Edited by
Nathan Brown

Scaffold Hopping in
Medicinal Chemistry
Methods and Principles in Medicinal Chemistry
Edited by R. Mannhold, H. Kubinyi, G. Folkers
Editorial Board
H. Buschmann, H. Timmerman, H. van de Waterbeemd, T. Wieland

Previous Volumes of this Series:

Hoffmann, Rémy D. / Gohier, Arnaud / Pospisil, Pavel (Eds.)
Data Mining in Drug Discovery
2014
ISBN: 978-3-527-32984-7
Vol. 57

Dömling, Alexander (Ed.)
Protein-Protein Interactions in Drug Discovery
2013
ISBN: 978-3-527-33107-9
Vol. 56

Kalgutkar, Amit S. / Dalvie, Deepak / Obach, R. Scott / Smith, Dennis A.
Reactive Drug Metabolites
2012
ISBN: 978-3-527-33085-0
Vol. 55

Brown, Nathan (Ed.)
Bioisosteres in Medicinal Chemistry
2012
ISBN: 978-3-527-33015-7
Vol. 54

Gohlke, Holger (Ed.)
Protein-Ligand Interactions
2012
ISBN: 978-3-527-32966-3
Vol. 53

Kappe, C. Oliver / Stadler, Alexander / Dallinger, Doris
Microwaves in Organic and Medicinal Chemistry
Second, Completely Revised and Enlarged Edition
2012
ISBN: 978-3-527-33185-7
Vol. 52

Smith, Dennis A. / Allerton, Charlotte / Kalgutkar, Amit S. / van de Waterbeemd, Han / Walker, Don K.
Pharmacokinetics and Metabolism in Drug Design
Third, Revised and Updated Edition
2012
ISBN: 978-3-527-32954-0
Vol. 51

De Clercq, Erik (Ed.)
Antiviral Drug Strategies
2011
ISBN: 978-3-527-32696-9
Vol. 50

Klebl, Bert / Müller, Gerhard / Hamacher, Michael (Eds.)
Protein Kinases as Drug Targets
2011
ISBN: 978-3-527-31790-5
Vol. 49

Sotriffer, Christoph (Ed.)
Virtual Screening
Principles, Challenges, and Practical Guidelines
2011
ISBN: 978-3-527-32636-5
Vol. 48
Scaffold Hopping in Medicinal Chemistry
Contents

List of Contributors XIII
Preface XVII
A Personal Foreword XIX

Part One Scaffolds: Identification, Representation Diversity, and Navigation 1

1 Identifying and Representing Scaffolds 3
   Nathan Brown
1.1 Introduction 3
1.2 History of Scaffold Representations 4
1.3 Functional versus Structural Molecular Scaffolds 7
1.4 Objective and Invariant Scaffold Representations 7
1.4.1 Molecular Frameworks 7
1.4.2 Scaffold Tree 8
1.5 Maximum Common Substructures 9
1.6 Privileged Scaffolds 11
1.7 Conclusions 11
References 12

2 Markush Structures and Chemical Patents 15
   David Anthony Cosgrove
2.1 Introduction 15
2.2 Encoding Markush Structures 18
2.2.1 The r_group Record 19
2.2.1.1 Exact R Groups 19
2.2.1.2 Inexact R Groups 19
2.2.1.3 Fused R Groups 19
2.2.2 The Menguin Program 20
2.2.3 Correspondence between the MIL File and the Markush Structure 21
2.3 The Search Algorithm 22
2.3.1 Matching R Groups 25
2.3.1.1 Exact R Groups 25
2.3.1.2 Inexact R Groups 27
2.3.1.3 Fused R Groups 28
2.3.1.4 Hydrogen Atoms 29
2.3.1.5 Managing Multiple Fragment/R Group Matches 30
2.4 Using Periscope for Scaffold Hopping 31
2.4.1 Substructure Searching 32
2.4.2 Free–Wilson Analysis 33
2.4.3 Fast Followers 36
2.5 Conclusions 36
References 37

3 Scaffold Diversity in Medicinal Chemistry Space 39
Sarah R. Langdon, Julian Blagg, and Nathan Brown
3.1 Introduction 39
3.1.1 Scaffold Representation 39
3.1.2 What Do We Mean by Scaffold Diversity? 41
3.2 Scaffold Composition of Medicinal Chemistry Space 41
3.2.1 Natural Products as a Source of Novel Medicinal Chemistry Scaffolds 45
3.2.2 Enumerating Potential Medicinal Chemistry Scaffolds 46
3.2.3 Using Scaffold Composition to Interpret Bioactivity Data 48
3.3 Metrics for Quantifying the Scaffold Diversity of Medicinal Chemistry Space 48
3.4 Visualizing the Scaffold Diversity of Medicinal Chemistry Space 53
3.5 Conclusions 56
References 57

4 Scaffold Mining of Publicly Available Compound Data 61
Ye Hu and Jürgen Bajorath
4.1 Introduction 61
4.2 Scaffold Definition 62
4.3 Selectivity of Scaffolds 63
4.3.1 Privileged Substructures 63
4.3.2 Target Community-Selective Scaffolds 64
4.3.3 Target-Selective Scaffolds 67
4.4 Target Promiscuity of Scaffolds 67
4.4.1 Promiscuous BM Scaffolds and CSKs 67
4.4.2 Scaffold–Target Family Profiles 70
4.4.3 Promiscuous Scaffolds in Drugs 70
4.5 Activity Cliff-Forming Scaffolds 71
4.5.1 Activity Cliff Concept 71
4.5.2 Multitarget Cliff-Forming Scaffolds 71
4.6 Scaffolds with Defined Activity Progression 73
4.6.1 Activity Profile Sequences 73
4.6.2 Conserved Scaffolds  75
4.7 Scaffold Diversity of Pharmaceutical Targets  76
4.7.1 Scaffold Hopping Potential  76
4.7.2 Structural Relationships between Scaffolds  76
4.7.3 Scaffold Hopping in Virtual Screening  78
4.8 Conclusions  79
References  80

5  Exploring Virtual Scaffold Spaces  83
William R. Pitt and Boris Kroeplien
5.1 Introduction  83
5.1.1 Virtual Chemistry  83
5.1.2 Chemical Space  83
5.1.3 Scaffold Definition  84
5.2 The Comprehensive Enumeration of Parts of Chemical Space  85
5.2.1 Fragments  85
5.2.2 Ring Systems  86
5.2.3 Reagents  87
5.3 The Iterative Generation of Virtual Compounds  88
5.3.1 Transformations  88
5.3.2 Manual Selection of Chemical Modifications  88
5.3.3 Analog Generators  89
5.3.4 Inverse QSAR  89
5.3.5 Multiple Objective Optimization  90
5.3.6 Structure-Based De Novo Design  90
5.4 Virtual Synthesis  92
5.4.1 Synthetic Tractability  92
5.4.2 Using Real-Life Reactions for in Silico Molecule Construction  93
5.4.3 Readily Synthesizable Compounds  94
5.4.3.1 Construction  94
5.4.3.2 Searching  95
5.4.3.3 Outside Big Pharma  96
5.4.4 Iterative Approaches  96
5.5 Visualizations of Scaffold Space  96
5.6 A Perspective on the Past and the Future  97
References  99

Part Two  Scaffold-Hopping Methods  105

6  Similarity-Based Scaffold Hopping Using 2D Fingerprints  107
Peter Willett
6.1 Fingerprints  107
6.2 Retrospective Studies of Scaffold Hopping Using 2D Fingerprints  109
6.3 Predictive Studies of Scaffold Hopping Using 2D Fingerprints 112
6.4 Conclusions 114
References 115

7 CATS for Scaffold Hopping in Medicinal Chemistry 119
Christian P. Koch, Michael Reutlinger, Nickolay Todoroff,
Petra Schneider, and Gisbert Schneider
7.1 Chemically Advanced Template Search 119
7.2 Retrospective Evaluation of Enrichment and Scaffold Hopping Potential 122
7.3 Prospective Scaffold-Hopping Applications 126
7.4 Conclusions 128
References 128

8 Reduced Graphs 131
Kristian Birchall
8.1 Introduction 131
8.2 Generating Reduced Graphs 133
8.2.1 Reduction Scheme 133
8.2.2 Node Labeling 134
8.2.3 Sheffield Implementations 135
8.2.4 Extended Reduced Graphs 136
8.3 Comparison and Usage of Reduced Graphs 137
8.3.1 Conventional Fingerprinting 138
8.3.2 RG-Specific Fingerprints 139
8.3.3 Augmenting Fingerprints with Edit Distance 140
8.3.4 Extended Reduced Graph Fingerprints 141
8.3.5 Graph Matching Approaches 143
8.3.6 Bioisostere Encoding 144
8.4 Summary 146
References 146

9 Feature Trees 149
Nathan Brown
9.1 Introduction 149
9.2 Feature Tree Generation 149
9.3 Feature Tree Comparison 150
9.4 Retrospective Validation 151
9.5 Implementations and Applications 152
9.5.1 MTree: Combinations of Query Molecules 152
9.5.2 Similarity Searching in Large Combinatorial Chemistry Spaces 152
9.6 Conclusions 153
References 154
10 Feature Point Pharmacophores (FEPOPS) 155  
Jeremy L. Jenkins

10.1 Similarity Searching in Drug Discovery 155
10.2 FEPOPS: An Analogy to Image Compression 157
10.3 Computing FEPOPS 159
10.4 Scaling and Correlations 162
10.5 Defining Scaffold Hopping 163
10.6 FEPOPS in Similarity Searching and Scaffold Hopping 164
10.7 Alternative Alignment 168
10.8 In Silico Target Prediction 170
10.9 Chemical Space Uniqueness 171
10.10 Perspective on FEPOPS’ 10 Year Anniversary 172
References 173

11 Three-Dimensional Scaffold Replacement Methods 175  
Nathan Brown

11.1 Introduction 175
11.2 Generic Three-Dimensional Scaffold Replacement Workflow 175
11.2.1 Molecule Databases 175
11.2.2 Fragment Generation and Filtering 177
11.2.3 Fragment Replacement Search and Scoring 178
11.3 SHOP: Scaffold HOPping by GRID-Based Similarity Searches 178
11.4 ReCore 179
11.5 BROOD 179
11.6 Conclusions 180
References 181

12 Spherical Harmonic Molecular Surfaces (ParaSurf and ParaFit) 183  
David W. Ritchie and Violeta I. Pérez-Nuena

12.1 Introduction 183
12.2 Spherical Harmonic Surfaces 183
12.3 Rotating Spherical Polar Fourier Expansions 185
12.4 Spherical Harmonic Surface Shape Similarity 186
12.5 Calculating Consensus Shapes and Center Molecules 187
12.6 The ParaSurf and ParaFit Programs 188
12.7 Using Consensus Shapes to Probe the CCR5 Extracellular Pocket 190
12.8 Conclusions 192
References 193

13 The XED Force Field and Spark 195  
Martin Slater and Andy Vinter

13.1 Pharmacological Similarity – More than Just Chemical Structure 195
13.2 Improving the Generation of Valid Molecular Fields 199
### Contents

13.3 The eXtended Electron Distribution (XED) Force Field 200
13.4 The XED Force Field Applied to Scaffold Hopping in Spark 202
13.5 How Spark Works 202
13.6 Application of Spark in Drug Discovery Scenarios 206
13.7 P38 Kinase Inhibitor Fragment Growing Using Spark 207
13.7.1 The Beauty of P38 207
13.8 Creating New Molecules 208
13.9 New Potential Inhibitors 210
13.10 The Far-Reaching Consequences of Using Molecular Fields as Measures of Similarity 212

References 213

14 Molecular Interaction Fingerprints 215
*Didier Rognan and Jérémie Desaphy*

14.1 Introduction 215
14.2 Target-Annotated Ligand Fingerprints 215
14.2.1 Interacting Atom/Fragment Fingerprints 216
14.2.2 Protein–Ligand Pharmacophores 217
14.3 Ligand-Annotated Target Fingerprints 217
14.4 True Target–Ligand Fingerprints 220
14.4.1 Association Fingerprints 220
14.4.2 Interaction Pattern Fingerprints 222
14.5 Conclusions 225

References 226

15 SkelGen 231
*Nikolay P. Todorov*

15.1 Introduction 231
15.2 Structure Generation and Optimization 232
15.2.1 Fragments and Fragment Sets 232
15.2.2 Structure Generation 234
15.2.3 Scoring and Optimization 234
15.2.4 Ligand-Based Design 234
15.3 Validation Studies 235
15.3.1 Retrospective Validation Study: CDK2, COX2, ER, MMP-3 235
15.3.2 Estrogen Receptor 235
15.3.3 Histamine H3 Inverse Agonists 236
15.4 Scaffold Hopping Using Fixed Fragments 237
15.5 Scaffold Hopping Using Site Points 238
15.6 Further Considerations for Scaffold Hopping 240
15.6.1 Receptor Flexibility 240
15.6.2 Water Molecules 241
15.6.3 Receptor Specificity 242
15.7 Conclusion 242

References 243
Part Three  Case Studies  245

16  Case Study 1: Scaffold Hopping for T-Type Calcium Channel and Glycine Transporter Type 1 Inhibitors  247
Leah C. Konkol, Timothy J. Senter, and Craig W. Lindsley
16.1 Introduction  247
16.2 T-Type Calcium Channel Inhibitors  248
16.3 Scaffold Hopping to Access Novel Calcium T-Type Channel Inhibitors  250
16.4 Scaffold Hopping to Access Novel Glycine Transporter Type 1 (GlyT1) Inhibitors  253
16.5 Conclusions  255
References  255

17  Case Study 2: Bioisosteric Replacements for the Neurokinin 1 Receptor (NK1R)  259
Francesca Perruccio
17.1 Introduction  259
17.2 Neurokinin 1 (NK1) Therapeutic Areas  259
17.3 The Neurokinin 1 Receptor (NK1R) and Its Mechanism  260
17.4 Neurokinin 1 Antagonists  261
17.5 NK1 Receptor: Target Active Site and Binding Mode  264
17.6 Bioisosteric Replacements in NK1 Receptor Antagonist  266
17.7 Bioisosteric Replacements in NK1 Receptor Antagonist: A Retrospective Study  270
17.8 Summary and Conclusions  274
References  274

18  Case Study 3: Fragment Hopping to Design Highly Potent and Selective Neuronal Nitric Oxide Synthase Inhibitors  279
Haitao Ji and Richard B. Silverman
18.1 Fragment-Based Drug Design  279
18.2 Minimal Pharmacophoric Elements and Fragment Hopping  281
18.3 Fragment Hopping to Design Novel Inhibitors for Neuronal Nitric Oxide Synthase  283
18.4 Fragment Hopping to Optimize Neuronal Nitric Oxide Synthase Inhibitors  288
18.5 Application of Neuronal Nitric Oxide Synthase Inhibitors to the Prevention of Cerebral Palsy  289
References  291

Index  297
List of Contributors

**Jürgen Bajorath**  
Rheinische Friedrich-Wilhelms-Universität  
Department of Life Science Informatics, B-IT  
LIMES Program Unit Chemical Biology and Medicinal Chemistry  
Dahlmannstr. 2  
53113 Bonn  
Germany

**Kristian Birchall**  
MRC Technology  
Centre for Therapeutics Discovery  
1-3 Burtonhole Lane  
London NW7 1AD  
UK

**Julian Blagg**  
The Institute of Cancer Research  
Division of Cancer Therapeutics  
Cancer Research UK Cancer Therapeutics Unit  
15 Cotswold Road  
Sutton, Surrey SM2 5NG  
UK

**Nathan Brown**  
The Institute of Cancer Research  
Division of Cancer Therapeutics  
Cancer Research UK Cancer Therapeutics Unit  
15 Cotswold Road  
Sutton, Surrey SM2 5NG  
UK

**David Anthony Cosgrove**  
AstraZeneca  
Discovery Sciences  
Chemistry Innovation Centre  
Mereside 30S391  
Alderley Park  
Macclesfield SK10 4TG  
UK

**Jérémy Desaphy**  
UMR 7200 CNRS/Université de Strasbourg  
Laboratoire d’Innovation Thérapeutique  
MEDALIS Drug Discovery Center  
74 route de Rhin  
Illkirch 67400  
France
List of Contributors

Ye Hu
Rheinische Friedrich-Wilhelms-Universität
Department of Life Science
Informatics, B-IT
LIMES Program Unit Chemical Biology and Medicinal Chemistry
Dahmannstr. 2
53113 Bonn
Germany

Jeremy L. Jenkins
Novartis Institutes for BioMedical Research
Developmental and Molecular Pathways
220 Massachusetts Avenue
Cambridge, MA 02139
USA

Haitao Ji
University of Utah
Department of Chemistry
Center for Cell and Genome Science
315 South 1400 East
Salt Lake City, UT 84112-0850
USA

Christian P. Koch
Eidgenössische Technische Hochschule (ETH)
Department of Chemistry and Applied Biosciences
Institute of Pharmaceutical Sciences
Wolfgang-Pauli-Str. 10
8093 Zürich
Switzerland

Leah C. Konkol
Vanderbilt University Medical Center
Department of Chemistry
2213, Garland Avenue
Nashville, TN 37232-6600
USA

Boris Kroeplien
UCB Celltech
208 Bath Road
Slough SL1 3WE
UK

Sarah R. Langdon
The Institute of Cancer Research
Division of Cancer Therapeutics
Cancer Research UK Cancer Therapeutics Unit
15 Cotswold Road
Sutton, Surrey SM2 5NG
UK

Craig W. Lindsley
Vanderbilt University Medical Center
Department of Chemistry
2213, Garland Avenue
Nashville, TN 37232-6600
USA

Violeta I. Pérez-Nuño
Harmonic Pharma
615 Rue du Jardin Botanique
54600 Villers-lès-Nancy
France

Francesca Perruccio
Francesca Perruccio
Novartis Pharma AG
Postfach
CH-4002 Basel
Switzerland

William R. Pitt
University of Cambridge
Department of Biochemistry
80 Tennis Court Road
Cambridge CB2 1GA
UK
List of Contributors

Michael Reutlinger
Eidgenössische Technische Hochschule (ETH)
Department of Chemistry and Applied Biosciences
Institute of Pharmaceutical Sciences
Wolfgang-Pauli-Str. 10
8093 Zürich
Switzerland

Petra Schneider
Eidgenössische Technische Hochschule (ETH)
Department of Chemistry and Applied Biosciences
Institute of Pharmaceutical Sciences
Wolfgang-Pauli-Str. 10
8093 Zürich
Switzerland

David W. Ritchie
Inria Nancy – Grand Est
Team Orpailleur
615 Rue du Jardin Botanique
54600 Villers-lès-Nancy
France

Timothy J. Senter
Vanderbilt University Medical Center
Department of Chemistry
2213, Garland Avenue
Nashville, TN 37232-6600
USA

Didier Rognan
UMR 7200 CNRS/Université de Strasbourg
Laboratoire d’Innovation Thérapeutique
MEDALIS Drug Discovery Center
74 route de Rhin
Illkirch 67400
France

Richard B. Silverman
Northwestern University
Department of Chemistry
Chemistry of Life Processes Institute
Center for Molecular Innovation and Drug Discovery
2145 Sheridan Road
Evanston, IL 60208-3113

Gisbert Schneider
Eidgenössische Technische Hochschule (ETH)
Department of Chemistry and Applied Biosciences
Institute of Pharmaceutical Sciences
Wolfgang-Pauli-Str. 10
8093 Zürich
Switzerland

Martin Slater
New Cambridge House
Bassingbourn Road
Litlington
Cambridgeshire
SG8 0SS
UK

Nickolay Todoroff
Eidgenössische Technische Hochschule (ETH)
Department of Chemistry and Applied Biosciences
Institute of Pharmaceutical Sciences
Wolfgang-Pauli-Str. 10
8093 Zürich
Switzerland
List of Contributors

**Nikolay P. Todorov**
Molscape Research Limited
145–157 St John Street
London EC1V 4PW
UK

**Andy Vinter**
New Cambridge House
Bassingbourn Road
Litlington
Cambridgeshire SG8 0SS
UK

**Peter Willett**
University of Sheffield
Information School
211 Portobello Street
Sheffield S1 4DP
UK
Preface

In 1999, Gisbert Schneider coined the term “scaffold hopping” for a systematic approach to modify the molecular skeleton of a lead structure [1]. Whereas in bioisosteric replacement atoms or small groups are substituted by other ones with identical or at least similar stereoelectronic features [2], scaffold hopping exchanges the central part of a molecule by a molecular frame of similar shape and pharmacophoric pattern [3]. Correspondingly, scaffold hopping may be considered as an extension of bioisosteric replacement. In this manner, it provides a conceptual and practical route for generating new chemistry and lead series with higher efficacy, better or modified selectivity, and/or improved pharmacokinetic properties, based on known active principles.

As often in science, this approach is also not completely new. The modification or exchange of a molecular scaffold was already applied in the chemical variation of morphine, quinine, some steroid hormones (e.g., estradiol), and β-blockers, to list only a few examples. Looking at naturally occurring β-lactams, that is, the penicillins, cephalosporins, and monobactams, we see that also nature sometimes uses this principle. Like in “fragment-based design,” where a breakthrough came only after the description of the advantages of this method, the definition “scaffold hopping” appealed medicinal chemists to use this strategy – and fueled its systematic application in lead structure search and optimization. Marketed analogs of celecoxib (Celebrex®), sildenafil (Viagra®), and several kinase inhibitors are recent examples of drugs and clinical candidates resulting from this approach.

The volume is logically organized in three parts. An introductory part deals with the representation, diversity, and navigation aspects of scaffold hopping. The next section is dedicated to topological methods, feature trees, shape-based methods, three-dimensional scaffold replacement methods as well as pharmacophore- and structure-based methods of scaffold hopping. Finally, some case studies demonstrate the value of scaffold hopping in all important target classes, exemplified by the design of ligands of the T-type calcium channel, the glycine transporter type 1, the neurokinin 1 receptor, and nitric oxide synthase.

The series editors highly appreciate that after editing the first monograph Bioisosteres in Medicinal Chemistry, Nathan Brown also undertook the effort to edit this monograph. We are very grateful that he organized this work, cooperating with so many excellent authors. Surely this book adds another fascinating new facet
to our book series on “Methods and Principles in Medicinal Chemistry.” Last but not least, we thank Wiley-VCH, in particular Frank Weinreich and Heike Nöthe, for their valuable contributions to this project and the entire series.

Düsseldorf
Weisenheim am Sand
Zurich
July 2013

Raimund Mannhold
Hugo Kubinyi
Gerd Folkers

References


A Personal Foreword

“…I want to stand as close to the edge as I can without going over. Out on the edge you see all the kinds of things you can’t see from the center.”

*Player Piano (1952)*

*Kurt Vonnegut, Jr.*

The foundation of a medicinal chemistry project is the determination and selection of the molecular scaffolds from which the potential drugs are grown. Therefore, it is essential that this fundamental core element be selected appropriately and with careful consideration. The selection of scaffolds and identification of ideal replacement scaffolds can be greatly assisted by computational analyses.

This book is the first to be dedicated to the analysis of molecular scaffolds in drug discovery and the discussion of the plethora of computational approaches that have been reported in scaffold hopping. Scaffold hopping is a subset of bioisosteric replacement where one tries to replace the core motif of a molecule while retaining important interaction potential, whether functionally or literally the necessary scaffolding for decorating with functional substituents.

There has been much published on what constitutes a molecular scaffold over the last century since the advent of Markush structures. A medicinal chemist tends to know it when they see it, whereas computational scientists apply various algorithms to identify what may be the scaffold of a chemical series or individual molecules. Part One of this book covers fully the many considerations in molecular scaffold identification, representation, diversity, and navigation. These are essential definitions and analyses that are prerequisites for application in scaffold hopping campaigns.

Part Two of this book, and the most substantial, covers a well-established subset of the many different computational methods that have been developed and applied in recent years. These range from ligand-based topological pharmacophores to abstracting three-dimensional structures in a variety of ways, including the use of protein structures.

Finally, and of key importance to the presentation of any approach, the book concludes with three chapters in Part Three in which scaffold hopping techniques and approaches have been applied prospectively in real projects. These case studies
consider scaffold hopping applied to designing ligands in four targets: nitric oxide synthase, the neurokinin 1 receptor, the T-type calcium channel, and the glycine transporter type 1.

I would like to extend my personal thanks to the contributors of all of the chapters in this book who have devoted so much time and effort in producing work that is of the high standard that we have come to expect in this book series “Methods and Principles in Medicinal Chemistry.” I would like to thank the series editors Raimund Mannhold, Hugo Kubinyi, and Gerd Folkers for commissioning to edit this book and also the previous book Bioisosteres in Medicinal Chemistry. Finally, I would like to thank the Wiley-VCH team for helping me pull this book together and making my life as editor a lot simpler in many ways; in particular, I would like to thank Frank Weinreich and Heike Nöthe for their invaluable efforts.

This book has been a labor of love for me and I am delighted that this book has formed so well through the duration of this project. I can only hope that you as the reader get as much out of reading it as I did in editing.

London, 2013

Nathan Brown
Part One
Scaffolds: Identification, Representation Diversity, and Navigation
Identifying and Representing Scaffolds

Nathan Brown

1.1 Introduction

Drug discovery and design is an inherently multiobjective optimization process. Many different properties require optimization to develop a drug that satisfies the key objectives of safety and efficacy. Scaffolds and scaffold hopping, the subject of this book, are an attempt to identify appropriate molecular scaffolds to replace those that have already been identified [1,2]. Scaffold hopping has also been referred to as lead hopping, leapfrogging, chemotype switching, and scaffold searching in the literature [3–6]. Scaffold hopping is an approach to modulating important properties that may contravene what makes a successful drug: safety and efficacy. Therefore, due consideration of alternative scaffolds should be considered throughout a drug discovery program, but it is perhaps more easily explored earlier in the process. Scaffold hopping is a subset of bioisosteric replacement that focuses explicitly on identifying and replacing appropriate central cores that function similarly in some properties while optimizing other properties. While bioisosteric replacement is not considered to a significant degree in this book, a sister volume has recently been published [7], many of the approaches discussed in this book are also applicable to bioisosteric replacement.

Some properties that can be modulated by judicious replacement of scaffolds are binding affinity, lipophilicity, polarity, toxicity, and issues around intellectual property rights. Binding affinity can sometimes be improved by introducing a more rigid scaffold. This is due to the conformation being preorganized for favorable interactions. One example of this was shown recently in a stearoyl-CoA desaturase inhibitor [8]. An increase in lipophilicity can lead to an increase in cellular permeability. The replacement of a benzimidazole scaffold with the more lipophilic indole moiety was recently presented as a scaffold replacement in an inhibitor targeting N5SB polymerase for the treatment against the hepatitis C virus [9]. Conversely, replacing a more lipophilic core with the one that is more polar can improve the solubility of a compound. The same two scaffolds as before were used, but this time the objective was to improve solubility, so the indole was replaced for the benzimidazole [10]. Sometimes, the central core of a lead molecule can have
pathological conditions in toxicity that needs to be addressed to decrease the chances of attrition in drug development. One COX-2 inhibitor series consisted of a central scaffold of diarylimidazothiazole, which can be metabolized to thiophene S-oxide leading to toxic effects. However, this scaffold can be replaced with diarylthiazolotriazole to mitigate such concerns [11,12]. Finally, although not a property of the molecules under consideration per se, it is often important to move away from an identified scaffold that exhibits favorable properties due to the scaffold having already been patented. The definition of Markush structures will be discussed later in this chapter and more extensively in Chapter 2.

Given the different outcomes that lead to what can be called a scaffold hop, one can surmise that there must be different definitions of what constitutes a scaffold hop and indeed the definition of a scaffold itself. This chapter particularly focuses on identifying and representing scaffolds in drug discovery. Markush structures will be introduced as a representation of scaffolds for inclusion in patents to protect intellectual rights around a particular defined core, which will also be discussed in Chapter 2. Objective and invariant representations of scaffolds are essential for diversity analyses of scaffolds and understanding the scaffold coverage and diversity of our screening libraries. Some of the more popular objective and invariant scaffold identification methods will be introduced later in this chapter. The applications of these approaches will be discussed in more detail later in this book, with particular reference to the coverage of scaffolds in medicinal chemistry space.

### 1.2 History of Scaffold Representations

Probably the first description, which is still in common use today, is the Markush structure introduced by Eugene A. Markush from the Pharma-Chemical Corporation in a patent granted in 1924 [13]. Markush defined a generic structure in prose that allowed for his patent to cover an entire family of pyrazolone dye molecules:

I have discovered that the diazo compound of unsulphonated amidobenzol (aniline) or its homologues (such as toluidine, xylidine, etc.) in all their isomeric forms such as their ortho, meta and para compounds, or in their mixtures or halogen substitutes, may be coupled with halogen substituted pyrazolones (such as dichlor-sulpho-phenyl-carboxlic-acid pyrazolone) to produce dyes which are exceptionally fast to light, which will dye wool and silk from an acidulated bath.

More specifically, Markush’s claims were as follows:

1) The process for the manufacture of dyes which comprises coupling with a halogen-substituted pyrazolone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of aniline and halogen substitution products of aniline.
2) The process for the manufacture of dyes which comprises coupling with a halogen-substituted pyrazolone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of aniline and halogen substitution products of aniline.

3) The process for the manufacture of dyes which comprises coupling dichlor-substituted pyrazolone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of aniline and halogen substitution products of aniline.

Interestingly, the careful reader will note that claims 1 and 2 in Markush’s patent are exactly the same. It is not known why this would have been the case, but it may be speculated that it was a simple clerical error with Markush originally intending to make a small change in the second claim as can be seen in the third claim. Therefore, Markush’s patent may not have been as extensive since it is possible one of his claims did not appear in the final patent.

Markush successfully defended his use of generic structure definitions at the US Supreme Court, defining a scaffold together with defined lists of substituents on that scaffold. Extending the chemistry space combinatorially from this simple schema can lead to many compounds being covered by a single patent. However, there remains a burden on the patent holders that although it may not be necessary to synthesize every exemplar from the enumerated set of compounds, each of the compounds must be synthetically feasible to someone skilled in the art. A patent may not be defendable if any of the compounds protected by a Markush claim cannot subsequently be synthesized.

An example of a possible Markush structure for the HSP90 inhibitor, NVP-AUY922 (Figure 1.1a) is given in Figure 1.1b. However, an example of a medicinal chemist may determine as the molecular scaffold is given in Figure 1.1c [14,15].

The Markush claim discussed above is clearly a mechanism for extending the protection of a single patent application to a multitude of related and defined compounds. The earliest reference to what we would now call a molecular scaffold definition that this author could identify was in 1969, in an article published in the *Journal of the American Chemical Society*, which provided the following definition [16]:

> The ring system is highly rigid, and can act as a scaffold for placing functional groups in set geometric relationships to one another for systematic studies of transannular and multiple functional group effects on physical and chemical properties.

Clearly, this is a simple description of what constitutes a molecular scaffold and is readily understandable to a scientist active in medicinal chemistry and a specific example of a structural scaffold. However, its simple definition belies an inherent challenge in the identification of molecular scaffolds. Quite often, a medicinal chemist can identify what they would refer to as a molecular scaffold. This often involves identification of synthetic handles. The challenge here though is to
Figure 1.1 The HSP90 inhibitor NVP-AUY922 depicted using different scaffold representations. (Reproduced from Ref. [20].)
understand how the scaffold has been determined, but this is a soft problem that is not capable of being reduced to an objective and invariant set of rules for scaffold identification. An expert medicinal chemist will bring to bear a wealth of knowledge from their particular research foci during their career and knowledge of synthetic routes: essentially, their intuition. Given a molecule, there are many ways of fragmenting that molecule that may render the key molecular scaffold of interest for the domain of applicability.

1.3 Functional versus Structural Molecular Scaffolds

Scaffolds can be divided roughly into two particular classes: functional and structural. A functional scaffold can be seen as a scaffold that contains the interacting elements with the target. Once defined, medicinal chemistry design strategies can concentrate on further improving potency while also optimizing selectivity and other properties, such as improving solubility. Conversely, a structural scaffold is one that literally provides the scaffolding of exit vectors in the appropriate geometries to allow key interacting moieties to be introduced to decorate the scaffold.

1.4 Objective and Invariant Scaffold Representations

It is important to be able define objective and invariant scaffold representations of molecules not only to permit rapid calculation of the scaffold representations but to also allow comparisons between the scaffolds of different molecules. Much research continues into objective and invariant scaffold representations, but here we summarize some of the methods that have seen significant utility. These scaffold representations use definitions of structural components of molecules: ring systems (Figure 1.1d), linkers (Figure 1.1e), side chains (Figure 1.1f), and the framework that is a connected set of ring systems and linkers (Figure 1.1g).

1.4.1 Molecular Frameworks

One of the first approaches to generating molecular scaffolds from individual molecules was the molecular framework (often referred to as Murcko frameworks) and graph framework representations [17]. Here, each molecule is treated independently; therefore, the method is objective and invariant.

The molecular framework is generated from an individual molecule by pruning all acyclic substructures that do not connect two cyclic systems (Figure 1.1h). The graph framework is a further abstraction in which the atom labels and bond orders are discarded to provide a simple abstraction of the general topology of the
molecule. The molecular (or Murcko) and graph framework representations of NVP-AUY-922 are given in Figure 1.1h and i, respectively.

This work was the first approach to classifying the crude shapes of molecules in terms of their cyclic frameworks. The inclusion of these topological representations and calculations of equivalences were suggested as being ripe for application to the \textit{de novo} design problem. The study also highlighted the lack of scaffold diversity based on these representations in drug-like molecules and concluded that this would be an area of interest for medicinal chemists to understand which frameworks are underrepresented. The framework definitions were also applied to analyze the scaffold diversity in the Chemical Abstracts Service registry of 24 282 284 compounds at the time of publication in 2008 [18]. This application will be discussed more thoroughly in Chapter 3.

1.4.2 Scaffold Tree

Schuffenhauer \textit{et al.} [19] defined the scaffold tree as a set of prioritization rules to systematically prune a given molecule. Starting from the molecular framework defined by Bemis and Murcko [17], rings are sequentially removed using the prioritization rules until only a single ring remains, the so-called level 0 scaffold. The prioritization rules defined for the scaffold tree are provided in Table 1.1.

By application of each of the prioritization rules defined by the scaffold tree method, each molecule in a data set is represented as a directed linear path of iteratively pruned fragments. The scaffold tree pruning strategy is data set independent: a given molecule will always result in the same result. However, the generation of the scaffold tree itself is a summary of a given data set. The pruning path of each molecule in a data set is analyzed and paths merged with one another to generate one or more scaffold trees. For a given data set, one scaffold tree will be the result if all of those molecules in the data set have the same common single

\begin{table}
\centering
\begin{tabular}{ll}
\hline
1 & Remove three-member heterocycles \\
2 & Retain macrocycles of greater than 11 members \\
3 & Remove rings first by longest acyclic linker \\
4 & Retain spiro, nonlinear, fused and bridged rings \\
5 & Retain bridged over spiro rings \\
6 & Remove rings of size 3, 5, and 6 first \\
7 & Fully aromatic rings should not be removed if remaining system is not aromatic \\
8 & Remove rings with fewest heteroatoms first \\
9 & If (8) is equal, use precedence relationship of N > O > S \\
10 & Remove smaller rings first \\
11 & Retain saturated rings \\
12 & Remove rings with a heteroatom connected to a linker \\
13 & Tiebreaking rule based on alphabetic ordering of a canonical SMILES representation \\
\hline
\end{tabular}
\caption{The prioritization rules defined to prune ring systems in the generation of the scaffold tree.}
\end{table}