleosides) and P2Y (binding nucleotides) receptors. Agonists and antagonists for these receptors have or are thought to have several therapeutic applications (Fredholm 2003). Examples of drugs targeting this GPCR subfamily are doxofylline, an antagonist of the adenosine A1 receptor, used as bronchodilator, and clopidogrel, an antagonist of the $P2Y_{12}$ receptor, a platelet aggregation inhibitor prescribed for thrombosis and for the prevention of stroke. In order to identify novel leads for receptors of this group of GPCRs, we have designed and synthesized proprietary scaffold libraries (Fig. 2a). Using known ligands of the P1 and P2Y family, we generated pharmacophore models describing the key chemical elements required for binding to the P1 and P2Y family. Figure 2a shows an example of the P2Y antagonist pharmacophore mapped onto AZD-6140, a known $P2Y_{12}$ antagonist. The 3D pharmacophore information was translated into 2D design principles for purinergic libraries (Fig. 2a, top right; example for P1 agonists). These design principles resemble a construction plan to guide the invention of novel scaffolds. The example shown illustrates the molecular requirements of a P1 agonist scaffold: (a) it reveals the positions of hydrogen bond donor and acceptor groups, (b) provides the tolerated

options for the attachment of side chains, and (c) gives guidance on how many atoms are tolerated in the bicyclic ring system. Following these constraints, chemists provided scaffold proposals meeting the molecular requirements. These were ranked by chemical feasibility, by novelty, and by the quality of their mapping on the 3D pharmacophore leading to ten scaffolds selected for synthesis. Fragments found to be privileged for the subfamily of purinergic GPCRs were used for "decoration" of the

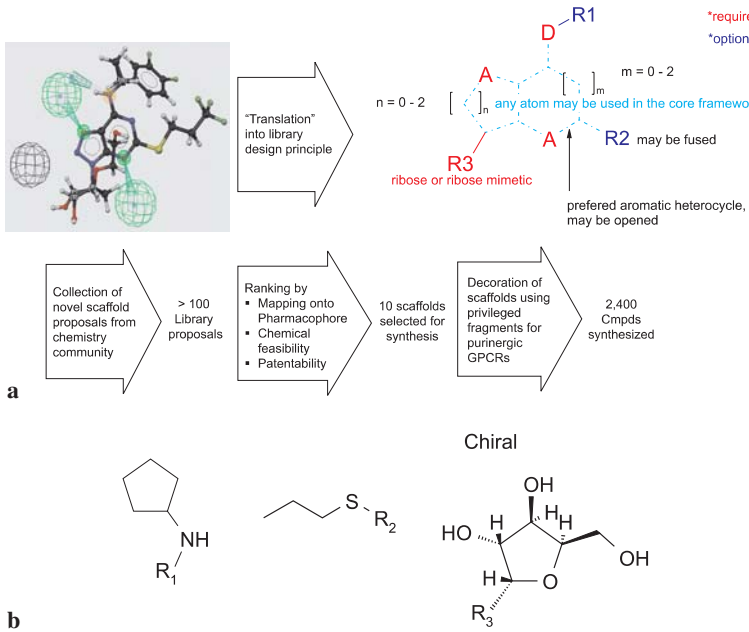
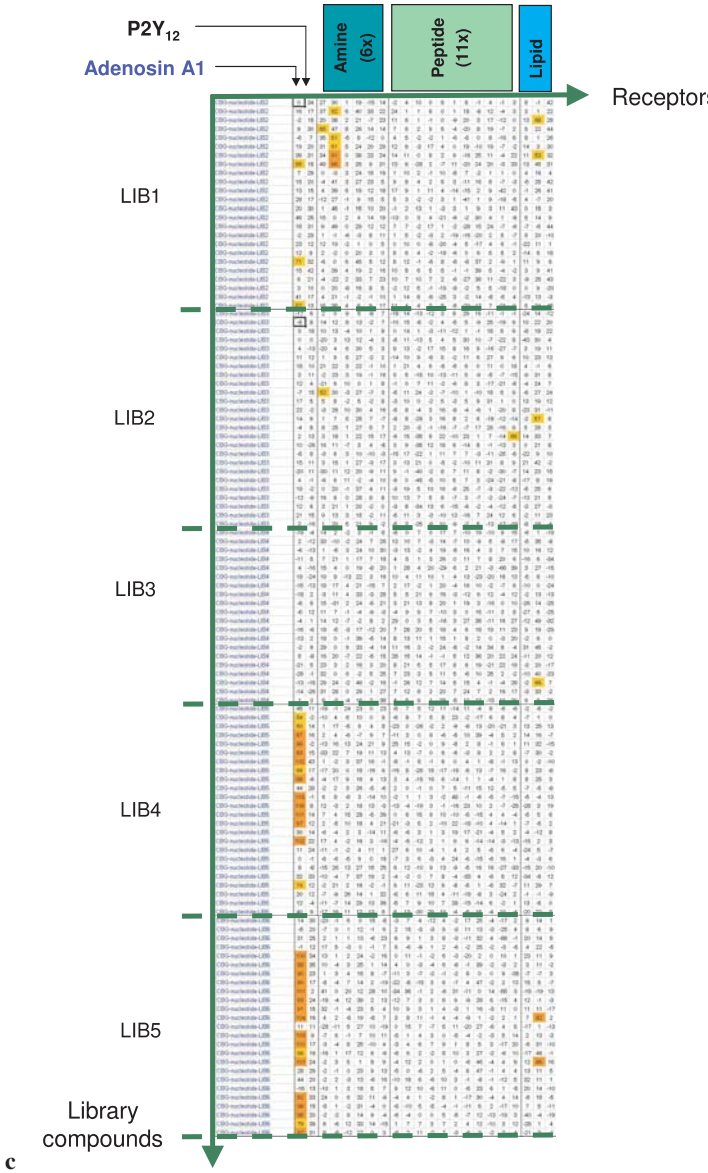


Fig. 2. a Design process for proprietary scaffold libraries targeting purinergic GPCRs. **b** Example of privileged fragments identified from known P2Y ligands and applied as building blocks for purinergic scaffold libraries. **c** Profiling of library sets targeting P1 receptors: along the *x-axis* the receptors are listed, including the adenosine A1 receptor, the $P2Y_{12}$ receptor, six biogenic amine receptors, 11 peptide-binding GPCRs and three lipid-binding GPCRs. Along the *y-axis* the library compounds are shown. Highlighted in *yellow* or *orange* are those compounds showing more than 50% or 80% inhibition of radioligand binding at a concentration of 10 μM, respectively


P2Y12
Adenosin A1
Amine (6x)
Peptide (11x)
Lipid
Receptors
LIB1
LIB2
LIB3
LIB4
LIB5
Library compounds

c

selected scaffolds: Ligands of the purinergic GPCR subfamily were extracted from the MDDR and Aureus ligand databases and were subjected to a computational retrosynthetic analysis using the RECAP algorithm (Lewell et al. 1998). The ligands were dissected into their fragments, which were sorted by the frequency of their occurrence. Fragments most commonly found among P2Yand P1 (ant)agonists were used as building blocks for library synthesis (examples of P2Y privileged fragments are shown in Fig. 2b). In total, 2,400 compounds were synthesized, all of which represent proprietary and drug-like compounds and thus offer excellent lead finding opportunities for the family of purinergic GPCRs.

The libraries designed to target the subfamily of P1 receptors were evaluated by extensive profiling against a set of 22 GPCRs using radioligand displacement assays. Of each of the five libraries, diverse subsets of 20–25 compounds were selected for profiling. The profiling result is depicted in Fig. 2c, supporting the applied design principle for P1 antagonists: (a) binding is mainly seen for the adenosine A1 receptor (some compounds show complete inhibition at a concentration of 10 μM), the only receptor of the P1 family among this receptor set; (b)

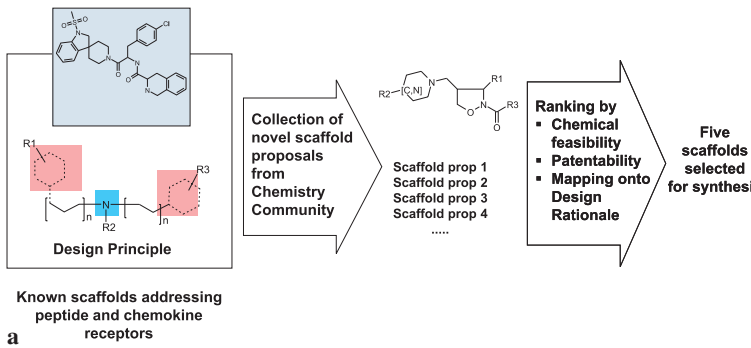
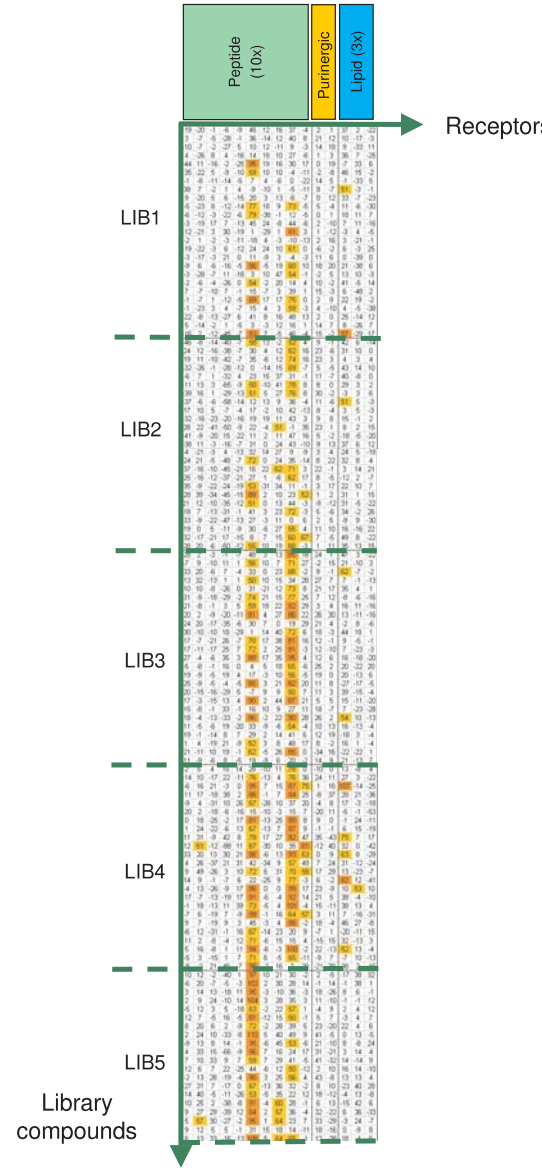


Fig. 3. a Design process for proprietary scaffold libraries targeting peptide-binding and chemokine GPCRs. **b** Profiling of library sets targeting a set of peptide-binding and chemokine receptors: along the *x-axis* the receptors are listed, including ten peptide-binding GPCRs, two purinergic GPCRs, and three lipid-binding GPCRs. Along the *y-axis* the library compounds are shown. Highlighted in *yellow* or *orange* are those compounds showing more than 50% or 80% inhibition of radioligand binding at a concentration of 10 μM, respectively


Peptide (10x)
Purinergic
Lipid (3x)
Receptors
LIB1
LIB2
LIB3
LIB4
LIB5
Library compounds

b

the affinity observed for other GPCR receptors is limited, resulting in excellent selectivity profiles for most of the compounds with A1 affinity; (c) three novel A1 antagonist scaffolds could thus be identified by profiling of subsets taken from five scaffold libraries. The complete set of ten libraries was also screened to identify A1 agonists and $P2Y_{12}$ antagonists. As a result, one novel scaffold with agonistic activity on the A1 receptor and two novel $P2Y_{12}$ antagonist scaffolds could be identified. These compounds are currently undergoing chemical optimization.

3.4 Scaffold Libraries Targeting Peptide-Binding and Chemokine Receptors

Another program to design and synthesize novel scaffold libraries targets the class of peptide- and chemokine-binding GPCRs. Looking at ligands of a subset of peptide-binding GPCRs and chemokine receptors, it becomes evident that on the one hand the capped amino acid scaffold appears to be an abundant GPCR privileged fragment (Fig. 3a). On the other hand, many peptide-binding and chemokine receptors have been shown to bind small-molecule ligands sharing a 2D pharmacophore described by a central positive ionizable group to which two hydrophobic groups are attached. As for the purinergic scaffold libraries, the idea was to take these known scaffolds as a starting point and to develop novel scaffold libraries. The provided scaffold proposals were ranked according to chemical feasibility, chances of deriving novel patentable structures, and the match on the design rationale. In the end, five scaffolds were implemented by synthesis using two to three variation points and using either solid- or solution-phase chemistry. These proprietary and lead-like scaffolds offer excellent lead-finding opportunities for chemokine- and peptide-binding GPCRs.

These libraries were used in a screening campaign on the chemokine CCR1. The primary indications for CCR1 antagonists are rheumatoid arthritis and multiple sclerosis. For both indications, the receptor is a highly validated literature target. Screening of the five scaffold libraries provided a chemical lead series with an excellent profile: (a) nanomolar functional activity in the functional in vitro assay, (b) nanomolar affinity in a radioligand binding assay, (c) submicromolar activity in a secondary

chemotaxis assay, (d) good solubility, and (e) excellent pharmacokinetic properties. Further chemical optimization of this series toward a clinical candidate is currently ongoing.

As for the purinergic GPCRs libraries, subsets of these libraries targeting peptide and chemokine receptors have been profiled against a set of 15 GPCRs using radioligand displacement assays. Of each of the five libraries, diverse subsets of 25–35 compounds were selected for profiling. The profiling result is depicted in Fig. 3b, validating the design principle for (a subset of) peptide receptors: (a) binding is mainly seen for the peptide receptors with five out of ten peptide-binding GPCRs (vasopressin V1, μ opiate, melanocortin MC3, neurokinin NK1, and angiotensin AT1 receptor), showing affinity toward these library subsets; (b) no activities are seen for the purinergic GPCRs and some unexpected affinity is found for some of the lipid-binding GPCRs; (c) two novel series with submicromolar binding affinities for the NK1 and the μ opiate receptor, and (d) one novel series with micromolar binding affinity toward the V1 receptor could be identified.

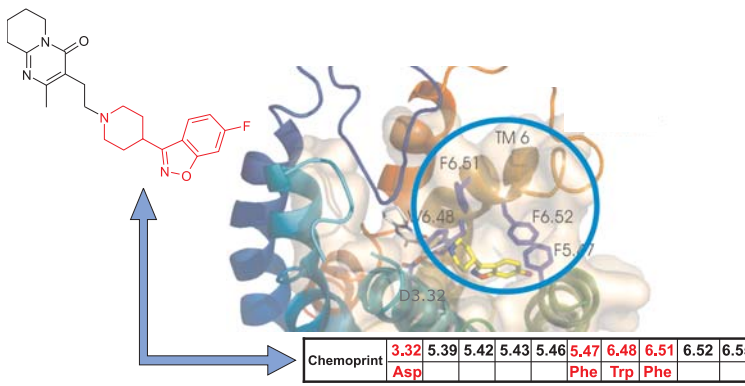
It is not surprising that the profile of the libraries directed against peptide-binding GPCRs is not as "clean" as found for the purinergic libraries. It needs to be considered that these libraries are based upon a design concept targeting a subset of peptide binding and chemokine receptors (providing a negatively charged anchoring group within the receptor) and are not meant to address every peptide receptor in general. Thus, obviously further information coming from the receptor sequences would be beneficial for a more precise clustering of these receptors (see Sect. 3.5).

3.5 Chemoprints for Recognition of Ligand Fragments

We have ongoing efforts using GPCR homology modeling, docking, and experimental profiling to identify motifs within GPCR sequences, termed chemoprints, that are linked to the recognition of fragments within GPCR ligands. These chemoprints allow for the identification of relevant receptor similarities and support the design of compound libraries tailor-made for the target GPCR. Like researchers at Novo Nordisk (Bondensgaard et al. 2004), we use rhodopsin-based homology

models and docking of compounds containing GPCR privileged fragments to derive these receptor–ligand interaction pairs, which we termed chemoprints. Figure 4 shows a homology model of the α_{1a} adrenergic receptor together with the putative binding mode of risperidone (Evers and Klabunde 2004), a $5HT_{2A}$ antagonist with nanomolar affinity toward the α_{1a} adrenergic receptor. The binding mode is in line with site-directed mutagenesis data supporting the interaction of the ligand with the aspartate residue in TM3 (Asp3.32) and the positioning of the 4-aryl-piperidine moiety into the subpocket formed by hydrophobic residues of helices TM5 and TM6. From this docking mode, the hypothesis can be derived that 4-aryl piperidines and piperazines, commonly found fragments within GPCR ligands, can bind to such GPCRs, which provide the chemoprint composed of the residue set Asp3.32, Phe5.47, Trp6.48, and Phe6.51 as recognition site.

A sequence comparison of several biogenic amine receptors (see Table 1) reveals the presence of this chemoprint in all biogenic amine receptors, suggesting that 4-aryl piperidines and piperazines are suited as building blocks for libraries targeting biogenic amine receptors. To test



Chemoprint	3.32	5.39	5.42	5.43	5.46	5.47	6.48	6.51	6.52	6.55
	Asp					Phe	Trp	Phe		

Fig. 4. Putative binding mode of risperidone in α_{1a} adrenergic receptor (Evers and Klabunde 2004). The model is in line with site-directed mutagenesis data and suggests that the chemoprint for recognition of 4-aryl piperidines and piperazines is formed by residues Asp3.32, Phe5.47, Trp6.48, and Phe6.51 (residue numbers according to the Weinstein nomenclature, Ballesteros and Weinstein 1995)

this hypothesis, we profiled 25 piperazines and piperidines commonly found within GPCR ligands against a set of eight biogenic amine receptors. Table 2 shows the results for three GPCR privileged fragments. The strong affinity of these fragments found for the biogenic amine GPCRs indicates that they are suitable building blocks for libraries targeting this class of GPCRs.

In order to further validate the correlation between the occurrence of the 4-aryl piperidine/piperazine recognizing chemoprint and experimental affinity, we profiled the fragments against a set of 11 peptide-binding

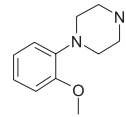
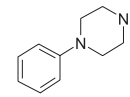
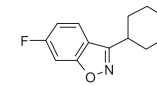
Table 1. Patchwork alignment of GPCRs sequences (biogenic amine and peptide-binding GPCRs)

	3.32	5.39	5.42	5.43	5.46	**5.47**	**6.48**	**6.51**	6.52	6.55	Chemoprint present
Adrenergic α_{1A}	**Asp**	Val	Ser	Ala	Ser	**Phe**	**Trp**	**Phe**	Phe	Met	Yes
Dopaminergic D2	**Asp**	Val	Ser	Ser	Ser	**Phe**	**Trp**	**Phe**	Phe	His	Yes
Histaminergic H1	**Asp**	Lys	Thr	Ala	Asn	**Phe**	**Trp**	**Tyr**	Phe	Phe	Yes
Histaminergic H3	**Asp**	Leu	Ala	Ser	Glu	**Phe**	**Trp**	**Tyr**	Thr	Met	Yes
Muscarinic M1	**Asp**	Thr	Thr	Ala	Ala	**Phe**	**Trp**	**Tyr**	Asn	Val	Yes
Muscarinic M3	**Asp**	Thr	Thr	Ala	Ala	**Phe**	**Trp**	**Tyr**	Asn	Val	Yes
Serotonin $5HT_{2C}$	**Asp**	Val	Gly	Ser	Ala	**Phe**	**Trp**	**Phe**	Phe	Asn	Yes
Serotonin $5HT_{2A}$	**Asp**	Val	Gly	Ser	Ser	**Phe**	**Trp**	**Phe**	Phe	Asn	Yes
Endothelin A	Glu	Met	Tyr	Phe	Tyr	**Phe**	**Trp**	Glu	Asn	Asn	No
Galanin 1	Phe	Val	Thr	Phe	Gly	**Phe**	**Trp**	His	His	His	No
Angiotensin 1	Leu	Ala	Lys	Asn	Gly	**Tyr**	**Trp**	**Phe**	His	Thr	No
Bradykinin 2	Ile	Asn	Leu	Asn	Gly	**Phe**	**Trp**	**Phe**	Gln	Thr	No
Melanocortin 3	Ile	Val	Cys	Leu	Met	**Phe**	**Trp**	**Phe**	Phe	Leu	No
Neurokinin 1	Pro	His	Val	Thr	Ile	**Tyr**	**Trp**	**Phe**	His	Phe	No
Chemokine CCR1	Tyr	Ala	Leu	Asn	Gly	Leu	**Trp**	**Tyr**	Asn	Ile	No
Choleocystokinin B	Met	Ser	Leu	Leu	Leu	**Phe**	**Trp**	Val	Tyr	Asn	No
Opiate	**Asp**	Lys	Val	Phe	Ala	**Phe**	**Trp**	Ile	His	Val	Modified
Neuropeptide Y	Gln	Thr	Leu	Leu	Gln	**Tyr**	**Trp**	Leu	Thr	Asn	No
Vasopressin 1a	Gln	Gly	Phe	Val	Val	Val	**Trp**	**Phe**	Phe	Gln	No
MCH 1	**Asp**	Thr	Gln	Phe	Ala	**Phe**	**Trp**	**Tyr**	Tyr	Gln	Yes
Somatostatin Sst2	**Asp**	Ile	Thr	Phe	Gly	**Phe**	**Trp**	**Phe**	Tyr	Asn	Yes
Urotensin II	**Asp**	Leu	Leu	Phe	Ser	Ile	**Phe**	**Phe**	Trp	Gln	Modified

The alignment reveals the presence or absence of the chemoprint composed of Asp3.32, Phe/Tyr5.47, Trp6.48, and Phe/Tyr6.51 (in bold), which is linked to the recognition of 4-aryl piperidines and piperazines. Residue numbers are given according to the Weinstein nomenclature (Ballesteros and Weinstein 1995)

GPCRs. The results for three fragments are depicted in Table 2. Indeed, some fragments were found to reveal moderate affinity toward the opiate receptor (greater than 50% inhibition at a concentration of 10 μM; binding affinities in the range of $K_i = 2–10$ μM). The patchwork alignment in Table 1 illustrates that the opiate receptor is the only receptor among the 11 receptors within the panel, that is offering an aspartate residue in position 3.32 and is providing a hydrophobic patch of residues from helices in TM 5 and TM 6 (with Ile in 6.51 instead of a Phe). Interest-

Table 2. Binding affinity profile of GPCR privileged fragments tested against a set of biogenic amine binding and peptide binding GPCRs

			
Adrenergic α_{1A}	**97**	**48**	**97**
Dopaminergic D2	**87**	3	**61**
Histaminergic H1	**76**	9	**61**
Histaminergic H3	18	6	19
Muscarinic M1	**65**	25	**45**
Muscarinic M3	**73**	12	39
Serotonin $5HT_{2C}$	**91**	**45**	**73**
Serotonin $5HT_{2A}$	**97**	**68**	**100**
Endothelin A	6	12	7
Galanin 1	0	0	−4
Angiotensin II, type 1	6	2	5
Bradykinin B2	−5	2	12
Melanocortin MC3	4	11	12
Neurokinin NK1	22	−2	5
Chemokine CXCR2	−3	−3	−12
Cholecystokinin CCK B	17	7	13
Opiate μ	**54**	23	9
Neuropeptide NPY	4	2	−3
Vasopressin V1a	11	18	16

The percent binding values of three fragments tested against a set of eight biogenic amine and eleven peptide-binding GPCR using radioligand displacement assays (at a concentration of 10 μM) are shown. Fragment–receptor pairs with significant affinity (greater than 40% of binding) are highlighted in bold

ingly, no significant affinity was seen for all other ten peptide-binding GPCRs lacking the corresponding chemoprint. This result confirms the chemoprint hypothesis linking a sequence motif to a privileged chemical fragment. It also suggests that 4-aryl piperidines/piperazines and their spiro derivatives not only represent excellent building blocks for the design of libraries targeting the family of biogenic amine receptors and the opiate receptor, but that they are also suited for libraries targeting other peptide-binding GPCRs sharing the respective chemoprint, like the somatostatin receptors, the melanocyte concentrating hormone (MCH) receptor as well as the urotensin UII receptor (see Table 1).

3.6 Summary and Conclusion

- Compilation of screening sets based on known GPCR ligands is a cost effective approach for identifying novel GPCR lead series.
- Designed libraries targeting GPCR subfamilies provide additional opportunities to identify patentable and drug-like scaffolds for competitive targets.
- Compilation of ligand recognition information (chemoprints) can be used to guide the design and synthesis of libraries tailor-made to target novel GPCRs.

References

Ballesteros JA, Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations of G protein-coupled receptors. Meth Neurosci 25:366–428

Bondensgaard K, Ankersen M, Thogersen H, Hansen BS, Wulff BS, Bywater RP (2004) Recognition of privileged structures by G-protein coupled receptors. J Med Chem 47:888–899

Bredel M, Jacoby E (2004) Chemogenomics: an emerging strategy for rapid target and drug discovery. Nat Rev Genet 5:262–275

Caron PR, Mullican MD, Mashal RD, Wilson KP, Su MS, Murcko MA (2001) Chemogenomic approaches to drug discovery. Curr Opin Chem Biol 5:464–470

Evers A, Klabunde T (2005) Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A receptor. J Med Chem 28:1088–1097

Fredholm, BB (2003) Adenosine receptors as targets for drug development. Drug News Perspect 16:283–289

Jacoby E, Fauchère J-L, Raimbaud E, Ollivier S, Michel A, Spedding M (1999) A three binding site hypothesis for the interaction of ligands with monoamine G protein-coupled receptors: implications for combinatorial ligand design. Quant Struct-Act Relat 18:561–571

Jimonet P, Jäger R (2004) Strategies for designing GPCR-focused libraries and screening sets. Curr Opin Drug Discov Devel 7:325–333

Klabunde T, Hessler G (2002) Drug design strategies for targeting G-protein-coupled receptors. Chembiochem 3:928–944

Kubinyi H, Müller G (2004) Chemogenomics in drug discovery. Wiley-VCH, Weinheim

Lewell XQ, Judd DB, Watson SP, Hann MM (1998) RECAP – retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Comput Sci 38:511–522

Shi L, Javitch JA (2002) The binding site of aminergic G Protein-coupled receptors: the transmembrane segments and second extracellular loop. Annu Rev Pharmacol Toxicol 42:437–467

Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351

Wise A, Gearing K, Rees S (2002) Target validation of G-protein coupled receptors. Drug Discov Today 7:235–246

4 Diversity Oriented Synthesis: A Challenge for Synthetic Chemists

A. Bender, S. Fergus, W.R.J.D. Galloway, F.G. Glansdorp, D.M. Marsden, R.L. Nicholson, R.J. Spandl, G.L. Thomas, E.E. Wyatt, R.C. Glen, D.R. Spring

Abstract. This article covers the diversity-oriented synthesis (DOS) of small molecules in order to generate a collection of pure compounds that are attractive for lead generation in a phenotypic, high-throughput screening approach useful for chemical genetics and drug discovery programmes. Nature synthesizes a rich structural diversity of small molecules, however, unfortunately, there are some disadvantages with using natural product sources for diverse small-molecule discovery. Nevertheless we have a lot to learn from nature. The efficient chemical synthesis of structural diversity (and complexity) is the aim of DOS. Highlights of this article include a discussion of nature's and synthetic chemists' strategies to obtain structural diversity and an analysis of molecular descriptors used to classify compounds. The assessment of how successful one diversity-oriented

synthesis is vs another is subjective; therefore we use freely available software (www.cheminformatics.org/diversity) to assess structural diversity in any combinatorial synthesis.

4.1 Introduction

Diversity-oriented synthesis (DOS) aims to synthesize a collection of compounds that differ substantially in their molecular structure (Burke and Schreiber 2004; Schreiber 2000; Spring 2003). This has application in aspects of chemical genetics and drug discovery.

Chemical genetics is the study of biological systems using small molecule (chemical) intervention, instead of only genetic intervention (Schreiber 1998, 2003; Spring 2005). Cell-permeable and selective small molecules can be used to perturb protein function rapidly, reversibly and conditionally with temporal and quantitative control in any biological system. Alternatively, biological tools can be used to study protein function such as gene knockouts/knockins, RNAi; but these tools act at the level of the gene, rather than protein and cannot be used in some situations (e.g. essential gene knockouts). Nevertheless they are general, fast, cheap and selective relatively. In order to exploit the advantages of the small molecule approach of chemical genetics, advances must be made in finding selective small molecules to any protein quickly, cheaply and with adequate selectivity. But we should be encouraged that even after the billions of years of evolution, nature still uses small molecules for signalling, protection and other essential functions. In drug discovery programmes in major pharmaceutical companies, there are teams of synthetic chemists whose roles involve adding new potential drug leads to the companies' compound collection. These libraries usually contain upwards of half a million compounds. But what should all these compounds look like?

The first point to appreciate is that chemical space is astronomic (Fig. 1). Chemical space is synonymous with multidimensional molecular descriptor space, where descriptors are characteristics of the compounds such as molecular weight. In the context of this chapter, chemical space is defined as the total molecular descriptor space that encompasses all organic compounds with a molecular weight less that 2,000 Da, i.e.

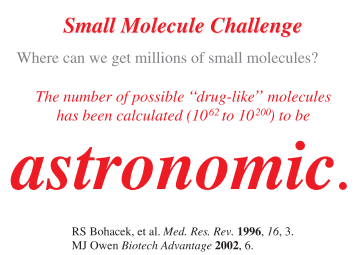


Fig. 1. Chemical space

most natural products and synthetic drugs. This chemical space is enormous. It has been estimated that the possible number of real organic compounds that are possible with a molecular weight less than 500 Da is over 10^{60} (Bohacek et al. 1996). To put this in context, the number of atoms on earth is approximately 10^{51}, so there are not enough atoms in the universe to explore all of chemical space, let alone the time it would take to make everything! Therefore, we cannot make everything, so we have to be selective.

The second point to appreciate is that biology survives with a surprisingly small number of small molecules, and for that matter a surprisingly small number of proteins. Simple life forms can function with a few hundred small molecules. Such life forms have genomes encoding less than a thousand proteins. The human proteome has been predicted to be around a quarter of a million proteins (O'Donovan et al. 2001). This is tiny in comparison to the number of proteins that are theoretically possible. The average size of natural proteins is 300 residues, and with the 20 proteinogenic amino acids this gives a staggering 10^{390} possibilities (20^{300}). Nature cannot have explored all these possibilities and therefore we can take heart that we can find a small molecule probe for a biological question, or a drug, without having to make everything! This is due to that fact that there is more than one answer to any (biological) question (Fig. 2). I am sceptical about biologically relevant chemical space being miniscule, as I would predict that the majority of the 10^{60} possible drug-like small molecules possible would have some biological activity, albeit often unwanted and unexploitable.

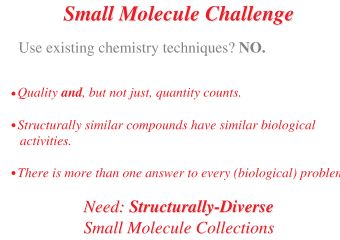


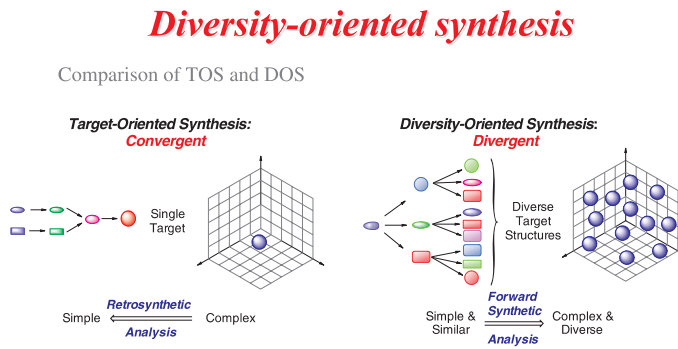
Fig. 2. Small molecule challenge

If chemical space is huge and we cannot make everything, then firstly, what should we make, and secondly, how should we make it? The first question is discussed in Sect. 4.2, and the second in Sects. 4.3 and 4.4 where nature's strategy is compared to strategies available to synthetic chemists.

4.2 What to Make?

Structural diversity is essential for lead generation in chemical genetics and drug discovery, as compounds that look the same structurally are likely to share similar physical and biological properties. The answer to the question "what to make?" is that it depends on what you want to use the compound for. If you are looking for an orally bioavailable drug with consideration of pharmacokinetics and the therapeutic index between efficacy and toxicity, then several observations have been made as to molecular characteristics that are desirable, such as size, shape, allowed functional groups and solubility in water and organic solvents. Such drug-like compounds have been evaluated in different ways, the most famous of which is Lipinski's analysis of the World Drug Index that led to the rule of five (Lipinski et al. 1997). Each pharmaceutical company will have its own criteria for what to make. In the realm of chemical genetics, there are many different situations where a small molecule may be required. If a small molecule were required for an *in vivo* animal model, then drug-like characteristics would be sensible. If cell-based assays or *in vitro* assays are being used, then a wider range of chemical space is exploitable than the restrictive chemical space de-

fined by Lipinski's rules; nevertheless, selectivity is always required for high-quality data. As regards allowed functional groups, we can be less prescriptive and even learn some lessons from nature. Nature makes an astonishing array of structural diversity in its secondary metabolites, and moreover they are often structurally complex. Complex structures are likely to interact with biology more selectively than flat, simple molecules. Therefore, structural complexity is desirable because it is simple to kill cells unselectively, e.g. with bleach. Unfortunately, there are some disadvantages with using natural product extracts. Firstly, nature does not make secondary metabolites in a pure form for us to screen; therefore, the extracts are usually screened as mixtures of many compounds, leaving the problem of purifying and identifying the active



Org. Biomol. Chem. **2003**, *1*, 3867-3870

Fig. 3. Target oriented synthesis (*TOS*) vs diversity-oriented synthesis (*DOS*). DOS concerns the efficient synthesis of structurally diverse (and complex) small molecules, i.e. where the molecules differ in their (a) attached groups, (b) stereochemistry, (c) functional groups and (d) molecular frameworks. TOS aims to synthesize a single target. Synthetic pathways in DOS are branched and divergent and the planning strategy extends simple and similar compounds to more complex and diverse compounds. Retrosynthetic analysis concepts focus on the existence of a defined target structure. In DOS, there is no single target structure and therefore retrosynthetic analysis cannot be used directly and a forward synthetic analysis algorithm is required. The three-dimensional grids of molecular descriptors illustrate the product(s) of the syntheses in chemical descriptor space

component(s). Secondly, the natural product extract may come from a limited source, leaving a supply problem if the active compound is desired. Thirdly, the active natural product may be so complex structurally, such as vancomycin, that making analogues to optimize activity is a formidable synthetic challenge. Fourthly, chemistry space encompassed by natural products (and drug-like compounds) is unlikely to be the only region useful for discovering physical or biological properties of interest, and moreover, may not be the most productive region. These complications have led organic chemists to take the complementary approach of *synthesizing* structurally diverse and complex small molecules directly (Fig. 3).

4.3 Nature's Strategies

The rich structural diversity and complexity of natural products have inspired all synthetic chemists. Many drugs in clinical use today are natural products or natural product derivatives. For example over the last 20 years, 5% of the 1,031 new chemical entities approved as drugs were natural products, and another 23% were natural product-derived (Newman et al. 2003). Natural products can be simple, such as serotonin and histamine, or complex structurally, such as vancomycin and taxol (Fig. 4). They occupy a greater volume of chemical space relative to drug-like compounds, but are still useful to the organisms that produce them at least. They tend to have less nitrogen, but more chiral centres and often have higher molecular masses (Clardy and Walsh 2004). Some natural products such as calicheamicin have highly reactive functional groups (ene diyne), yet are selective (Fig. 4). Most of the rich diversity of secondary metabolites appears to come from organisms such as bacteria or plants (Clardy and Walsh 2004). But how do they make such as diverse range of compounds?

Biosynthetic routes to secondary metabolites are usually linear using simple building blocks usually from primary metabolism (such as amino acids for nonribosomal peptides, acyl-CoA thioesters for polyketides, isoprenyl diphosphates for terpenes). Unusual monomers are synthesized at the same time as the secondary metabolite, with the biosynthetic machinery being encoded in the same gene cluster. Once the monomer

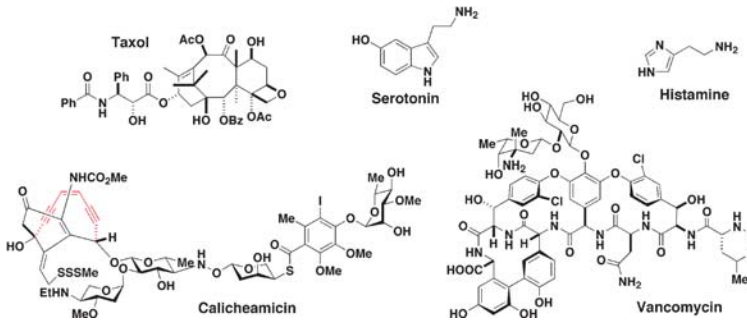


Fig. 4. Natural product structures

units have been added together in a linear fashion to give the scaffold, appendage diversification steps can be taken, for example oxidation (e.g. taxadiene to taxol, reticuline to morphine) or glycosidation (e.g. vancomycin). Nature has the advantage over present-day synthetic chemists in that it can use enzymes to conduct synthetic chemistry with usually complete chemo-, regio- and stereoselectivity. It should also be pointed out that we have identified only a small percentage of natural products to date. Improvements in culturing bacteria, combinatorial biosynthesis and secondary metabolite expression will undoubtedly lead to the discovery of new and exciting leads.

4.4 Synthetic Chemist's Strategies

Synthetic chemists have the advantage over nature with respect to a wider selection of building blocks and chemical reactions (nature does not seem to have discovered alkene metathesis, at least not via Ru, Mo or W catalysis). A collection of compounds with the highest level of structural diversity will consist of molecules that have incorporated different building blocks, stereochemistries, functional groups and molecular frameworks (Spring 2003). Consider a coupling reaction that involves a substrate, a building block (or more than one building block for multicomponent coupling reactions), and a reagent to give the product. In simple terms, strategies to generate structural diversity would involve

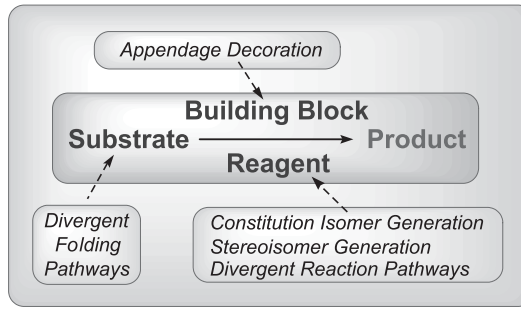


Fig. 5. Diversity generation strategies. Skeletal diversity can be generated by constitutional isomer and stereoisomer generation, divergent reaction pathways and divergent folding pathways

varying the building block [(a) appendage decoration], reagent [(b) constitutional isomer generation, (c) stereoisomer generation, (d) divergent reaction pathways] or substrate [(e) divergent folding pathways] (Fig. 5). The most successful syntheses of structural diversity incorporate multiple strategies.

Appendage decoration is the most straightforward diversity-generating processes and a central feature in combinatorial chemistry, particularly to improve the biological activity of a drug lead; it involves the use of coupling reactions to attach different building blocks to a common molecular framework (cf. nature's strategy). Many examples are available from the literature of this approach to combinatorial synthesis. If only appendage decoration is used in the library synthesis, then all the products will have the same molecular frameworks, which is ideal if a focussed library is required. Nevertheless, if a very diverse range of building blocks is used, then although the scaffold is the same, the overall structural diversity can be very high. In order to generate an even greater degree of structural diversity in the molecular scaffold, other strategies need to be incorporated into the synthesis too.

Constitutional isomer generation involves using chemoselective and/or regioselective reactions to synthesize different product isomers. Stereoisomer generation involves using reactions that proceed with diastereoselectivity and/or enantioselectivity.

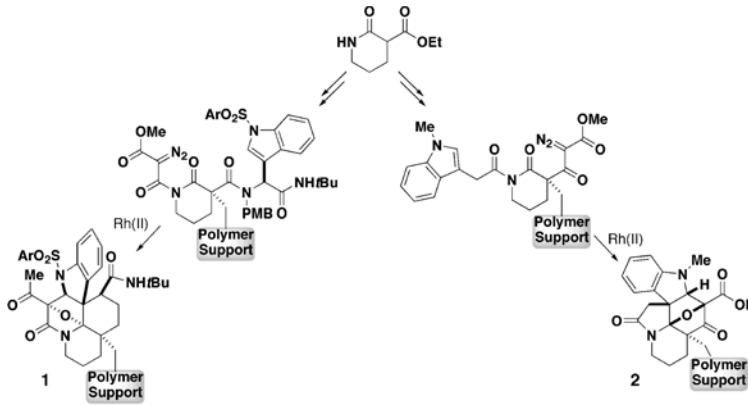


Fig. 6. Strategies to give structurally diverse molecular frameworks by divergent folding pathways

Divergent reaction pathways are a very efficient way of generating structural diversity, particularly diverse molecular frameworks and functional groups. Skeletal diversity is generated by using different reagents to change a common substrate into a collection of products having varied molecular skeletons.

Divergent folding pathways utilizes substrates with different appendages that pre-encode skeletal information into a collection of products having different molecular skeletons using common reaction conditions. Most DOS libraries use several strategies to generate structural diversity. For example, Oguri and Schreiber have elegantly demonstrated that six structurally diverse indole alkaloid-like frameworks can be generated by shifting the relevant functionality around three points on a starting scaffold (Fig. 6). A rhodium-catalysed tandem cyclization-cycloaddition reaction was used to efficiently generate distinct frameworks (**1** and **2**) with complete diastereocontrol (Oguri and Schreiber 2005).

But how do you assess the *degree* of structural diversity that is created? Intuition? It is clear that a less subjective method of assessment is required to assess diversity.

4.5 Assessment of Molecular Diversity

In order to assess the molecular diversity of a collection of compounds on a large scale, it is necessary to use computer algorithms that, generally speaking, consist of two operations. Firstly, the structures are put into chemical descriptor space using molecular descriptors, and secondly, diversity in chemical descriptor space is calculated (Xue and Bajorath 2000). The calculation of molecular descriptors creates an abstract representation of the molecule (Bender and Glen 2004; Brown and Martin 1996). The representations of molecules can be classified according to their dimensionality (Willett et al. 1998):

1. One-dimensional (1D) were bulk properties such as volume, molecular weight and log P (Downs et al. 1994).
2. Two-dimensional descriptors (2D) are derived from the connectivity table of a molecular structure (Estrada and Uriarte 2001).
3. Three-dimensional descriptors (3D) use geometrical information from points in 3D space.

Since binding of a ligand to a target is an event in space, the geometry of the ligand in relation to that of the binding pocket is critical. Therefore, is it still advisable to use a 2D method over a 3D method in certain situations? Molecules are not rigid entities, they are conformationally flexible, especially if many single bonds are present in a molecule, this leads to a "curse of dimensionality" when dealing with 3D information. In addition, since the active (binding) conformation of a structure is usually unknown, most of the possible conformations cannot be excluded. Dealing with the complete conformational ensemble results in an increase in noise, since virtually every spatial arrangement can be assigned to the ligand. Two-dimensional methods, on the other hand, do not explicitly capture shape; shape is implicitly contained in the connectivity table. Therefore the information required is greatly reduced, eliminating noise. This leads to a much faster generation of results while usually retaining their validity. Atom environment descriptors are employed as a molecular representation (Bender et al. 2004), as shown in Fig. 7. For diversity assessment, we can calculate the average number of atom environments per molecule. The absolute number of features necessarily increases if nonidentical structures are added, but here we are interested in a diver-

sity measure relative to the size of the library. This software is freely available via a web interface at www.cheminformatics.org/diversity.

To test this computational assessment of structural diversity, a range of combinatorial libraries was chosen from the literature, and an ideal diverse library consisting of 40 diverse natural products. The diversity values of each library are shown in Fig. 8.

The diverse libraries generally result in a higher value of diversity than the focussed libraries; however, certain limitations require highlighting when evaluating the diversity of a collection of compounds. The diversity value is dependent on the number of compounds in the collection; therefore, very small libraries (fewer than ten library members) give illogical results that should be utilized with caution. Also, since the programme compares compounds using two factors, (a) the hybridization of the atoms and (b) the variation of atoms, a focussed library using a common scaffold with varying appendages that contain a wide variety of elements and different degrees of hybridization will give a higher value than perhaps expected. This programme is a useful tool in assessing the diversity of a collection of compounds; however, it should be employed with due care upon understanding some of its limitations, as outlined above.

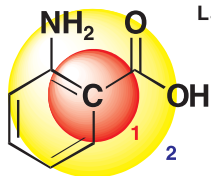


Fig. 7. Illustration of descriptor generation step, applied to an aromatic carbon atom. The distance (layers) from the central atom is shown in *brackets*. Every heavy atom in the hydrogen-depleted structure of the molecule is assigned its Sybyl atom types. Sybyl atom types are used to classify atoms according to the element type and hybridization state. An individual atom fingerprint is calculated for each heavy atom in the molecule capturing its local environment at a distance of n bonds. Frequencies of atom types at a given distance ($n = 0, 1, 2$) are recorded

Library	Ref.	Number of Molecules	Number of Features	Features per Molecule[a]	**Diversity Value[b]**
1	[11]	168	38	0.226	**2**
2	[12]	88	30	0.341	**3**
3	[13]	49	50	1.02	**10**
4	[14]	62	101	1.63	**16**
5	[18]	24	56	2.33	**23**
6	[17]	15	46	3.07	**30**
7	[16]	18	72	4.00	**39**
8	[15]	60	288	4.80	**46**
9	[19]	10	55	5.50	**53**
Ideal[c]	–	40	414	10.4	**100**

Increasing Molecular Diversity

Fig. 8. Diversity value of nine collections of compounds. The diversity value is calculated on a scale from 0 to 100 incorporating the number of features per molecule. [a] To 3 significant figures. [b] Nearest integer value. [c] The ideal diverse library consists of acetic acid, alliin, ampicillin, bee pheromone, benzene, bergenin, beta carotene, blebbistatin, caffeine, catechin, cinnamic acid, ciprofloxacin, cocaine, cortisone, cyclosporin, cysteine, D-glucose, dopamine, erythromycin, fluzanim, fumiquinazoline G, genistein isoflavonoid, glucosamine, L-DOPA, methane, methanol, morphine, nandrolone, omega-6 fatty acid, phenylalanine, quinine, rapamycin, serotonin, streptomycin, sucrose, taxol, testosterone, vitamin A, vitamin E and vitamin K

4.6 Technology Aspects

If chemical genetics is going to become more accessible, then the synthesis and screening of diverse compound collections needs to be done in a much smaller, faster and cheaper way. These considerations are also attractive to the drug discovery industry where profit margins are being squeezed. Synthesis using microwaves has accelerated compound production to a degree, but really order of magnitude step changes are required to make chemical genetics more competitive relative to biological techniques. Microarray and microfluidics technologies have the potential to make such a step change.

4.7 Conclusion

The diversity-oriented synthesis of small molecules is a challenge to synthetic chemists, requiring new strategies to generate appendage and skeletal diversity. Progress has been made recently and we have assessed the structural diversity achieved by using a free computer programme (www.cheminformatics.org/diversity) that utilizes fragment-based molecular descriptors to quantify the structural diversity of collections of small molecules. If DOS is to be more useful generally the process of selective small-molecule discovery to modulate the function of a given protein will need to be more efficient and economical.

Acknowledgements. We thank BBSRC, EPSRC, GSK, AstraZeneca, Pfizer, Syngenta, Lilly and the Gates Cambridge Trust for funding.

References

Bender A, Mussa HY, Glen RC (2004) Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. J Chem Inf Comput Sci 44:1708–1718

Bender A, Glen RC (2004) Molecular similarity: a key technique in molecular informatics. Org Biomol Chem 2:3204–3218

Bohacek RS, McMartin C, Guida WC (1996) The art and practice of structure-based drug design: a molecular modelling perspective. Med Res Rev 16:3–50

Brown RD, Martin YC (1996) Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. J Chem Inform Comput Sci 36:572–584

Burke MD, Schreiber SL (2004) A planning strategy for diversity-oriented synthesis. Angew Chem Int Ed 43:46–58

Clardy J, Walsh C (2004) Lessons from natural molecules. Nature 432:829–837

Downs GM, Willett P, Fisanick W (1994) Similarity searching and clustering of chemical-structure databases using molecular property data. J Chem Inf Comput Sci 34:1094–1102

Estrada E, Uriarte E (2001) Recent advances on the role of topological indices in drug discovery research. Curr Med Chem 8:1573–1588

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25

Newman DJ (2003) Natural products as sources of new drugs over the period 1981–2002. J Nat Prod 66:1022–1037
O'Donovan C, Apweiler R, Bairoch A (2001) The human proteomics initiative (HPI). Trends Biotechnol 19:178–181
Oguri H, Schreiber SL (2005) Skeletal diversity via a folding pathway: synthesis of indole alkaloid-like skeletons. Org Lett 7:47–50
Schreiber SL (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. Bioorg Med Chem 6:1127–1152
Schreiber SL (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. Science 287:1964–1969
Schreiber SL (2003) The small molecule approach to biology. Chem Eng News 81:51–61
Spring DR (2003) Diversity-oriented synthesis; a challenge for synthetic chemists. Org Biomol Chem 1:3867–3870
Willett P, Bernard JM, Downs GM (1998) Chemical similarity searching. J Chem Inf Comput Sci 38:983–996
Xue L, Bajorath J (2000) Molecular descriptors in chemoinformatics, computational combinatorial chemistry and virtual screening. Comb Chem High Throughput Screen 3:363–372

5 Probing Protein Function with Small Molecules

J.D. Gough, C.M. Crews

Abstract. The interface of chemistry and biology offers many opportunities to explore different aspects of cell biology. The emerging field of chemical genetics is providing the chemical means to understand biological systems not easily accessible using classical genetic manipulations. In this article, we will discuss how natural product mode of action studies and novel bio-organic manipulation of intracellular protein levels are proving useful in the exploration of cell biology.

5.1 Natural Product Mode-of-Action Studies: Following Natures Lead

Nature provides a plethora of structurally diverse and biologically active small molecules. These natural products are attractive probes of cell

biology; many are cell-permeable, potent inhibitors, or perturbers of intracellular processes. Over the last few decades, natural products have played an important role not only in the development of drugs but also in our understanding of cell biology. The diverse areas that have significantly benefited from this approach include:

1. Cytoskeletal research
2. Immune cell signaling
3. Gene expression
4. Protein translation
5. Protein phosphorylation and dephosphorylation

While individual natural products have proven useful as cellular probes, it has only been in the past few years that labs have taken a more systematic approach to natural product mode-of-action studies with the goal of developing new biological tools (Crews and Splittgerber 1999; Koh and Crews 2002; Schreiber 1998; Yeh and Crews 2003).

The chemical genomic approach uses a wide range of disciplines to accomplish its goal: total synthesis for the generation of usable quantities of the natural products and related affinity reagents, biochemistry and genetics for isolating and determining a given target protein, and cell biology for validating and investigating target proteins. This interdisciplinary crosstalk has yielded the identification of new target proteins and insights into cell signaling events.

5.1.1 Epoxyketones

Our laboratory's investigation of biologically active natural products and their mode-of-actions studies have led us in many interesting directions. An early project focused on elucidating the mode of action of the anti-angiogenic natural product fumagillin. We identified methionine aminopeptidase 2 as the fumagillin target protein (Sin et al. 1997; Yeh et al. 2000). Following these studies, we sought other natural products reported to exhibit anti-angiogenic activity. Thus, we next focused on a group of structurally related α'-, β'-epoxyketone natural products, epoxomicin and eponemycin (Fig. 1), which had been shown to display anti-tumor and anti-angiogenic activities (Hanada et al. 1992; Sugawara et al. 1990).

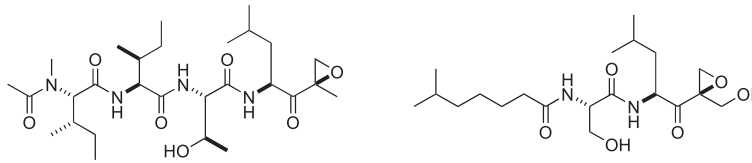


Fig. 1. The structures of epoxomicin (*left*) and dihydroeponemycin (*right*)

Using a multidisciplinary approach, we set out to identify the target protein(s) of these potent natural products. Comparison of the structural motifs present in the two molecules led us to hypothesize that the terminal epoxyketone was an important aspect of the functional pharmacophore. Using this moiety as our focus, we designed and executed divergent total syntheses that yielded the biologically active natural products and their N-terminal biotinylated affinity analogs (Sin et al. 1999; Sin et al. 1998).

With these biological probes now in hand, we focused our attention on identifying the target proteins of each of these compounds. Due to the potent biological activity, we hypothesized that the epoxyketone could be sensitive to nucleophilic attack, thus forming a covalent adduct with the target protein(s). To test our hypothesis, the affinity reagents were incubated with cultured endothelial cells, and cell lysates were prepared and analyzed for the presence of new biotinylated proteins. Additionally, to test the reversibility and potency of the inhibitors, cells were pretreated with natural product and subsequently challenged with their corresponding biotin-natural product analogs (Fig. 2). These data reveal the specificity and irreversibility of the inhibitors.

Mass spectrometric analyses identified these epoxomicin- and dihydroeponemycin-binding proteins as catalytic subunits of the 20S proteosome (Meng et al. 1999a, b), the proteolytic chamber of the 26S proteasome complex. The 26S proteasome is responsible for diverse activities such as the ubiquitin-dependent proteolysis of misfolded or aged housekeeping proteins, cell cycle regulation via cyclin degradation, and cellular immune response via antigenic peptide processing. Both biochemical and cell-based assays showed that epoxomicin and dihydroeponemycin are potent irreversible proteasome inhibitors. Interestingly,

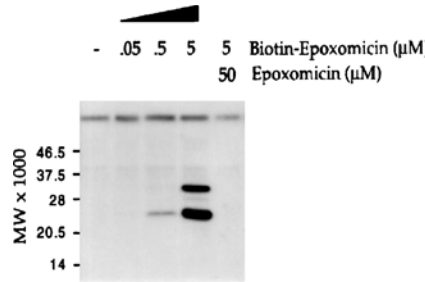


Fig. 2. Biotinylated epoxomicin protein binding pattern

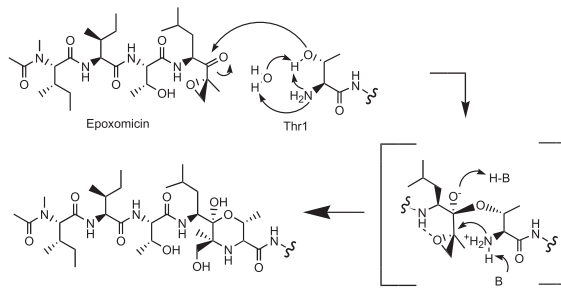


Fig. 3. Proposed mechanism for morpholino adduct formation

unlike other proteasome inhibitors that also had cross-inhibitory activity against various intracellular proteases, the epoxyketone peptide-based inhibitors are very selective for the proteasome. X-ray crystallographic analysis of epoxomicin complexed with the 20S proteasome performed in collaboration with Prof. Robert Huber (MPI, Martinsried) explained the unique selectivity of the epoxyketone pharmacophore. These structural studies showed that a morpholino ring was generated upon coupling Thr1 of the proteasome with the epoxyketone pharmacophore of epoxomicin (Fig. 3) (Groll et al. 2000).

Given that epoxomicin effectively inhibits two of the three catalytic activities of the proteasome, we next sought to design and synthesize a more specific proteasome inhibitor. We chose to target the chymotrypsin-like site of the 20S proteasome because no individual

inhibitor specifically targets that site. Taking a medicinal chemical approach, we generated a series of epoxyketone peptide-based proteasome inhibitors. One in particular, YU101, was found to be several-fold more potent and more selective for the chymotrypsin-like activity than the parent compound epoxomicin (Elofsson et al. 1999).

YU101 and epoxomicin subsequently proved useful in the analysis of the role that the proteasome plays in regulating the transcription factor NF-κB. In the absence of extracellular stimuli, NF-κB remains in the cytoplasm as a complex bound with unphosphorylated inhibitor protein, IκB. In response to certain pro-inflammatory cytokines such as tumor necrosis factor-α (TNF-α), a signal transduction cascade is initiated. This leads to the activation of a multisubunit IκB kinase (IKK), which phosphorylates IκB. Phosphorylation of IκB serves as signal for it to be ubiquitinated via the action of the E3 ubiquitin ligase, β-TrCP, and subsequently destroyed by the 26S proteasome. Using these epoxyketone peptides, we showed that proteasome inhibition leads to loss of NF-κB activation. Thus, we showed that, in addition to their anti-tumor activities, epoxomicin and YU101 also possess anti-inflammatory activities in cells and *in vivo* (Elofsson et al. 1999; Meng et al. 1999b).

5.1.2 Parthenolide

To determine if other anti-inflammatory natural products inhibited the proteosome, we focused on the sesquiterpene lactone parthenolide (Fig. 4), which is the active natural product in the medicinal herb Feverfew (Heptinstall et al. 1992). This medicinal herb has been used for 2,000 years to treat fevers, headaches, and inflammation.

Initial studies showed that parthenolide did not inhibit the proteasomes but was a potent inhibitor of NF-κB nuclear translocation as well as IκB phosphorylation. Structurally, the sesquiterpene lactone family shares a exocyclic methylene moiety, as part of a γ-lactone. Assuming that this moiety was important part of the molecule's pharmacophore, we designed and synthesized a biotinylated analog of parthenolide, biotin-parthenolide, for use in affinity chromatography experiments (Kwok et al. 2001).

The biologically active analog biotin-parthenolide was incubated with intact cells, and the biotinylated proteins were isolated using affinity

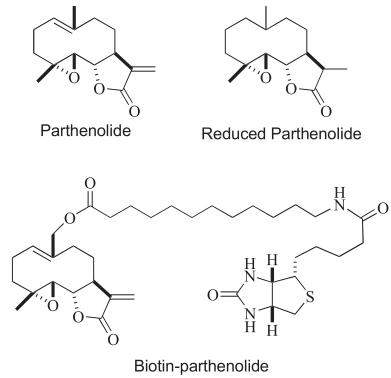


Fig. 4. Structure of parthenolide, reduced parthenolide (RP) and biotinylated parthenolide

chromatography. Immunoblot analysis with anti-IKKβ antibodies revealed that parthenolide formed a covalent adduct with IKKβ in a dose-dependent manner. By initially challenging cells with parthenolide before challenging with biotin-parthenolide, we were also able to show that parthenolide binds to IKKβ in a specific manner.

Having identified the target protein using our biotin natural product analog, we then sought use the natural product to further specify its mode of action in the cell-signaling cascade. Our starting point was the IKK complex, which acts as a meeting point for a number of other upstream activating kinases. To discern the specificity of parthenolide, a constitutively activated mutant of IKKβ was created and tested for its sensitivity. Cells transfected with the constitutively activated (SS/EE) IKKβ were found to induce IκB phosphorylation in kinase assays, and more importantly the activated IKK activity was found to be parthenolide-sensitive. These data indicate that parthenolide specifically targets IKKβ directly, independent of other upstream kinases. Mass spectrometric analysis of parthenolide-binding residues revealed that cysteine 179 of IKKβ, which lies between the two phosphorylated serines in the kinase activation loop, is covalently modified by parthenolide. A constitutively activated protein with a C179A point mutation was constructed and subsequently found to be insensitive to 40 μM parthenolide, indicating

that parthenolide inhibits IKKβ via Michael addition by Cys179 in the kinase activation loop (Kwok et al. 2001).

With the knowledge that a nucleophilic cysteine residue was responsible for the covalent modification, we turned our efforts to better understanding the structure activity relationship in parthenolide. Other biologically active members of the sesquiterpene lactone family share a γ-lactone that possesses an exocyclic methelene moiety. This moiety is a likely electrophilic candidate for a Michael addition as it is in parthenolide.

5.2 Proteolysis Targeting Chimeric Molecules

Whereas our work with epoxomicin and dihydroeponemycin reveals a novel and specific pharmacophore capable of *inhibiting* the proteasome, we next sought to develop a means to *induce* selective protein degradation using small molecules. PROteolysis TArgeting Chimeric (PROTACs) are heterobifunctional molecules that were developed as a potential proteomic tool for selectively "mutagenizing the proteome." Fundamentally, a PROTAC facilitates the destruction of a targeted protein by subverting the function of a given E3 ligase of the ubiquitin-proteosome pathway. We believe that this strategy will provide a new narrative for studying proteins at the post-translational level in whole cells, a method that potentially does not require any biological manipulation of a cellular system.

5.2.1 Targeting Proteins for Degradation Using Small Molecules

As described earlier E3 ubiquitin ligases, which recognize proteins that need to be degraded, play important roles in the ubiquitin/proteasome pathway (Jentsch 1992). E3 ligases, in conjunction with ubiquitin-conjugating enzymes (E2), facilitate the labeling of their target proteins with a polyubiquitin tag. The resulting polyubiquitinated target protein is then recognized by the 26S proteasome, whereupon ubiquitin is removed and the substrate protein threaded into the proteolytic chamber of the proteasome, where it is degraded into short peptides. The substrate specificity of the ubiquitin proteosome pathway is conferred by the E3 ligase (Seufert and Jentsch 1992).

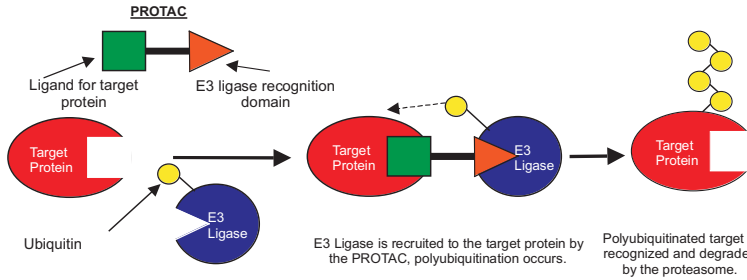


Fig. 5. Overview of PROTAC technology. Proteins are targeted for degradation via their recruitment to an E3 ubiquitin ligase by means of a PROTAC

Functionally, a PROTAC is a bifunctional molecule, designed such that one end binds to an E3 ligase, and the other binds to a target protein (Fig. 5). By binding to an E3 ligase and a target protein, the protein of interest is recruited to the E3 ligase to be ubiquitinated. Once ubiquitinated, the target protein is then destroyed by the 26S proteasome.

Initial proof-of-concept experiments took advantage of the known interactions between IκB and β-TrCP. Once phosphorylated, IκB is recognized by its cognate E3 ubiquitin ligase β-TRCP, which promotes its ubiquitination and subsequent degradation. The phosphopeptide recognition sequence of IκB that conveys recognition by β-TrCP was used in designing our first PROTAC.

We chose to target methionine aminopeptidase 2 (MetAP-2) with our first PROTAC molecule. Our lab, in collaboration with John Clardy, determined the crystal structure of human MetAP-2 complexed with the natural product fumagillin (Liu et al. 1998). The crystal structure of the protein revealed a number of lysine residues on the surface of the protein near the fumagillin binding site, thereby providing a number of viable ubiquitination sites to be presented to the degradation machinery. We designed and synthesized the first PROTAC using ovalicin, a derivative of fumagillin, linked to the IκB diphosphopeptide. In *Xenopus* extracts, the degradation of MetAP-2 was induced upon the addition of the PROTAC within 15 min (Fig. 6) (Sakamoto et al. 2001).

Second-generation PROTACs were designed to be membrane-permeable. Given the difficulty of diphophorylated peptides to transverse the

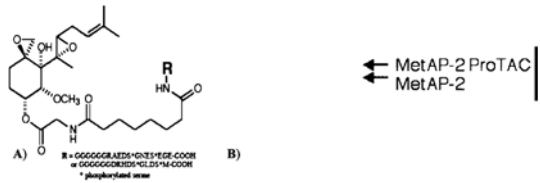


Fig. 6. (**A**) Structure of MetAP-2 targeting PROTAC consisting of the MetAP-2 Ligand ovalicin tethered to the diphosphorylated IkB peptide recognized by β-TrCP. (**B**) Western blot of PROTAC mediated MetAP-2 degradation

cell membrane, we looked to a different E3 ligase–ligand pair: hypoxia inducible factor 1α (HIF1α) as the E3 ubiquitin ligase recognition element and its cognate E3 ligase, the Von Hippel Lindau protein complex (VHL) (Cockman et al. 2000; Tanimoto et al. 2000). Under normoxic conditions, proline 564 of HIF1α is oxidized by an intracellular proline hydroxylase (Hon et al. 2002; Min et al. 2002). VHL selectively binds the oxidized form of HIF1α, and induces its ubiquitination and subsequent degradation. To improve upon the membrane permeability, we chose to use a polyarginine molecular transporter, which mimics the HIV-Tat and *Antennaepedia* proteins (Wender et al. 2000). Combining the seven-amino-acid recognition domain of HIF1α and the polyarginine, we designed and synthesized the next generation of PROTACs (Fig. 7).

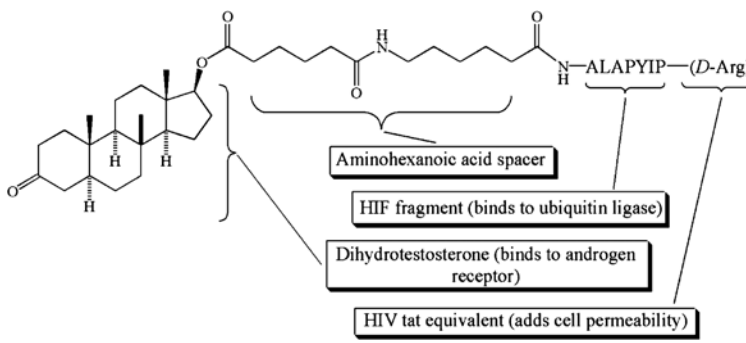


Fig. 7. Design of a testosterone-based cell permeable PROTAC

This generation of PROTACs was designed to use natural ligands of intracellular proteins to induce those receptors to be degraded. Specifically, we targeted the androgen receptor (AR) using a dihydrotestosterone-based PROTAC (Fig. 8). Dihydrotestosterone was tethered to the seven-amino-acid HIF peptide and the poly-D-Arg moiety. The $(Arg)_8$ tag facilitates entry into HEK-293 cells stably expressing AR fused to green fluorescent protein (GFP). Once inside the cell, the seven-amino-acid HIF peptide binds VHL and thereby, via DHT, recruits the GFP-AR fusion protein to be ubiquitinated and subsequently destroyed by the 26S proteasome. This results in the loss of cellular fluorescence within 60 min (Fig. 8) (Schneekloth et al. 2004).

The PROTAC technology offers several advantages. First, a PROTAC can act catalytically. As long as the small molecule targeting moiety does not covalently modify the target protein, the PROTAC is free to recruit multiple proteins to be ubiquitinated. Second, unlike traditional small molecule enzyme inhibitors, a PROTAC need not bind to the active site of a protein in order to function. By merely binding to a unique protein surface domain, a PROTAC can recruit a protein to an E3 ligase. These advantages offer the possibility to manipulate many different protein classes using small molecules. For example, a PROTAC could target nonenzymatic proteins, structural proteins, and regulatory proteins.

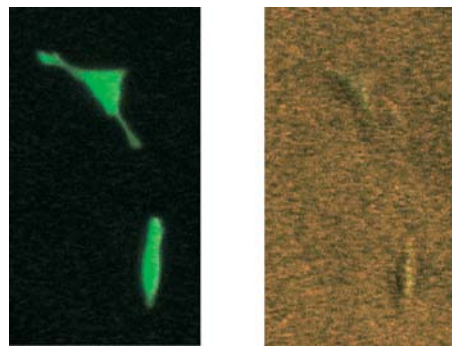


Fig. 8. PROTAC-mediated degradation of intracellular androgen receptor-green fluorescent protein. *Left:* Before addition of 25 μM PROTAC; *right*: 1 h after PROTAC addition

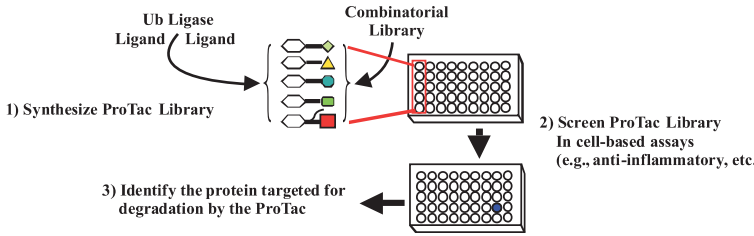


Fig. 9. A PROTAC library-based chemical genetic screen

Several other applications for this technology can be envisioned. First, PROTACs could be used to control a desired cellular phenotype, for example via the induced degradation of a crucial regulatory transcription factor, which is difficult to target pharmaceutically. A "chemical knockout" of a protein could prove viable as an alternative for a genetic knockout, which would be extremely valuable in the study of protein function. *This strategy also provides significantly more temporal or dosing control than gene inactivation at the DNA (i.e., genetic mutation) or RNA level (i.e., RNAi).*

5.2.2 Chemical Genetic Screens

The future of the PROTAC technology is its potential use in a chemical genetic screen. We propose a chemical genetic screen whereby we will mutagenize the proteome, as one mutagenizes the genome in a genetic screen, by selectively inducing the loss of individual proteins using a library of PROTAC molecules (Fig. 9). By scoring for the ability of individual PROTAC molecules to perturb specific intracellular processes, one could identify those proteins necessary for a particular intracellular process.

5.3 Conclusions and Outlook

Natural product mode-of-action studies and the PROTAC technology can serve the needs of both basic science and drug discovery. The latter, in particular, is an area in search of new technologies. The human

genome sequencing project has provided a plethora of potential drug target candidates. However, given the limited resources to devote to the development of any given drug, an emerging bottleneck in the pharmaceutical industry is now the identification and validation of new drug targets from among the many thousand possibilities. *The strategies described here have the potential to identify those proteins in cells that are pharmaceutically vulnerable and thus viable drug target candidates.*

Acknowledgements. The authors are grateful to Jay Schneekloth, Jr. for his helpful comments on the manuscript. Financial support for the work of our group cited in this review was generously provided by the NIH (GM062120).

References

Cockman ME, Masson N, Mole DR, Jaakkola P, Chang GW, Clifford SC, Maher ER, Pugh CW, Ratcliffe PJ, Maxwell PH (2000) Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein. J Biol Chem 275:25733–25741

Crews CM, Splittgerber U (1999) Chemical genetics: exploring and controlling cellular processes with chemical probes. Trends Biochem Sci 24:317–320

Elofsson M, Splittgerber U, Myung J, Mohan R, Crews CM (1999) Towards subunit specific proteasome inhibitors: synthesis and evaluation of peptide α′,β′-epoxyketones. Chem Biol 6:811–822

Groll M, Kim KB, Huber R, Crews CM (2000) Crystal structure of epoxomicin:20S proteasome reveals molecular basis for selectivity of α′,β′-epoxyketone proteasome inhibitors. J Am Chem Soc 122:1237–1238

Hanada M, Sugawara K, Kaneta K, Toda S, Nishiyama Y, Tomita K, Yamamoto H, Konishi M, Oki T (1992) Epoxomicin, a new antitumor agent of microbial origin. J Antibiot 45:1746–1752

Heptinstall S, Awang DV, Dawson BA, Kindack D, Knight DW, May J (1992) Parthenolide content and bioactivity of feverfew (Tanacetum parthenium (L) Schultz-Bip.) Estimation of commercial and authenticated feverfew products. J Pharm Pharmacol 44:391–395

Hon WC, Wilson MI, Harlos K, Claridge TD, Schofield CJ, Pugh CW, Maxwell PH, Ratcliffe PJ, Stuart DI, Jones EY (2002) Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. Nature 417:975–978

Jentsch S (1992) The ubiquitin-conjugation system. Annu Rev Genet 26:179–207

Koh B, Crews CM (2002) Chemical genetics: a small molecule approach to neurobiology. Neuron 36:563–566

Kwok BH, Koh B, Ndubuisi MI, Elofsson M, Crews CM (2001) The anti-inflammatory natural product parthenolide from the medicinal herb Feverfew directly binds to and inhibits IkappaB kinase. Chem Biol 8:759–766

Liu S, Widom J, Kemp CW, Crews CM, Clardy J (1998) Structure of human methionine aminopeptidase-2 complexed with fumagillin. Science 282:1324–1327

Meng L, Kwok BHB, Sin N, Crews CM (1999a) Eponemycin exerts its antitumor effect through inhibition of proteasome function. Cancer Res 59:2798–2801

Meng L, Mohan R, Kwok BHB, Elofsson M, Sin N, Crews CM (1999b) Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo anti-inflammatory activity. Proc Natl Acad Sci U S A 96:10403–10408

Min JH, Yang H, Ivan M, Gertler F, Kaelin WG Jr, Pavletich NP (2002) Structure of an HIF-1alpha-pVHL complex: hydroxyproline recognition in signaling. Science 296:1886–1889

Sakamoto KM, Kim KB, Kumagai A, Mercurio F, Crews CM, Deshaies RJ (2001) Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. Proc Natl Acad Sci U S A 98:8554–8559

Schneekloth JS Jr, Fonseca FN, Koldobskiy M, Mandal A, Deshaies R, Sakamoto K, Crews CM (2004) Chemical genetic control of protein levels: selective in vivo targeted degradation. J Am Chem Soc 126:3748–3754

Schreiber SL (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. Bioorg Med Chem 6:1127–1152

Seufert W, Jentsch S (1992) In vivo function of the proteasome in the ubiquitin pathway. EMBO J 11:3077–3080

Sin N, Meng L, Wang MQW, Wen JJ, Bornmann WG, Crews CM (1997) The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. Proc Natl Acad Sci U S A 94:6099–6103

Sin N, Meng L, Auth H, Crews CM (1998) Eponemycin analogs: synthesis and use as probes of angiogenesis. Bioorg Med Chem 6:1209–1217

Sin N, Kim K, Elofsson M, Meng L, Auth H, Kwok BHB, Crews CM (1999) Total synthesis of the potent proteasome inhibitor epoxomicin: a useful tool for understanding proteasome biology. Bioorg Med Chem Lett 9:2283–2288

Sugawara K, Hatori M, Nishiyama Y, Tomita K, Kamei H, Konishi M, Oki T (1990) Eponemycin, a new antibiotic active against B16 melanoma. I. Production, isolation, structure and biological activity. J Antibiot 43:8–18

Tanimoto K, Makino Y, Pereira T, Poellinger L (2000) Mechanism of regulation of the hypoxia-inducible factor-1 alpha by the von Hippel-Lindau tumor suppressor protein. EMBO J 19:4298–4309

Wender PA, Mitchell DJ, Pattabiraman K, Pelkey ET, Steinman L, Rothbard JB (2000) The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. Proc Natl Acad Sci U S A 97:13003–13008

Yeh JR, Crews CM (2003) Chemical genetics: adding to the developmental biology toolbox. Dev Cell 5:11–19

Yeh JR, Mohan R, Crews CM (2000) The antiangiogenic agent TNP-470 requires p53 and p21CIP/WAF for endothelial cell growth arrest. Proc Natl Acad Sci U S A 97:12782–12787

6 Tamoxifen-Based Probes for the Study of Estrogen Receptor-Mediated Transcription

J.P. Trebley, E.L. Rickert, P.T. Reyes, R.V. Weatherman

Abstract. The nuclear receptors are ideal targets to control the expression of specific genes with small molecules. Estrogen receptor can activate or repress transcription though a number of different pathways. As part of an effort to develop reagents that selectively target specific transcriptional regulatory pathways, analogs of 4-hydroxytamoxifen were synthesized with variations in the basic side chain. In vitro binding assays and cell-based luciferase reporter gene

assays confirm that all the derivatives have high affinity for the receptor and high potency at repressing direct estrogen receptor-mediated transcription.

6.1 Introduction

One of the ultimate goals of chemical genomics is to study the role of a specific protein by directly altering its activity with a small molecule. This could be performed either at the protein level by direct binding or at the transcriptional level by modulating the expression of its gene. Reagents such as small interfering RNA (siRNA) that block the production of protein have great utility, but small molecules that could either block or activate transcription of specific genes at specific time points would have a dramatic impact on discerning the role of a specific protein in cellular processes (Wang et al. 2004). One necessary component for developing these tools is a better understanding of the molecular mechanisms of transcriptional regulation and how small molecules can affect this complex process (Weatherman 2003).

Nuclear receptors such as the estrogen receptor (ER) represent an ideal system in which to study the effect of small molecules on the modulation of gene expression. Most nuclear receptors are ligand-dependent modulators of transcription, thus providing a tool to study the molecular mechanisms by which gene transcription is regulated. Nuclear receptors can activate or repress transcription upon ligand binding depending on the structure of the ligand, the nature of the promoter and the cell type (Katzenellenbogen et al. 1996). The estrogen receptor is a particularly interesting member of the nuclear receptor family because its effects on transcription can vary greatly depending on the ligand structure and the cellular context. For example, estradiol (**1**) has been shown to activate the expression of the *c-Myc* gene in breast cell lines and the breast cancer drug tamoxifen (**2**) antagonizes this activation (Shang and Brown 2002) (Fig. 1). In a uterine cell line, however, tamoxifen and estradiol both activate *c-Myc* expression. Other ER ligands with very similar structures to tamoxifen antagonize *c-Myc* expression in both types of cell lines. This tissue-dependent response profile of tamoxifen has therapeutic importance because the ER-agonist effects of tamoxifen in the uterus and in tamoxifen-resistant

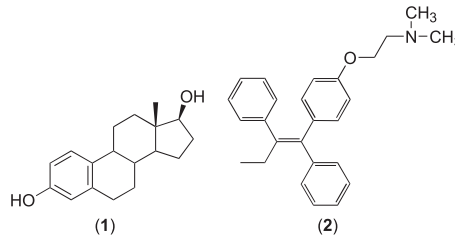


Fig. 1. Estradiol (**1**) and tamoxifen (**2**)

breast tumors are major obstacles to improving the success of tamoxifen therapy. These different response profiles allow for comparison of the different transcriptional states to help elucidate the molecular mechanisms underpinning the selective modulation of specific subsets of genes.

It is well known that estrogen receptor regulates gene transcription by binding to specific DNA sequences in the promoter region, but ER can also regulate gene transcription through indirect means. Estrogen receptor can directly interact with other transcription factors such as AP-1 and alter their activity, but it can also rapidly activate signal transduction proteins such as ERK and Akt, which can then activate downstream transcription factors such as Elk-1 and serum response factor (SRF) (Chen et al. 2004; Cheung et al. 2005). The activation of some of this rapid signaling occurs more prominently in cells in which tamoxifen acts as an estrogen receptor agonist, suggesting that the overall response profile of tamoxifen is tied to its ability to stimulate estrogen receptor crosstalk with other signal transduction pathways (Shah and Rowan 2005). Some evidence suggests that these rapid signaling events are initiated from the plasma membrane (Losel et al. 2003). Molecules that could selectively target only these crosstalk pathways would be very useful in delineating their role in the overall responses to tamoxifen. The work detailed here describes the synthesis and testing of tamoxifen analogs suitable for conjugation to other molecules such as fluorophores, affinity tags, and cell-impermeable polymer scaffolds in order to better understand the role of crosstalk signaling in the control of estrogen receptor-mediated transcription.

6.2 Results and Discussion

6.2.1 Synthesis of 4-Hydroxytamoxifen Analogs

The key issue in making tamoxifen analogs suitable for conjugation to other moieties is the placement of the attachment point. One obvious location for attachment is the amine on the basic side chain. Based on the structure of 4-hydroxytamoxifen, the most potent form of tamoxifen, bound to the ligand-binding domain of estrogen receptor alpha (ERα), the basic side chain extends out away from the interior of the binding pocket (Shiau et al. 1998). It has also previously been shown that endoxifen (**5**), a primary, bioactive metabolite of tamoxifen, can bind to the estrogen receptor both in vitro and in cells, with only small decreases in affinity compared to 4-hydroxytamoxifen (Johnson et al. 2004). Based on this evidence, a number of analogs of 4-hydroxytamoxifen with different lengths of alkylamine side chains were synthesized (Fig. 2).

The compounds were synthesized by using a modification of a previously reported synthesis of 4-hydroxytamoxifen (Yu and Forman 2003). The triphenylethylethylene scaffold can be synthesized as the diphenol (**3**) in a single step from commercially available starting materials and then monoalkylated with dibromoethane. The resulting compound (and every compound hereafter) is generated as a mixture of *E* and *Z* isomers, but the two forms readily interconvert at room temperature. Previous work with 4-hydroxytamoxifen has shown that despite this interconversion, the *Z* isomer is almost exclusively bound by the receptor both in vitro and in vivo (Katzenellenbogen et al. 1985).

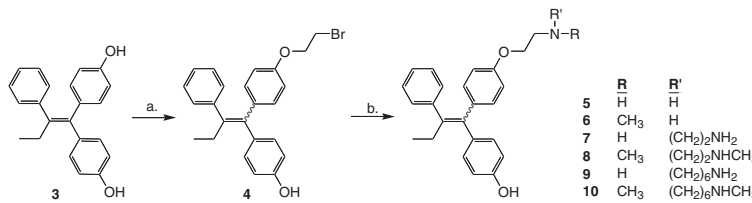


Fig. 2. a Cs_2CO_3, DMF, 60 °C; 1,2 dibromoethane, 16 h. **b** RNHR′, THF, 60 °C, sealed tube, 12 h

Coupling to different amines provided the different compounds for testing. Since the optimal distance between the tamoxifen scaffold and any conjugate is not known, alkyldiamines with two and six methylene unit spacers were synthesized. Previous work has indicated that the methylation state of the amines could also be important in increasing the affinity of ligands for the estrogen receptor, so analogs with methylated amines were also synthesized.

6.2.2 In Vitro Binding Assays

The binding affinity of the compounds for estrogen receptor alpha was measured using a fluorescence polarization-based competition assay using purified full-length human estrogen receptor alpha. Displacement of a fluorescent ER ligand from the receptor by the competitor results in a decrease in the fluorescence polarization of the fluorophore. As shown in Fig. 3 and summarized in Table 1, all of the analogs had submicromolar affinities for the receptor. The only two compounds showing

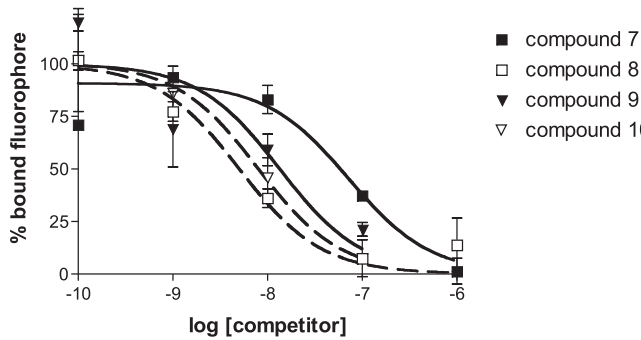


Fig. 3. Relative ER binding affinity of tamoxifen analogs **7–10**. The ability of various concentrations of different compounds to displace a synthetic fluorescent estrogen from recombinant preparations of ERα was evaluated as described in the material and methods section. *100* represents no displacement of fluorescent ligand, *0* represents total displacement. Each point represents the mean and standard error of the mean of three different samples. The *lines* represent the best fit to a single binding-site competition model. *Dashed lines* represent the fit for the methylated compounds

significantly different affinity for the receptor were the compounds with short extensions from the side chain terminating in primary amines (**5** and **7**). This could perhaps be due to some somewhat unfavorable interaction between the polar amine group and some nonpolar residues at the outer boundary of the binding pocket. A comparison of compounds **5** and **7** to compound **9** seems to indicate that pushing the primary amine further out of the binding pocket appears to be sufficient to overcome this unfavorable interaction.

6.2.3 Cell-Based Reporter Assays

The ability of the compounds to modulate estrogen receptor-mediated gene transcription was tested using a luciferase reporter gene assay. The ER-negative HeLa cervical cell line was transiently transfected with a plasmid expressing human ERα and a plasmid containing the luciferase gene under the control of the vitellogenin promoter. This promoter contains two consensus estrogen receptor binding sites and is activated strongly in the presence of ER and estradiol. None of the compounds showed any agonist activity (data not shown), so antagonist activity was determined by performing competition assays in the presence of 10 nM estradiol. As shown in Fig. 4 and summarized in Table 1, the compounds were all antagonists of estradiol-induced ER activation at the vitellogenin promoter at relatively low concentrations. Although the variability between assays is much greater with cell-based

Table 1. Summary of K_i values for compounds calculated from the receptor competition experiments and IC_{50} values vs 10 nM estradiol calculated from the reporter gene assays

Compound	K_i (nM)	IC_{50} (nM)
Estradiol (**1**)	6.3 ± 0.2	N.D.
5	48 ± 5	800 ± 400
6	8.5 ± 3.9	40 ± 10
7	32 ± 10	150 ± 50
8	3.4 ± 2.1	39 ± 12
9	9.8 ± 6.2	85 ± 55
10	6.2 ± 4.6	126 ± 33

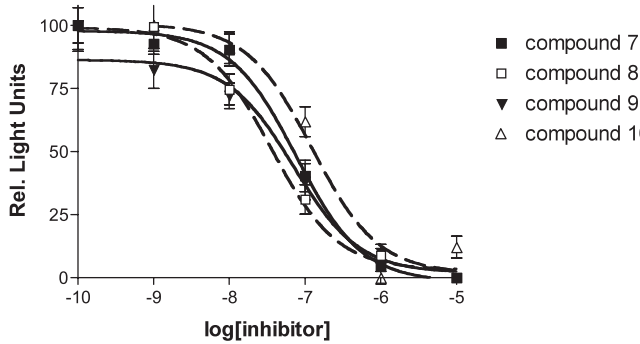


Fig. 4. Competition of the compounds **7–10** vs 10 nM estradiol in transient transfection assay of HeLa cells with ERα and the vitellogenin A2 ERE-tk-driven luciferase reporter gene. The *curve* represents the best fit to a single-site competition binding model. 100% activation represents the activation with 10 nM estradiol alone. Each point represents the mean and standard error of the mean of three different samples. *Solid lines* represent the best fit to a single binding-site competition model. *Dashed lines* represent the fit for the methylated compounds

assays than with the in vitro binding assay, compound **5** showed a significant decrease in antagonist potency compared to the other compounds. Whether this decrease is due to weaker binding affinity for the receptor or diminished cell uptake is unknown. Overall, however, all of the tamoxifen analogs inhibited ER-mediated transcriptions at concentrations that are low enough to allow for future derivatization studies.

6.2.4 Conclusion

In summary, a novel set of tamoxifen analogs has been made using a relatively simple synthetic scheme. Receptor affinity assays and reporter gene assays indicate that many of the analogs have potencies similar to tamoxifen and would make suitable analogs to conjugate to other moieties in order to study roles of the different pathways leading to estrogen receptor-mediated transcriptional regulation. These moieties will include fluorescent molecules that will allow for the visualization of binding either inside the cell or on the cell surface. The analogs

will also be conjugated to cell-impermeable polyacrylate polymers that should allow for selective targeting of membrane-initiated responses of estrogen receptor. It is envisioned that these tools will help elucidate the pleiotropic behavior of tamoxifen and could be used in the future to help engineer novel transcription factors that could either activate or repress the transcription of specific genes.

6.3 Materials and Methods

6.3.1 General Methods

All reagents were purchased from Sigma-Aldrich. The expression plasmids used in this study, pSG5-ERα and ERE-luciferase, were generously provided by Thomas Scanlan (UCSF) and have been described elsewhere (Weatherman et al. 2001; Weatherman and Scanlan 2001). The ERE-driven luciferase reporter gene consists of two repeats of the upstream region of the vitellogenin ERE promoter from −331 to −289, followed by region −109 to +45 of the thymilidate kinase upstream region and the luciferase gene. Proton and ^{13}C nuclear magnetic resonance spectra (1H NMR, 13C NMR) were obtained on a Bruker ARX300 (300 MHz) instrument; 1H NMR chemical shifts are reported as δ values in parts per million (ppm) downfield from internal tetramethylsilane. The 13C NMR chemical shifts are reported as δ values with reference to the solvent peak. Mass spectrometry (MS) and NMR instruments were provided by the Shared Resource center of the Purdue Cancer Center.

6.3.2 Synthesis of Tamoxifen Analogs

E* and *Z
4-{1-[4-(2-Bromo-ethoxy)-phenyl]-2-phenyl-but-1-enyl}-phenol (4)

Diphenol (**3**) (0.5 g, 1.59 mmol) (Yu and Forman 2003) was dissolved in DMF (10 ml) and then cesium carbonate (2.07 g, 6.4 mmol, 4 equiv.) was added and the solution was heated at 60 °C for 15 min. The 1,2 dibromoethane (0.5 ml, 5.7 mmol, 4.5 equiv.) was then added all at once and the reaction was allowed to stir for 16 h at 60 °C. Water (30 ml) was then added to the reaction mixture and the compounds were extracted

with ethyl acetate twice. The organic layer was washed with brine, dried with magnesium sulfate, and then the solvent was evaporated under reduced pressure. Purification by flash silica gel chromatography using 30% ethyl acetate in hexane as the eluent provided 0.25 g of desired product (0.59 mmol, 37% yield) as a mixture of interconverting *E* and *Z* isomers. 1H NMR (300 MHz) (CDCl3) δ 7.15 (7H, m) δ 6.94 (2H, d) δ 6.83 (1H, dd) δ 6.78 (1H, d) δ 6.62 (1H, d) δ 6.56 (1H, d) δ 4.69 (1H, t) δ 4.57 (1H, t) δ 4.12 (1H, t) δ 4.01 (1H, t) δ 3.10 (2H, q) δ 1.77 (3H, t); 13C NMR (300 MHz) (CDCl3) δ 157.18, δ 153.83, δ 142.95, δ 141.704, δ 138.01, δ 137.48, δ 136.29, δ 132.55, δ 131.15, δ 130.13, δ 128.24, δ 126.42, δ 115.43, δ 114.75, δ 114.00, δ 68.30, δ 29.68, δ 14.06. MS (CI) m/z 423/425 $(M + H)^+$.

General Synthesis of Amine Analogs

The bromide (**5**) (50 mg, 0.12 mmol) was dissolved in THF (2 ml) and 0.5 g of the appropriate diamine (as described below) was then added and the solution was heated at 60 °C for 12 h in a sealed tube. The solvent was then removed, evaporated under reduced pressure, and then purified by silica gel flash chromatography using 5.5/4/0.5 $CHCl_3/CH_3OH/NH_4OH$ as the eluent provided the product as a mixture of interconverting *E* and *Z* isomers. Below is information for each compound:

E* and *Z
4-{1-[4-(2-Aminoethoxy)-phenyl]-2-phenyl-but-1-enyl}-phenol (5)

NH_4OH was used as the amine and 43 mg of purified product was isolated (0.11 mmol, 92% yield). 1H NMR (300 MHz) (CDCl3) δ 7.15 (7H, m) δ 6.88 (1H, d) δ 6.81 (2H, dd) δ 6.72 (1H, d) δ 6.58 (1H, d) δ 6.52 (1H, d) δ 4.51 (1H, t) δ 4.37 (1H, t) δ 3.58 (1H, t) δ 3.49 (1H, t) δ 3.12 (5H, m) δ 2.02 (1H, s) δ 1.76 (3H, t). MS (CI) m/z 360 (M + H).

***E* and *Z* 4-{1-[4-(2-Methylaminoethoxy)-phenyl]-2-phenyl-but-1-enyl}-phenol (6)**

2 M methylamine in THF was used as the amine and 35 mg of purified product was isolated (0.094 mmol, 78% yield). 1H NMR (300 MHz) (CDCl3) δ 7.15 (7H, m) δ 6.88 (1H, d) δ 6.81 (2H, dd) δ 6.72 (1H, d) δ 6.58 (1H, d) δ 6.52 (1H, d) δ 5.76 (2H, s) δ 4.51 (1H, t) δ 4.37 (1H, t)

δ 3.58 (1H, t) δ 3.49 (1H, t) δ 3.12 (5H, m) δ 2.02 (1H, s) δ 1.76 (3H, t); 13C NMR (300 MHz) (CDCl3) δ 157.61, δ 156.77, δ 156.26, δ 155.34, δ 143.18, δ 141.20, δ 138.43, δ 137.28, δ 136.78, δ 135.47, δ 135.10, δ 132.43, δ 131.13, δ 130.16, δ 128.26, δ 126.27, δ 115.71, δ 115.04, δ 114.40, δ 113.65, δ 66.56, δ 50.81, δ 36.14, δ 29.50, δ 14.11. MS (CI) m/z 374 (M + H).

E and *Z* 4-(1-{4-[2-(2-Aminoethylamino)-ethoxy]-phenyl}-2-phenyl-but-1-enyl)-phenol (7)

Ethylenediamine was used as the amine and 32 mg of purified product was isolated (0.087 mmol, 73% yield). 1H NMR (300 MHz) (CD_3OD) δ 7.15 (7H, m) δ 6.88 (1H, d) δ 6.81 (2H, dd) δ 6.72 (1H, d) δ 6.58 (1H, d) δ 6.52 (1H, d) δ 4.51 (1H, t) δ 4.37 (1H, t) δ 3.58 (3H, t) δ 3.49 (3H, t) δ 3.12 (5H, m) δ 2.02 (1H, s) δ 1.76 (3H, t); 13C NMR (300 MHz) (CD_3OD) δ 159.4, δ 158.5, δ 157.9, δ 157.0, δ 144.6, δ 142.4, δ 142.2, δ 140.2, δ 138.43, δ 137.6, δ 136.3, δ 133.47, δ 132.43, δ 131.13, δ 130.16, δ 128.26, δ 126.27, δ 115.71, δ 115.04, δ 114.40, δ 113.65, δ 66.56, δ 42.13, δ 31.2, δ 29.50, δ 14.11. MS (CI) m/z 403 (M + H).

E and *Z* 4-[1-(4-{2-[Methyl-(2-methylaminoethyl)-amino]-ethoxy}-phenyl)-2-phenyl-but-1-enyl]-phenol (8)

N,N′ dimethylethylenediamine was used as the amine and 15 mg of purified product was isolated (0.035 mmol, 29% yield). 1H NMR (300 MHz) ($CDCl_3$) δ 7.15 (7H, m) δ 6.88 (1H, d) δ 6.81 (2H, dd) δ 6.72 (1H, d) δ 6.58 (1H, d) δ 6.52 (1H, d) δ 4.37 (1H, t) δ 4.12 (3H, t) δ 3.95 (3H, t) δ 3.6 (5H, m) δ 2.58 (3H, s), δ 2.50 (3H, s), δ 2.02 (1H, s) δ 1.76 (3H, t).

E and *Z* 4-(1-{4-[2-(6-Amino-hexylamino)ethoxy]-phenyl}-2-phenyl-but-1-enyl)-phenol (9)

1,6-diaminohexane was used as the amine and 40 mg of purified product was isolated (0.092 mmol, 77% yield). 1H NMR (300 MHz) ($CDCl_3$) δ 7.15 (7H, m) δ 6.88 (1H, d) δ 6.81 (2H, dd) δ 6.72 (1H, d) δ 6.58 (1H, d) δ 6.52 (1H, d) δ 3.6 (5H, m) δ 2.58 (2H, t), δ 2.50 (2H, t), δ 2.02 (1H, s) δ 1.6 (3H, t), δ 1.3 (8H, m).

***E* and *Z* 4-[1-(4-{2-[Methyl-(6-methylaminohexyl)-amino]-ethoxy}-phenyl)-2-phenyl-but-1-enyl]-phenol (10)**

N,N′ dimethyl-1,6-diaminohexane was used as the amine and 18 mg of purified product was isolated (0.037 mmol, 31% yield). 1H NMR (300 MHz) ($CDCl_3$) δ 7.15 (7H, m) δ 6.88 (1H, d) δ 6.81 (2H, dd) δ 6.72 (1H, d) δ 6.58 (1H, d) δ 6.52 (1H, d) δ 3.2 (2H, t) δ 3.1 (2H, t), δ 2.55 (2H, t), δ 2.45 (6H, s), δ 2.22 (2H, t) δ 1.6 (3H, m), δ 1.3 (8H, m).

6.3.3 Fluorescence Polarization Assay

Fluorescent polarization-based competition binding assays were conducted to determine the relative affinity of the 4-hydroxytamoxifen analogs for ERα using a commercially available kit (PanVera Corp., Madison, WI). Briefly, serial dilutions of the different compounds were prepared in ES2 screening buffer (100 mM potassium phosphate, pH7.4, 100 μg/ml bovine gamma globulin) and 50 μl of each concentration was aliquoted into three wells of a black 96-well assay plate. Fifty microliters of a solution containing 20 nM recombinant ERα, and 2 nM of a proprietary fluorescent ER ligand (Fluormone-ES2) was added to each well. The plate was shaken on a plate mixer and incubated for 2 h in the dark at room temperature. Fluorescence polarization signals were then measured using a Packard Fusion fluorimeter. The data were then fit to a single binding site competition curve by nonlinear regression analysis (Prism 3 software package). K_i values were determined from the average of three different experiments and calculated using a $K_D = 4$ nM for Fluormone binding to ERα.

6.3.4 Cell Culture and Transient Transfection Experiments

Cell Culture

HeLa cells were obtained from the American Type Culture Collection (ATCC). HeLa cells were maintained in DME media without phenol red (Sigma) supplemented with 4.5 g/l glucose, 0.876 g/l glutamine, 100 mg/l streptomycin sulfate, 100 units/ml of penicillin G, and 10% FBS at 37 °C in an air/carbon dioxide (95:5) atmosphere. Transfection

assays were run with the same media conditions except the FBS was treated for 24 h with dextran-coated charcoal.

Transient Transfection Assays

HeLa cells were plated in 24-well plates and grown to approximately 70%–80% confluency. Transfections were performed according to the protocol for Lipofectamine 2000 (Invitrogen). In order to normalize for the transfection efficiency in each well, the dual luciferase system was used in which a constitutively expressed, chemically orthogonal luciferase expression vector was also transfected. The total amount of DNA/well for each plasmid was as follows: pSG5-ERα 0.25 μg/well, ERE-luciferase 0.5 μg/well, and *Renilla*-luciferase 0.25 μg/well. The ratio of total DNA/Lipofectamine 2000 was 1:5. After transfection, the plates incubated at 37 °C for 6 h before dosing with drug. All drugs were delivered in DMSO or ethanol and the total concentration of organic solvent in each was 0.1%. For competition experiments, the drug was added to media already containing 10 nM estradiol. After 18–24 h, the cells were lysed and assayed for dual luciferase activity in a TopCount luminometer according to the protocol provided by Promega. The relative light units (RLU) were then calculated by dividing the output of the ERE-driven luciferase in each well by the output of the *Renilla* luciferase. Each drug concentration was tested in triplicate.

Acknowledgements. We acknowledge the financial support of the American Cancer Society (IRG-58-006-41A), the Indiana Elk's Charities, the Army Breast Cancer Research Program (BC030507), and the American Association of Colleges of Pharmacy New Investigator Program. J.P.T. was supported by an IRL Fellowship from the National Science Foundation. P.T.R. was supported by a MARC predoctoral fellowship from the National Institutes of General Medical Sciences (F31-GM67595–01).

References

Chen CC, Lee WR, Safe S (2004) Egr-1 is activated by 17beta-estradiol in MCF-7 cells by mitogen-activated protein kinase-dependent phosphorylation of ELK-1. J Cell Biochem 93:1063–1074

Cheung E, Acevedo ML, Cole PA, Kraus WL (2005) Altered pharmacology and distinct coactivator usage for estrogen receptor-dependent transcription through activating protein-1. Proc Natl Acad Sci U S A 102:559–564

Johnson MD, Zuo H, Lee KH, Trebley JP, Rae JM, Weatherman RV, Desta Z, Flockhart DA, Skaar TC (2004) Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. Breast Cancer Res Treat 85:151–159

Katzenellenbogen JA, Carlson KE, Katzenellenbogen BS (1985) Facile geometric isomerization of phenolic non-steroidal estrogens and antiestrogens: limitations to the interpretation of experiments characterizing the activity of individual isomers. J Steroid Biochem 22:589–596

Katzenellenbogen JA, O'Malley BW, Katzenellenbogen BS (1996) Tripartite steroid hormone receptor pharmacology: interaction with multiple effector sites as a basis for the cell- and promoter-specific action of these hormones. Mol Endocrinol 10:119–131

Losel RM, Falkenstein E, Feuring M, Schultz A, Tillmann HC, Rossol-Haseroth K, Wehling M (2003) Nongenomic steroid action: controversies, questions, and answers. Physiol Rev 83:965–1016

Shah YM, Rowan BG (2005) The Src kinase pathway promotes tamoxifen agonist action in Ishikawa endometrial cells through phosphorylation-dependent stabilization of estrogen receptor (alpha) promoter interaction and elevated steroid receptor coactivator 1 activity. Mol Endocrinol 19:732–748

Shang Y, Brown M (2002) Molecular determinants for the tissue specificity of SERMs. Science 295:2465–2468

Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL (1998) The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell 95:927–937

Wang S, Sim TB, Kim YS, Chang YT (2004) Tools for target identification and validation. Curr Opin Chem Biol 8:371–377

Weatherman RV (2003) Chemical approaches to studying transcription. Org Biomol Chem 1:3257–3260

Weatherman RV, Scanlan TS (2001) Unique protein determinants of the subtype-selective ligand responses of the estrogen receptors (ERalpha and ERbeta) at AP-1 sites. J Biol Chem 276:3827–3832

Weatherman RV, Clegg NJ, Scanlan TS (2001) Differential SERM activation of the estrogen receptors (ERalpha and ERbeta) at AP-1 sites. Chem Biol 8:427–436

Yu DD, Forman BM (2003) Simple and efficient production of (Z)-4-hydroxytamoxifen, a potent estrogen receptor modulator. J Org Chem 68:9489–9491

7 Protein Structure Similarity Clustering and Natural Product Structure as Guiding Principles for Chemical Genomics

M.A. Koch, H. Waldmann

Abstract. The majority of all proteins are modularly built from a limited set of approximately 1,000 structural domains. The knowledge of a common protein fold topology in the ligand-sensing cores of protein domains can be exploited for the design of small-molecule libraries in the development of inhibitors and ligands. Thus, a novel strategy of clustering protein domain cores based exclusively on structure similarity considerations (protein structure similarity clustering, PSSC) has been successfully applied to the development of small-molecule inhibitors of acetylcholinesterase and the 11β-hydroxysteroid dehydrogenases based on the structure of a naturally occurring Cdc25 inhibitor. The efficiency of making use of the scaffolds of natural products as biologically prevalidated

starting points for the design of compound libraries is further highlighted by the development of benzopyran-based FXR ligands.

7.1 Introduction

The rapidly growing knowledge related to possible new targets for chemical biology and medicinal chemistry research has been paralleled by the development of combinatorial chemistry techniques that allow for rapid synthesis of compound libraries.

In this context diversity (Golebiowski et al. 2001; Mason and Hermsmeier 1999; Schreiber 2000), drug-likeness (Walters et al. 1999; Ajay et al. 1998; Sadowski and Kubinyi 1998; Ghose et al. 1999; Lee and Schneider 2001), and biological relevance (Breinbauer et al. 2002; Koch et al. 2003; Koch and Waldmann 2004, 2005) have been identified as being central in designing compound libraries for protein ligand development. The last criterion may be fulfilled if so-called privileged (Evans et al. 1988) structures or natural products that evolved to bind to biological macromolecules are taken as guiding structural principles. Indeed, hit rates increase significantly if natural compounds or analogs thereof are included in high-throughput screenings (Breinbauer et al. 2002; Koch et al. 2003; Brohm et al. 2002a, b). Considering potential target proteins, several concepts mainly based on a clustering of target proteins according to evolutionary relatedness and conserved molecular recognition have been developed and applied to steer ligand development. Since, for proteins, spatial structure is more conserved in evolution than amino acid sequence (Grishin 2001), a focus on the principal architecture and structure of proteins does provide alternative guiding principles for the development of biologically relevant compound collections.

The structural conservatism of nature in the design of proteins that can be regarded as modularly built biomolecules assembled from individual domains as building blocks may provide opportunities to develop guidelines for the identification of biologically relevant ligand structures. Thus, the core structure of protein domains, i.e., the catalytic or the ligand-sensing cores, are widely reused in different functional contexts in a more or less modified form. Very often, the binding or catalytic

sites in proteins are diverse, whereas the domain cores are structurally conserved.

The ligand-binding or catalytic sites are the most relevant subsets of a protein domain from the point of view of development of small-molecule binders. These are – more precisely – located within the so-called ligand-sensing core of the domain where the actual catalytic conversion in the case of enzymes or the binding event of small-molecule ligands takes place (Koch et al. 2004; Koch and Waldmann 2005). To define these topologically distinct domain parts is all the more important when ligand binding domains are quite large compared to the structural subset relevant for binding. Thus, we suggest confining structural similarity considerations to these distinct parts of a protein domain and to group ligand-sensing cores instead of whole domains according to 3D similarities into so-called protein structure similarity clusters (PSSCs). This further broadens the structural view on proteins as it also allows consideration of structural parts of a previously defined fold. Thus, some structural fold characteristics may be reused within a different fold. When these similar fold parts describe the ligand-sensing cores of a domain, then a comprehensive clustering according to purely structural arguments may be a viable abstracting rationale in the context of ligand development, irrespective of more or less arbitrarily assigned fold types.

7.2 Protein Structure Similarity and Natural Product Frameworks as Guiding Principles for Compound Library Development

Classically, potential protein targets are clustered into target families on the basis of functional relatedness and amino acid sequence homology alignments reflecting their evolutionary relationship. This categorization is then used to pool known ligands of a target family and to take them as starting points for compound library design. This strategy constitutes a rationale that allows the direct conversion of genetic information and relatedness into actual chemical ligand design. A further, analogous principle was outlined as the structure-activity relationship homology concept. Potential drug discovery targets are grouped into families based on the relatedness of the structure-activity relationship of their ligands

(Frye 1999). It is assumed that the conservation of binding-site architectures and thus the relatedness of molecular recognition within a target family or a subfamily thereof translate into a conservation of ligand scaffolds that bind to these targets (Jacoby et al. 2003). The major limitation of these concepts is that usually only close sequence homologs can be considered because target proteins and their ligands are predominantly categorized on the basis of function and sequence similarity. Family assignment derived from sequence information alone in the absence of structural information usually requires sequence identities greater than 30% (Gerlt and Babbitt 2001).

As protein spatial structures are typically more conserved in evolution than amino acid sequences (Grishin 2001), these should first and foremost be considered for the clustering of target proteins. Usually nature's structural conservatism is confined to the domain cores of proteins, whereas the binding sites for ligands may be structurally diverse yet topologically similarly located. This concurrence of binding sites rarely indicates any obvious functional similarity, since substrates can show high variability concerning their chemical structure. This observation may be attributed to a general tendency of certain folds to bind substrates at similar locations, at so-called supersites, despite little evidence of a common ancestor for the proteins considered. Possible explanations for this phenomenon may be certain principles of protein structure or chemical constraints that may lead to common "optimal" binding sites even when proteins do not share a common ancestor (Russell et al. 1998).

These observations indicate that a clustering of ligand-sensing or catalytic cores of proteins exclusively based on structural considerations may be a valuable tool for compound library development thereby taking into consideration nature's structural conservatism. Thus, we propose to group protein domain cores in a protein structure similarity cluster (PSSC) because of 3D similarity and regardless of significant sequence similarity. A substance that binds to one member protein of a PSSC may be used as a starting point for the development of small-molecule modulators of the other members of the cluster. The biological diversity occurring in the ligand binding sites has to be addressed with an appropriate chemical diversity that has to be generated around the identified structural framework in a library approach in order to find ligands with

an acceptable frequency in the first place and also to achieve potency and selectivity. Thus, focusing first on the conserved architecture of protein domains or domain cores is used as an abstracting guiding principle that leads to biologically relevant frameworks to be used as cornerstones of a compound library with significantly enhanced hit rates (Koch et al. 2003, 2004; Koch and Waldmann 2004) (see Fig. 1).

The PSSC approach, synergistically applied together with natural product-inspired combinatorial chemistry, may provide a strategy for the development of compound libraries for chemical biology and medicinal chemistry research. Natural products can be regarded as small molecules evolutionarily selected for binding to protein domains. They did interact with multiple proteins in the course of their biosynthesis and they target further proteins when they fulfill their biological functions, e.g.,

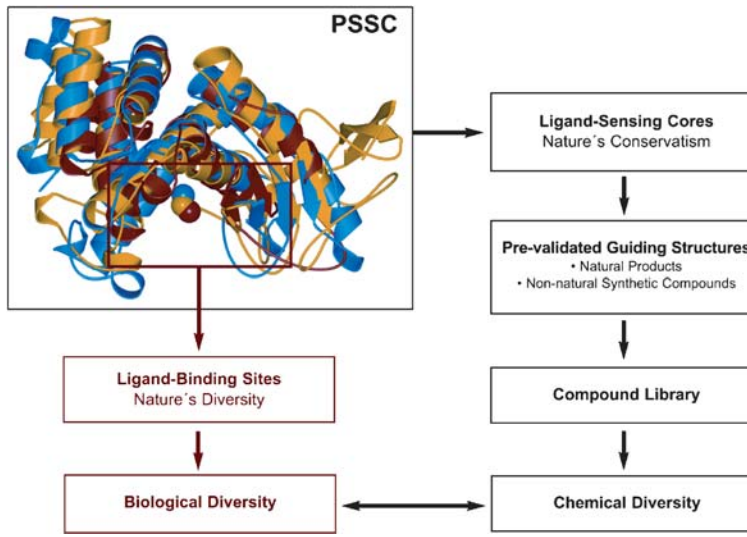


Fig. 1. Protein fold conservatism and binding site diversity – implications for compound library development. Structural conservatism in the ligand-sensing cores of proteins leads to the identification of guiding structures for ligand development. In order to address the biological diversity in the ligand binding sites, however, an appropriate chemical diversity has to be generated around the identified guiding scaffolds in a library approach

in communication or chemical defense. This is further supported by the finding that major classes of natural products show multiple biological activities. Therefore, due to their biological prevalidation and evolutionary proving, they are particularly well suited for compound library development. However, not only natural products fulfill the criterion of biological prevalidation to serve as promising guiding structures for library design. Various non-natural product classes, discovered in particular in medicinal chemistry programs, have proven to incorporate this property as well.

PSSC-guided small molecule binder development in principle (although this may not necessarily be desirable) could be initiated exclusively using bioinformatics tools without further knowledge about the target's biological functions, binding partners, and so on, which are usually obtained by laborious biochemical and cell biological techniques. In fact, the evolved ligands can be used for further characterization of the physiological role of the target protein, which is of outstanding importance in the target validation process and in chemical biology research.

In this chapter, evidence for the applicability of the PSSC concept in compound library development is given. First, an example from the literature that led to the discovery of small-molecule modulators of protein function is analyzed in light of the PSSC concept. Finally, the first successful application of the PSSC approach as a rationale for actual *de novo* compound library design is discussed.

7.3 A Retrospective Case Study: Development of Nuclear Hormone Receptor Modulators

Nuclear receptors (NRs) are ligand-inducible transcription factors consisting of a ligand binding domain (LBD) and a DNA-binding domain (DBD). Structural comparison of the moderately conserved NR LBDs reveals that these domains exhibit a canonical fold, consisting of 12 α-helices, which is better conserved than the primary sequence. In the hydrophobic core of the LBD, the fully buried ligands are bound. NRs comprise receptors for hydrophobic molecules such as steroid hormones, retinoic acids, thyroid hormones, fatty acids, leukotrienes, and prostaglandins (Robinson-Rechavi et al. 2003). As they are naturally

switched on and off by small molecule hormones bearing physicochemical properties that are very similar to therapeutic chemical entities, they intrinsically represent very attractive and promising targets in terms of therapeutic applications. Examples for the current therapeutic exploitation of NRs are, among others, the use of estrogen receptor-α (ERα) antagonists (for example, tamoxifen) for the treatment of breast cancer and the clinical use of the structural class of thiazolidinediones (the so-called glitazones), which are agonists of peroxisome proliferator-activated receptor γ (PPARγ), and therefore insulin sensitizers, as antidiabetic drugs (Schapira et al. 2000; Schapira 2002).

The farnesoid X receptor (FXR) has recently been identified as a bile acid-activated NR that plays a regulatory role in cholesterol metabolism. Recent advances in FXR biology suggest that FXR may represent a valuable and pharmacologically interesting target for the identification of novel drugs to treat dyslipidemia and cholestasis (Claudel et al. 2004). For further validation of FXR as a potential drug target, however, it is necessary to understand its physiological role precisely. A selective, cell-permeable high-affinity agonist as a tool compound would be helpful in this context to elucidate FXR-mediated effects in a combined chemical and biological approach. In order to find a starting point in chemical structural space for compound library development, a clustering approach based on protein structure similarity occurring in the ligand binding domains of the NRs would also have been successful. Thus, the ligand-binding domains of ERβ, PPARγ, and FXR may be grouped into a PSSC despite low sequence homology (sequence identities amount to 20%) (see the superimposition of these NRs' LBDs in Fig. 2A). The natural product genistein, an isoflavone phytoestrogen (**1**, Fig. 2B), is found in significant levels in soy beans and soy products. Genistein binds to both estrogen receptor (ER) isoforms α and β with moderate affinity but exhibits a preference for ERβ, acting as a partial agonist (Pike et al. 1999). Additionally, genistein is found to be a PPARγ agonist (Dang et al. 2003). Another known synthetic PPARγ agonist is troglitazone (**2**, Fig. 2B), which was in clinical use as an antidiabetic agent but withdrawn from market due to its liver toxicity (Van Gaal et al. 2002). Genistein and troglitazone have in common a benzopyran core moiety. The benzopyran framework represents a privileged motif and occurs in many natural products that cover a broad spectrum of bio-

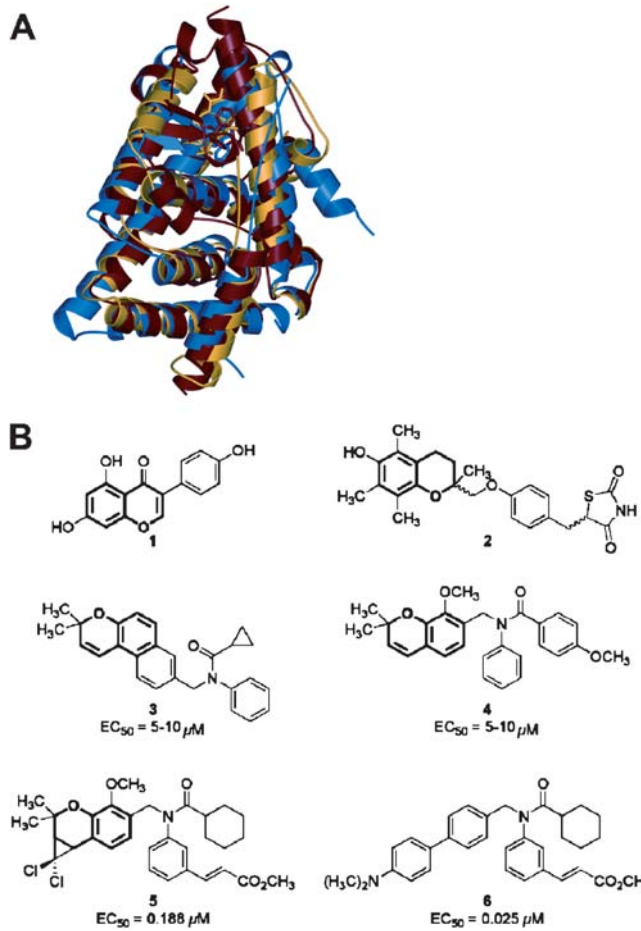


Fig. 2. A Superimposition of the X-ray structures of the ligand binding domains of ERβ, PPARγ and FXR, each with bound ligand. These proteins exhibit the same fold (SCOP: NR LBD) and their ligand-binding sites share a common conserved location. ERβ with genistein (**1**, *blue*), PPARγ with rosiglitazone (*red*), FXR with **6** (*yellow*). **B** Genistein (**1**) and troglitazone (**2**) bear a benzopyran moiety and bind to ERβ and PPARγ, respectively. Compounds **3**–**5** are benzopyran-inspired FXR agonists generated by combinatorial solid- and solution-phase chemistry

logical activities such as antitumor, antibacterial, and estrogenic effects, to name but few. Thus, in light of structural conservation of the LBD fold, a compound library inspired by the structure of a natural product modulator of one member of the NR class, here genistein, may also yield hits for the targeting of FXR.

Indeed, in an initial screening of a combinatorial natural product-like and diversity-orientated library of 10,000 benzopyran-based small molecules built up by Nicolaou and co-workers (Nicolaou et al. 2000a, b) using a cell-based assay for FXR activation, afforded several lead compounds (see **3** and **4**, Fig. 2B). Further elaboration of the identified lead structures yielded FXR binders with EC_{50} values in the low nanomolar range (see **5** and **6**, Fig. 2B). In compound **6**, the benzopyran moiety was further deconstructed to the privileged biaryl motif (Downes et al. 2003; Nicolaou et al. 2003). These findings convincingly support the idea that the PSSC approach in conjunction with natural product (genistein)-inspired compound library design would have been successful for the design of FXR agonists as well.

The above analysis and reinterpretation of the example for ligand development extracted from the literature in the light of the PSSC approach suggests the applicability of the new concept to compound library design and demonstrates that such ligand development can be successfully initiated and guided by the structure of a natural product that is known to bind to one member protein of a PSSC. This analysis and similar examples extracted from the literature encouraged us to use the PSSC approach in a real *de novo* ligand design situation. The following example represents the first successful forward application of the PSSC concept in conjunction with natural product-guided compound library development.

7.4 PSSC Cdc25A Phosphatase–Acetylcholinesterase–11β-Hydroxysteroid Dehydrogenases

The concept of analyzing protein domain cores, purely with respect to structural similarity without regard to functional or evolutionary arguments, and to group the respective ligand-sensing cores into a protein structure similarity cluster (PSSC) proposed by us (Koch et al. 2004;

Koch and Waldmann 2004, 2005) in conjunction with natural product guided compound library development, was successfully applied *ab initio*, starting with the phosphatase Cdc25A as the initial protein of interest. A data mining and analysis strategy was developed that allowed for identification of structurally similar protein cores from large data sets (see Fig. 3).

Applying this strategy, Cdc25A phosphatase (Cdc25A), acetylcholinesterase (AChE), and 11β-hydroxysteroid dehydrogenases type 1 and type 2 (11βHSD1 and 11βHSD2) were identified as sharing significant

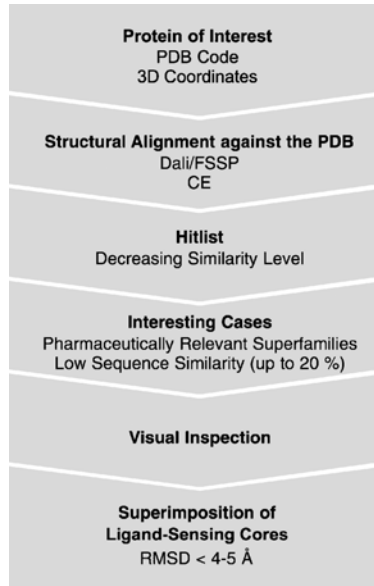


Fig. 3. Database search strategy and procedure developed for the identification of protein structure similarity clusters (PSSCs). Database searches, e.g., in the Dali/FSSP (Holm et al. 1996; Holm et al. 1997) (http://www.ebi.ac.uk/dali/) and the CE (Shindyalov et al. 1998) (http://cl.sdsc.edu/ce.html) databases, using the 3D coordinates of a query protein may provide insights into their structural neighborhood. For compound library development, since the ligand-sensing cores of the proteins are of paramount importance, one has to make sure that these relevant parts of the protein domains share structural similarity

structural resemblance in their catalytic cores (see Fig. 4A). Consequently, they were grouped into a PSSC. Despite low sequence similarity (sequence identities amount to 5%–8%), the ligand-sensing cores of all three enzymes could be aligned structurally with RMSD (root mean square deviation for aligned C^{α} positions) values ranging from 3 Å to 4 Å.

All three enzymes represent known or viable targets for the treatment of various diseases. Cdc25A, which regulates progression of cell division at the G1→S checkpoint by dephosphorylating Cdk2/cyclin complexes (Fauman et al. 1998), may be a valuable target for the development of novel anti-tumor drugs. AChE hydrolyzes the neurotransmitter acetylcholine and thereby terminates impulse transmission at cholinergic synapses (Sussman et al. 1991) and is currently a major target in the treatment of myasthenia gravis, glaucoma, and Alzheimer's disease (Ibach et al. 2004). 11βHSD1 is essential for the local and tissue-specific activation of glucocorticoid receptors, since it catalyzes the oxo-reduction of cortisone to cortisol, and it may be a promising therapeutic target for the antagonization of glucocorticoid actions (Chrousos 2004; Walker et al. 2003). Its inhibition is considered to be a promising approach to the treatment of obesity (Masuzaki et al. 2001; Schweizer et al. 2003), the metabolic syndrome (Masuzaki et al. 2003; Paterson et al. 2004), diabetes type 2 (Alberts et al. 2002; Ross et al. 2004), and cognitive dysfunction (Sandeep et al. 2004). The 11βHSD2 isoenzyme catalyzes exclusively the oxidation of cortisol, and inhibition of 11βHSD2 causes sodium retention resulting in hypertension (New et al. 1999). Therefore isoenzyme specificity is a major prerequisite for the clinical use of 11βHSD1 inhibitors.

In light of this structural similarity, a compound collection was synthesized based on a naturally occurring inhibitor of one of the enzymes.

The sesterterpene Dysidiolide (**7**, see Fig. 4B) is an inhibitor of Cdc25A. Based on earlier investigations (Brohm et al. 2002a, b) and literature reports on the phosphatase-inhibiting activity of related natural products (Lyon et al. 2002), it was hypothesized that the γ-hydroxybutenolide group incorporated into the natural product is a major determinant of its phosphatase inhibiting activity. Consequently, a 147-member compound collection of γ-hydroxybutenolides, and closely related α,β-unsaturated five-membered lactones was synthesized and subjected to

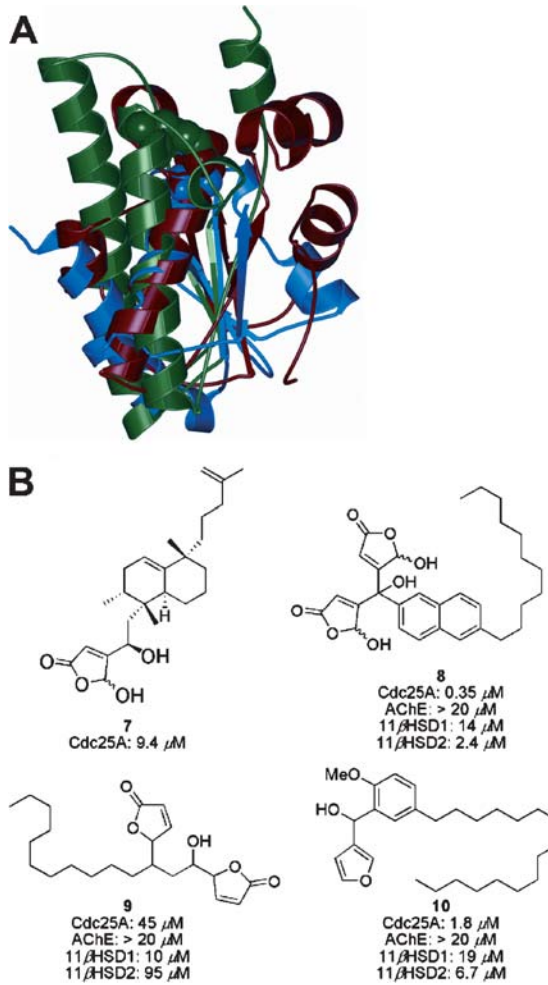


Fig. 4. A Superimposition of the catalytic cores of Cdc25A (*red*), 11βHSD1 (*green*, homology model) and AChE (*blue*). The key catalytic residues, Cys^{430} (Cdc25A), Tyr^{183} (11βHSD1), and Ser^{200} (AChE), shown in CPK representation, are located similarly. **B** Analogs of the naturally occurring Cdc25A inhibitor dysidiolide (**7**) profiled against Cdc25A, AChE, and 11βHSD1/2 (IC_{50} values are given), the PSSC member proteins

biochemical investigation for possible inhibition of Cdc25A, AChE, or 11βHSD1/2. Compounds displaying IC_{50} values 10 μM were considered as hits (see Fig. 4B). Of the 147 compounds investigated, 42 qualified as hits in the Cdc25A assay. The most potent compound **8** had an IC_{50} value of 350 nM, which is significantly lower than the reported IC_{50} value for dysidiolide (9.4 μM; Gunasekera et al. 1996). Three compounds inhibited AChE with IC_{50} values of 1.3–4.5 μM. The collection contained three 11βHSD1 inhibitors with IC_{50} values of 7.8–10 μM and four 11βHSD2 inhibitors with IC_{50} values of 2.4–6.7 μM. Thus, the hit rates for the enzymes identified as being similar to Cdc25A are approximately 2%–3%.

Even at this comparably small library size, the hits displayed a pronounced degree of selectivity for individual enzymes and also for the isoenzymes 11βHSD1 and 11βHSD2. Thus, compound **8** was a significantly more potent inhibitor for Cdc25A than for the other enzymes. Most remarkably, the α,β-unsaturated lactone **9** inhibited only the therapeutically relevant 11βHSD1, but not or only very weakly the other enzymes investigated. Also, a furan derivative (**10**) was identified as an inhibitor for Cdc25A and for 11βHSD2. A selective inhibitor for AChE could not be discovered.

This example gives evidence for the *de novo* applicability of the PSSC concept. It clearly demonstrates that also 3D protein structures generated using homology modeling techniques can be considered as for both 11βHSD1 isoforms no crystal structures were available at the time of investigation. It is, however, expected that the reliability of the PSSC approach will increase with the growing number of experimentally determined protein structures becoming available.

7.5 A New Guiding Principle for Chemical Genomics?

The examples detailed above demonstrate that PSSC can serve as conceptually new principle guiding the development of compound libraries, in particular for medicinal chemistry research. However, beyond this, PSSC may open up new opportunities for research in the currently developing field of chemical genomics. In a general sense, chemical genomics may be defined as the genomic response to chemical compounds, i.e.,

chemistry is employed to probe a biological system. A more focused, workable definition appears to be the identification of small molecule lead-like compounds for a member of a gene family product and the subsequent use of these compounds to elucidate the function of other (disease-associated) members of the gene family. Currently, in this approach the gene family products are predominantly classified on the basis of sequence similarities and function, i.e., into kinases, phosphatases, proteases, etc. (see Fig. 5).

A protein domain core-centered approach that considers domain organization and architecture, however, may provide a new guiding principle for the combinatorial development of compounds that will pave the way to a new series of chemical proteomic and genomics experiment. Accordingly, a family of gene products (proteins) of interest would be dissected in structural terms, i.e., into domains. After domain assignment, structural comparison of the ligand-sensing cores with known domains/folds would be carried out, leading to a cluster of structurally related domain cores that may share little sequence homology (see Fig. 5). This pool of structurally similar ligand-sensing cores with their respective ligands may be employed for the generation of potent *and* selective small molecule modulators of protein function of the PSSC member proteins. The structures of known ligands for a spatially similar reference domain core constitute biologically validated starting points in chemical structural space for the design of focused libraries yielding comparably high hit rates when screened against the PSSC member proteins. Selectivity and enhanced potency can be achieved by generating diversity around the small-molecule binder core structure, thus taking into account the requirements of the individual binding pockets possibly harboring substantially different amino acid residues. This strategy initially reduces complexity and focuses on the 3D similarity of protein domain cores. It leads to structural frameworks guiding compound library development.

As such, non-natural substances or natural products known to bind to one member protein of the PSSC may serve as leitmotifs. For natural products, we postulate that their evolutionarily selected scaffolds represent biologically prevalidated structures providing basic affinity to the protein domain cores, with which they were evolved to interact. Natural products can thus be regarded as inherently promising guiding compounds for the design of domain-selective small-molecule modula-

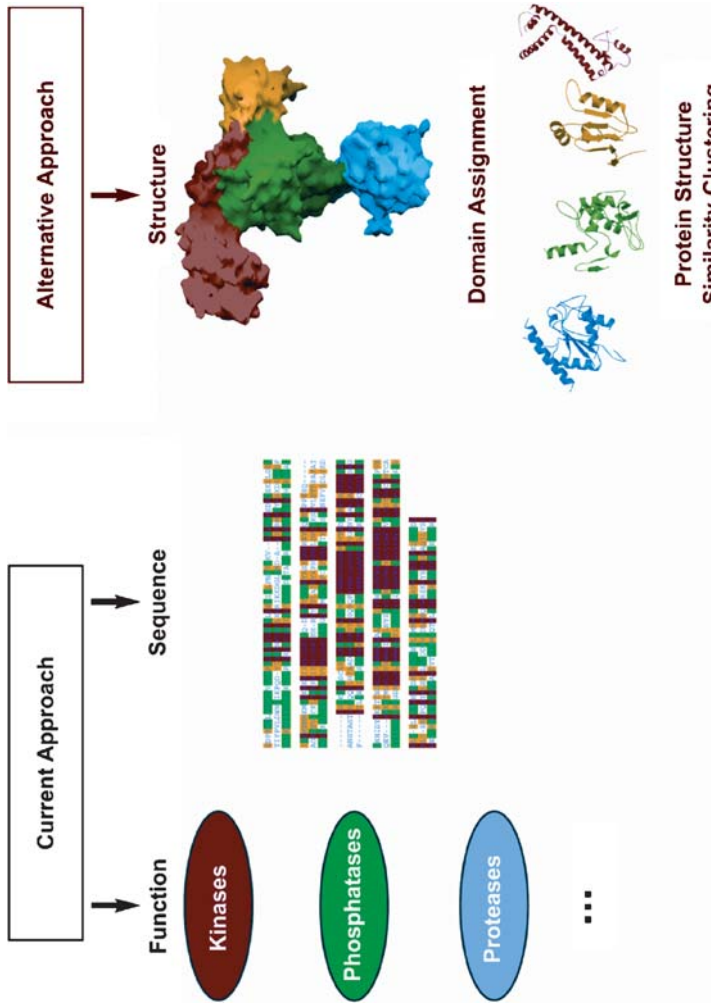


Fig. 5. Approaches for protein categorization. The currently predominating approach in chemical genomics based on the clustering of target proteins according to their sequence and function may be complemented by an alternative approach based on a purely structural view of protein domains or cores

tors of protein function. Of course, this prerequisite is not only fulfilled by natural products. Also, non-natural synthetic small-molecule ligands with known biological relevance can be regarded as valuable starting points. In a sense, their binding properties have evolved in the course of an accelerated artificial evolutionary process.

Once a biologically relevant structural framework has been found, the varying requirements of the different binding sites can be addressed by generating diversity around this core structure, thus allowing for evolution of potent and selective binders. This can be accomplished using a library approach that may be supported by molecular modeling techniques. Often, due to the structural complexity of natural products and analogs thereof, the synthesis of natural product-derived compound libraries may require the development of demanding multi-step synthesis sequences including, for instance, enantioselective transformations. Thus, the initial investment in the synthesis of such libraries may be high. However, compound libraries generated following this approach may be small compared to classical combinatorial libraries primarily developed on the basis of chemical feasibility and accessibility. Thus, a higher developmental investment for the generation of such focused libraries due to the structural complexity of the natural products used as sources of inspiration is justified, since such libraries will yield high hit rates and – most importantly – biologically prevalidated hits. In light of this argument, further development of chemical methods enabling the rapid and efficient synthesis of complex molecules in library format has to be regarded as being of utmost importance (Abreu and Branco 2003; Arya et al. 2002; Barun et al. 2004; Breinbauer et al. 2002; Brohm et al. 2002a, b; Nicolaou et al. 2001).

To summarize, in the initial step of the PSSC approach the overall structural resemblance of protein domain cores is employed as a guiding principle to choose possible small-molecule binder scaffolds. In a second step, the structural diversity occurring in the binding sites of the PSSC member proteins is addressed by synthesizing a compound library, thereby ideally identifying chemical entities that efficiently address the biological diversity found in the binding sites and yielding selective and potent binders. The advantage of a certain initial indeterminateness when comparing overall domain core structures is that predicted and modeled protein structures with a certain tolerance with

respect to the binding sites can also be considered because, finally, the indeterminateness is overcome by the combinatorial approach.

The PSSC concept should be particularly helpful in the initial stages of compound development and screening when little may be known about the function of a newly discovered protein. Thus, PSSC serves as an abstracting principle that allows for identification of novel compound classes for a given target. The identified structural frameworks then have to be refined (!) in a medicinal chemistry program to optimize selectivity and reduce unwanted activities.

Acknowledgements. We thank Dr. Ingrid Vetter (Max Planck Institute of Molecular Physiology, Dortmund) for stimulating discussions. This work was supported by the Max-Planck-Gesellschaft, the Deutsche Forschungsgemeinschaft, and the Studienstiftung des deutschen Volkes (doctoral scholarship for M.A.K.).

References

Abreu PM, Branco PS (2003) Natural product-like combinatorial libraries. J Braz Chem Soc 14:675–712

Ajay Walters WP, Murcko MA (1998) Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? J Med. Chem 41:3314–3324

Alberts P, Engblom L, Edling N, Forsgren M, Klingstrom G, Larsson C, Ronquist-Nii Y, Ohman B, Abrahmsen L (2002) Selective inhibition of 11beta-hydroxysteroid dehydrogenase type 1 decreases blood glucose concentrations in hyperglycaemic mice. Diabetologia 45:1528–1532

Arya P, Joseph R, Chou DTH (2002) Toward high-throughput synthesis of complex natural product-like compounds in the genomics and proteomics age. Chem Biol 9:145–156

Barun O, Sommer S, Waldmann H (2004) Asymmetric solid-phase synthesis of 6,6-spiroketals. Angew Chem Int Ed 43:3195–3199, and references therein

Bogan AA, Cohen FE, Scanlan TS (1998) Natural ligands of nuclear receptors have conserved volumes. Nat Struct Biol 5:679–681

Breinbauer R, Vetter IR, Waldmann H (2002) From protein domains to drug candidates – natural products as guiding principles in the design and synthesis of compound libraries. Angew Chem Int Ed 41:2879–2890

Brohm D, Metzger S, Bhargava A, Muller O, Lieb F, Waldmann H (2002a) Natural products are biologically validated starting points in structural space for compound library development: solid-phase synthesis of dysidiolide-derived phosphatase inhibitors. Angew Chem Int Ed 41:307–311

Brohm D, Philippe N, Metzger S, Bhargava A, Muller O, Lieb F, Waldmann H (2002b) Solid-phase synthesis of dysidiolide-derived protein phosphatase inhibitors. J Am Chem Soc 124:13171–13178

Chrousos GP (2004) Is 11beta-hydroxysteroid dehydrogenase type 1 a good therapeutic target for blockade of glucocorticoid actions? Proc Natl Acad Sci U S A 101:6329–6330

Claudel T, Sturm E, Kuipers F, Staels B (2004) The farnesoid X receptor: a novel drug target? Expert Opin Investig Drugs 13:1135–1148

Dang ZC, Audinot V, Papapoulos SE, Boutin JA, Lowik C (2003) Peroxisome proliferator-activated receptor gamma (PPAR gamma) as a molecular target for the soy phytoestrogen genistein. J Biol Chem 278:962–967

Evans BE, Rittle KE, Bock MG, Dipardo RM, Freidinger RM, Whitter WL, Lundell GF, Veber DF, Anderson PS, Chang RSL et al (1988) Methods for drug discovery – development of potent selective orally effective cholecystokinin antagonists. J Med Chem 31:2235–2246

Downes M, Verdecia MA, Roecker AJ, Hughes R, Hogenesch JB, Kast-Woelbern HR, Bowman ME, Ferrer JL, Anisfeld AM, Edwards PA et al (2003) A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. Mol. Cell 11:1079–1092

Fauman EB, Cogswell JP, Lovejoy B, Rocque WJ, Holmes W, Montana VG, Piwnica-Worms H, Rink MJ, Saper MA (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase Cdc25A. Cell 93:617–625

Frye SV (1999) Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. Chem. Biol 6:R3–R7

Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu Rev Biochem 70:209–246

Ghose AK, Viswanadhan VN, Wendoloski JJ (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem 1:55–68

Golebiowski A, Klopfenstein SR, Portlock DE (2001) Lead compounds discovered from libraries. Curr Opin Chem Biol 5:273–284

Grishin NV (2001) Fold change in evolution of protein structures. J Struct Biol 134:167–185

Gunasekera SP, McCarthy PJ, KellyBorges M, Lobkovsky E, Clardy J (1996) Dysidiolide: a novel protein phosphatase inhibitor from the Caribbean sponge Dysidea etheria de Laubenfels. J Am Chem Soc 118:8759–8760

Holm L, Sander C (1996) The FSSP database: fold classification based on structure-structure alignment of proteins. Nucleic Acids Res 24:206–209

Holm L, Sander C (1997) Dali/FSSP classification of three-dimensional protein folds. Nucleic Acids Res 25:231–234

Ibach B, Haen E (2004) Acetylcholinesterase inhibition in Alzheimer's disease. Curr Pharm Des 10:231–251

Jacoby E, Schuffenhauer A, Floersheim P (2003) Chemogenomics knowledge-based strategies in drug discovery. Drug News Perspect 16:93–102

Koch MA, Waldmann H (2004) Natural product-derived compound libraries and protein structure similarity as guiding principles for the discovery of drug candidates. In: Kubinyi H, Müller G (eds) Chemogenomics in drug discovery: a medicinal chemistry perspective. Wiley-VCH, Weinheim, pp 377–403

Koch MA, Waldmann H (2005) Protein structure similarity clustering and natural product structure as guiding principles in drug discovery. Drug Discov Today 10:471–483

Koch MA, Breinbauer R, Waldmann H (2003) Protein structure similarity as guiding principle for combinatorial library design. Biol Chem 384:1265–1272

Koch MA, Wittenberg L-O, Basu S, Jeyaraj DA, Gourzoulidou E, Reinecke K, Odermatt A, Waldmann H (2004) Compound library development guided by protein structure similarity clustering and natural product structure. Proc Natl Acad Sci U S A 101:16721–16726

Lee ML, Schneider G (2001) Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. J Comb Chem 3:284–289

Lyon MA, Ducruet AP, Wipf P, Lazo JS (2002) Dual-specificity phosphatases as targets for antineoplastic agents. Nat Rev Drug Discov 1:961–976

Mason JS, Hermsmeier NA (1999) Diversity assessment. Curr Opin Chem Biol 3:342–349

Masuzaki H, Flier JS (2003) Tissue-specific glucocorticoid reactivating enzyme, 11 beta-hydroxysteroid dehydrogenase type 1 (11beta-HSD1) – a promising drug target for the treatment of metabolic syndrome. Curr Drug Targets Immune Endocr Metabol Disord 3:255–262

Masuzaki H, Paterson J, Shinyama H, Morton NM, Mullins JJ, Seckl JR, Flier JS (2001) A transgenic model of visceral obesity and the metabolic syndrome. Science 294:2166–2170

New MI, Wilson RC (1999) Steroid disorders in children: congenital adrenal hyperplasia and apparent mineralocorticoid excess. Proc Natl Acad Sci U S A 96:12790–12797

Nicolaou KC, Pfefferkorn JA, Roecker AJ, Cao GQ, Barluenga S, Mitchell HJ (2000a) Natural product-like combinatorial libraries based on privileged structures 1. General principles and solid-phase synthesis of benzopyrans. J Am Chem. Soc 122:9939–9953

Nicolaou KC, Pfefferkorn JA, Mitchell HJ, Roecker AJ, Barluenga S, Cao GQ, Affleck RL, Lillig JE (2000b) Natural product-like combinatorial libraries based on privileged structures 2. Construction of a 10,000-membered benzopyran library by directed split-and-pool chemistry using NanoKans and optical encoding. J Am Chem Soc 122:9954–9967

Nicolaou KC, Pfefferkorn JA (2001) Solid phase synthesis of complex natural products and libraries thereof. Biopolymers 60:171–193

Nicolaou KC, Evans RM, Roecker AJ, Hughes R, Downes M, Pfefferkorn JA (2003) Discovery and optimization of non-steroidal FXR agonists from natural product-like libraries. Org Biomol Chem 1:908–920

Paterson JM, Morton NM, Fievet C, Kenyon CJ, Holmes MC, Staels B, Seckl JR, Mullins JJ (2004) Metabolic syndrome without obesity: hepatic overexpression of 11beta-hydroxysteroid dehydrogenase type 1 in transgenic mice. Proc Natl Acad Sci U S A 101:7088–7093

Pike ACW, Brzozowski AM, Hubbard RE, Bonn T, Thorsell AG, Engstrom O, Ljunggren J, Gustafsson JK, Carlquist M (1999) Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. EMBO J 18:4608–4618

Robinson-Rechavi M, Escriva Garcia H, Laudet V (2003) The nuclear receptor superfamily. J Cell Sci 116:585–586

Ross SA, Gulve EA, Wang M (2004) Chemistry and biochemistry of type 2 diabetes. Chem Rev 104:1255–1282

Russell RB, Sasieni PD, Sternberg MJE (1998) Supersites within superfolds. Binding site similarity in the absence of homology. J Mol Biol 282:903–918

Sadowski J, Kubinyi H (1998) A scoring scheme for discriminating between drugs and nondrugs. J Med Chem 41:3325–3329

Sandeep TC, Yau JL, MacLullich AM, Noble J, Deary IJ, Walker BR, Seckl JR (2004) 11Beta-hydroxysteroid dehydrogenase inhibition improves cognitive function in healthy elderly men and type 2 diabetics. Proc Natl Acad Sci U S A 101:6734–6739

Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I (1991) Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. Science 253:872–879

Schapira M (2002) Pharmacogenomics opportunities in nuclear receptor targeted cancer therapy. Curr Cancer Drug Targets 2:243–256

Schapira M, Raaka BM, Samuels HH, Abagyan R (2000) Rational discovery of novel nuclear hormone receptor antagonists. Proc Natl Acad Sci U S A 97:1008–1013

Schreiber SL (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. Science 287:1964–1969

Schweizer RA, Atanasov AG, Frey BM, Odermatt A (2003) A rapid screening assay for inhibitors of 11beta-hydroxysteroid dehydrogenases (11beta-HSD): flavanone selectively inhibits 11beta-HSD1 reductase activity. Mol Cell Endocrinol 212:41–49

Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 11:739–747

Spencer RW (1998) High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. Biotechnol Bioeng 61:61–67

Todd AE, Orengo CA, Thornton JM (1999) Evolution of protein function, from a structural perspective. Curr Opin Chem Biol 3:548–556

Urizar NL, Liverman AB, Dodds DT, Silva FV, Ordentlich P, Yan YZ, Gonzalez FJ, Heyman RA, Mangelsdorf DJ, Moore DD (2002) A natural product that lowers cholesterol as an antagonist ligand for FXR. Science 296:1703–1706

Van Gaal L, Scheen AJ (2002) Are all glitazones the same? Diabetes Metab Res Rev 18 [Suppl 2]:S1–S4

Walker BR, Seckl JR (2003) 11beta-hydroxysteroid dehydrogenase type 1 as a novel therapeutic target in metabolic and neurodegenerative disease. Expert Opin Ther Targets 7:771–783

Walters WP, Murcko A, Murcko MA (1999) Recognizing molecules with drug-like properties. Curr Opin Chem Biol 3:384–387

8 Tackling the Chemogenomic Space by Novel Screening Technologies

L.M. Mayr

Abstract. Drug discovery in the chemogenomic space has seen some tremendous changes over the last decade. Compared to previous times, not only the number of available chemical compounds for screening, but also the number of molecular targets used for screening has increased significantly. This has triggered the need for very fast, efficient, and effective novel readout technologies for compound testing. Novartis has developed two novel high-throughput screening (HTS) technologies for that purpose – NanoScreen and SpeedScreen. NanoScreen is a highly miniaturized and fully automated HTS/uHTS test system with confocal single-molecule as well as non-confocal detection capabilities and is used for functional screening in the range of 1–5 μl per sample. The integration of the single-molecule readout technologies into the system enables highly sophisticated biochemical test systems with multi-parameter readout for very high data quality. SpeedScreen is a highly miniaturized and automated screening system for high-throughput affinity-selection of compounds. In practice, pools

of compounds are incubated with the target protein and the unbound chemical compounds are removed from the target-compound complex via very fast, multiparallel size-exclusion-chromatography. The holoenzyme is disintegrated and analyzed via microbore reversed-phase high performance liquid chromatography (microbore RP-HPLC). Both systems have been developed and implemented with great success at the Novartis Lead Discovery Center (LDC) in Basel. These technologies have enabled us to access targets that would otherwise not have been possible, e.g., very expensive targets, "orphan" drug targets, or targets that are "non-tractable" by conventional screening technologies. Taken together, these novel screening technologies enable novel approaches for chemogenomic research that would have not been possible in the past.

8.1 Introduction

Current drug discovery relies on massive screening of chemical libraries against various extra- and intracellular molecular targets to find compounds with the desired mode of action. In recent years, various technologies used in combinatorial synthesis, automation technologies used in medicinal chemistry, and natural product isolation, but also the availability of large compound collections from commercial providers, have increased the compound collections of pharmaceutical and biotech companies up to several hundreds of thousands, in some cases even up more than one million, distinct chemical entities. At the same time, sequencing of the human genome as well as sequencing the genome of various pathogens, such as microbes, bacteria, and viruses, has delivered hundreds or thousands of novel molecular targets for pharmaceutical intervention.

The strong increase in both the number of available compounds as well as molecular targets has caused a fundamental change in the drug discovery process in pharmaceutical and biotech companies in recent years. Various technologies for miniaturization, lab automation, and robotics enable testing of chemical compounds in biological systems by means of high-throughput screening (HTS) and ultra-high-throughput screening (uHTS). Whereas HTS is defined by the number of compounds tested to be in the range of 10,000–100,000 per day, uHTS is defined by screening numbers that exceed 100,000 compounds tested per day. Taken together, the technologies of HTS and uHTS are seen as key

elements for filling the drug discovery pipeline in industry with new chemical compounds and new modes of action.

Filling the drug discovery pipeline with novel research projects (targets and/or compounds) is seen as one of the major challenges for the industry in the first decade of the twenty-first century. A detailed analysis of the productivity numbers for pharmaceutical R&D has revealed that the process is very difficult, time-consuming, and cost-intense (Fig. 1). Average timelines for pharmaceutical R&D are about 14.8 years from basic research to clinical application, while average costs per new molecular entity (NME) are about US $850 million to $1 billion. Despite the strong increase in R&D expenditures of the last few decades, the output of these investments (as measured by the number of NMEs per year) has remained constant or may even have declined in recent years (Fig. 2). This trend is commonly described as the "productivity gap" of pharmaceutical R&D over the last few years.

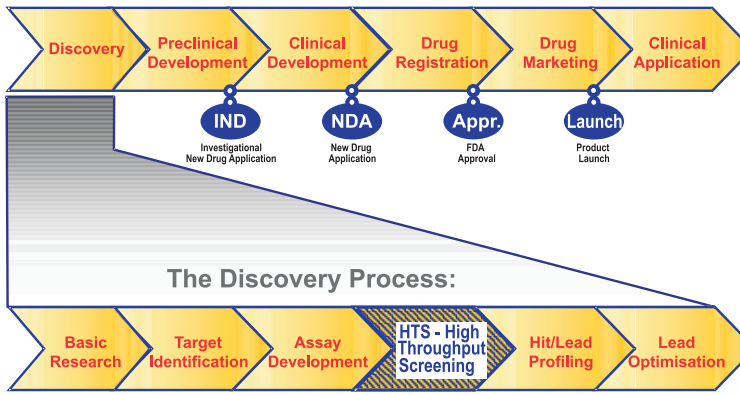


Fig. 1. Overview of the pharmaceutical research and development (R&D) process. The overall process of pharmaceutical R&D consists of various elements arranged in a sequential manner all the way from initial discovery research toward application of the drug on the market. The discovery process consists of various elements ranging from basic research toward target identification, target validation, assay development, high-throughput screening (HTS), and hit/lead profiling toward lead optimization

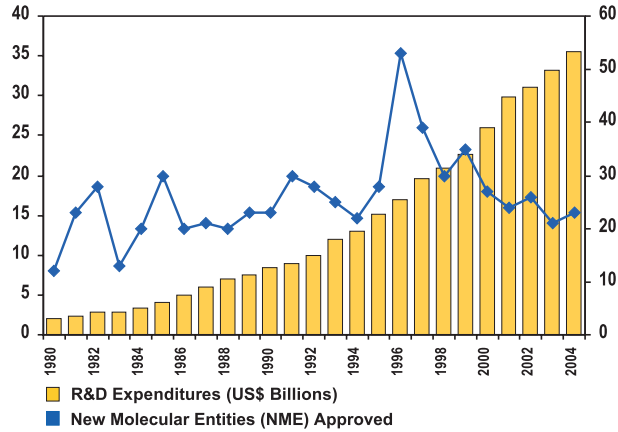


Fig. 2. Productivity analysis of pharmaceutical R&D process, showing the increase of R&D expenditures among pharmaceutical companies in the years 1980–2004, as taken from the PhRMA Annual Membership Survey 2004 (source: http://www.phrma.org/publications/), and the number of new molecular entities (NMEs) approved in the years 1980–2004 by the U.S. Food and Drug Administration (FDA) (source: http://www.fda.gov/cder/)

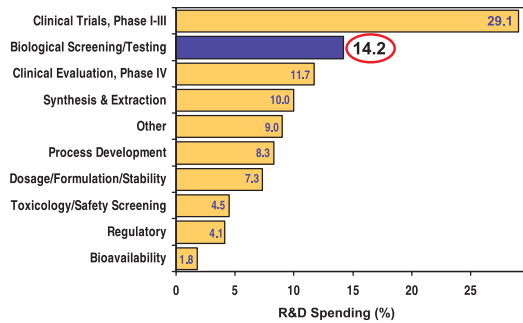


Fig. 3. Allocation of pharmaceutical R&D spending by function, showing the allocation of R&D costs toward various functions throughout the whole R&D process. Biological screening and testing is the second-largest cost element of the pharmaceutical R&D process, as taken from the PhRMA Annual Membership Survey 2001 (source: http://www.phrma.org/publications/)

It is clear that the main cost drivers in pharmaceutical R&D have been the ever-increasing costs for phase I–III clinical studies. This increase in study costs is partially related to more complex diseases under investigation for new drugs, but also due to higher safety demands from the various international regulatory authorities. Despite the high costs for clinical studies, the second-largest cost factor in pharmaceutical R&D is allocated to the fields of biological screening and compound testing (Fig. 3). These factors point directly toward the need for a careful analysis of the main cost factors in biological assays, both in the HTS phase as well as in compound profiling and compound testing during the hit-to-lead (H2L) and lead optimization (LO) phase.

8.2 The Novartis Lead Discovery Center

Novartis Pharma AG has started various measures to fight the trend of increasing costs and decreasing productivity throughout the pharmaceutical R&D process. Among many other factors, Novartis implemented a novel organizational structure called the Novartis Lead Discovery Center (LDC) in the years 2001 and beyond. This organizational unit within Novartis Pharma Research forms a global support function for all research activities with the Novartis Pharma Research organization (NIBR, Novartis Institutes of BioMedical Research) with its various disease areas. The keystone of the Novartis LDC is formed by the elements of centralized assay development in combination with industrialized, multiparallel HTS facilities (Fig. 4). This organizational structure is able to run the various formats of biochemical and cell-based assays commonly used in pharmaceutical research for finding biologically active chemical compounds on molecular targets. The chemical space is covered by compounds from various sources: medicinal chemistry (internal and external sources), combinatorial chemistry, and natural products. This compound archive consists of more than 1 million distinct chemical entities accessible for HTS. Storage and retrieval of these compounds is only possible by sophisticated systems from liquid handling, lab automation, and robotics. Information technology is an essential part of these systems in order to optimize both content as well as operation of the Novartis company-wide compound repository. The biological di-

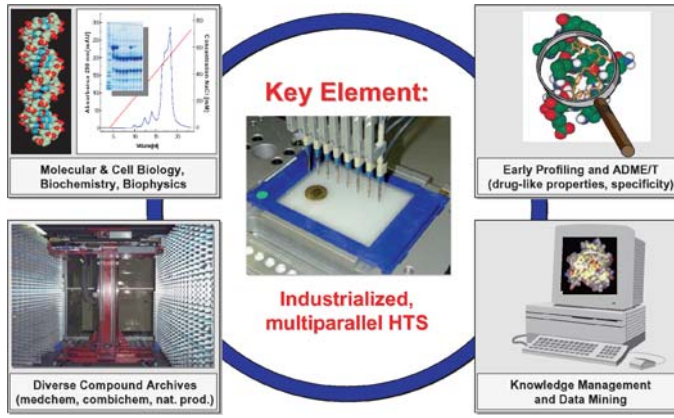


Fig. 4. The main elements of the Novartis Lead Discovery Center, a global organizational unit for industrialization of the hit/lead finding process. Various classes of highly diverse and drug-like chemical compounds are derived from the efforts of medicinal chemistry, combinatorial chemistry, and from natural products ("chemical space"). The biological assay systems are configured according to the needs of the particular target by various technologies from molecular and cell biology, biochemistry, biophysics, enzymology, and readout technology/spectroscopy. The keystone of the process is formed by industrial and multiparallel HTS via various biochemical and cell-based screening technologies at high density and fast speed. Fully automated screening stations are used to perform the process of compound testing/screening of compound collections with more than 1 million distinct entities. The results of these screening campaigns are analyzed and stored in highly specific databases for assay interpretation and assay-to-assay cross comparison. Finally, hitlists from HTS are further analyzed in various profiling and early ADME/T (administration, distribution, metabolism, excretion, and toxicology) assays in order to sort out compounds with bad compound profiles or characteristics as early as possible in the process

mension of the Novartis LDC is covered by a broad expertise ranging from molecular and cell biology to biophysics and chemistry.

The Novartis LDC has its own facilities for cloning and expression of the molecular targets which are used for compound testing. Many of these steps have been industrialized and automated whenever possible and feasible (Figs. 4–6). The process of HTS is done mostly by fully

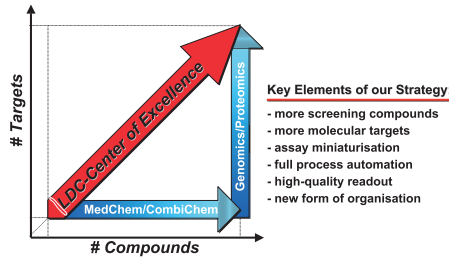


Fig. 5. The key strategies of the Novartis Lead Discovery Center (LDC). The figure shows the impact of increased compound collections ("chemical space") and increased numbers of molecular targets ("biological space") on the lead discovery process. LDC is a dedicated Center of Excellence (CoE) for industrialized lead finding in the post-genome era. Key elements of this organizational structure are miniaturization and automation of processes, high-quality readouts, and a new organizational structure dedicated toward the unique needs of lead discovery

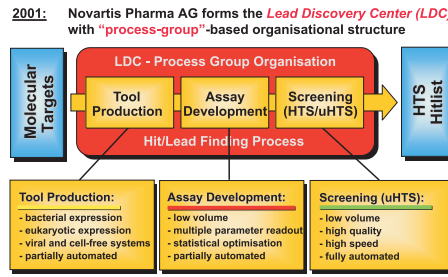


Fig. 6. Process group concept at the Novartis Lead Discovery Center (LDC). The figure shows the concept of a process group organizational structure for lead finding activities with biochemical assays ("Biochemical Assay Process Group"). The process covers all activities between a molecular target nominated for lead finding and the final outcome of the lead finding process, the HTS hitlist. The main elements are divided into tool production, assay development, and HTS. All key managerial elements in this process, e.g., goal setting, performance appraisal, steering meetings, priority settings, etc., are performed within a single group which is compiled with experts from various disciplines

automated screening stations that handle the processes of compound and reagent dispensing, plate handling, and readout technologies. Among the various readout technologies, various spectroscopic detection methods (fluorescence, luminescence, absorbance), but also radiometric, mass-spectrometric, and electrochemical methods are applied for testing the effect of chemical compounds in biological test systems. Key elements for operation of these screening systems are the various technologies for instrument control, process control, and data analysis. This is also used to compare lists of active compounds ("hitlists") from one HTS assay with the repository of hitlists from previous HTS campaigns. Upon doing these target-by-target comparisons of active compounds from HTS, one can easily sort out highly unspecific compounds or compounds which are active either on a broad range of targets within one target family or even on many targets across various target families ("frequent hitters"). Furthermore, these in silico tools are frequently used for identification of active pharmacophores for particular targets, and one can use these tools also for "hit explosion" on active scaffolds to obtain a first structure-activity relationship (SAR) for a particular target. Beyond data analysis and data interpretation, the tools of information technology are used for in-process quality control of the robotic screening systems. This enables the masking of plates where liquid handling, plate handling, or plate readout devices did not function properly. Subsequently, whole plates or single compounds from these plates can be measured again in the same screening process and thereby one can ensure that all compounds from a compound collection are screened with sufficient quality. One additional element in the process of high-throughput hit finding is the early characterization of compounds (hitlists) toward their properties regarding early ADME/T (administration, distribution, metabolism, excretion and toxicology). If compounds with poor parameters in ADME/T can be ruled out early in the drug discovery process, one can save a lot of precious resources and time needed for the overall process. Therefore, the Novartis LDC has implemented dedicated teams for compound profiling and early ADME/T of hitlists. Profiling also covers physicochemical characterization of compounds and biochemical profiling of these compounds against a broad panel of biochemical and cell-based assays. Of particular interest for these activities are panels of well-known cell surface receptors, ion channels, and various classes of enzymes.

Recent data also suggest the benefit of analyzing the effect of chemical compounds from these HTS hitlists on the hERG channel as early as possible.

Taken together, the Novartis Lead Discovery Center (LDC) is able to deal with the dramatic increase in the number of available chemical compounds for screening ("chemical space") as well as with the increase in the number of molecular targets for screening ("biological space"). This strong expansion of numbers in both dimensions has triggered the need for formation of a dedicated Center of Excellence (CoE) for lead discovery. Key characteristics of the LDC are assay miniaturization and full process automation, including assay development and reagent (tool) production whenever possible. Despite the quantitative expansion in lead discovery, the Novartis LDC has a clear mission toward constant quality increase in order to maximize the results of high-throughput compound testing. This implies that a number of novel high-quality readout technologies had to be implemented for the HTS process. The formation of the Novartis LDC also triggered the formation of a new organizational structure. Organizational units were formed around specific tasks and defined according to specific processes, e.g., the formation of the process group "biochemical assays" (Fig. 6). This process group structure at the Novartis Center of Excellence for lead discovery enables efficient management of the particular processes for lead discovery and differs from the organizational structure currently applied in isolated disciplines of pharmaceutical research and development. Similar concepts were implemented in the automotive industry about 10 years ago in order to foster collaboration among team members and to increase productivity in largely automated and industrialized processes. The same management principles can also be applied for industrialization of the lead discovery process in the pharmaceutical and biotech industry. The process group is largely defined by the common process for all team members, e.g., lead finding on biochemical assays.

All steps in the lead finding process between a molecular target nominated for lead discovery and, ultimately, delivery of a hitlist for a particular molecular target are covered by the same organizational unit. This process group covers all areas of tool production, assay development, and HTS. All activities within that group are closely linked and managed according to the same goals. This implies that team members with

very different fields of expertise are forming one single group focused on a single common task, i.e., lead finding with biochemical assays. This also implies that all managerial elements in the process group are closely interdigitated with common goal setting, project steering, and performance appraisal for all team members according to the same, high-level goals for all team members. This organizational structure minimizes the interface problem among various disciplines necessary for the same process, maximizes teamwork and information flow among team members, and ensures very high productivity and coworker motivation with regards to the common goal of a particular process group (Fig. 6).

For the process group "biochemical assays", the main element within each particular process step has always been the idea of massive automation and miniaturization of tasks, as maintaining or even increasing the quality of each particular process step has also been very important to us. This means that each working step is carefully analyzed towards its potential benefits, while the risks for doing more automation and miniaturization also need analysis, as sometimes the benefits of these technologies are more than offset by the disadvantages of setting up, running, and controlling automated systems.

It is well known in the field that HTS can easily be automated and miniaturized with regards to the various high-density screening plates and automated liquid and plate handling systems. Similar approaches can be applied at least in part to assay development, i.e., assays can be developed already within high-density, low-volume plates upon using automated liquid handling systems and the concepts of statistical optimization and design-of-experiment (DOE) software. Production of the necessary screening reagents ("tool production") is much harder to be automated. Nevertheless, the production of tools such as proteins, cell membranes, and cell lines can be automated in part with the use of modern liquid handling and robotics equipment. The process group "biochemical assays" has automated various steps for the tool-production part of the lead discovery process. Cloning and protein expression has been standardized and automated by the use of generic, but highly flexible, processes that enable a large and standardized variation of the expressed protein in order to find the best-suited construct for fast and efficient production of the desired assay reagents.

8.3 The NanoScreen HTS Platform

A strong increase in the number of chemical compounds for testing and the concomitant increase in the number of molecular targets for lead finding can be accommodated only via substantial miniaturization of HTS assays. In the past, microtiter plates with 96 wells per plate (96w-MTP) have been the main compound handling and screening format among most pharmaceutical and biotech companies (Fig. 7). This plate format and derivatives thereof are clearly defined by the MTP standards of the Society for Biomolecular Sciences, SBS (http://www.sbsonline.org/). For this as well as the plate formats with higher density, various types of plate materials exist, depending on the desired application, readout technology, and interference with the biological test system. The typical working volume for 96w-MTP is in the range of about 100–200 µl total volume with a standard volume of about 150 µl per well. This volume range can be handled routinely with manual and automated liquid handling systems by pipetting or dispensing of the appropriate volumes with sufficient precision and accuracy. Depending on the applied read-

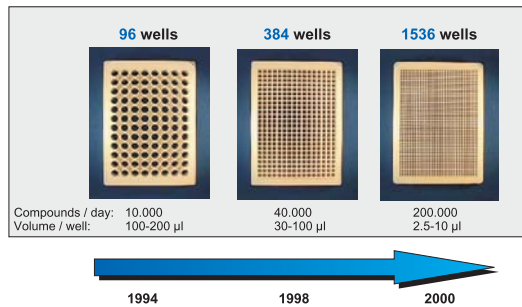


Fig. 7. The history of microtiter plates (MTP) used for HTS. The different types of microtiter plates used for compound storage and HTS are shown. In the early and mid 1990s, the main screening format was the microtiter plates with 96 wells per plate (96w-MTP). This plate type has been largely replaced in recent years with 384w- and in some cases even with 1,536w-plates (384w-MTP, 1,536w-MTP). Despite the fact that all these plate types have the same footprint, they can carry different numbers of distinct reaction wells, thereby enabling miniaturization and higher throughput in the screening process

out technology and the measurement time per well, this volume range equals about 10,000 compounds or compound concentrations tested per 8-h working day at a tick rate of about 4–5 min per plate (combined plate readout and plate handling time). This tick rate equals about 2–3 s average readout time (including plate loading and plate unloading from the plate reader) per well in the 96w-MTP.

The last decade has seen a strong trend toward plate types with higher densities at the same footprint. The first development was the 384w-MTP which accommodates four times more samples than a 96w-MTP. The typical working volume for 384w-MTP is in the range of about 30–100 μl total volume with a standard volume of about 50 μl per well. This volume range can still be handled by most of the routinely used liquid handling systems and can also be filled still by manual pipetting. Depending on the applied readout technology and the measurement time per well, this volume range equals about 40,000–50,000 compounds or compound concentrations tested per 8-h working day at a tick rate of about 4–5 min per plate (combined plate readout and plate handling time). This tick rate equals about 0.6–0.7 s average readout time (including plate loading and plate unloading from the plate reader) per well in the 384w-MTP via a sequential readout technology (well by well is read out by the MTP reader). Alternatively, various imaging technologies can read out the whole plate simultaneously via CCD (charged coupled device) technology and can achieve readout times independent of plate format of a few minutes. Despite the benefits of volume reduction on the amount of used assay reagents and chemical compounds, assay quality as defined by signal over background (S/B) or signal over noise (S/N) can be affected by reduction of volume and the statistical quality of the assay can be hampered. The majority of all assays, biochemical or cell-based, can be adapted toward 384w-MTP, and this plate format has been established as the format of choice for compound storage and screening assays among most pharmaceutical and biotech companies.

Several companies have managed to adapt their processes in part or in total toward 1,536w-MTP formats for compound handling and screening. The typical working volume for 1,536w-MTP is in the range of about 2.5–10 μl total volume with a standard volume of about 5 μl per well. Depending on the complexity of biological test systems (number of volume addition steps), this volume range needs special equipment to

handle addition of small volumes in the range of 0.1–1 μl with sufficient accuracy and precision. Depending on the applied readout technology and the measurement time per well, this volume range equals about 100,000–200,000 compounds or compound concentrations tested per 8-h working day at a tick rate of about 4–8 min per plate (combined plate readout and plate handling time). This tick rate equals about 0.1–0.3 s average readout time (including plate loading and plate unloading from the plate reader) per well in the 1,536w-MTP via a sequential readout technology (well by well is read out by the MTP reader). These fast reading times per well are only feasible with a few readout technologies and some distinct MTP readers. Therefore, in most cases 1,536w-MTPs are read out via CCD-based imaging of the whole plate and typical readout times are in the range of several minutes. Special care has to be taken that volume/surface ratios in 1,536w-MTP do not affect the biological test system. In many cases, lids are used on 1,536w-MTPs to minimize evaporation during extended incubation times or incubation times at higher temperatures, e.g., 37 °C. The main limitation of the 1,536w-MTP format, however, is the negative effect of volume reduction on the readout signal.

Further trends toward miniaturization are still ongoing. Several examples with biological assays in 3,456w-MTPs have been reported (total assay volume, 1–2 μl). An interesting alternative of ever-increasing importance is becoming the 384w low-volume MTP (384w lv-MTP). These plates use a 384w format of the plate, as conical shapes of the wells and can be used to run biological assays at fairly low volume (10–20 μl total volume). The benefit of the 384w lv-MTP is, however, the applicability of 96/384 liquid handling and dispensing technology. Current industry trends point toward the 384w standard, 384w low volume and the 1,536w MTP as the main plate formats for compound testing in the future.

Almost 10 years ago, Novartis (formerly Sandoz) entered into a technology development partnership with Evotec OAI toward novel HTS readout technologies and strong miniaturization of HTS systems. This concept implies the development of a readout technology almost independent of the assay volume as well as the development of liquid handling technologies in the nanoliter range. As a result of this development, the Novartis/Evotec NanoCarrier plate (NTP) was developed for HTS measurements via confocal fluorescence spectroscopy. Each

NTP holds 2,080 wells (2,080w-NTP) for compound testing and is surrounded by wells filled with water to minimize evaporation of liquid from the array of screened wells (Fig. 8). In addition to the compounds for testing, various wells of the 2,080w-NTP are filled with controls (high/low control of the assay, dye control, etc.) necessary for productive screening and in-process control. The typical working volume for 2,080w-NTP is in the range of about 0.7–1.5 μl total volume with a standard volume of about 1.2 μl per well. Depending on the complexity of biological test systems (number of volume addition steps), this volume range implies that all volume additions are done in the range of only tens to hundreds of nanoliters per single step of reagent addition. These small volumes require unique liquid handling technologies not necessarily available for routine compound testing in 96w- or 384w-MTP formats. These small volumes require also unique detection methods with sufficient sensitivity for assay readouts at ultra-low volumes, like confocal fluorescence spectroscopy with single molecule spectroscopy capabilities and detection volumes below 1 fl.

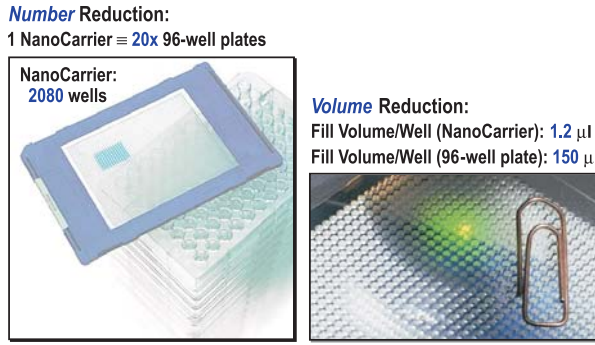


Fig. 8. The Novartis/Evotec NanoCarrier (NTP) concept. The *left panel* shows a standard NanoCarrier plate with 2,080 wells (2,080w-NTP) surrounded by wells loaded with water to minimize evaporation of liquid. For demonstration purposes, 96 wells of the NTP are filled with blue dye solution and the plate was placed on top of a pile of 20 96w-MTPs to demonstrate the high density format of the 2,080w-NTP. The *right panel* shows for demonstration purposes a paper clip placed on top of the 2,080w-NTP. Each well of the 2,080w-NTP is filled with a total volume of 1.2 μl per well

Most pharmaceutical and biotech companies run approximately 50–200 HTS campaigns per year with compound collections in excess of 0.5–1.0 million compounds. The importance of volume reduction via various microtiter plate formats is exemplified in Fig. 9. This table compares assays at the various plate formats from 96w-MTP standard plates as a reference, all the way to miniaturized 2,080w NanoCarrier plates. The example is based on a protease assay with a commercial fluorescent substrate and 1.0 million compounds for screening. With regards to the needs of automation and robotics, a full screening campaign equals either 11,364 plates in a 96w-MTP format or just 569 plates in a 2,080w-NTP format (the plate format, including control wells per plate, is either a 88/96w-, 352/384w-, 1,408/1,536w-, or 1,760/2,080w-plate format). The difference in assay volume equals either 150 l of assay reagents per

	96w Plate (120-250 µl)	384w Plate (30-100 µl)	1536w Plate (2.5-10 µl)	2080w Plate (0.5-1.4 µl)
Volume/Well	150 µl	50 µl	8.5 µl	1.2 µl
Number of Plates	11'364	2'841	711	569
Reaction Mix	150.1 Liters	50.1 Liters	8.5 Liters	1.2 Liters
Compound/Well (MW 500 Da; 10µM)	750 ng	250 ng	42.5 ng	6 ng
Enzyme (MW 80 kDa; 20nM)	240.2 mg	80.2 mg	13.6 mg	1.94 mg
Substrate Amount (MW 1000 Da; 100 µM)	15.01 g	5.01 g	851 mg	121 mg
Substrate Costs (1 mg = 100 $)	1'501'000 $	501'000 $	85'100 $	12'100 $

Example: protease HTS assay with fluorescence read-out and 1.0 Mio compounds tested:
Library size 1'000'000 compounds, control wells on each plate (88/96w-, 352/384w-, 1408/1536w- and 1760/2080w-plate format); In addition, excellent sensitivity for the confocal detection setup with single-molecule detection in NanoScreening will allow even lower substrate concentrations.

Fig. 9. Comparison of various microtiter plate formats for screening, showing the effect of volume reduction for various types of plates used in routine high-throughput screening (96w-, 384w-, 1,536–2,080w plates). Volume reduction has a pronounced effect on the number of plates and the amount of liquid to be handled per full screening campaign, the amount of chemical compound necessary for testing, the amount of biological assay reagent and, ultimately, the reagent costs associated with screening of large compound collections. The numbers are based on a commercially available protease assay with a fluorogenic peptide as a substrate and 1.0 million compounds on the screening deck. The exact plate formats are (including the controls): 88/96w-, 252/384w-, 1,408/1,536w- and 1,760/2,080w plate format

full screening campaign or just 1.2 l, which has a strong impact on lead discovery campaigns where waste proposal can become a major cost issue. With regards to precious chemical compounds or isolates from rare natural products, the difference per well is either 750 ng per data point (MW = 500; screening concentration = 10 μM) in a 96w-MTP or just 6 ng per data point in a 2,080w-NTP. One element of utmost importance is the difference in amount of biological assay components, like recombinant proteins and cell lines. For biochemical targets in HTS, the majority of the proteins have to be produced in eukaryotic expression systems like insect cells or mammalian cells. Under the same enzymological conditions, a conventional 96w-MTP based assay would need more than 200 mg of recombinant protein for screening the full compound deck. The highly miniaturized 2,080w-NTP would require less than 2 mg of recombinant protein (higher sensitivity due to confocal single molecule detection methods not even included). For some target proteins, the request for very large amounts of recombinant protein can even become prohibitive for a lead finding campaign due to the high costs of eukaryotic protein expression, with mammalian cell lines in particular. The next factor to consider is the costs for the detection reaction in biological assay systems, for example fluorescently labeled substrates for proteases and kinases, antibodies for the detection of phosphorylation/dephosphorylation reactions, etc. The current example shows the difference in costs for the substrate based on a commercially available peptide substrate for a protease. Just the costs for the substrate would be either more than US $1,500,000 in a standard 96w-MTP format or just slightly more than US $12,000 in a 2,080w-NanoCarrier format. Taken together, despite higher initial investment costs for setting up miniaturized HTS systems, a detailed fully loaded cost analysis among the various assay technologies at Novartis Pharma AG has shown that the overall costs for screening clearly has been diminished mostly driven by assay miniaturization (data not shown).

Detection of biochemical reactions almost independent of the detection volume can be achieved by confocal fluorescence spectroscopy (Fig. 10). This technology uses high-energy laser beams focused with confocal optics toward illumination volumes of about 0.24 fl (10–15 l). The illumination volume is shaped like a cylindrical element of about 1.9 μM in height and about 0.4 μM in diameter. For chemical solutions

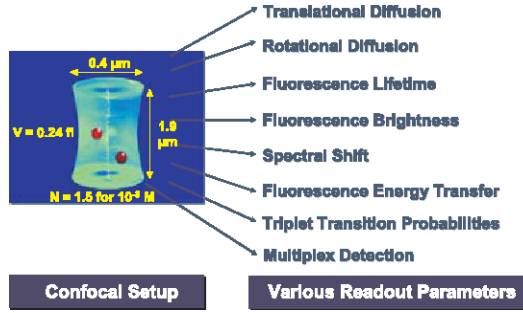


Fig. 10. The principles of confocal fluorescence spectroscopy. The figure shows the confocal illumination field of about 1.9 µm in height and about 0.4 µm in diameter. This equals a detection volume of about 0.24 fl (10–15 l). Only fluorophores in this illumination field are excited by the confocal setup in fluorescence spectroscopy via excitation by high-energy laser light of various sources. For 10-nM solutions of fluorescent detection molecules, on average only 1.5 fluorescent molecules are present over time in the confocal detection field. This enables multiparametric readout technologies at very high quality

containing a fluorophore at a concentration of 10 nM, this implies that on average only 1.5 molecules are populated at any time in the confocal illumination volume and one can apply the measurement principles of confocal single molecule fluorescence spectroscopy. One can measure translational diffusion, rotational diffusion, and molecular brightness, among many other factors. This detection setup has two main benefits: First, one can measure multiple parameters of chemical and biochemical reactions with very high precision. This enables multidimensional readout technologies in contrast to conventional readout technologies that generally measure only one readout signal at a time and where this readout signal is just the average of the whole ensemble of molecules. Applying these technologies one can characterize chemical molecules and biochemical reactions at much higher detail compared to any of the other spectroscopic readout technologies in the detection volume. Second, this readout technology is largely insensitive to the volume used for the biochemical reaction in a microtiter plate. Even for the high-density, low-volume NanoCarrier plate with a typical well volume

of 0.8–1.2 μl total, the detection volume of 0.24 fl represents just one billionth of the total well volume. In essence, the biochemical reaction in the microscopic detection volume is largely insensitive to the macroscopic volume of the microtiter (MTP) or NanoCarrier (NTP) plate (see Fig. 11, right panel). This implies that the same readout technology could also be applied to other plate formats beyond the NanoCarrier plate, e.g., microfluidic devices with volumes for reaction chambers in the nanoliter range and below. In summary, the fluorescence spectroscopy with confocal detection technology is the only readout technology suited for current and future small-volume, high-density plate-based and non plate-based screening technologies.

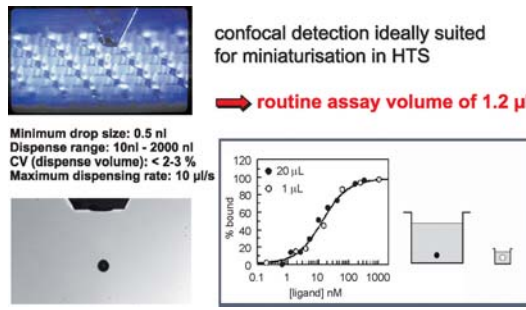


Fig. 11. The principles of NanoDispensing technology. The *left panel* shows tips of piezoelectric dispensers (NanoPen). The tip of the NanoPen is placed above the 1.2-μl wells of the NanoCarrier plate and fills the wells with samples of biochemical stock solutions in the range of 10–1,000 nl (*upper left panel*). Sample reservoirs hold the stock solutions of the assay components in the range of 10–50 ml. Electric stimulation of the piezoelectric quartz within the NanoPen causes reagent dispensing at high speed with the appropriate frequency for droplet deposition. Each droplet has a volume in the range of nanoliters. A picture from a high-speed camera shows a nanoliter droplet dispensed from the tip of a piezoelectric dispenser (*lower left panel*). The figure on the *right panel* shows that confocal detection of biochemical reactions (detection volume: 0.2 fl) is generally insensitive to the actual volume (1–100 μl total volume) used for biochemical reactions. This implies that confocal detection methods can be applied to volumes of 1 μl and below

A total fill volume of 1.2 µl per well also implies that individual pipetting steps in biochemical reactions are carried out with volumes in the volume range of 10–1,000 nl. In general, the plates are loaded first with chemical compounds from stock solutions in buffer/dimethylsulfoxide (DMSO); afterwards, protein in the appropriate assay buffer is added and the compound-target complex is incubated for the requested period of time to reach equilibrium. Subsequently, substrate for the enzyme is added and the biochemical reaction is incubated under the appropriate conditions for some time, typically more than 30 min and less than 6 h. Similar approaches are applied for binding reactions in equilibrium such as protein-ligand interaction assays. Depending on the target, shorter as well as longer incubation times are feasible on most robotic screening systems. Finally, in the case of enzymatic reactions, the biochemical assay is stopped by the addition of a stop solution and the amount of product is detected upon addition of the appropriate detection reagents. According to this regime, most biochemical assays (enzymatic reactions, binding reactions) comprise at least 3–5 pipetting steps. Conventional liquid handling devices with disposable tips or needles are not suited for these tasks. Piezoelectric dispensing technology, however, fits well to the specific requirements of liquid handling in the nanoliter range. Dispensing of reagents with piezoelectric technology is modulated by the electric field applied to the quartz in these tips (Fig. 11, upper left panel). This causes the generation of small droplets in the nanoliter range, and – depending on the frequency applied to the piezoelectric chip – relatively fast dispensing rates (Fig. 11, lower left panel) can be achieved for volumes in the nanoliter volume range.

Compound testing with biological assays in microtiter plates requires addition of chemical compounds and biological reagents from the appropriate stock solutions. Chemical compounds are routinely stored in large compound repositories such as dry powders and/or stock solutions at either 100% DMSO (v/v) or 90%/10% DMSO/H_2O (v/v) under controlled conditions of temperature and humidity. Depending on the setup of a particular compound storage facility and the concentration used for the stock solutions, source plates for screening are prepared at compound concentrations of 0.5–2.0 mM per compound per well. These plates are then fed into automated compound reformatting stations for a particular screening system or technology. For the Nano-

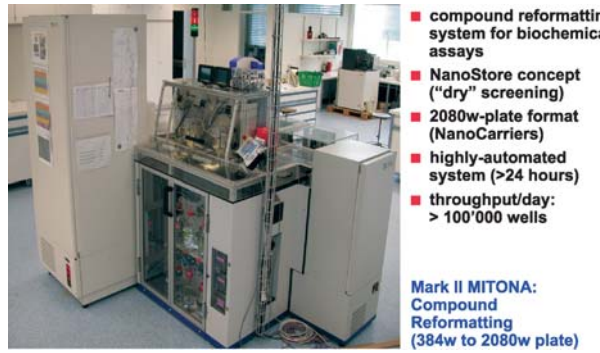


Fig. 12. The Evotec Mark II MITONA (micro-to-nano) compound reformatting system for transfer of compound stock solutions (10–100 nl). Compound stock solutions in 384w source plates are loaded from the Novartis compound archive into the compound stores of the MITONA and reformatted on the automated system into 2,080w- or 1,536w-NanoCarrier plates

Screen platform, source plates from the Novartis compound repository are loaded into the Mark II MITONA (micro-to-nano) compound reformatting station (Fig. 12). This system is used for transfer of 50–100 nl of chemical compound into 2,080w-NanoCarrier plates that are then used routinely for confocal NanoScreening. This system is highly automated and can reformat more than 100,000 distinct chemical compounds per day from source plates into appropriate screening plates. After transfer of small samples of chemical compounds into screening plates coated with hydroxypropyl-beta-cyclodextrin (HBC), plates are dried down and can be stored for extended periods of time. During usage of the plates in screening, each well is then filled with buffer solution to ensure proper re-dissolution of chemical compounds into the assay reaction volume.

Addition of biological assay reagents to compound plates is done for the HTS NanoScreen platform either on the Mark II SCARINA (screening and readout in NanoCarriers) or the Mark III SCARINA systems (Figs. 13, 14). The Mark II SCARINA system contains all components necessary for reagent dispensing, assay incubation, and plate readout for biochemical assays with confocal readout technology. Screening plates (2,080w- and 1,536w-NanoCarrier plates) with chemical compounds are

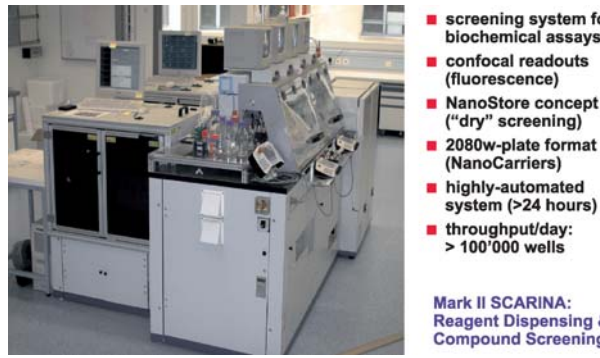


Fig. 13. The Evotec Mark II SCARINA (screening and readout in NanoCarriers) compound testing and plate handling NanoScreen system. The system can handle both 1,536w- and 2,080w-NanoCarrier plates and is equipped with four reagent dispensing tables and with two fully automated confocal fluorescence microscopes for detection. Compound storage devices, plate handling devices, and system control are integrated into one screening station

loaded into the plate storage devices of the Mark II SCARINA. Reagent dispensing is done with four tables of piezoelectric NanoDispensers with temperature-controlled devices. The plate-handling part of the system ensures appropriate transport of the plates from the various plate storage devices to the dispensing and readout devices of the system. After incubation of the plates loaded with compounds and biological assay systems, the screening plates are transferred to one of the two confocal microscopes attached to the system (Fig. 13, left side). For the Mark II SCARINA, only biochemical assays and only confocal fluorescence readout technologies are applied on that screening system. The daily throughput of the screening system very much depends not only on the applied spectroscopic readout technology and measurement time per well, but also on the number of pipetting steps, pipetting volume, and the incubation time for the enzymatic or binding assay. On average, the highly automated Mark II SCARINA system achieves daily throughputs in excess of 100,000 wells per day.

The latest addition to the NanoScreen HTS platform has been the Mark III SCARINA (Fig. 14). This system has a highly modular setup

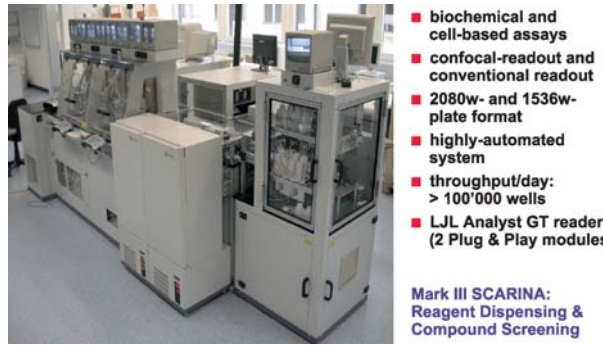


Fig. 14. The Evotec Mark III SCARINA (screening and readout in NanoCarriers) compound testing and plate handling system. The system can handle both 1,536w- and 2,080w-NanoCarrier plates as well as conventional 1,536w-MTP plates. The system is equipped with eight tables equipped with NanoDispenser technology as well as low-volume and high-speed liquid dispensing technology based on solenoid valves (SynQuad technology). Both confocal fluorescence spectroscopy and non-confocal multi-mode readout technology are implemented into one modular system. Compound storage devices, plate handling devices, and system control are integrated into one screening station

and can be used for HTS of both biochemical as well as cell-based assays (with suspension cells). The system has a specific cell-dispensing module with temperature- and CO_2-controlled incubation of cell lines in suspension. Dispensing of cells as well as dispensing of 'bulk' amounts (several hundred nanoliters to several microliters) of biochemical reagents is done via solenoid valves on Cartesian SynQuad liquid handling systems. Reagent dispensing in the range of tens to several hundreds of nanoliters is done via the piezoelectric NanoDispensers. The system can handle both the confocal readout technologies as well as almost all of the commercially available conventional readout technologies, e.g., fluorescence intensity, fluorescence polarization, TR-FRET (time-resolved fluorescence resonance energy transfer), luminescence, absorbance, etc. The system is capable of handling both the Evotec 2,080w- and 1,536w-NanoCarriers, as well as standard SBS-type 1,536w-MTPs from various vendors. Daily throughput of the screening system depends on the applied spectroscopic readout technology and

the specific reagent and incubation protocol of that assay. On average, the highly automated Mark III SCARINA system achieves daily throughputs between 100,000 and 250,000 wells per day. Due to the modular nature of the system with two dedicated docking positions on the system, almost all commercially available HTS readers can be attached sequentially or in parallel to the Mark III SCARINA HTS system.

Most of the assays currently run on the Mark II and Mark III SCARINA HTS systems at Novartis are biochemical assays. Among biochemical assays, one can distinguish mainly between the enzymatic and the binding assays. Whereas enzymatic assays include a kinetic component in the assay (i.e., signal increase by prolonged incubation time), the binding assays are routinely performed under non-kinetic conditions with a biochemical binding reaction at equilibrium. Whereas the incubation time in binding assays can be done under controlled conditions in the plate storage devices and does not directly influence scheduling time for the readout, the incubation time in enzymatic assays has a pronounced effect on the scheduling time for the readout of the plates. Ideally, all plates in enzymatic biochemical assays have obtained the same incubation time to achieve the same amount of substrate conversion in all plates and even so in all individual wells per plate.

Biochemical binding assays measure the binding of a fluorescently labeled tracer to the target molecule (Fig. 15). Typically, the target molecule is a protein of interest either as a full-length entity or defined by a particular domain of that target protein. The labeled tracer can be either a peptide or a low-molecular-weight compound with an attached fluorophore. The nature of the used fluorophore very much depends on the characteristics of the available light sources, i.e., wavelengths of the laser beam, filters/gates for conventional light sources, as well as on the compatibility with a biochemical assay setup, availability of appropriate substrates, and also the assay developer's preferences for the various fluorescence dyes. Characteristic examples for the biochemical binding reaction are protein-peptide interactions, protein-nucleic acid interactions, or well-known binding reactions for low-molecular-weight chemical compounds onto molecular targets (i.e., binding of cyclosporin to cyclophilin A). HTS with binding assays intends to find novel chemical scaffolds that interfere (inhibit or stimulate) the bind-

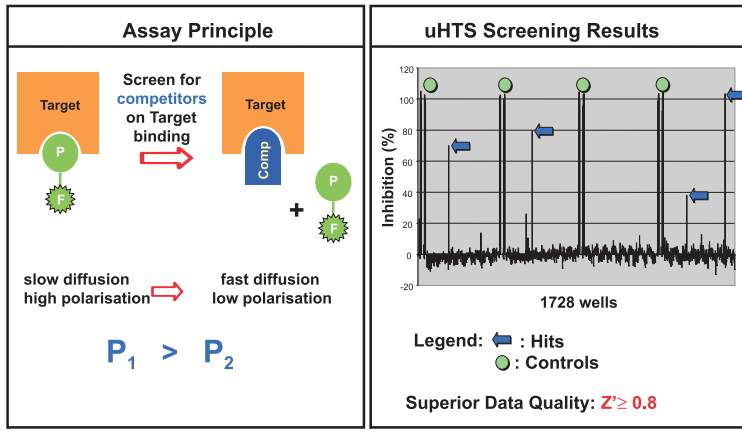


Fig. 15. Binding assay on the NanoScreen HTS platform. Binding of a fluorescently labeled tracer (peptide, oligonucleotide or chemical compound labeled with fluorophore) to the biological target of interest (protein) is shown. Upon displacement of the tracer from the target via chemical compounds in screening, the polarization of the fluorescent tracer changes and can be used to follow the binding/displacement reaction (*left panel*). This readout technology can be used to screen chemical compound collections with very high statistical quality ($Z' > 0.8$) (*right panel*)

ing of a fluorescently labeled tracer molecule to the target protein. This approach was used at Novartis to find chemical compounds to displace a peptidic ligand for a particular target protein (Fig. 15, left panel). The most convenient readout technology for this assay setup is fluorescence polarization that uses the difference in anisotropy of the fluorophore in the bound and in the free state of peptide. Whereas most compounds screened in that assay are fully inactive, just a few wells (= compounds) per plate might contain active compounds that are able to displace the fluorescently labeled compound off the target (Fig. 15, right panel). For reasons of online in-process quality control, each screening plate also contains the appropriate number and type of control wells. With the currently used setup at the Novartis NanoScreen HTS platform, only 1,728 out of a total of 2,080 wells per plate actually contain chemical compounds for testing, the rest are used for the various types of quality

control. The statistical quality of an HTS assay is expressed by the Z- and Z'-value (Zhang et al. 1999) defined by:

$$Z = 1 - \left[\frac{(3 \times \text{SD of sample} + 3 \times \text{SD of control})}{(\text{mean of sample} - \text{mean of control})} \right] \tag{8.1}$$

Whereas the Z-value describes the statistical quality of the assay with regards to values from screening samples vs control, the Z'-value describes the statistical quality of the assay for the values from the high vs the low controls. The Z'-value of an HTS assay can approximate an upper theoretical limit of 1.0 and it is generally accepted among experts in the field that HTS assays with Z'-values exceeding 0.5 per plate are regarded as being of sufficient statistical quality for lead discovery projects. Excellent HTS assays with regards to statistical quality of the readout signals and standard deviations of the controls (which does not include biological and chemical quality of the assay) are typically described with Z'-values in the range of 0.80–0.95. Visual inspection of the controls and the noise for inactive compounds in the biochemical assay of Fig. 15 as well as scientific analysis by Z'-value shows that this assay was running at very high statistical quality ($Z' > 0.80$).

Biochemical enzymatic assays measure the effect of chemical compounds on the enzymatic turnover of substrate toward product. In general, inhibition of the enzymatic activity is the prime focus of biochemical enzymatic assays. Typical examples for this assay setup are from the various target classes such as proteases, kinases/phosphatases, hydrolases, transferases, isomerases, metabolic enzymes, etc. For the Novartis NanoScreen HTS platform, either confocal fluorescence spectroscopy or any of the conventional readout technologies (e.g., fluorescence intensity, fluorescence polarization, TR-FRET, luminescence, absorbance, etc.) can be applied for these target classes. This decision toward the applied assay readout technology very much depends on the particular requirements of each assay toward data quality, assay sensitivity, and assay costs. The confocal fluorescence spectroscopy needs a fluorescent label attached to the substrate. A representative example for that assay setup is given in Fig. 16 with a protease assay measured by the change in fluorescence polarization (2D-FIDA, two-dimensional fluorescence intensity distribution analysis) of the cleaved vs the uncleaved peptide.

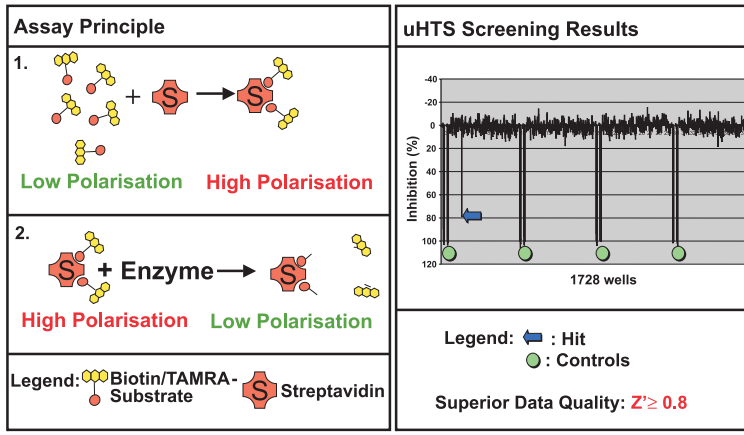


Fig. 16. An enzymatic assay on the NanoScreen HTS platform followed by the change in fluorescence polarization due to the change in the rotational correlation coefficient for the free vs the bound peptide. The peptide is labeled with both fluorophore (for detection) and biotin (for increase in mass upon binding to streptavidin). Upon proteolytic cleavage of the peptide, the polarization of the fluorophore gets changed and can be used to follow the enzymatic reaction (*left panel*). The good statistical quality of the assay ($Z' > 0.8$) is reflected by the clear separation of inactive compounds from active hits and positive controls, respectively (*right panel*)

The peptide is labeled with a fluorophore for detection on one end and with biotin on the other end to enlarge the molecular weight of the peptide via formation of a biotin-streptavidin complex (Fig. 16, left panel). A representative example for one of the screening plates in this HTS campaign shows the good statistical quality of the assay ($Z' > 0.8$; Fig. 16, right panel). One can see that the hit rate for this assay is fairly low, just about 0.058% (1 hit out of 1,728 compounds tested). Based on this hit rate, one would find only 580 hits out of a screening collection of 1 million compounds. This number is realistic for assays with very high quality (low number of false positives, low number of false negatives) and targets with medium to low drugability, i.e., ease of finding low-molecular-weight compounds interfering with biological effects.

On average, hit rates for biochemical assays range vary around 0.05%–0.1%. This strongly depends on the used assay technology, the particular target and its drugability, the statistical quality of the assay (Z'-value), the biological quality (stability) of the assay reagents, and the operational quality of the HTS screening system. This implies that greater than 98%–99% of all of the tested compounds either have no activity at all in the assay or at least no significant activity in that particular assay (e.g., less than 20% inhibitory effect). In general, only compounds with inhibitory effects greater than three times the standard deviation of the control can be regarded as potential hits. These potential hits could still be screening artifacts due to unspecific compound interference (e.g., protein precipitation) with the biological system or compound interference with the readout technology (e.g., fluorescence quenching, auto-fluorescence, inner filter effect, etc.). Only compounds identified as hits in primary screening (e.g., 0.3% in Fig. 17) are then taken for subsequent hit confirmation and validation by dose-response-curve (DRC) measurements. Various attempts are ongoing to minimize compound interference and to increase statistical quality of HTS primary hitlists. The confocal multi-dimensional readout technology of

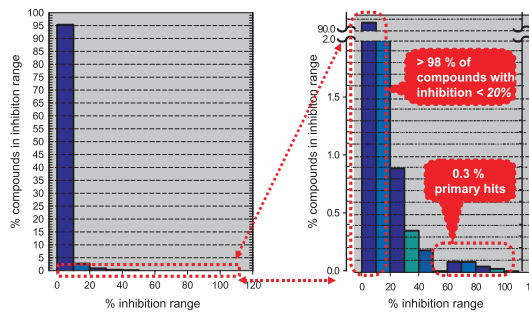


Fig. 17. Statistical analysis of NanoScreen HTS assays, showing the high percentage of inactive compounds in standard HTS campaigns (>95%–98% with no or almost no activity). Based on the standard deviation of the inactive compounds, typical activity thresholds are set at three times the standard deviation of the controls. In the current example, about 0.3% of all compounds in primary screening can be classified as primary hits. On average, the hit rate in biochemical screening assays is in the range of 0.05%–0.1%

the confocal NanoScreen HTS platform at Novartis is certainly among the most advanced screening setups of biochemical screening stations. Due to the power of high-quality readouts based on single molecule detection (SMD) spectroscopy applied to HTS, one can largely minimize false positives generated by fluorescent properties of the chemical compounds already in primary screening and can run assays at higher quality with regards to false-negatives (see above).

A recent addition to the various readout technologies already in place on the Mark II and Mark III SCARINA systems at Novartis has been the implementation of confocal SMD fluorescence lifetime spectroscopy assays (cFLA). This novel readout technology measures the decay of a particular fluorescent intensity signal (fluorescence lifetime) after excitation of the fluorophore with a pulsed energy source. The fluorescent lifetime (τ) of an excited fluorophore depends on the electromagnetic environment of a fluorophore and, in general, does not depend on the concentration of the fluorophore used in the assay system (Fig. 18). This means that the assay system is largely insensitive to inner filter effects and/or concentration variations in the assay system. Contrary to

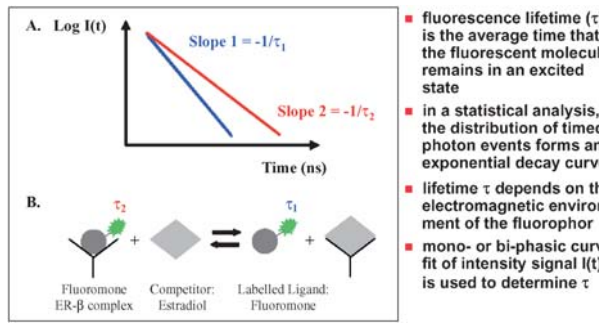


Fig. 18. A,B The principles of fluorescence lifetime. Fluorescence lifetime is defined as the average time that a fluorescent molecule remains in the excited state after excitation with a pulsed source of light. The lifetime is defined as the inverse of the rate constant for the fluorescence decay on a single exponential curve (**A**). The lifetime of a fluorophore is, in general, independent of concentration and molecular weight of the fluorophore and depends only on the electromagnetic environment of the fluorophore (**B**)

fluorescence polarization, fluorescence lifetime does not depend on the mass (rotational correlation coefficient) of the fluorophore. This enables experimental readouts for fluorescence lifetime with assay setups that would not have been feasible by using just fluorescence polarization. The main advantage, however, of fluorescence lifetime is the fact that the measured lifetime of the fluorophores typically used for detection of biological effects is generally in a different time range compared to the lifetime caused by those chemical compounds that have fluorescent properties. In other words, after excitation of a sample containing both a fluorescent chemical compound and a fluorescent probe of a biochemical reaction, the fluorescence decay occurs via a biphasic curve. Intelligent curve fitting according to a biphasic exponential decay curve enables a clear distinction between fluorescence properties of the readout signal and an interfering chemical compound with fluorescence properties. In summary, fluorescence lifetime measurements have not only all of the benefits of a conventional fluorescence measurement (such as high assay sensitivity, feasibility for assay miniaturization, low costs of goods), but also the highly advantageous effect of being almost insensitive to compound interference.

Confocal fluorescent lifetime is implemented at the Novartis NanoScreen HTS platform with pulsed lasers both in several of the off-line assay development readers, as well as on the screening readers integrated into the robotic screening systems (Mark II, Mark III). A typical experimental setup consists of readout times of 250 ms per well, excitation with laser pulses on the picosecond time scale and readout times of about 20 ns (cycle time). Extensive mathematical analysis of the fluorescence decay via automated software tools enables precise determination of the dye's fluorescence lifetime (Fig. 19). This readout technology has been applied with great success in various biochemical assays setups such as screening campaigns on proteases, kinases, and protein-peptide interaction assays.

In order to proof the effect of fluorescent lifetime on the quality of HTS assays, we compared the same assay plate with different readout technologies (Fig. 20). We compared an assay plate measured at the same time with both fluorescence polarization and fluorescence lifetime readout technology in a confocal setup. Whereas the polarization readout was giving an assay with average quality ($Z' = 0.45$), the same plate

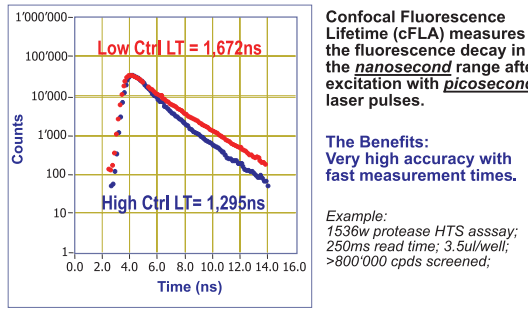


Fig. 19. Protease assay with confocal fluorescence lifetime readout. The figure shows a single cycle (16 ns) out of a total of 250 ms readout time per well for a protease assay with a fluorescence lifetime readout. The *curve* shows the burst in fluorescence upon excitation with a pulsed laser in the picosecond time range and the subsequent increase and decrease in fluorescence. Subtle differences in fluorescence lifetime can be used to measure biochemical reactions (low control in *red*, high control in *blue*)

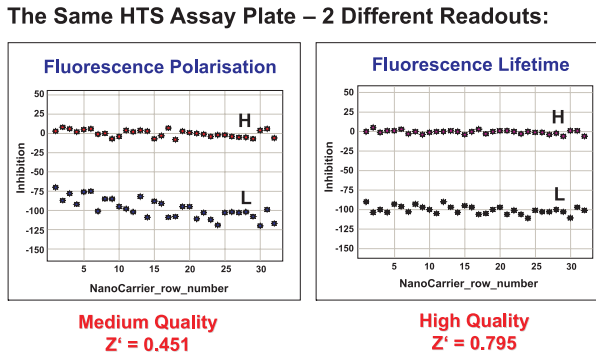


Fig. 20. Superior data quality by confocal fluorescence lifetime readout. Results of two different readout technologies for the same plate are shown. For clarification, only experimental data of controls but no screening compounds are displayed in the figures. One screening plate was measured simultaneously by confocal fluorescence polarization (*left panel*) and confocal fluorescence lifetime (*right panel*). One can clearly see the effect of fluorescence lifetime on the statistical quality of the HTS assay plate

analyzed via fluorescence lifetime gave us an assay with excellent quality for the controls ($Z' = 0.79$). This is confirmed by the lower number of hits in primary screening due to fewer artifacts by compound interference in primary screening (data not shown).

In summary, the Novartis NanoScreen HTS platform is a state-of-the-art technology for miniaturized and high-quality HTS. This technology required substantial investments in development of appropriate readout and fluidics technologies. These past investments are today more than compensated by the high quality of the screening data and the strong savings on the reagent costs. The system currently in place at Novartis is an open, modular system that also includes non-confocal readout technologies next to the classical confocal readout technologies. This enables selection of the most appropriate readout technology depending on the unique requirements of each biological test system. Furthermore, the platform is equipped with biochemical and cell-based assay capabilities that further enhance the breadth of the screening platform toward many target classes and almost all currently available readout technologies, except radioactivity. Recent advancements of the Novartis NanoScreen platform toward confocal fluorescence lifetime measurements have put this screening platform in a prime position for state-of-the-art industrial lead finding activities. This technological leadership is corroborated by careful full-cost analyses for screening campaigns on various HTS platforms with benchmarking studies on external screening efforts. Due to the strong savings on assay reagents, the Novartis NanoScreen platform is not only among the most high-quality, but also among the most cost-efficient screening technologies currently available in the pharmaceutical and biotech industry.

8.4 The SpeedScreen HTS Platform

In recent years, the average number of chemical compounds used for HTS has steadily increased and already exceeds more than 1 million distinct chemical entities for most pharmaceutical and biotech companies. In some cases, even more than 2–3 million compounds are applied for HTS. Despite numerous approaches to cut costs for HTS campaigns via automation and miniaturization, the costs of HTS efforts are one of the

main factors in the ever-increasing costs of lead discovery. Furthermore, assay development for fully automated screening systems but also data analysis and data validation for very large screening campaigns can still last quite some time and might demand a huge quantity of resources. Despite all the success of the HTS approach for modern drug discovery, the dilemma of ever-increasing resources for finding actives molecules has stimulated the search for faster and more cost-effective alternatives compared to conventional HTS as it has been implemented over the last decade.

Second, the number of pharmaceutical targets has dramatically increased over the last few years, mostly due to the sequencing of the human (and microbial) genomes. Whereas only about 480–520 different molecular targets have been investigated by the pharmaceutical industry from its infancy in the eighteenth century until the year 2000, the sequencing of the human genome and the concomitant "post-genomic" revolution is claiming to deliver several thousands new distinct molecular targets for pharmaceutical intervention (Fig. 21). Taking into account that pharmaceutical targets can comprise also the combination of two and more individual gene sequences but also splice variants of partic-

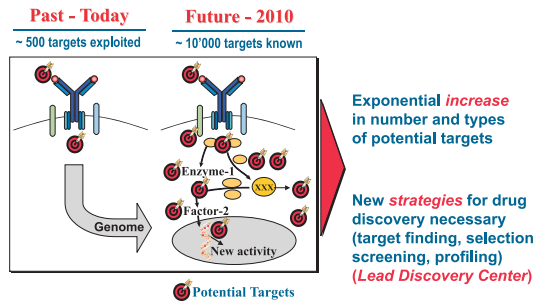
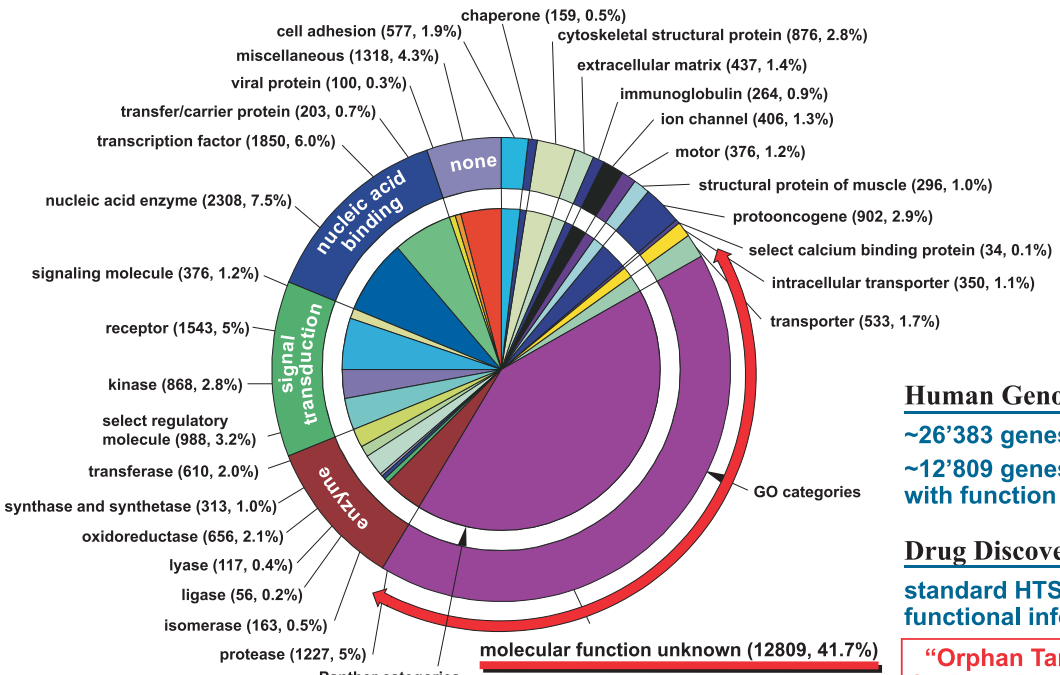


Fig. 21. Post-genomic era: a "burst" of new targets. The figure shows the difference in the number of exploited targets for drug discovery before (~500) and after (~10,000) the "genomic revolution" with sequencing of the human genome. Deciphering of the genome and the cause of human disease on a molecular level will generate a much larger number of potential targets for pharmaceutical intervention. Besides an increase in number, there might also be other or different types of molecular targets, e.g., protein–protein interactions

ular genes, the number of potential drug targets might be in the order of tens of thousands. This enlargement of the biological dimension for drug discovery has even further accelerated the need for novel screening technologies running at high speed and low cost.

A pivotal achievement for biomedical research and development has been the deciphering of the human genome via sequencing of expressed sequence tags (ESTs) as well as shot-gun sequencing of the human genome on the genomic level. This has enabled a bioinformatic characterization and classification of all of the expressed human genes into various families according to sequence information (Fig. 22). Despite the various different protein families expressed in the human genome, this analysis has also revealed that about 40% of the human genome comprises open reading frames (ORFs) with unknown function. Neither bioinformatics nor any other technology can reveal the molecular function of these targets. This part of the genome has not been tackled in greater detail by conventional drug discovery efforts, since our conventional approaches rely on functional knowledge of a particular target in order to develop hits/leads in lead discovery. In summary, about 40% of the human genome comprises targets not yet applicable for conventional drug discovery ("orphan targets" for drug discovery). If one could develop a methodology that would approach screening of those targets without knowing the function of these targets, one would be able to discover not only the biological function of these targets, but might also discover the chemical drugs to modulate the function of these targets. The most straightforward and the easiest possible approach would be to test all available chemical compounds simply for those compounds that bind to these biological targets (binders). These compounds can be used afterwards to link the chemical space with the biological space in drug discovery (chemogenomic approach). Such a technology should be simple, fast, and efficient for the identification of chemical binders to biological targets and must be able to enrich/select only for binders to the target. This concept was proposed several years ago and is commonly termed "affinity-selection of binders".

However, even targets with known function can often cause substantial problems for conventional drug discovery by HTS. Some molecular targets catalyze biochemical reactions that are not feasible for con-



J.C. Venter et al. (2001), *Science* 291, 1304-1354

Human Genome:

~26'383 genes total

~12'809 genes (41.7%) with function **unknown**

Drug Discovery:

standard HTS requires functional information

"Orphan Targets" for Drug Discovery ?

Fig. 22. The various types of human proteins classified by families according to similar function. Out of a total of approx. 27,000 human genes, more than 12,000 genes (41%) encode for proteins with unknown function. Without any knowledge about the function of the proteins, these molecules cannot be assessed by any of the existing, conventional drug discovery technologies and might remain unused for modern drug discovery ("orphan targets for drug discovery")

ventional HTS since the chemical reaction is either too fast (e.g., prolyl *cis/trans* isomerases) or too slow for measurements on highly automated screening systems. Another possibility is that some enzymatic reactions cannot be followed by conventional readout technologies such as fluorescence, luminescence, and radioactivity. In some cases, metabolic enzymes transfer functional groups that cannot be tracked by any of the existing assay technologies in a homogeneous format (e.g., transfer of methylene groups by serine hydroxymethyltransferase, SHMT) without making use of tedious and error-prone indirect readouts. This area of research has been largely unexploited in the past, since many of these targets were not feasible using conventional HTS approaches ("non-tractable targets").

Even further, some of the molecular targets for drug discovery are highly complex by nature. Some target proteins consist of large polypeptides with several autonomous folding units (domains) that can comprise different functional activities on the same polypeptide chain (e.g., BRCA2 gene product). Deciphering a clear causal link between various functional activities on a single polypeptide chain and a particular pathophysiological disease state can be become very challenging, in some cases almost impossible, by the typical project timelines for drug discovery in pharmaceutical and biotech companies. Furthermore, some molecular targets might even consist of several distinct subunits and are only formed in an active conformation via the simultaneous presence of all of the various subunits (e.g., bacterial RNA polymerases). Interfering with one such component or functional activity is a priori not any better than interfering with another. Therefore, having a screening methodology independent of the single components or activities and free of any bias toward the experimental setup of the assay would be highly advantageous for such targets.

Finally, even for some of the well-studied target families among the enzymes, like proteases, kinases, transferases, hydrolases, phosphatases, etc., finding of the appropriate substrate can be a tedious process that can take up a lot of resources (in particular, time). Even with an identified substrate, one cannot exclude the formal possibility that different substrates would have given rise to different or even higher enzymological activity (catalytic efficiency) of a particular target protein. In summary, if one had a screening technology independent of the functional activity of the target protein at hand, one could make those targets better and more rapidly approachable for modern drug discovery (Fig. 23).

With all these requests for alternative lead discovery technologies, one should rethink the fundamental concepts of drug discovery. Interference of chemical compounds with a biological target always comprises transient or permanent binding of a chemical compound to a biological target. This implies that a universal screening technology should be capable of identification of all binding events of chemical compounds

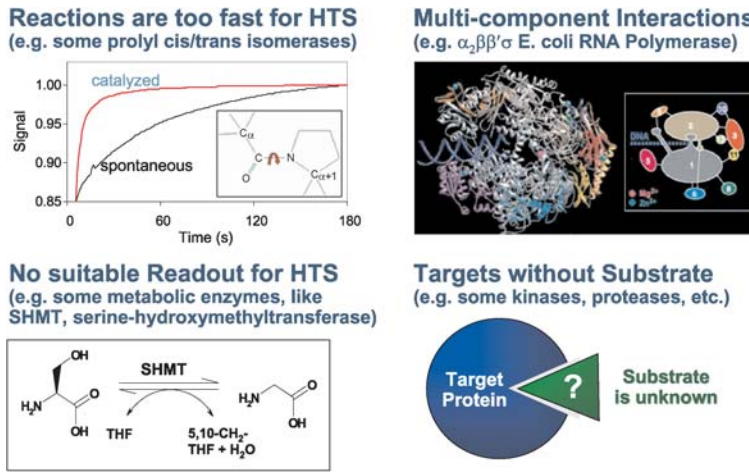


Fig. 23. Conventional, but "non-tractable" targets for HTS. The figure shows the various types of conventional targets that are not easily tractable or not tractable at all by the conventional screening technologies despite the fact that the molecular function of these target proteins is described in the literature

to biological targets in order to support a real "chemogenomic" approach for lead finding. In this concept, chemistry would provide the chemical space by the various means of medicinal chemistry, combinatorial chemistry, and natural products. Genomics would provide the biological space with targets of sufficient evidence to start lead finding and screening efforts. Finally, selection of the appropriate technology would enable the identification of true binders, i.e., separation of bound vs unbound chemical compounds for a particular target protein (Fig. 24). This process is commonly called "affinity-selection of binders". Only those compounds that bind with sufficient affinity to the target would be identified via such an approach. This also implies that one fine-tunes the system in such a way that only binders of the desired properties, like affinity constants, on-rate, off-rate, etc., are identified by such a technology. Furthermore, such a technology should be cheap and very fast for the identification of lead compounds

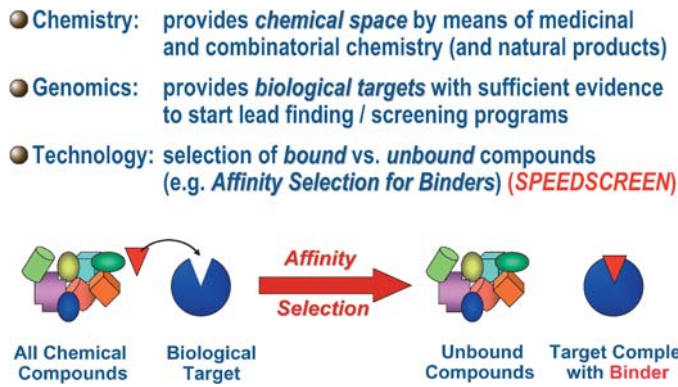


Fig. 24. Chemogenomics: the combination of chemistry and genomics. Chemogenomics combines the fields of chemistry and genomics toward a joint effort for discovery of active drug molecules. Affinity-selection of binders confers the most direct link between the scientific disciplines of chemistry and genomics by simply making use of modern analytical technologies to identify chemical compounds binding to biological targets. Whereas the chemical space is covered by the various sources of chemical libraries, the biological space is covered by the various targets with a variable degree of target validation, i.e., causal link of the target toward a particular disease

on a broad range of target families. Such a technology was developed in-house at Novartis Pharma Research over recent years and is called SpeedScreen.

It should be noted that the concept of "affinity-selection of binders" was proposed at various occasions by numerous people already more that 10 years ago. Numerous technologies for affinity selection were proposed and the most popular idea was based on a covalent coupling of the target protein to the solid phase and subsequent incubation of single compounds or even compound pools on those targets. In most cases, detection of the bound chemical compound was achieved by mass spectrometric readout of the compounds after dissociation from the target protein (Fig. 25). However, such an approach is compromised by the fact that some compounds are notoriously difficult for chemical coupling to the solid phase by any of the standard coupling chemistries. Furthermore, chemical coupling can affect the functional activity of the target protein in part or in total by unspecific reactivity of the coupling

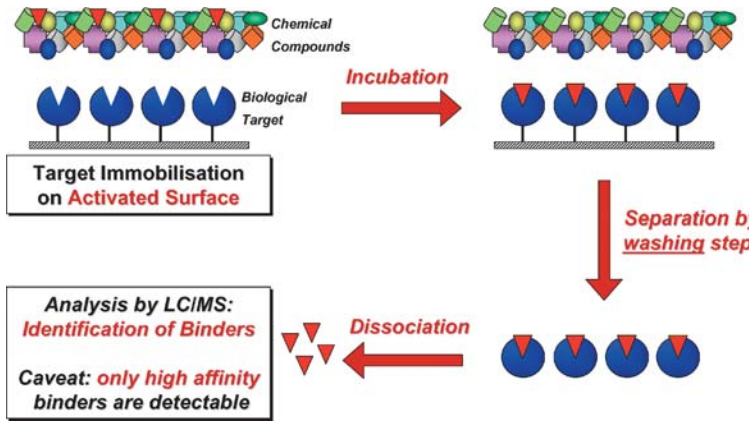


Fig. 25. Affinity-selection of compounds in a heterogeneous, plate-based format. The basic principles for a plate-based, heterogeneous, format for affinity-selection of binders to the target molecule are shown. The target molecule is coupled to the solid phase and mixtures of chemical compounds are incubated with the target protein. After (extensive) washing of the plates, the bound compounds are eluted and analyzed by mass spectrometry

reagent to all functional groups of the same nature at one target protein (e.g., ε-amino groups of all lysine residues). Moreover, covalent attachment of a target protein to the solid phase can cause unfavorable steric and/or electrostatic effects on the accessibility of the target protein for the substrate. The main disadvantage of the conventional affinity-selection methodologies based on a heterogeneous format (i.e., target bound to the solid phase) is, however, that these technologies routinely include various washing steps to remove the unbound from the bound fraction of chemical compounds. These procedures can be quite labor-intense and time-consuming and they can also have a quite detrimental effect on the outcome of these experiments. In general, repetitive washing steps to remove the unbound fraction from the incubation mixture cause dissociation of "weak" binders off the target protein. Depending on the off-rate of the compounds and the number and nature of the washing steps, only compounds with relatively high affinity (nanomolar affinity) are generally retained. All compounds with relatively high off-rates, i.e., often compounds with binding affinities in the range of tens to hundreds of micromolars, can not be retained in sufficient amounts to the immobilized target protein in order to be detected by the appropriate technologies. These limitations of conventional affinity-selection of binders were obvious to us when we decided to implement novel technologies for affinity-selection at Novartis Pharma AG.

The SpeedScreen technology comprises affinity-selection in solution without any covalent attachment of the target protein to the solid phase. This procedure not only eliminates the risk of chemical modification of a target protein due to the use of coupling agents, but also minimizes the risk for inactivation of a target protein due to steric problems upon binding of a target to the solid phase material. This setup also enables generic application of the technology independent of the presence of particular functional groups necessary and used for chemical coupling. The main prerequisite for this approach, however, is the fast separation of unbound chemical compounds from the complex of chemical compounds bound to target protein. After incubation of compounds with target for sufficient time in order to reach the binding equilibrium, the necessary separation of the bound vs the unbound fraction should be done in a very short time. If the separation time for affinity selection is shorter than the dissociation time of chemical compounds, compounds

bound to the target protein can be identified by this methodology. If one can separate the unbound compounds from the target complex in a few seconds, also low affinity compounds with association constants in the range of tens to hundreds of micromolars can be detected by this method (Fig. 26). This also implies that conventional size-exclusion chromatography (SEC) columns with separation times in the range of minutes should not be applied for this purpose. Furthermore, whenever possible, the SEC step to remove unbound compounds from the binder-target complex should also be feasible for multi-parallel processing of the samples in order to meet the requirements of HTS. And finally, the costs of consumables for affinity selection, as well as the costs for the detection method, should be kept low to achieve the goal of a low-cost, fast, and effective screening methodology.

Through extensive development work at the Department of Discovery Technologies at the Novartis Institutes of BioMedical Research (NIBR),

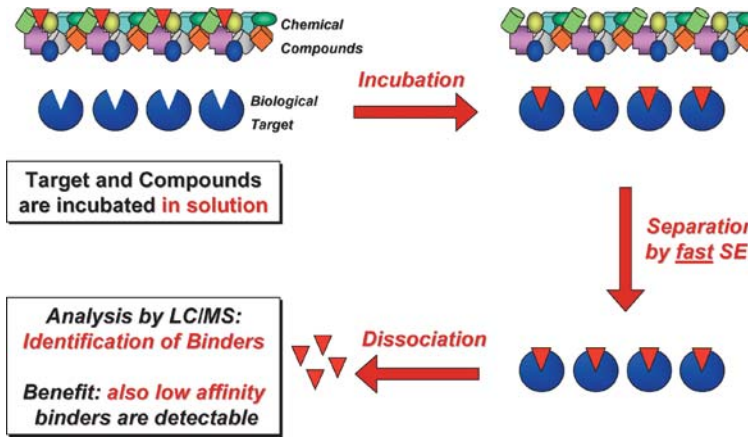


Fig. 26. Affinity-selection of compounds in a homogeneous, in-solution format. The figure shows the basic principles for a homogeneous, in-solution format for affinity-selection of binders to the target molecule. The target molecule is incubated with mixtures of chemical compounds and the separation of the unbound compounds from the compound-binder complex is done by a very fast size-exclusion chromatography (*SEC*) step in a multi-parallel way. Compounds bound to the target protein are subsequently analyzed by mass spectrometry

we were able to develop a novel screening methodology for label-free, high-throughput affinity selection of binders. Due to the fast nature of the separation technology, this screening technology was termed Speed-Screen by its developers (Muckenschnabel et al. 2004; Zehender et al. 2004; Brown et al. 2006).

The basic setup (Fig. 27) is as follows: Pools of compounds (400 distinct chemical entities) are incubated with a particular target of interest under the appropriate assay conditions to achieve binding of chemical compounds. This process is well suited for automation by automated liquid handling stations and does not need manual intervention. Incubation time and conditions can be adjusted to the particular needs of the assay. This process step is done in 96w MTPs with small pinholes at the bottom of each well ("loading plate"). Upon loading of the compound-target mixtures, surface tension of the aqueous assay solution will serve like a very tight seal for the 96 small pinholes in the plate. These plates can also be incubated under any condition (buffer, temperature, oxidative/reductive conditions, light, etc.) and for any time

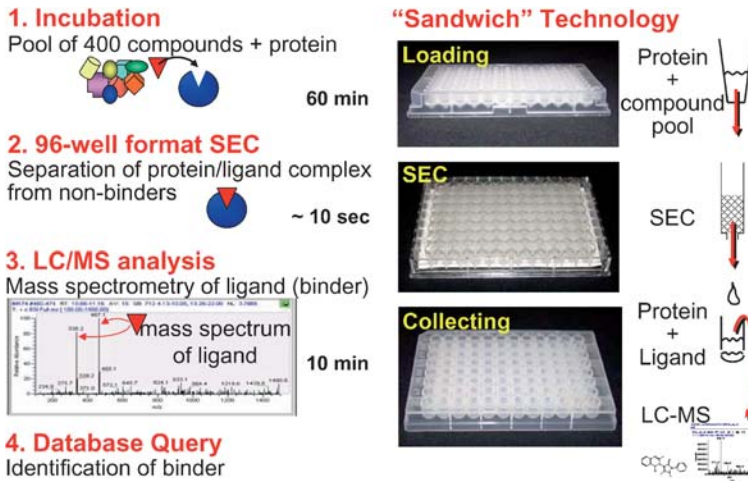


Fig. 27. The basic principles of SpeedScreen technology. The *left panel* describes the four process steps of incubation, 96w-SEC, LC/MS-analysis, and database query. The *right panel* depicts the material used for these process steps

required for a particular affinity selection assay. At the next step, multiples of these microtiter plates can be loaded into special microtiter plate centrifuges with a loading capacity for spinning several microtiter plates at the same time. Upon centrifugation, centrifugal force will drive the affinity selection incubation mixture from the loading plate through the pinholes into the gel matrix of the multi-well SEC plate. Depending on the dimensions of the bed volume in the SEC plate, one can achieve ultra-fast separation times of target-binder complexes from the unbound molecules with a time frame of just a few seconds. The target-binder complex is then eluted toward a collection plate that holds denaturing conditions for the target-binder complex and is placed underneath the SEC plate ("collection plate"). The whole process of affinity selection of thousands of compounds is done in just a few moments. Finally, the loading plate is removed from the centrifuge and transferred into a temperature-controlled auto-sampler for sequential injection into the liquid-chromatography/mass spectrometry (LC/MS) system. Finally, the mass chromatogram is analyzed toward the mass spectrum of the compounds present in the starting pool of compounds.

The individual steps of the SpeedScreen procedure can be described in more detail: The compounds are pooled into mixtures of 400 compounds per well by the Novartis Compound Archive (NCA) and stored as stock solutions at 350 μM per compound in 100% DMSO solution. Aliquots of 0.5 μl per well are transferred by liquid handling robots from the source plates (stock solution) into the 96w-plate screening plates and kept frozen until usage. Protein and incubation buffer are added with 24.5 μl per well to fill up the incubation volume toward a total volume of 25 μl per well with 10 μM of target protein and 7 μM per chemical compound in the pool of compounds. If necessary, control compounds are spiked into the incubation solutions at the appropriate concentration. With pools of 400 compounds per well and 96 wells per microtiter plate, the total number of chemical compounds in one plate used for affinity selection-based screening equals 38,400 compounds. Careful studies with variation of compound and target concentration, as well as variation of pool size, have been used to optimize the experimental setup toward our needs (data not shown). With the use of a standard LC auto-sampler, more than 600,000 distinct chemical compounds are kept with temperature control in the storage device and are ready for

fully automated injection into the LC/MS system. The incubation mixture of protein and pool of compounds can be kept at any temperature, pH-value, oxidative, or buffer conditions necessary for each particular binding assay. The only limitation is the careful selection of the appropriate buffer conditions feasible for MS detection. Neither the protein nor the compounds carries any type of label or are attached to any solid support; therefore, this screening technology can be regarded as a truly "label-free" HTS technology (Fig. 28). This fact gives us a significant advantage over many of the other screening technologies, since labeling of compounds and/or targets can interfere either with the applied readout technology (e.g., fluorescence) or with the biological assay system (e.g., change in binding affinity, etc.). In-solution screening methods without the use of any label certainly constitute the most inert way of measuring the effect of chemical compounds on biological targets. The beauty of the SpeedScreen technology is that one can use the chemical nature of a compound itself – as expressed by the molecular mass of a chemical compound – as the molecular identifier for detection of a biological ef-

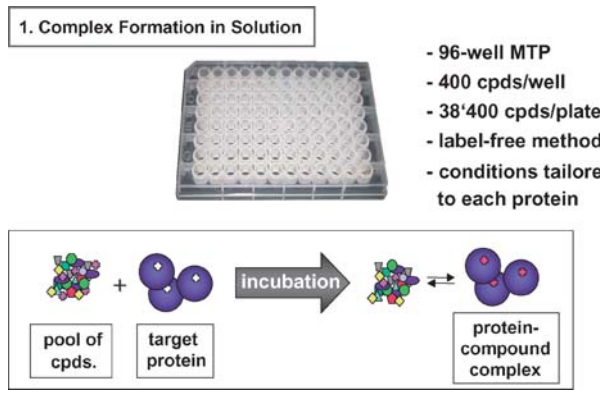


Fig. 28. Step 1 of SpeedScreen: complex formation in solution. The key elements of the complex formation step in the SpeedScreen procedure before SEC-based separation are shown. Each 96w-plate contains 96 wells with pools of 400 compounds each/well, the whole plate contains 38,400 compounds ready for affinity-selection. It should be noted that buffer conditions for this step can be tailored specifically to the needs of a particular target protein

fect, in particular binding to the target molecule. A further benefit of the SpeedScreen process is the fact that all liquid handling steps for compound transfer, compound and protein incubation, etc. can be handled via automated liquid handling systems already in place at compound logistics, analytics, and screening facilities at major pharmaceutical and biotech companies.

After incubation of compounds with the target molecule, the next step comprises the separation of unbound chemical compounds from the binder-target protein complex (Fig. 29). This is done by fast SEC via centrifugation of the loaded SpeedScreen "sandwich" plates in benchtop centrifuges suited for spinning microtiter plates. The SpeedScreen "sandwich" consists of three plates stacked on top of each other (Fig. 29, for details see Fig. 32). The upper plate contains the 96-well incubation plate with the pinholes, the middle plate contains the 96-well SEC plate, and the lower plate contains the 96-well collection plate. Upon centrifugation, centrifugal force drives the transfer of the compound

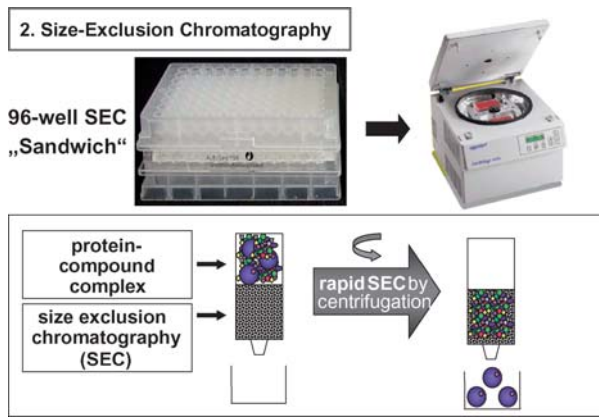


Fig. 29. Step 2 of SpeedScreen: size-exclusion chromatography (*SEC*). The figure shows the stack of incubation, separation, and collection plates as they are used in the SpeedScreen technology (SpeedScreen "sandwich"). Multiples of these plates are taken into Eppendorf benchtop centrifuges. Application of centrifugal force drives the flow of liquid through the column, giving rise to the separation of the ligand-receptor complex vs the free ligand

pool-target protein complex through the pinholes of the loading plate onto and through the matrix of the SEC gel in the separation plate. This needs a centrifuge with sufficient rpm to achieve the appropriate centrifugal force and a system with fast acceleration to achieve the critical rpm in short time. Furthermore, this needs SEC plates of the right volume and matrix material in order to achieve separation of the target-binder complex from the unbound compounds. After various attempts and optimization steps, we currently apply 96-well SEC plates loaded with Sephadex G-50 as a gel filtration material. The SEC step of SpeedScreen is one of the critical steps in the procedure: First, it has to be very fast to achieve fast removal of the unbound compounds from the complex. Second, the separation procedure has to be tuned in such a way that the fraction of the high-molecular weight complex (binder-target protein complex) elutes almost quantitatively from the SEC plate, whereas the fraction of low-molecular-weight compounds (pool of unbound compounds) remains trapped into the SEC matrix. Furthermore, caution has to be taken in order to minimize any interference of chemical compounds or target protein with the gel matrix of even the solid support of the SEC plate. In addition, the costs of the SEC plate should be low, since this device is regarded as a consumable in the process and screening of 1 million compounds would need 26 individual SEC plates. In addition, the plate should be of sufficient throughput (96w-, 384w- plate, or beyond) in order to achieve sufficient sample throughput per working step. Independent of our needs, exactly the same requirements were addressed several years ago in a successful manner by technology developments for high-throughput DNA-sequencing needed for deciphering of the human genome. During the sequencing reaction, dye-primer (or dye-terminator, depending on the technology) is used in excess over the sequencing template and has to be removed from the reaction products of the PCR-cycle sequencing reaction. This is routinely done by 96w- and/or 384w-plates filled with SEC material. Due to the high demand for large numbers of plates and due to low production costs for these plates, that material has become a fairly cheep commodity (<US $50 per plate) for use at the high-throughput sequencing facilities around the world. With this knowledge at hand, we decided to build up a high-throughput affinity selection method based on fast SEC as a separation method to remove the unbound compounds (Fig. 29).

In the third step of the SpeedScreen approach, all binder-target complexes from the collection plate have to be analyzed by high-throughput LC/MS readout technology. All plates from the SEC procedure are loaded into the temperature-controlled auto-sampler of the LC/MS system, which has a loading capacity of 16 plates (614,400 compounds) per system. Screening of larger compound collections is easily achieved by reloading of the LC/MS auto-sampler with new sets of collection plates obtained from the SEC step of the SpeedScreen technology. The auto-sampler opens and closes the drawers of the microtiter plate storage device and also performs sample injection into the LC/MS detection system. Samples of 5 μl each are taken from each well of the microtiter plate and injected onto a reversed-phase high-performance chromatography (RP-HPLC) column running at fairly low flow rates. Separation of compounds by the RP-HPLC column and subsequent compound identification by mass spectrometry with electrospray ionization (ESI) is obtained by a dual micro-bore HPLC system running with two independent columns at low flow rates and injection cycles of 10 min per analysis

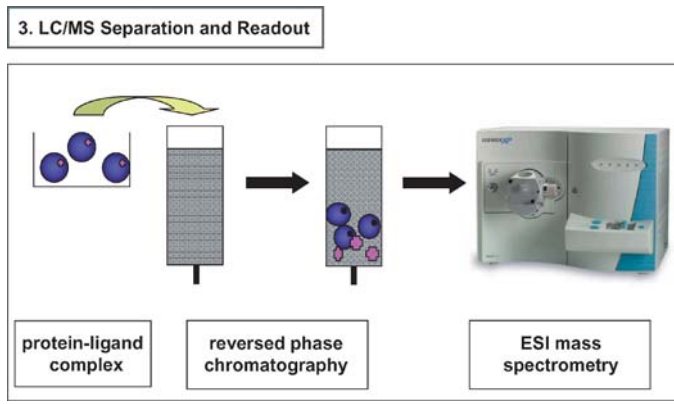


Fig. 30. Step 3 of SpeedScreen: LC/MS separation and readout. The separation and analysis of the protein-ligand complex by reversed phase high-pressure liquid chromatography (RP-HPLC) are shown. Samples from the column are analyzed by electrospray ionization mass spectrometry (ESI MS) on highly sensitive MS instruments with two-dimensional and three-dimensional ion traps for maximum sensitivity

(Fig. 30). By combining two independent HPLC systems with interrogated injection cycles between the two systems, one can achieve fairly short injection intervals for the LC/MS detection system. The currently used mass spectrometry system used electrospray-ionization mass spectrometry (ESI-MS) with three-dimensional or two-dimensional ion traps (Thermo Finnigan LCQ-Deca XPPlus and/or LTQ systems) for enrichment of the charged particles. These systems and the current setup of the hardware are specifically configured for the needs of the SpeedScreen technology. We have seen a very high reliability of these systems with very long running times of uninterrupted use of up to several months. This is only possible due to the fact that there are barely any moving parts on the system and all the other parts of the system, like the injector, auto-sampler, HPLC-pump, degasser, etc. are clearly designed for the long-term uninterrupted use typically required for standard laboratory equipment in an analytical chemistry setup.

In the fourth step of the SpeedScreen procedure, binder identification is done by the analysis of the trace and mass spectrum generated upon each single injection (Fig. 31). This part of the process required a lot of careful thought toward making it a reliable, fast, and automated

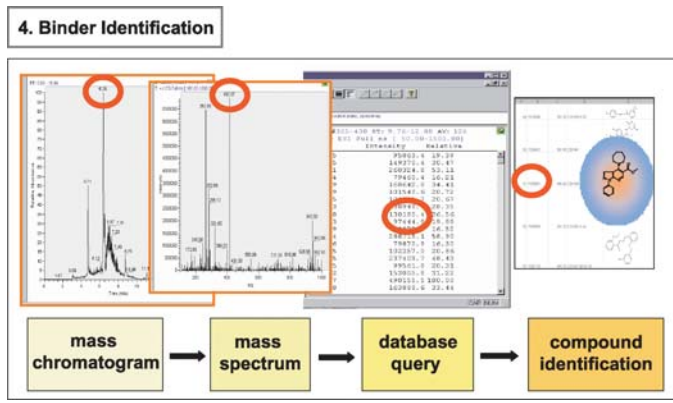


Fig. 31. Step 4 of SpeedScreen: binder identification. The figure shows the key elements of the procedure used for data analysis for potential binders arising from a SpeedScreen project. The mass chromatogram and the mass spectrum are used for identification of the proper chemical compound in the starting pool

process. Originally, we started by doing tedious manual data analysis by the operators. Based on the very high data load per single screening campaign with several gigabytes of data and a lot of visual information from the chromatograms with peak patterns and distinct mass ranges for analysis, manual data analysis of just small-to-medium sized screening campaigns needed up to several weeks of manual data inspection and data analysis for a single operator. This was causing a severe problem for further throughput increase toward larger number of screened compounds and larger numbers of molecular targets. We were able to overcome these obstacles by the development of custom-made software with analysis algorithms for the detection of binders in a SpeedScreen campaign. Current analysis time on largely automated systems is in the range of hours to days with only minimal interaction by the data analyst during the quality controls of the SpeedScreen data analysis. As an outcome, the software delivers a list of identified compounds ("binders") with confidence intervals per data set.

Various attempts were made to develop the ideal setup for SEC, since this is one of the key steps of the procedure and can have a tremendous effect on the outcome of a screening campaign. It is obvious that changing the setup of the SEC step will have a direct effect on the detection threshold, data reliability, and data quality. The current setup with a 96w-MTP "SpeedScreen sandwich" is depicted in Fig. 32. Throughout the process of technology development, we have seen profound differences for the various vendors, plate types, column materials, and SEC column materials (data not shown). After several rounds of optimization, we have been able to standardize the process to such an extent that it has become very reliable and fully compliant to the needs of an industrialized, HTS environment.

During method development for SpeedScreen, we first started with several model systems with well-known binders from the literature. Among the various target proteins tested, protein kinase A (PKA, EC 2.7.1.37) proved to be extremely helpful. Various compounds with well-known binding affinities toward PKA are described in the literature or are available from in-house studies at Novartis Pharma Research. We started to optimize the SpeedScreen procedure with the reference compounds olomoucine, staurosporine, and CHC 12844708 (Novartis internal), among several others. All these compounds can be detected very

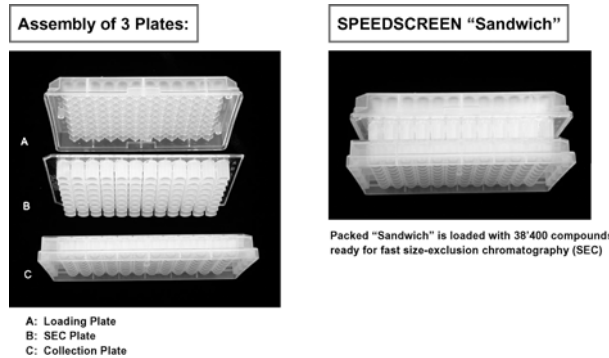
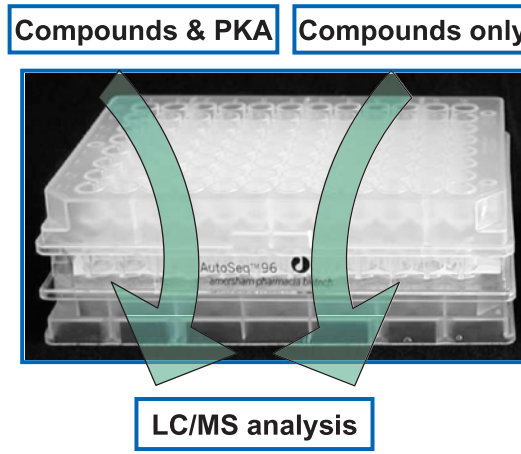


Fig. 32. Details of a "SpeedScreen sandwich", showing three separate plates used in the SpeedScreen approach. Each bottom of a loading plate contains a short pinhole that will open upon centrifugal force. Each "sandwich" contains 38,400 compounds and is, in general, ready for use

easily with a SpeedScreen setup and PKA as a target protein (Fig. 33, right panel). These compounds are unambiguously identified by their molecular mass as shown with the ion trace chromatograms in the presence of target protein. It is obvious that the signal-to-noise ratio of the experimental setup is very pronounced which will allow easy peak allocation by either manual or automated peak determination methods (as described above). As expected, absence of the target protein PKA gives rise to no detectable peak for low-molecular compounds, i.e., binders to the target protein PKA (Fig. 33, right panel). Taken together, these data from known compounds with various affinities toward PKA clearly demonstrate the feasibility of the SpeedScreen approach for detection of low-molecular-weight chemical compounds via affinity-selection based on the principles of fast SEC and LC-MS compound detection.

During development of the SpeedScreen methodology it was obvious that one has to develop the single compound per well affinity-selection method toward application with pooled compound collections. This is required due to the sequential detection method of an MS-based screening technology and due to the requirements for protein consumption, plate handling, etc. Throughout the development phase, we tested various concentrations of target protein, different compound concentrations,



PKA: Protein Kinase A
(E.C. 2.7.1.37)

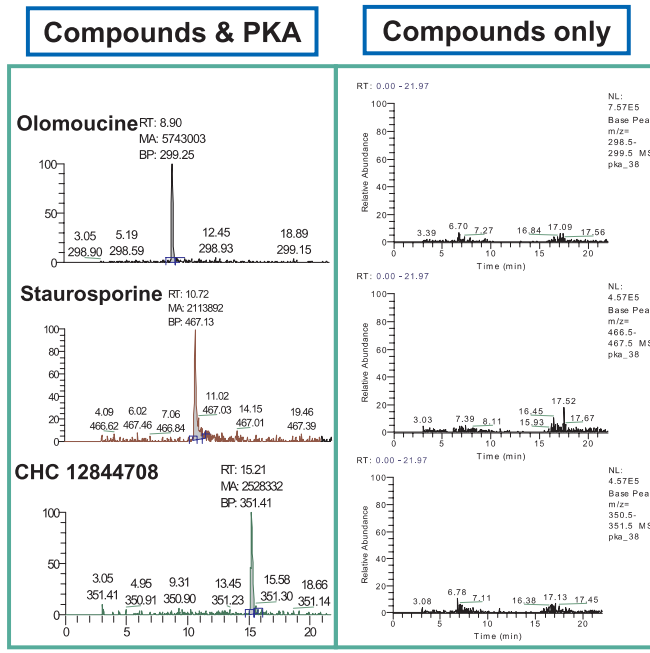


Fig. 33. SpeedScreen technology for single compounds. The figure shows the results from studies with single compounds on the model target protein PKA (EC 2.7.1.37). Various reference compounds, such as olomoucine, staurosporine, and CHC 12844708, are detected on the presence of target protein, but not in the absence of the target protein

and we varied the pool size between 10 and 2,000 compounds per well (data not shown). In our experimental setup, a pool size of 400 compounds per well under the given conditions of target and protein concentration has shown to give us the best compromise with regards to cost, time, and efficiency of the process. With PKA as a model system and a selection of 400 arbitrarily chosen chemical compounds from the Novartis compound collection, we performed "spiking" experiments with a known binder to PKA into the pool of random compounds. Upon analysis of this sample by the standard LC/MS detection method prior to the SEC step in SpeedScreen, one sees the expected broad range of irresolvable peaks for the pool of 400 compounds in the mixture (Fig. 34, top panel). During the SpeedScreen affinity-selection process, one would expect enrichment of staurosporine as a known binder to PKA. With the current setup, one can clearly identify just one binder to PKA and the corresponding mass unambiguously identifies staurosporine as the appropriate binder to this target (Fig. 34, lower right panel). It should be noted that the peak in the ion trace chromatogram of that experiment shows two peaks that can be clearly separated and clearly separated from background noise due to the very high signal-to-background ratio of the peaks. Upon performing the appropriate control with loading the same 400 compounds and staurosporine onto the SpeedScreen SEC setup, albeit in the absence of target protein, one would expect no enrichment for staurosporine. It should be noted that exactly this result has been obtained for that control setup (Fig. 34, lower left panel). This also shows that only a matrix peak is carried over from the setup in the presence and absence of target protein. Taken together, these data nicely show that well-known binders to a particular target protein can be detected even in the presence of a large excess of non-binders to that protein in pools of 400 compounds in a well.

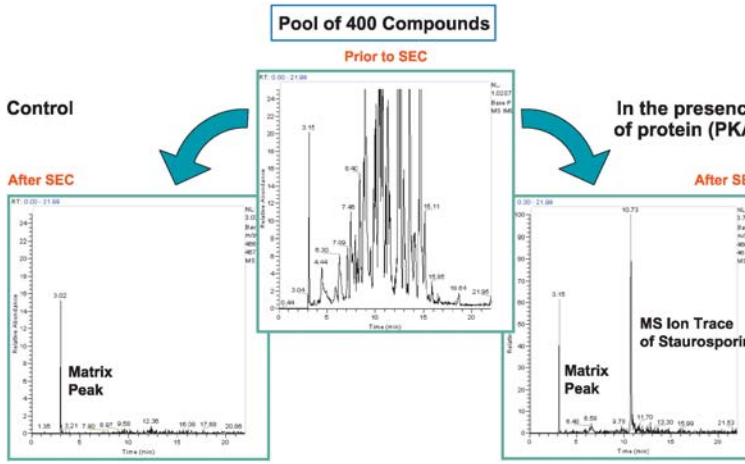


Fig. 34. SpeedScreen technology for pools of compounds. The figure shows the results from studies with 400 compounds on the model target protein PKA (EC 2.7.1.37). Staurosporine has been spiked into the pool of 400 compounds, but the system is still able to detect binders

If the SpeedScreen procedure is used for HTS efforts of binders to target proteins, it will be important to define the key requirements for a particular screening campaign in an HTS environment of a pharmaceutical or biotech company or at an academic research facility. Commonly, the screening collections range between 100,000 and 1 million distinct chemical entities, in some cases up to 4 million chemical compounds. Most of the major pharmaceutical companies use around 1 million chemical compounds from various sources for their screening efforts. Based on the SpeedScreen setup (Fig. 35), one can define the key requirements for the SpeedScreen affinity-selection process as follows: Each well in a 96w-MTP contains the target protein and a pool of 400 distinct chemical compounds at a total volume of 25 μl per well. The final protein concentration of the sample is adjusted to 10 μM, the final concentration of chemical compound is 7 μM per single compound. All compounds and the protein are dissolved in aqueous assay buffer with a final concentration of 1.4% DMSO in order to enhance solubility of

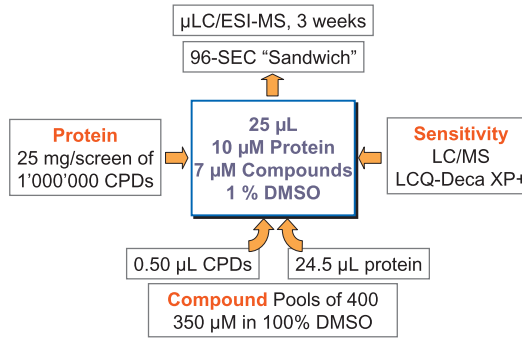


Fig. 35. The key requirements for the SpeedScreen technology, such as protein concentration, compound concentration, sensitivity of the used ESI-detector, and size of compound test deck, are shown

the chemical compounds without changing the biological activity of the target protein. This incubation mixture is obtained by the combination of 0.7 µl of compound stock solution at a concentration of 250 µM per compound with pools of 400 compounds and the addition of 24.3 µl protein stock solution with the appropriate biological assay conditions. It is recommended to use freshly prepared solutions of chemical compound pools in order to minimize potential chemical cross-reactivity between the various compounds in the compound mixture of 400 compounds at high concentration. Based on the molecular weight of the target protein, defined amounts of target protein are needed for screening of a compound collection of 1 million chemical compounds. With a hypothetical mass of 25 kDa for the target protein, one needs 25 mg of target protein for the screening of this compound collection at the given concentrations. If more sensitive detectors or other variations of the SpeedScreen affinity-selection setup are done, this number for protein consumption in a single screening campaign will go up or down accordingly. With the previously described setup and a daily throughput of 57,600 compounds per day (403,200 compounds/week), a primary screening campaign for 1 million compounds will last about 18 days (2.5 weeks) of continuous operation. Current efforts are underway to further shorten the amount of screening time per campaign and to re-

duce the amount of required protein for the full SpeedScreen campaign (data not shown).

Whereas data analysis has been taking about 2–3 weeks of full-time manual inspection of the mass chromatograms and mass spectrums by the SpeedScreen operators, newer systems can be used to automate the full data analysis procedure (Fig. 36). The current system is an expert system which analyses the mass chromatogram and the mass spectrum from the experimental setup and performs a comparison to the expected data based on the known input with sample sizes of 400 compounds per mixture. During assay development of a particular campaign, the parameters for automated data analysis of the SpeedScreen data are adjusted to maximize the sensitivity of the software system for automated compound detection in the dataset and to minimize the amount of false positives and false negatives from the analysis. This requires that mass and chemical structures of all the compounds fed into the SpeedScreen process are known and can be used for the computational methods. The proprietary SpeedScreen data analysis software collects all information on putative

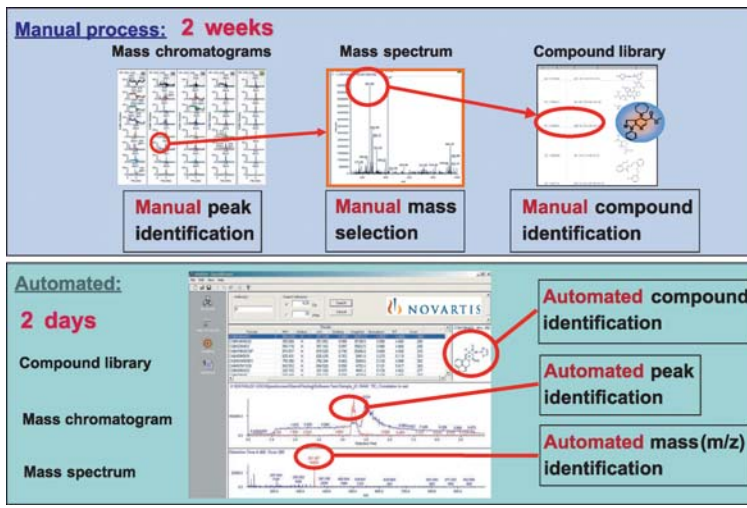


Fig. 36. The quite labor-intense work necessary for manual data analysis (2–3 weeks) is compared to the fully automated data analysis expert system

binders from the screening campaign, performs data warehousing of the results and enables export of the result files for further processing of the data, e.g., cherry-picking of compounds for subsequent conformation studies or data storage of the results in company-wide databases for chemical compounds. It should be noted that the same software is also used during confirmation and validation studies where all active compounds from a primary screening campaign are measured again with single compounds per well in the presence and absence of target protein. Upon this procedure, false positives of the procedure such as compound oligomers, compound aggregates, or compounds adhering as unspecific binders to the column material and plate surface are eliminated from the SpeedScreen hitlist (data not shown). This procedure offers some tremendous benefits over alternative screening technologies, since the SpeedScreen technology can address the issue of unspecific compound oligomers and removes these entities from the SpeedScreen hitlist.

As pointed out before, the technology is broadly applicable to various different target classes (Fig. 37). Over the last few years, various types of pharmaceutical targets have been tested with great success by the SpeedScreen technology in an affinity-selection approach. Targets include various members from the families of transferases, isomerases, dehydrogenases, kinases, proteases, phosphatases, oxidoreductases, transcription factors, adapter molecules, regulatory subunits, heat shock

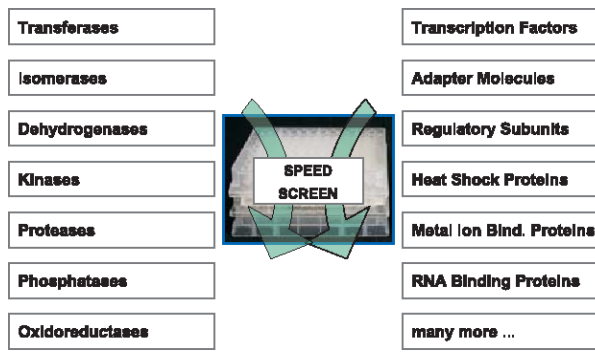


Fig. 37. Summary of various target classes feasible for SpeedScreen, showing various of the target classes we have already tested in the SpeedScreen setup

proteins, metal ion binding proteins, and many more. In principle, all types of soluble homogenous target proteins should be feasible for the SpeedScreen approach. Screening of transmembrane proteins might also become feasible in the future (data not shown). It is noteworthy that a recent cheminformatic analysis of a set of 25 different, randomly chosen SpeedScreen campaigns has clearly demonstrated the broad applicability of the technology to various target classes and the lack of enrichment for unspecific and/or promiscuous binders.

In summary, the SpeedScreen technology offers a variety of benefits over other screening technologies currently applied in the field (Fig. 38):

1. The technology is proprietary to Novartis. The technology was fully developed at Novartis Pharma Research with in-house efforts.
2. The technology enables HTS of chemical binders at a rate that has not been possible before. A single SpeedScreen system has a daily capacity of 57,600 compounds per working day.
3. The SpeedScreen method is a truly label-free, in-solution method, since neither any label on protein or compound, nor any type of chemical coupling of the target protein to the solid phase is required.
4. The SpeedScreen technology has proved its value for the Novartis Lead Discovery Center (LDC) with regards to number and quality of the SpeedScreen lead finding campaigns.

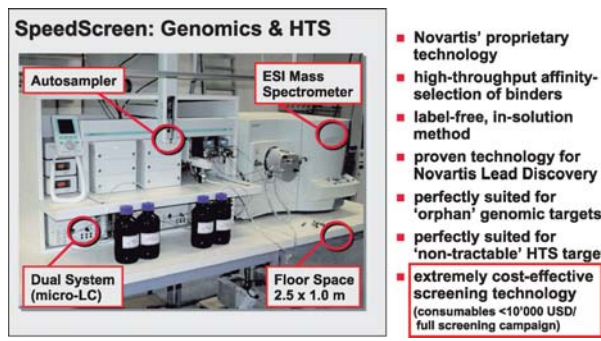


Fig. 38. SpeedScreen: system setups and essentials. The figure describes the key essentials for the SpeedScreen technology. Each station has a fairly small footprint and consists of autosampler, dual-microbore HPLC and LC/MS

5. The SpeedScreen setup is perfectly suited for "orphan" genomic targets, i.e., targets with unknown biological activity.
6. The SpeedScreen technology is perfectly suited for "non-tractable" targets, i.e., targets that cannot be tackled by the existing, conventional readout technologies for HTS.
7. The SpeedScreen technology is extremely cost-effective, since the price of consumables for a screening campaign of 1 million compounds is around US $10,000 or less. This is a number that cannot be rivaled by any other screening technology commonly applied in modern drug discovery efforts.

In order to show the application of the SpeedScreen technology, we have chosen the ubiquitin pathway with its various enzymes as a model system (Hershko and Ciechanover 1998). Several members of the family of E2 proteins, a family of ubiquitin-conjugating enzymes, might be attractive targets for pharmaceutical intervention (Fig. 39, left panel). Due to the particular enzymatic activity, conjugation of ubiquitin, the enzymatic activity of these proteins, is hard to detect in conventional HTS detec-

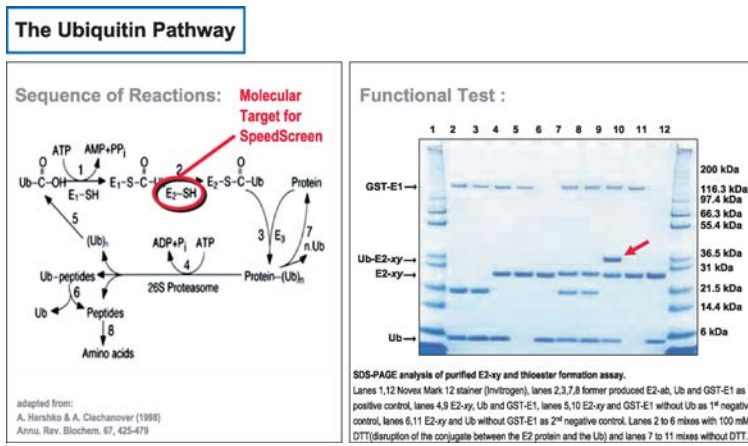


Fig. 39. The ubiquitin pathway with all the involved enzymes, such as members of the E2 family (ubiquitin-conjugating enzymes; *left panel*). The figure also shows the SDS-PAGE with the functional band-shift assay to follow the enzymatic activity of the protein

tion systems, whereas the functional activity can easily be monitored by SDS-PAGE analysis (Fig. 39, right panel). It becomes immediately evident that such band-shift assays for measuring the functional enzymatic activity cannot be applied for routine testing of 500,000–1 million compounds in an HTS mode. It was therefore decided to test the applicability of the SpeedScreen technology on several members of the E2 family of ubiquitin-conjugating enzymes. In these affinity-selection screening campaigns for binders to E2 proteins, numerous chemical compounds with binding activities to the target proteins were identified. About 1/3 of all binders also showed an inhibitory effect in band-shift assays for the E2 protein. In order to confirm these data by independent setup, we also investigated several of these compounds by a BiaCore S51 surface-plasmon-resonance (SPR) readout technology. This technology enables the determination of on- and off-rates for biochemical reactions and can be used to determine the binding constant for biochemical binding reactions (Fig. 40). One of the chemical compounds ("binder")

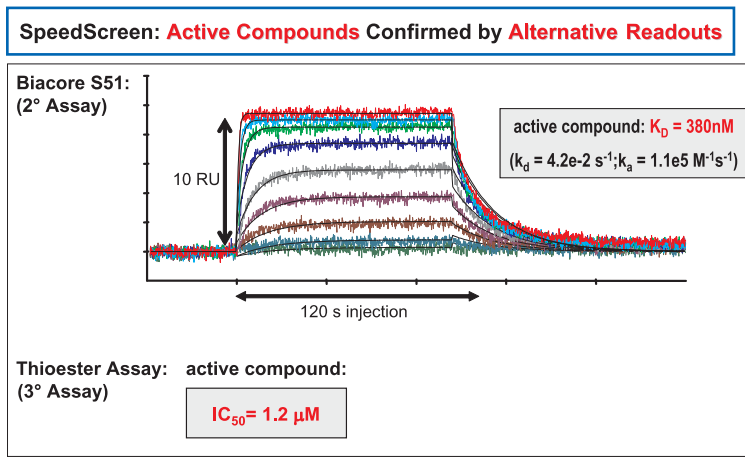


Fig. 40. Confirmation of active SpeedScreen binders by independent readouts. The figure shows the results of a SpeedScreen compound that binds to a member of the E2 family of proteins. The *upper panel* shows data from a Biacore instrument with the appropriate k_{on}- and k_{off}-values. The *lower panel* specifies the activity of the same compound in a tedious multi-lane band-shift assay

from the SpeedScreen assay showed on-rates of $1.1 \times 10^5\ M^{-1}s^{-1}$ and off-rates of $4.2 \times 10^{-2}\ s^{-1}$ giving rise to an equilibrium dissociation constant of 340 nM. Similar data where obtained with a thioester band shift assay, where the same compound confirmed with an IC_{50} of 1.2 μM. These data clearly prove the value of the SpeedScreen approach for the identification of low-molecular binders to defined target proteins via the means of high-throughput affinity-selection screening technologies.

8.5 Summary

Independent of the precise nature of the applied screening technology, lead discovery efforts can always be analyzed and optimized along the sample fundamental principles of performance management ("the magic triangle of HTS"): time, costs and quality of the process (Fig. 41). Since HTS always deals with large amounts of samples to be analyzed, the measurement time for a single well in the screening campaign is a key performance parameter. The same holds true for the number of wells that can be run per working day and the number of screens that can be run per year. Changing the time component of the HTS efforts among the types just described will have a direct and indirect effect on the costs of a screening campaign since instruments and other resources

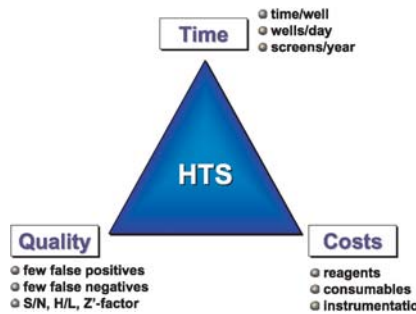


Fig. 41. The "magic triangle of HTS". The key success factors for modern lead discovery via HTS are time, costs, and quality. As can seen on the slide, all three factors are closely interdigitated and every change on any of these factors changes the setup for all the other factors as well

will be needed for either shorter or longer periods of time. A major cost driver for screening is not only the technological hardware for screening, like robotic plate handling and readout systems, but also the costs for reagents and consumables. Typical consumable costs are the costs for plates, tips, vials, etc. Typical reagent costs are the costs for the biological test samples, mostly protein, cells, substrate, etc., which can become detrimental in cases with some sensitive or hard-to-produce biological assay reagents. Ultimately, however, the most important factor in lead discovery efforts is the quality of the process. This is particularly important for screening that deals with very large data sets, since only assays of high statistical quality can be used for proper data analysis. It is important to consider that the number of false positives, the number of false negatives, and other statistical terms such as S/N (signal/noise), H/L (high/low) or Z'-factor can be used to optimize the statistical quality of an HTS campaign. It should be noted that the quality of an assay is not only expressed by the statistical quality of an assay (exemplified by Z'-value), but also by the biochemical or biological sensitivity of a particular assay setup. In other words, assays should be developed not only toward maximum statistical quality, but also for sensitive detection of weak inhibitors in a lead finding campaign to offer a potential new avenue for drug discovery with some new chemical series or scaffold. It is important that all three main elements of successful HTS – time, costs, quality – are closely linked and interdigitated in reality. Every lead finding effort, but also every lead finding technology, can be evaluated according to these generic criteria for success.

With the help of the various automated screening systems, in particular NanoScreen and SpeedScreen, our organization was able to strongly increase the number of screened entities per year in the last 5–10 years (Fig. 42). This clearly indicates that the organization was capable of taking up the challenge with ever-increased compound collections and the increase in number of screenable targets. This was of course only possible due to the constant decrease in costs per data point in HTS. It has to be shown in the future whether the trend toward the strong increase in number of data points and the strong increase in fully loaded costs per screening data point can continued. Despite that, it has become clear that automation and miniaturization must have been capable of more than over-seeding the initial setup

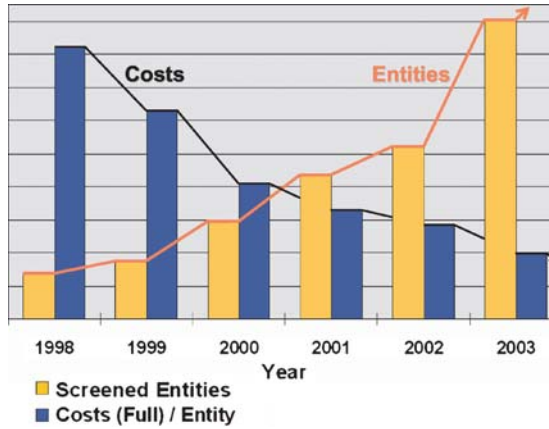


Fig. 42. Costs/productivity development over time. The developments and trends for previous years in terms of the total number of screened entities and the concomitant decrease in costs are shown

costs for large automation and miniaturization systems. It should be pointed out that the cost analysis already contains a 2- to 5-year depreciation period for instruments and robotic equipment. We therefore can conclude form Fig. 42 that Novartis has been very successful at validating and implementing novel screening and liquid handling technologies and that this strategy has already given a remarkably good return on investment as expressed by the strong and continuous decrease in costs per data point. Another way of looking at this is given by the comparison of the Novartis costs per screen with external benchmarks for screening exactly the same amount of chemical compounds (Fig. 43). This figure clearly shows the variance of total costs needed for a full screening campaign at external and internal service providers. This is a natural consequence of the heterogeneity and complexity of biological systems that cannot always be standardized and run with just a single experimental setup. This figure also shows that not only the ranges, but also the costs for the internal screening efforts at Novartis Pharma are significantly lower than the costs at external service providers.

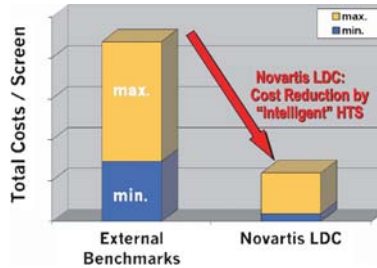


Fig. 43. Benchmarking studies. The figure shows the broad variation among the costs for internal and external service providers. Internal numbers include all costs except value of a chemical compound; external numbers are derived from various benchmarking studies with regards to outsourcing of screening campaigns

8.6 Outlook

We have developed two very powerful technological solutions for the ever-demanding tasks in industrial lead finding. The NanoScreen approach is mostly a very low volume reformatting and screening system for biochemical and cell-based assays, it uses both confocal as well as non-confocal readout technologies, it is highly automated, and it has a focus on high-quality, cost-efficient screening with functional binding and enzymatic assays. The SpeedScreen approach, however, makes use of a proprietary high-throughput affinity-selection process for identification of chemical binders to proteins. This method is completely label-free with no need for chemical coupling and/or labeling of either the compounds or the target protein. The SpeedScreen technology has been very successful with regards to application to "orphan" genomic targets and to "non-tractable" targets, i.e., targets that are either not tractable at all by conventional technologies or can be worked on only with major efforts and compromises like indirect readout via coupling to secondary and tertiary enzymatic reactions. SpeedScreen also has the benefit that the technology is very cheap with regards to the necessary consumables and reagents.

Taken together, the NanoScreen and SpeedScreen technologies at Novartis Pharma Research have proved to be very successful for No-

vartis during tackling the chemogenomic space with novel screening technologies.

Acknowledgements. I am very grateful to all my colleagues at NIBR (Novartis Institute of Biomedical Research), Department of Discovery Technologies, who have contributed to the successful implementation and application of the NanoScreen and SpeedScreen technologies, in particular René Amstutz, Manfred Auer, Jutta Blank, Ireos Filipuzzi, Michael Forstner, Felix Freuler, Peter Fuerst, Fraser Glickman, Martin Klumpp, Lukas Leder, Ingo Muckenschnabel, Reto Naef, Johannes Ottl, and Hartmut Zehender, and I fully appreciate their continuous support, encouragement, and enthusiasm for novel technologies in drug discovery.

References

Brown N, Zehender H, Azzaoui K, Schuffenhauer A, Mayr LM, Jacoby E (2006) A chemoinformatics analysis of hitlists obtained from high-throughput affinity-selection screening. J Biomol Screen 10 (in press)

Hershko A, Ciechanover A (1998) The ubiquitin system. Annu Rev Biochem 67:425–479

Muckenschnabel I, Falchetto R, Mayr LM, Filipuzzi I (2004) SpeedScreen: label-free liquid chromatography-mass spectrometry-based HTS for the discovery of orphan protein ligands. Anal Biochem 324:241–249

Zehender H, Le Goff F, Lehmann N, Filipuzzi I, Mayr LM (2004) SpeedScreen: the 'missing link' between genomics and lead discovery. J Biomol Screen 9:498–505

Zhang JH, Chung TD, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4:67–73

Ernst Schering Research Foundation Workshop

Editors: Günter Stock
Monika Lessl

Vol. 1 (1991): Bioscience ⇌ Societly Workshop Report
Editors: D.J. Roy, B.E. Wynne, R.W. Old

Vol. 2 (1991): Round Table Discussion on Bioscience ⇌ Society
Editor: J.J. Cherfas

Vol. 3 (1991): Excitatory Amino Acids and Second Messenger Systems
Editors: V.I. Teichberg, L. Turski

Vol. 4 (1992): Spermatogenesis – Fertilization – Contraception
Editors: E. Nieschlag, U.-F. Habenicht

Vol. 5 (1992): Sex Steroids and the Cardiovascular System
Editors: P. Ramwell, G. Rubanyi, E. Schillinger

Vol. 6 (1993): Transgenic Animals as Model Systems for Human Diseases
Editors: E.F. Wagner, F. Theuring

Vol. 7 (1993): Basic Mechanisms Controlling Term and Preterm Birth
Editors: K. Chwalisz, R.E. Garfield

Vol. 8 (1994): Health Care 2010
Editors: C. Bezold, K. Knabner

Vol. 9 (1994): Sex Steroids and Bone
Editors: R. Ziegler, J. Pfeilschifter, M. Bräutigam

Vol. 10 (1994): Nongenotoxic Carcinogenesis
Editors: A. Cockburn, L. Smith

Vol. 11 (1994): Cell Culture in Pharmaceutical Research
Editors: N.E. Fusenig, H. Graf

Vol. 12 (1994): Interactions Between Adjuvants, Agrochemical and Target Organisms
Editors: P.J. Holloway, R.T. Rees, D. Stock

Vol. 13 (1994): Assessment of the Use of Single Cytochrome P450 Enzymes in Drug Research
Editors: M.R. Waterman, M. Hildebrand

Vol. 14 (1995): Apoptosis in Hormone-Dependent Cancers
Editors: M. Tenniswood, H. Michna

Vol. 15 (1995): Computer Aided Drug Design in Industrial Research
Editors: E.C. Herrmann, R. Franke

Vol. 16 (1995): Organ-Selective Actions of Steroid Hormones
Editors: D.T. Baird, G. Schütz, R. Krattenmacher

Vol. 17 (1996): Alzheimer's Disease
Editors: J.D. Turner, K. Beyreuther, F. Theuring

Vol. 18 (1997): The Endometrium as a Target for Contraception
Editors: H.M. Beier, M.J.K. Harper, K. Chwalisz

Vol. 19 (1997): EGF Receptor in Tumor Growth and Progression
Editors: R.B. Lichtner, R.N. Harkins

Vol. 20 (1997): Cellular Therapy
Editors: H. Wekerle, H. Graf, J.D. Turner

Vol. 21 (1997): Nitric Oxide, Cytochromes P 450,
and Sexual Steroid Hormones
Editors: J.R. Lancaster, J.F. Parkinson

Vol. 22 (1997): Impact of Molecular Biology
and New Technical Developments in Diagnostic Imaging
Editors: W. Semmler, M. Schwaiger

Vol. 23 (1998): Excitatory Amino Acids
Editors: P.H. Seeburg, I. Bresink, L. Turski

Vol. 24 (1998): Molecular Basis of Sex Hormone Receptor Function
Editors: H. Gronemeyer, U. Fuhrmann, K. Parczyk

Vol. 25 (1998): Novel Approaches to Treatment of Osteoporosis
Editors: R.G.G. Russell, T.M. Skerry, U. Kollenkirchen

Vol. 26 (1998): Recent Trends in Molecular Recognition
Editors: F. Diederich, H. Künzer

Vol. 27 (1998): Gene Therapy
Editors: R.E. Sobol, K.J. Scanlon, E. Nestaas, T. Strohmeyer

Vol. 28 (1999): Therapeutic Angiogenesis
Editors: J.A. Dormandy, W.P. Dole, G.M. Rubanyi

Vol. 29 (2000): Of Fish, Fly, Worm and Man
Editors: C. Nüsslein-Volhard, J. Krätzschmar

Vol. 30 (2000): Therapeutic Vaccination Therapy
Editors: P. Walden, W. Sterry, H. Hennekes

Vol. 31 (2000): Advances in Eicosanoid Research
Editors: C.N. Serhan, H.D. Perez

Vol. 32 (2000): The Role of Natural Products in Drug Discovery
Editors: J. Mulzer, R. Bohlmann

Vol. 33 (2001): Stem Cells from Cord Blood, In Utero Stem Cell Development, and Transplantation-Inclusive Gene Therapy
Editors: W. Holzgreve, M. Lessl

Vol. 34 (2001): Data Mining in Structural Biology
Editors: I. Schlichting, U. Egner

Vol. 35 (2002): Stem Cell Transplantation and Tissue Engineering
Editors: A. Haverich, H. Graf

Vol. 36 (2002): The Human Genome
Editors: A. Rosenthal, L. Vakalopoulou

Vol. 37 (2002): Pharmacokinetic Challenges in Drug Discovery
Editors: O. Pelkonen, A. Baumann, A. Reichel

Vol. 38 (2002): Bioinformatics and Genome Analysis
Editors: H.-W. Mewes, B. Weiss, H. Seidel

Vol. 39 (2002): Neuroinflammation – From Bench to Bedside
Editors: H. Kettenmann, G.A. Burton, U. Moenning

Vol. 40 (2002): Recent Advances in Glucocorticoid Receptor Action
Editors: A. Cato, H. Schaecke, K. Asadullah

Vol. 41 (2002): The Future of the Oocyte
Editors: J. Eppig, C. Hegele-Hartung

Vol. 42 (2003): Small Molecule-Protein Interaction
Editors: H. Waldmann, M. Koppitz

Vol. 43 (2003): Human Gene Therapy:
Present Opportunities and Future Trends
Editors: G.M. Rubanyi, S. Ylä-Herttuala

Vol. 44 (2004): Leucocyte Trafficking:
The Role of Fucosyltransferases and Selectins
Editors: A. Hamann, K. Asadullah, A. Schottelius

Vol. 45 (2004): Chemokine Roles in Immunoregulation and Disease
Editors: P.M. Murphy, R. Horuk

Vol. 46 (2004): New Molecular Mechanisms of Estrogen Action and Their Impact on Future Perspectives in Estrogen Therapy
Editors: K.S. Korach, A. Hillisch, K.H. Fritzemeier

Vol. 47 (2004): Neuroinflammation in Stroke
Editors: U. Dirnagl, B. Elger

Vol. 48 (2004): From Morphological Imaging to Molecular Targeting
Editors: M. Schwaiger, L. Dinkelborg, H. Schweinfurth

Vol. 49 (2004): Molecular Imaging
Editors: A.A. Bogdanov, K. Licha

Vol. 50 (2005): Animal Models of T Cell-Mediated Skin Diseases
Editors: T. Zollner, H. Renz, K. Asadullah

Vol. 51 (2005): Biocombinatorial Approaches for Drug Finding
Editors: W. Wohlleben, T. Spellig, B. Müller-Tiemann

Vol. 52 (2005): New Mechanisms for Tissue-Selective Estrogen-Free Contraception
Editors: H.B. Croxatto, R. Schürmann, U. Fuhrmann, I. Schellschmidt

Vol. 53 (2005): Opportunities and Challenges of the Therapies Targeting CNS Regeneration
Editors: D. Perez, B. Mitrovic, A. Baron Van Evercooren

Vol. 54 (2005): The Promises and Challenges of Regenerative Medicine
Editors: J. Morser, S.I. Nishikawa

Vol. 55 (2006): Chronic Viral and Inflammatory Cardiomyopathy
Editors: H.-P. Schultheiss, J.-F. Kapp

Vol. 56 (2006): Cytokines as Potential Therapeutic Target for Inflammatory Skin Diseases
Editors: R. Numerof, C.A. Dinarello, K. Asadullah

Vol. 57 (2006): The Histone Code and Beyond
Editors: S.L. Berger, O. Nakanishi, B. Haendler

Vol. 58 (2006): Chemical Genomics
Editors: S. Jaroch, H. Weinmann

Supplement 1 (1994): Molecular and Cellular Endocrinology of the Testis
Editors: G. Verhoeven, U.-F. Habenicht

Supplement 2 (1997): Signal Transduction in Testicular Cells
Editors: V. Hansson, F.O. Levy, K. Taskén

Supplement 3 (1998): Testicular Function:
From Gene Expression to Genetic Manipulation
Editors: M. Stefanini, C. Boitani, M. Galdieri, R. Geremia,F. Palombi

Supplement 4 (2000): Hormone Replacement Therapy
and Osteoporosis
Editors: J. Kato, H. Minaguchi, Y. Nishino

Supplement 5 (1999): Interferon: The Dawn of Recombinant
Protein Drugs
Editors: J. Lindenmann, W.D. Schleuning

Supplement 6 (2000): Testis, Epididymis and Technologies
in the Year 2000
Editors: B. Jégou, C. Pineau, J. Saez

Supplement 7 (2001): New Concepts in Pathology
and Treatment of Autoimmune Disorders
Editors: P. Pozzilli, C. Pozzilli, J.-F. Kapp

Supplement 8 (2001): New Pharmacological Approaches
to Reproductive Health and Healthy Ageing
Editors: W.-K. Raff, M.F. Fathalla, F. Saad

Supplement 9 (2002): Testicular Tangrams
Editors: F.F.G. Rommerts, K.J. Teerds

Supplement 10 (2002): Die Architektur des Lebens
Editors: G. Stock, M. Lessl

Supplement 11 (2005): Regenerative and Cell Therapy
Editors: A. Keating, K. Dicke, N. Gorin, R. Weber, H. Graf

Supplement 12 (2005): Von der Wahrnehmung zur Erkenntnis –
From Perception to Understanding
Editors: M. Lessl, J. Mittelstraß