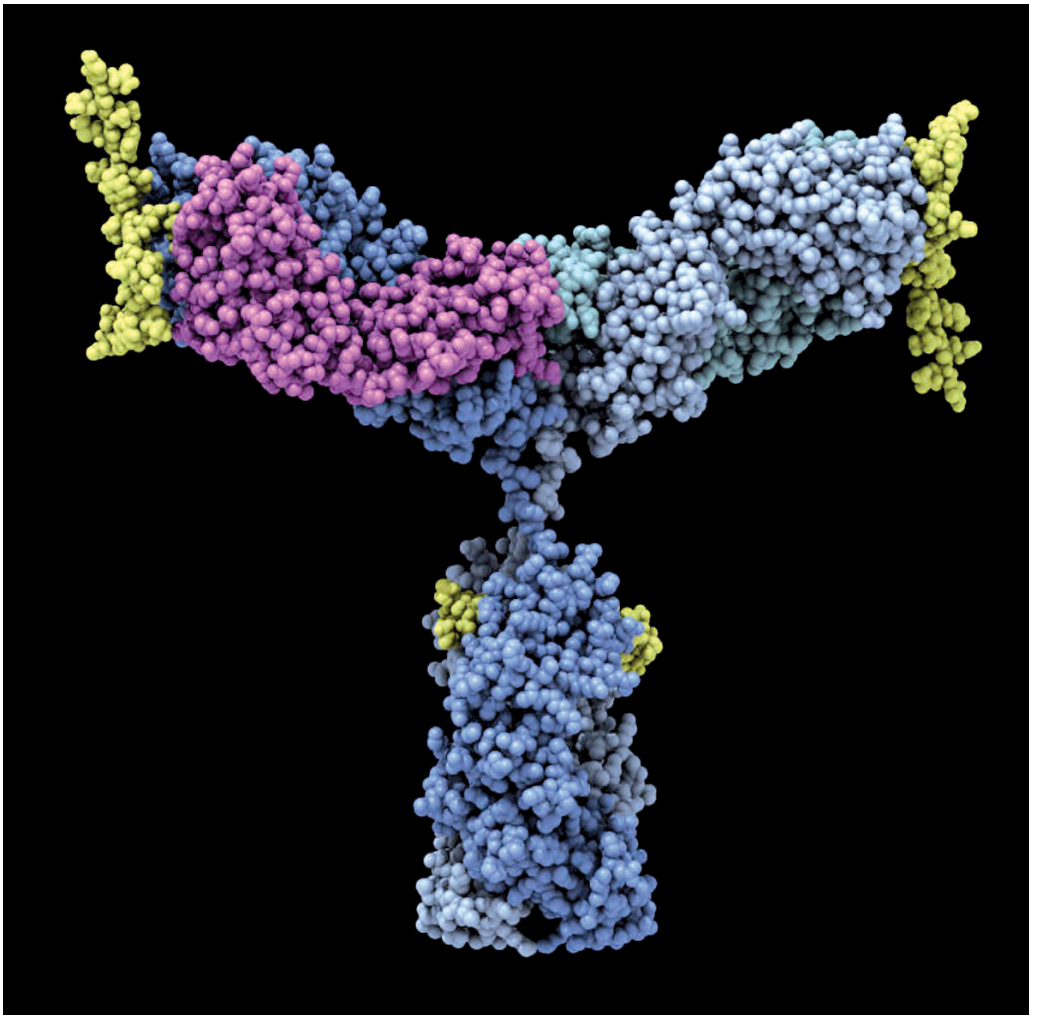
RSC Drug Discovery

Edited by Mark E Bunnage

New Frontiers in Chemical Biology

Enabling Drug Discovery



RSC Publishing

New Frontiers in Chemical Biology
Enabling Drug Discovery

RSC Drug Discovery Series

Titles in the Series:
1: Metabolism, Pharmacokinetics and Toxicity of Functional Groups: Impact of Chemical Building Blocks on ADMET
2: Emerging Drugs and Targets for Alzheimer's Disease; Volume 1: Beta-Amyloid, Tau Protein and Glucose Metabolism
3: Emerging Drugs and Targets for Alzheimer's Disease; Volume 2: Neuronal Plasticity, Neuronal Protection and Other Miscellaneous Strategies
4: Accounts in Drug Discovery: Case Studies in Medicinal Chemistry
5: New Frontiers in Chemical Biology: Enabling Drug Discovery

How to obtain future titles on publication:
A standing order plan is available for this series. A standing order will bring delivery of each new volume immediately on publication.

For further information please contact:
Book Sales Department, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK
Telephone: +44 (0)1223 420066, Fax: +44 (0)1223 420247, Email: books@rsc.org
Visit our website at http://www.rsc.org/Shop/Books/

New Frontiers in Chemical Biology
Enabling Drug Discovery

Edited by

Mark E. Bunnage
Pfizer Global Research and Development, Sandwich, Kent, UK

RSC Publishing

RSC Drug Discovery Series No. 5

ISBN: 978-1-84973-125-6
ISSN: 2041-3203

A catalogue record for this book is available from the British Library

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our website at www.rsc.org

Preface

"Few scientists acquainted with the chemistry of biological systems at the molecular level can avoid being inspired."

Donald J. Cram

Despite ever-increasing investment in biomedical research, there has been a significant decline in the number of new drug approvals in recent years. In an effort to improve productivity, and expand accessible drug space, the drug discovery community has increasingly looked to advance alternative therapeutic paradigms, such as biotherapeutics (antibodies, vaccines, nucleic acids, peptides, *etc*.), in addition to more classical small molecule medicinal chemistry approaches. However, independent of chosen therapeutic modality, improving future drug discovery productivity also critically depends on reducing the high levels of attrition currently seen in "proof-of-concept" Phase II clinical trials. This will require the development of a much deeper knowledge of biological systems in order to identify and validate those biomolecular targets for which there is the highest possible confidence of disease-relevance in humans.

Chemical biology is an emerging field at the interface between chemistry and biology that utilises the tools and techniques of chemical synthesis to study and influence biological systems. Recent developments in chemical biology have great potential to help address the drug discovery productivity challenges expressed above. For example, chemical biology studies have already led to the identification of novel targets with exciting therapeutic potential and it is clear that the field will prove a key enabler of improved target discovery and validation in the future. Moreover, the precise synthetic manipulation of biological molecules germane to many chemical biology approaches is also now fuelling a new wave of chemically modified biologics with unique properties ("chemologics") that emerge from the union of medicinal chemistry design with biotherapeutics research. In these, and many other ways, chemical biology is

RSC Drug Discovery Series No. 5
New Frontiers in Chemical Biology: Enabling Drug Discovery
Edited by Mark E. Bunnage

Published by the Royal Society of Chemistry, www.rsc.org

beginning to emerge as a key discipline within 21st century drug discovery and the purpose of this book is to highlight a number of the important developments in this regard. Since the field of chemical biology is both vast and growing, it should be stressed that this book is not intended to be exhaustive. Instead, the combination of general overviews in this book, coupled with the more in depth treatments of selected topics, is intended as an introduction to the field for the reader.

I would like to thank all of the contributors to the chapters in this book for their outstanding efforts and support of this project. I would also like to thank the staff at RSC, especially Gwen Jones and Katrina Harding, for their patience and support in bringing the book to completion. It is very much hoped that this book will provide a useful resource for scientists, both in industry and academia, who are looking to build their awareness of chemical biology and how this exciting field of science may enable future drug discovery.

Mark E. Bunnage
Executive Director
Worldwide Medicinal Chemistry
Pfizer Inc., Sandwich, Kent, UK

Contents

Chapter 1 The Chemical Genetic Approach: The Interrogation of Biological Mechanisms with Small Molecule Probes **1**
Martin Fisher and Adam Nelson

1.1 Introduction to Chemical Genetics 1
1.1.1 Forward and Reverse Chemical Genetics 2
1.1.2 Screening Small Molecule Libraries: Some General Considerations 4
1.2 Synthetic Strategies for Exploring Biologically Relevant Chemical Space 4
1.2.1 Synthesis of Compound Libraries with High Substitutional Diversity 6
1.2.2 Synthesis of Compound Libraries with High Stereochemical Diversity 7
1.2.3 Synthesis of Compound Libraries with High Scaffold Diversity 7
1.2.4 Emerging Approaches for Ligand Discovery Using Large Small Molecule Libraries 14
1.3 Discovery of Small Molecule Probes 16
1.3.1 Discovery of Small Molecule Probes Using a Reverse Chemical Genetic Approach 16
1.3.2 The Forward Chemical Genetic Approach 20
1.4 Case Study: Interrogation of the Sonic Hedgehog Pathway Using Complementary Small Molecule Probes 25
1.5 Summary and Outlook 27
References 28

RSC Drug Discovery Series No. 5
New Frontiers in Chemical Biology: Enabling Drug Discovery
Edited by Mark E. Bunnage

Published by the Royal Society of Chemistry, www.rsc.org

Chapter 2 Applications for Activity-based Probes in Drug Discovery 33
L.E. Edgington and M. Bogyo

2.1 Background 33
2.1.1 Introduction 33
2.1.2 Activity-based Probes 34
2.2 Applications of Activity-based Probes to Drug Discovery 48
2.2.1 Identification and Validation of Drug Targets Using Activity-based Probes 48
2.2.2 Use of Activity-based Probes for Drug Lead Identification and Assessment of Selectivity 51
2.2.3 Use of Activity-based Probes for Assessment of *In vivo* Pharmacodynamic Properties and Efficacy of Lead Compounds 52
2.2.4 *In vivo* Imaging Applications 53
2.3 Outlook 57
Acknowledgements 58
References 58

Chapter 3 Targeted Intracellular Protein Degradation as a Potential Therapeutic Strategy 64
A.R. Schneekloth and C.M. Crews

3.1 Background 64
3.1.1 Introduction 64
3.1.2 The Ubiquitin–Proteasome System 65
3.1.3 Modulation of Protein Levels in Cells 69
3.2 Development of the PROTACs Approach 71
3.2.1 Initial Proof-of-Concept 71
3.2.2 Targeting Tumor-relevant Proteins *In vitro* and *In vivo* 74
3.2.3 Development of Cell Permeable PROTACs 77
3.2.4 Development of an All-small Molecule PROTAC 83
3.2.5 Further Exploration of the PROTAC Technique 86
3.3 Outlook 90
3.3.1 Strengths of the PROTAC Technique 90
3.3.2 Potential Applications of the PROTAC Technique 92
3.3.3 Continued Development of the PROTAC Technique 93
Acknowledgement 94
References 94

Chapter 4 Chemical Biology of Stem Cell Modulation **97**
Stephen G. Davies and Angela J. Russell

4.1 Introduction 97
4.2 Stem Cells: Ontogeny, Characterisation and Clinical Utility 99
4.3 Stem Cell Manipulation *In Vitro* 102
4.3.1 Small Molecules to Promote Stem Cell Expansion 102
4.3.2 Small Molecules to Direct Stem Cell Differentiation 113
4.3.3 Small Molecules to Promote Dedifferentiation and Transdifferentiation 126
4.4 Stem Cell Manipulation *In Vivo* 132
4.4.1 Exogenous Stem Cells: Homing, Engraftment and Activation 134
4.4.2 Endogenous Stem Cell Manipulation 134
4.4.3 Stem Cells in Dysplasia: Cancer and Teratogenicity 142
4.5 Conclusions and Future Prospects 143
References 143

Chapter 5 Chemical Biology of Histone Modifications **151**
Nathan R. Rose, Christopher J. Schofield and Tom D. Heightman

5.1 Histones, Epigenetics and Chromatin 151
5.1.1 DNA Modifications 154
5.1.2 Euchromatin and Heterochromatin 154
5.1.3 Correlation of Histone Modifications with Transcriptional States 155
5.1.4 Mechanisms by Which Histone Modifications Achieve Transcriptional Regulation 156
5.1.5 Epigenetics 158
5.2 The Lysine Acetylation System 159
5.2.1 Histone Acetyltransferases 160
5.2.2 Histone Deacetylases 164
5.2.3 Bromodomains 169
5.3 The Lysine and Arginine Methylation System 173
5.3.1 Histone Methyltransferases 174
5.3.2 Histone Demethylases 179
5.3.3 Methyl Binding Domains 185
5.4 Serine/Threonine/Tyrosine Phosphorylation of Histones 190
5.5 Lysine Ubiquitylation/SUMOylation/Biotinylation of Histones 191

5.6 Appendix: Abbreviations Used in this Chapter 192
Acknowledgements 192
References 193

Chapter 6 Chemologics 204
Lyn H. Jones

6.1 Introduction 204
6.2 Synthetic Vaccines 204
6.2.1 Carbohydrate-based Vaccines 206
6.2.2 Purely Synthetic Vaccines 208
6.2.3 Synthetic Virus-like Particles 209
6.2.4 Expanding the Genetic Code and Breaking Self-tolerance 209
6.2.5 Chemically Programmed Vaccination 210
6.3 Antibodies 210
6.3.1 Antibody-directed Therapies 210
6.3.2 Improved Pharmacokinetics 213
6.4 Oligonucleotides 213
6.4.1 Aptamers – Macugen 213
6.4.2 siRNA 214
6.4.3 Locked Nucleic Acids 218
6.5 Conclusions 219
References and Notes 220

Chapter 7 Antibody–Drug Conjugates in Oncology 224
Philip R. Hamann and Russell G. Dushin

7.1 Introduction and General Considerations 224
7.2 Calicheamicin Conjugates 229
7.3 Maytansine Conjugates 232
7.4 Auristatin Conjugates 238
7.5 Duocarmycin Conjugates 244
7.6 Miscellaneous Conjugates 246
7.7 Conclusions 247
References 248

Chapter 8 Drug Discovery by DNA-encoded Libraries 258
Yizhou Li, Zheng Zhu and Xiaoyu Li

8.1 Introduction 258
8.1.1 The Lack of Discovery Tools Interrogating Undrugged Targets 258
8.1.2 Applying Nature's Strategy in Discovering Functional Molecules 259
8.1.3 Topics in this Chapter 261

8.2 DNA-templated library 261
8.2.1 Basic Principles of DNA-templated Organic Synthesis 261
8.2.2 The Development of DNA-templated Libraries 268
8.2.3 The Application of DNA-templated Libraries in Drug Discovery 277
8.3 DNA-recorded Library 289
8.3.1 GSK/Praecis 289
8.3.2 Neri Group/Philochem 291
8.3.3 A DNA-recorded Library by DNA Assembly: ESAC Library 292
8.4 Discussion 295
8.4.1 The Advantages of DNA-encoded Library 296
8.4.2 Prospects and Outlook 297
Acknowledgements 298
References 298

Subject Index **303**

CHAPTER 1

The Chemical Genetic Approach: The Interrogation of Biological Mechanisms with Small Molecule Probes

MARTIN FISHER[a,b] AND ADAM NELSON[a,b,*]

[a] School of Chemistry, University of Leeds, Leeds, UK, LS2 9JT; [b] Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK, LS2 9JT

1.1 Introduction to Chemical Genetics

The chemical genetic approach[1] exploits small molecule probes to modulate the function of biological macromolecules, and hence to reveal insights into biological mechanisms. The term ''chemical genetics'' stems from the analogy of the approach to classical genetics. Chemical genetics exploits small molecule probes to modulate protein function; in contrast, classical genetics modulates protein function indirectly through mutation of the corresponding gene. The features of the chemical genetics and classical genetics are rather different, making the two approaches broadly complementary.

The use of small molecule probes can confer a number of advantages over conventional genetic manipulations. First, the effects of small molecules on the function of a target protein are conditional and, in addition, are usually rapid (usually diffusion-controlled) and reversible.[2] Second, by varying the

RSC Drug Discovery Series No. 5
New Frontiers in Chemical Biology: Enabling Drug Discovery
Edited by Mark E. Bunnage

Published by the Royal Society of Chemistry, www.rsc.org

concentration of a small molecule probe, it may be possible to tune the activity of a target protein in a refined way.[3] Third, the chemical genetic approach can sometimes allow multiple functions of an individual protein to be independently modulated.[4,5] Fourth, by using more than one small molecular probe in combination, the function of multiple proteins may be independently modulated.[6] Finally, the approach is useful when classical genetics is difficult to employ: for example, in organisms (especially mammals) with diploid genomes and a slow reproductive rate.[7] The chemical genetic approach is particularly useful for investigating tightly coordinated, dynamic biological mechanisms, especially when classical genetics may render an organism inviable.[8]

One of the obvious features of classical genetics is its incredible specificity: mutation of a specific gene can allow a single protein (from a family of closely related proteins) to be targeted. In addition, classical genetics is highly general, and can be used to study the function of many proteins. However, classical genetics does not generally allow the *conditional* modulation of a specific protein (unless, for example, a temperature-sensitive[9] mutant is used). RNA interference (RNAi) is a popular alternative to classical genetics which works by inhibiting the synthesis of a specific protein by targeting a specific RNA transcript.[10] However, RNAi is also poorly suited to time-sensitive studies because the modulation is made at the mRNA level: degradation of the protein target needs to occur before the effects of modulation are observed.

The chapter is not an exhaustive review of chemical genetics: the field, and its impact on our understanding of fundamental biological mechanisms, has already been comprehensively reviewed elsewhere.[1] Rather, the chapter will emphasise the approaches that may be used to discover small molecule probes, the solutions that address the challenges raised by chemical genetics, and the insights into fundamental biological mechanisms that may be revealed.

1.1.1 Forward and Reverse Chemical Genetics

The chemical genetic approach comes in two broad guises that are contrasted in Figure 1.1: forward chemical genetics and reverse chemical genetics.[1] In the forward chemical genetic approach, the target of the small molecule modulator is not known at the outset of the investigation. The active ligand is, therefore, identified on the basis of the observation of a specific phenotypic change that it induces. A major challenge of the forward chemical genetic approach is the subsequent identification of the protein that the active ligand targets.[11] The interrogation of biological mechanisms using the forward chemical genetic approach (and the associated challenges associated with target identification[11]) are described in Section 1.3.2. The forward chemical genetic and the forward genetic approaches are broadly analogous in that a phenotypic change to a biological system is induced in both cases: through modulation using either small molecule probes (sometimes known as "perturbogens"[12]) (forward chemical genetics) or random mutagenesis (forward genetics).

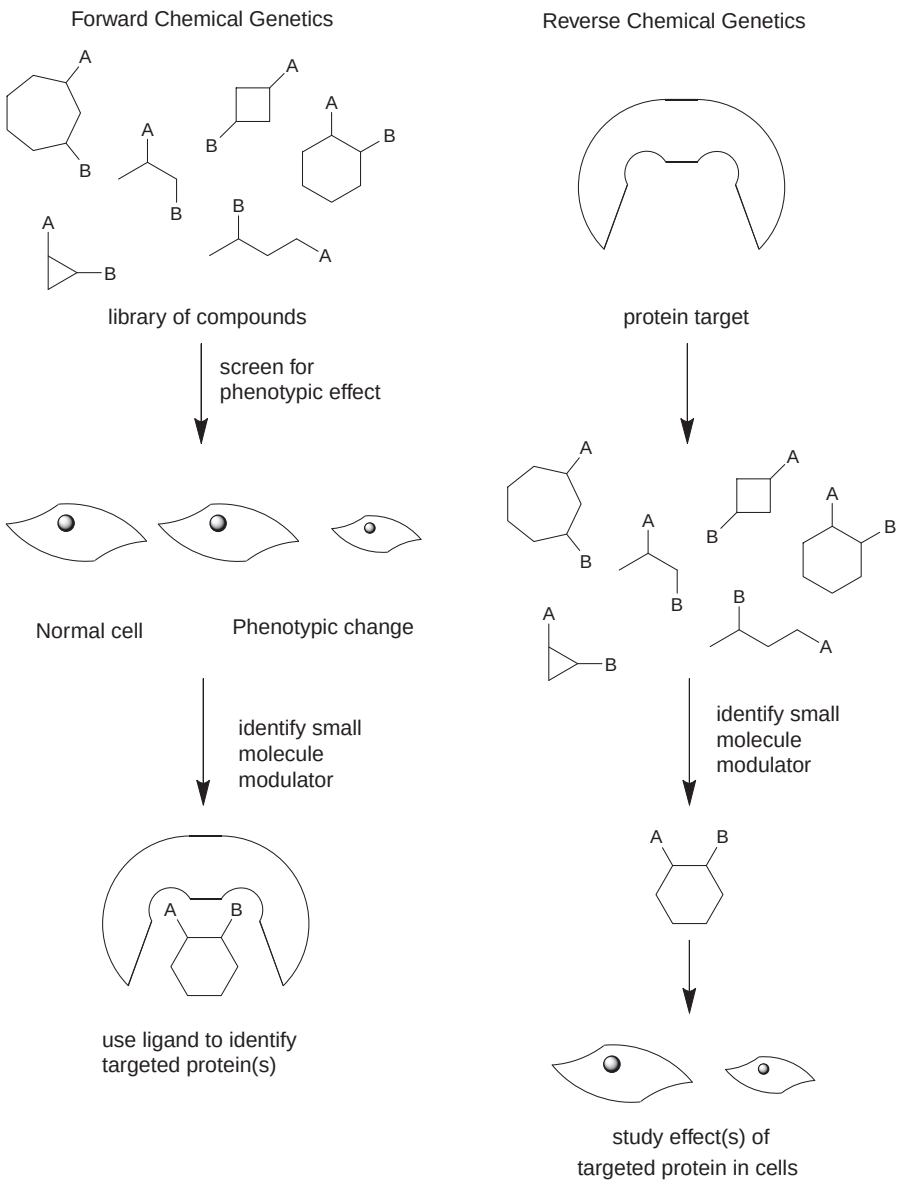


Figure 1.1 Schematic overview of forward and reverse chemical genetics. In forward chemical genetics (left panel), a small molecule probe is discovered on the basis of its phenotypic effect; the small molecule probe may then be subsequently used to identify the protein responsible for the phenotypic effect. In contrast, in reverse chemical genetics (right panel), a small molecule probe is discovered on the basis of its ability to modulate a specific protein; the small molecule probe may then be used to modulate the cellular function of that specific protein, and the phenotypic effects that result.

In contrast, the reverse chemical genetic approach begins with the discovery of a small molecule modulator of a specific target protein. Most usually, the small molecule is identified on the basis of the modulation of the activity of the purified protein; however, assays of the cellular activity of a specific protein are also possible. Both reverse chemical genetics and reverse genetics modulate the function of a specific protein: either directly by the binding of a ligand (reverse chemical genetics), or indirectly through targeted mutagenesis of the corresponding gene. The discovery of ligands for use in reverse chemical genetics, and some of the insights into biological mechanism that have thereby been revealed, are described in Section 1.3.1.

1.1.2 Screening Small Molecule Libraries: Some General Considerations

The forward and reverse chemical genetic approaches both require the identification of an appropriate small molecule modulator. Most usually, active ligands are initially discovered using a high-throughput primary screen. In such cases, it is, of course, imperative that active molecules can be reliably distinguished from inactive molecules, and the performance of an assay must be assessed using appropriate statistical parameters (such as plate-based Z factors).[13] In addition, it is necessary to correct screening results for systematic error by normalisation to controls (for example, to controls on each screening plate).[14] Although guidelines for reporting data from high-throughput screening of small molecule libraries have been suggested, appropriate standards have not yet been formalised.[15]

Having identified a "positive" in a primary screen, further investigation is necessary to confirm its validity.[15,16] Greater confidence is gained if an active compound is related structurally to other active compounds identified in the assay (allowing preliminary structure-activity relationships to be formulated). It is important to confirm that the active compound did not interfere with the primary screen: possible problems can include inhibition of a coupling enzyme in an assay (rather than the target protein), or the interference of a fluorescent compound with a fluorescence-based readout. The chemical structure of the active compound must be verified, and then, ideally, resynthesised, repurified and retested. Ideally, the activity of the compound should be confirmed using one or more independent assays.

1.2 Synthetic Strategies for Exploring Biologically Relevant Chemical Space

A challenge in chemical biology is to design small molecule libraries that span large tracts of biologically relevant chemical space.[17] The chemical space defined by small molecules is vast: it has been estimated that there are $>10^{60}$

compounds with molecular weight less than 500 Da (only a tiny fraction of which have been prepared by chemical synthesis).[18] How then, can the biologically relevant sub-fractions of chemical space be identified and targeted?

One approach is to design libraries around known scaffolds that have been biologically validated such as those found in drug molecules[19] or natural products.[20] In biology-oriented synthesis (BIOS),[20] for example, natural product scaffolds are regarded as pre-validated starting points for ligand design since they have been selected through evolution to interact with protein binding sites. In a related approach, libraries may be designed around scaffolds that are closely related to those that are already known to be biologically relevant.[21]

Alternatively, a fragment-based approach may be used in ligand design.[22] Essentially, a fragment-based approach dramatically reduces the chemical space that needs to be searched. Emphasis is placed on ligands with high "ligand efficiency" (binding energy per heavy atom) rather than high absolute affinity. The approach allows a much higher proportion of the relevant chemical space to be explored, whilst leaving ample scope (and providing a better starting point) for ligand optimisation.

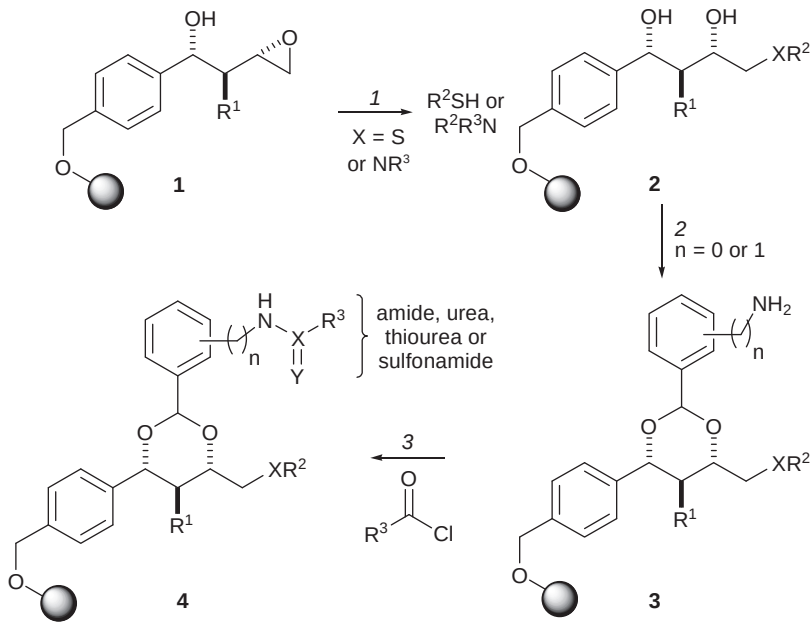
The diversity of chemical libraries may derive from the high substitutional, stereochemical and scaffold diversity of its members. Combinatorial chemistry is focused on the synthesis of chemical libraries with only high substitutional diversity. In general, combinatorial syntheses exploit a rather limited palette of chemical transformations to vary the substitution of library members. The synthesis of compound libraries with high substitutional diversity is described in Section 1.2.1.

Varying the configuration of library members is more challenging, but has been hugely facilitated by the many reliable asymmetric transformations that have been developed in the past 30 years. In many cases, the high enantioselectivity of these transformations can dominate over substrate-based diastereoselectivity, and allow the configuration of compounds to be varied systematically. The synthesis of a compound library with both high substitutional and stereochemical diversity is described in Section 1.2.2.

Charting chemical space systematically, however, does require the preparation of libraries with high scaffold diversity. Historically, chemists' exploration of chemical space has been highly uneven and unsystematic, and much emphasis has been placed on a small number of molecular scaffolds.[23] (Note that the $\sim$30 million cyclic organic compounds in the CAS registry are overwhelmingly dominated by a remarkably small number of molecular scaffolds:[1] thirty (of the 2.5 million!) molecular scaffolds are found in 17% of the compounds, and 0.25% of the scaffolds are found in 50% of the compounds.) A key goal of diversity-oriented synthesis[24] (DOS) is to address this problem through the preparation of libraries of small molecules that populate broad tracts of chemical space. In Section 1.2.3, the progress that has been made towards developing synthetic strategies that allow the systematic variation of scaffolds of ligands will be described.

1.2.1 Synthesis of Compound Libraries with High Substitutional Diversity

The combinatorial chemistry approach has been extended to the diversification of complex scaffolds, including natural product[25] and natural product-like[26] scaffolds. For example, a library of conformationally restricted 1,3-dioxanes **4** with high substitutional diversity has been prepared (Scheme 1.1).[26] A split-pool approach was used to diversify three substituents in the final compounds **4**. The supported epoxides **1** were opened with either a thiol or an amine nucleophile to yield 1,3-diols **2**, which were converted into Fmoc-substituted 1,3-dioxanes. Some of the nucleophiles used in the first step were hydroxy-substituted, leading to the formation of mixed acetals in the second step: these acyclic acetals were, therefore, hydrolysed by treatment of the resins with PPTS in THF-MeOH. Following removal of the Fmoc group (→ **3**), the resulting free amines were converted into amides, ureas, thioureas or sulfonamides (→ **4**), and the final compounds cleaved from the beads. Using material cleaved from a single bead, and with knowledge of the substituent added into the final step, it was possible to use mass spectrometry to identify possible combinations of substituents added in the first two diversification steps. The library has been exploited in the discovery of small molecule tools: uretupamine



Scheme 1.1 Schreiber's split–pool synthesis of 1,3-dioxanes with high substitutional diversity. *Reagents and conditions*: (1) R^2SH or R^2R^3N; (2) (i) $ArCH(OMe)_2$, Me_3SiCl, HCl, dioxane; (ii) PPTS, THF–MeOH; (iii) piperidine; (iv) Me_3SiCl; (3) amide, urea, thiourea or sulfonamide formation.

(see Section 1.3.1.1),[27a,b] which targets Ure2p, a repressor of metabolic genes, and tubacin,[27c,d] a class II histone deacetylase inhibitor. In addition, derivatisation of the scaffold of the alkaloid, galanthamine, yielded secramine, a potent inhibitor of protein trafficking from the Golgi apparatus to the plasma membrane;[25] crucially, although secramine's structure was inspired by galanthamine, it had an entirely distinct biological function.

1.2.2 Synthesis of Compound Libraries with High Stereochemical Diversity

The synthesis of stereochemically diverse compound libraries has been facilitated by the development of a wide range of reliable asymmetric transformations. For example, an asymmetric hetero-Diels-Alder reaction was used to prepare a library of dihydropyrancarboxamides in which both substitution and configuration was varied (structures **7** and their enantiomers) (Scheme 1.2).[28] The library of 4320 structures was encoded using a binary encoding protocol which involved the attachment of tags to individual macrobeads. The asymmetric step (**5** → **6**) served a number of purposes: it generated the molecular scaffolds; it introduced some stereochemical diversity; and it incorporated the variable substituent, R^1. Deprotection of the ester, amide formation, and release from solid support gave the final compounds. The library enabled the discovery of haptamide A, a ligand which targeted a subunit of the transcription factor, Hap3p.[28]

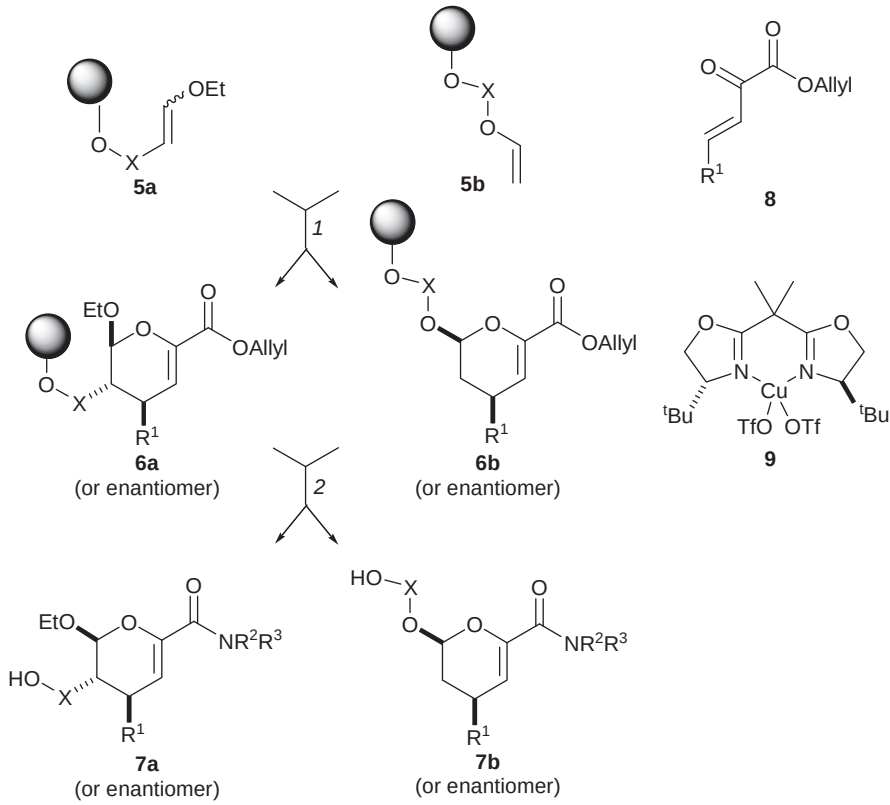
1.2.3 Synthesis of Compound Libraries with High Scaffold Diversity

The development of robust and reliable synthetic methods that yield a range of diverse small molecule scaffolds has proved extremely demanding. In this section, we describe the progress that has been made towards developing synthetic strategies that allow the scaffolds of small molecules to be systematically varied: in particular, the development of "folding" pathways (Section 1.2.3.1), "branching" pathways (Section 1.2.3.2) and oligomer-based approaches (Section 1.2.3.3).

1.2.3.1 Use of "Folding Pathways" to Introduce Scaffold Diversity

The "folding pathway" approach uses common reaction conditions to transform a range of substrates into products with alternative molecular scaffolds. The substrates are encoded to "fold" into the alternative scaffolds through strategically placed appending groups (sometimes known as σ-elements[29]). This strategy leads to scaffold diversification under substrate control.

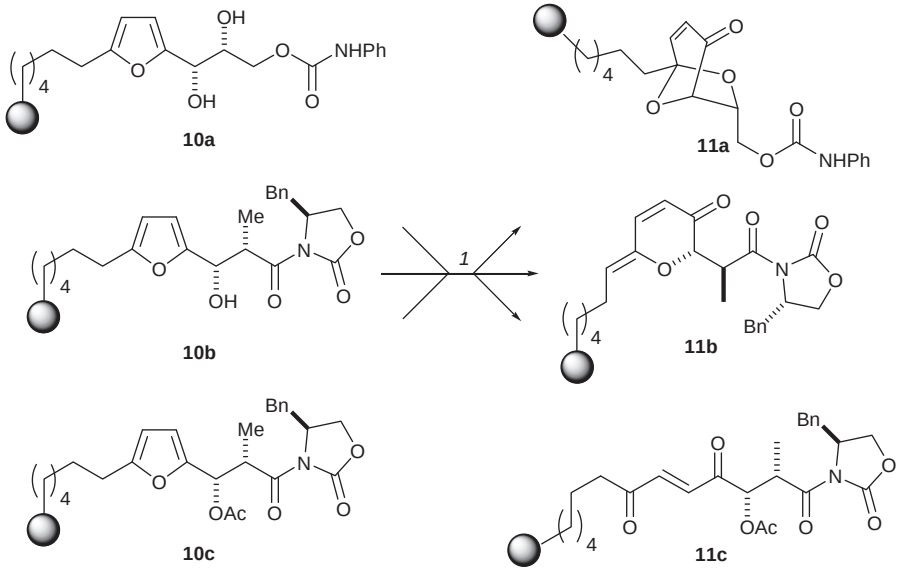
Schreiber has developed a folding pathway which exploits the Achmatowicz reaction (Scheme 1.3).[29] The fate of oxidation of the furan substrates **10**



Scheme 1.2 Schreiber's synthesis of dihydropyrans with high substitutional and stereochemical diversity. *Reagents and conditions*: (1) **8**, 20 mol% **9** (or enantiomer), THF, rt; (2) (i) $Pd(PPh_3)_4$, thiosalicylic acid, THF, rt; (ii) R^2R^3NH, PyBOP, iPr_2NEt, CH_2Cl_2–DMF, rt; (iii) HF · pyridine, THF, rt then Me_3SiOMe.

depends on the functionalisation of groups elsewhere in the molecule (the σ-elements). Hence, oxidation of the furan **10c**, which does not bear any free hydroxy groups, simply generated the ene-dione **11c**. With suitably positioned nucleophilic groups in the starting material, however, a *cis*-enedione intermediate could be intercepted. Hence, upon oxidation, the furyl alcohol **10b** folded, and eliminated water, to yield the alkylidene-pyran-3-one **11b**. In contrast, with two free hydroxyl groups, **10a** folded to yield the bicyclic ketal **11a**. The combination of solid phase chemistry and common reaction conditions made the strategy amenable to split-pool synthesis, which significantly increased the efficiency of library generation.

Schreiber used substrate-based control to create alternative indole alkaloid-like scaffolds (Scheme 1.4).[30] An α-diazo-β-keto-carbonyl group and an indole ring were strategically placed at different positions on the scaffolds of the starting materials (**12**). The densely functionalised starting materials were



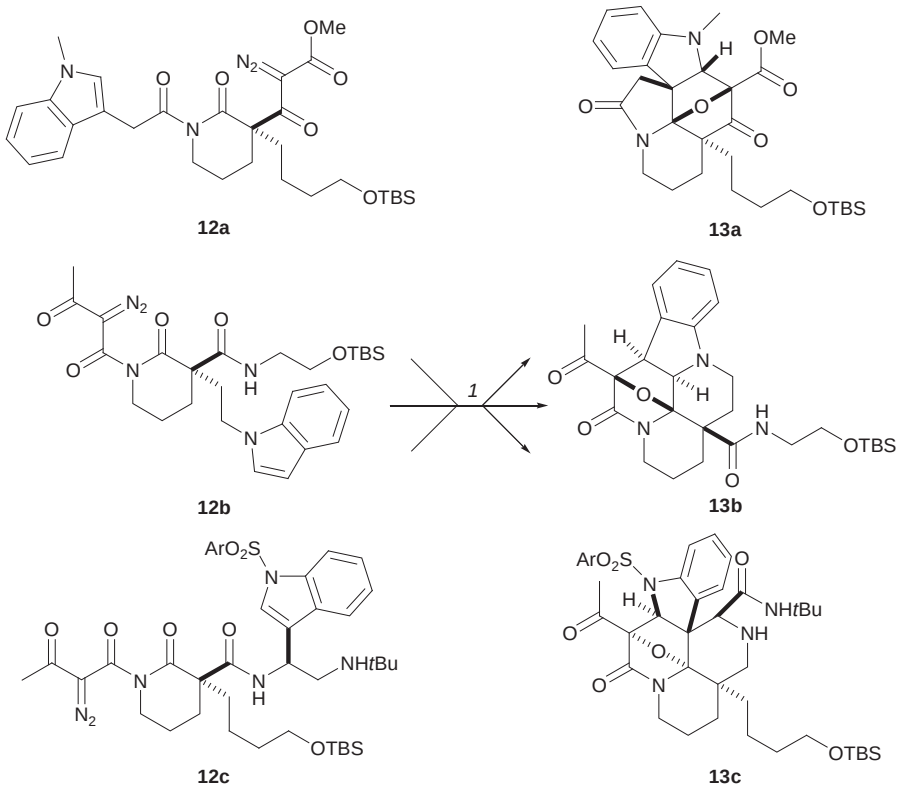
Scheme 1.3 Schreiber's approach to pre-encoding skeletal information using the Achmatowicz reaction. *Reagents and conditions*: (1) *N*-bromosuccinimide, $NaHCO_3$, NaOAc, THF–H_2O (4:1), rt, 1 h; PPTS, CH_2Cl_2, 40–45 °C, 20 h, **11a**: 33% (64%); **11b**: 35% (86%); **11c**: 81% (>90%). Numbers in parentheses are purities determined by LC-MS.

assembled either using an Ugi four-component coupling reaction or by alkylation of a common scaffold; subsequent conversion into a diazo compound yielded the cyclisation precursors. The substrates, **12**, were treated with a catalytic amount of rhodium(II) octanoate dimer in benzene (80 °C). Presumably, generation of a carbonyl ylid was followed by intramolecular 1,3-dipolar cycloaddition with the indole ring to give the alternative polycyclic scaffolds, **13**.

The "folding" pathway strategy has been shown to be useful in the preparation of small molecule libraries based on a few diverse scaffolds. In addition to the examples highlighted here, radical chemistry has been harnessed to "fold" precursors into a range of fused and bridged amine scaffolds.[31] The scope of the "folding" pathway strategy has, however, been enormously expanded by using an oligomer-based approach to prepare precursors for a "folding" step (see Section 1.2.3.3).

1.2.3.2 Use of "Branching Pathways" to Introduce Scaffold Diversity

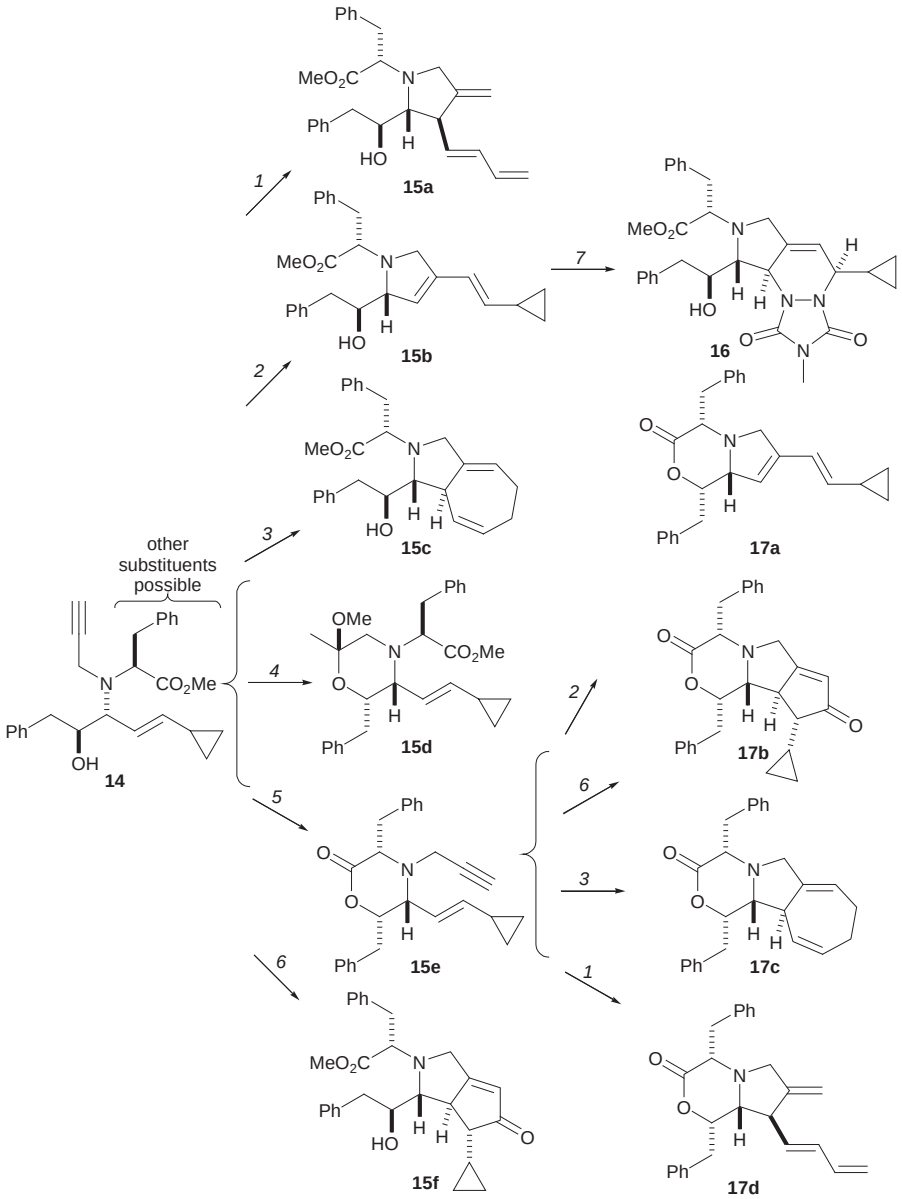
The "branching pathway" strategy involves the conversion of common precursors into a range of distinct molecular scaffolds through careful choice of the



Scheme 1.4 Schreiber's approach to pre-encoding skeletal information via a rhodium-catalysed tandem cyclisation–cycloaddition strategy. *Reagents and conditions*: (1) $Rh_2(O_2CC_7H_{15})_4$, benzene, 50 °C; **13a**: 74%; **13b**: 73%; **13c**: 57%.

reaction conditions. Ideally, a range of flexible precursors would be designed which could participate in complementary reactions leading to alternative scaffolds.

An outstanding example of a "branching pathway" exploited complementary cyclisation reactions to yield alternative molecular scaffolds (Scheme 1.5).[32] A four-component Petasis condensation reaction was used to assemble flexible cyclisation precursors (*e.g.* **14**). Alternative cyclisation reactions were then used to yield products with distinct molecular scaffolds: Pd-catalysed cyclisation (→ **15a**); enyne metathesis (→ **15b**); Ru-catalysed cycloheptatriene formation (→ **15c**); Au-catalysed cyclisation of the alcohol onto the alkyne (→ **15d**); base-induced cyclisation (→ **15e**); Pauson–Khand reaction (→ **15f**); and Miesenheimer [2,3]-sigmatropic rearrangement (not shown). Four of these cyclisation reactions could be used again to convert the enyne **15e** into molecules with four further scaffolds (**17a–d**). In addition, Diels–Alder reactions with 4-methyl-1,2,4-triazoline-3,5-dione converted the dienes **15b**, **17a** and **17d**



Scheme 1.5 Schreiber's branching pathway exploiting complementary cyclisation reactions. *Reagents and conditions*: (1) 10 mol% $Pd(PPh_3)_2(OAc)_2$, benzene, 80 °C, [**15a**: single diastereomer, 81%; **17d**: single diastereomer, 70%]; (2) 10 mol% Hoveyda–Grubbs 2nd gen. cat., CH_2Cl_2, reflux, [**15b**: *trans*:*cis* 85:15, 89%; **17a**: *trans*:*cis* 85:15, 87%]; (3) 10 mol% $CpRu(MeCN)_3PF_6$, acetone, rt, [**15c**: single diasteromer, 85%; **17c**: single diasteromer, 91%]; (4) 10 mol% $NaAuCl_4$, MeOH, rt, single diasteromer, 80%; (5) NaH, toluene, rt, 88%; (6) $Co_2(CO)_8$, Et_3NO, NH_4Cl, benzene, rt, [**15f**: dias. >90:10, 85%; **17b**: single diasteromer, 85%]; (7) 4-methyl-1,2,4-triazoline-3,5-dione, CH_2Cl_2, rt, single diasteromer from *trans* diene, 72%.

into the corresponding polycyclic products (*e.g.* **16**; other products not shown). The key to this powerful synthetic approach lay in the design of precursors (*e.g.* **14**) which were effective substrates in a wide range of efficient and diastereoselective cyclisation reactions.

The "branching pathway" strategy has been shown to allow the preparation of small molecule libraries based on a range of diverse scaffolds. The strategy requires the careful design of flexible precursors that may be converted selectively into different scaffolds under alternative reaction conditions. The development of effective branching pathways may also require considerable optimisation of the individual steps leading to alternative scaffolds. It is probably unlikely that the approach can be extended such than many scores of alternative scaffolds may be generated from individual precursors.

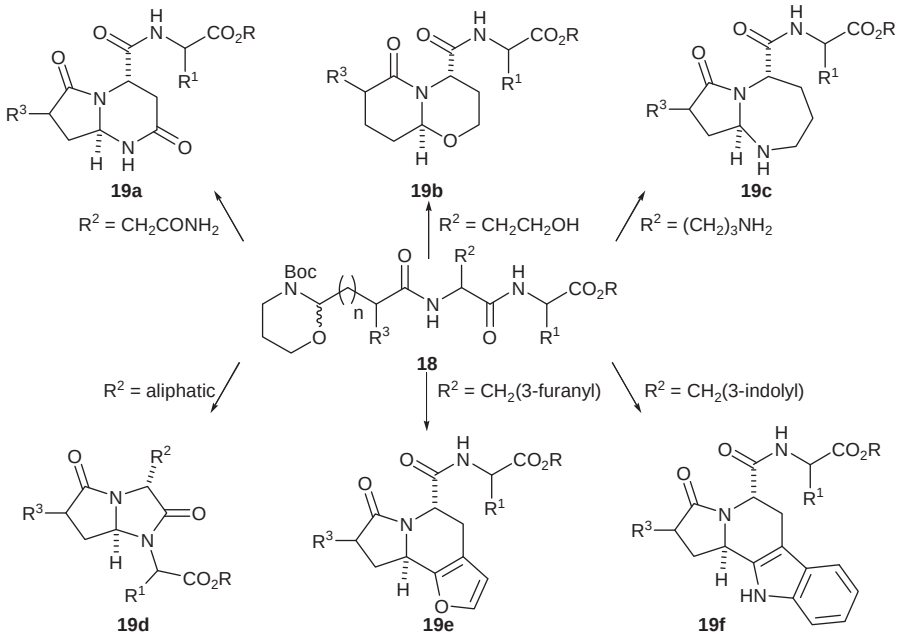
1.2.3.3 Introducing Scaffold Diversity by Combining Building Blocks: the "Build–Couple–Pair" Strategy

An extremely powerful strategy for preparing small molecule libraries with high scaffold diversity exploits simple building blocks in combination. The approach requires the building blocks to be prepared ("built") and then connected ("coupled"). Finally, pairs of functional groups are reacted ("paired") intramolecularly to yield new ring systems in the final scaffolds. The so-called "build–couple–pair" strategy has been reviewed.[24b]

The scope of the approach is extremely broad and, indeed, some folding pathways[30] (*e.g.* Scheme 1.4; Section 1.2.3.1) and branching pathways[32] (*e.g.* Scheme 1.5; Section 1.2.3.2) can be considered to exemplify the "build–couple–pair" strategy. For example, the four component Petasis reaction illustrated in Scheme 1.4 allowed simple building blocks to be combined; complementary cyclisation reactions were then use to "pair" functional groups to yield a diverse range of product scaffolds.[32]

The cyclisation reactions of *N*-acyl iminium ions have yielded a wide range of distinct small molecule scaffolds (Scheme 1.6).[33] Initially, peptide synthesis was used to connect a range of functionalised amino acid building blocks. Each of the resulting peptides, **18**, was designed to contain a masked aldehyde, a suitably positioned secondary amide, and a pendant nucleophile. Treatment of the peptides with acid triggered the release of an aldehyde, the formation of an *N*-acyl iminium ion, and cyclisation to yield a final scaffold (*e.g.* **19a–f**).

Metathesis cascades have underpinned the synthesis of diverse small molecule libraries.[34,35] Metathesis is a superb "pairing" reaction for the "build–couple–pair" approach: first, it can yield many different ring systems and, second, alkenes (and alkynes) are compatible with the many reactions that may be used to connect building blocks. Metathesis has been used to prepare a library of natural product-like molecules (Scheme 1.7).[35] Initially, unsaturated building blocks were attached iteratively to fluorous-tagged linker to yield metathesis precursors **20**. Crucially, alternative attachment reactions were used such that the building blocks were connected through bonds that either did, or

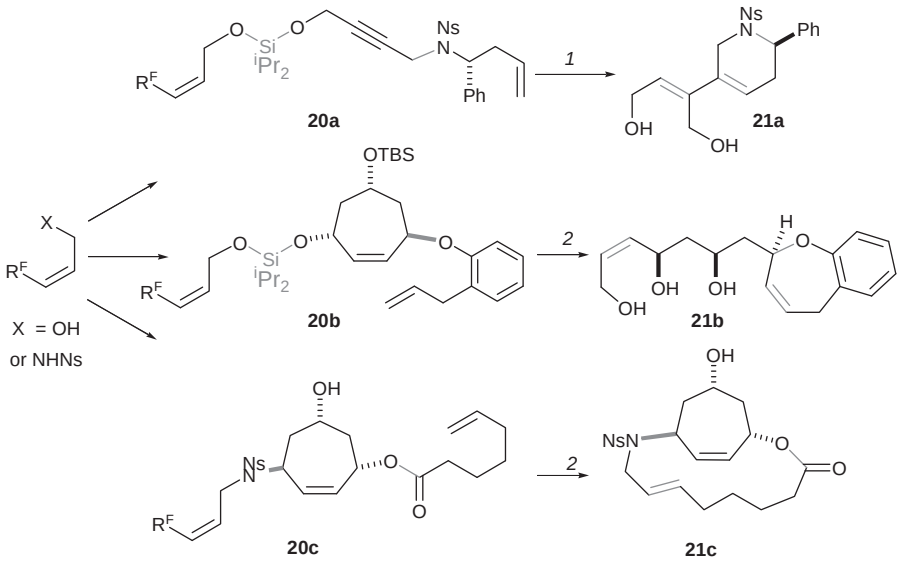


Scheme 1.6 Cyclisation of *N*-acyl iminium ions in the synthesis of a range of small molecule scaffolds. *Reagents and conditions*: H^+.

did not, remain as a vestige in the final products. Finally, metathesis cascades were used to ''reprogramme'' the scaffolds and to release the final molecules, **21**, from the fluorous tag. The resulting library had unprecedented scaffold diversity (over eighty distinct scaffolds).[36]

The ''build–couple–pair'' is an extremely powerful and flexible approach for preparing small molecule libraries based on a diverse range of scaffolds. The approach is highly general, and many different reactions have been used to connect (''couple'') building blocks. Variants of the Ugi multi-component reaction are, however, still over-represented amongst the reactions that have been used to connect building blocks.[32,37]

A wide range of other reactions have been exploited in the final cyclisation (''pairing'') step: in addition to the examples illustrated here, lactamisations,[38,39] metal-catalysed cyclisations[37,39,40] and cycloadditions[37] have, for example, been exploited to yield final product scaffolds. At its most powerful, the ''build–couple–pair'' strategy can allow the combinatorial variation of the scaffolds of small molecules. However, a significant challenge will be to identify reactions other than olefin metathesis that have the broad scope and chemoselectivity needed to yield scores of different ring systems. It is certainly possible that the overall approach may, in the future, be used to prepare small molecule libraries based on hundreds, or even thousands, of distinct molecular scaffolds.



Scheme 1.7 Synthesis of natural product-like molecules with unprecedented scaffold diversity. Initially, building blocks were added iteratively to a fluorous-tagged linker, with intermediates purified by fluorous–solid phase extraction. Metathesis cascades were used to reprogramme the scaffolds and to release final products from the fluorous-tagged linker. *Reagents and conditions*: (1) Grubbs' first-generation catalyst, **21a**: 23%; **21b**: 56%; (2) fluorous-tagged Hoveyda–Grubbs second-generation catalyst, **21c**: 33%.

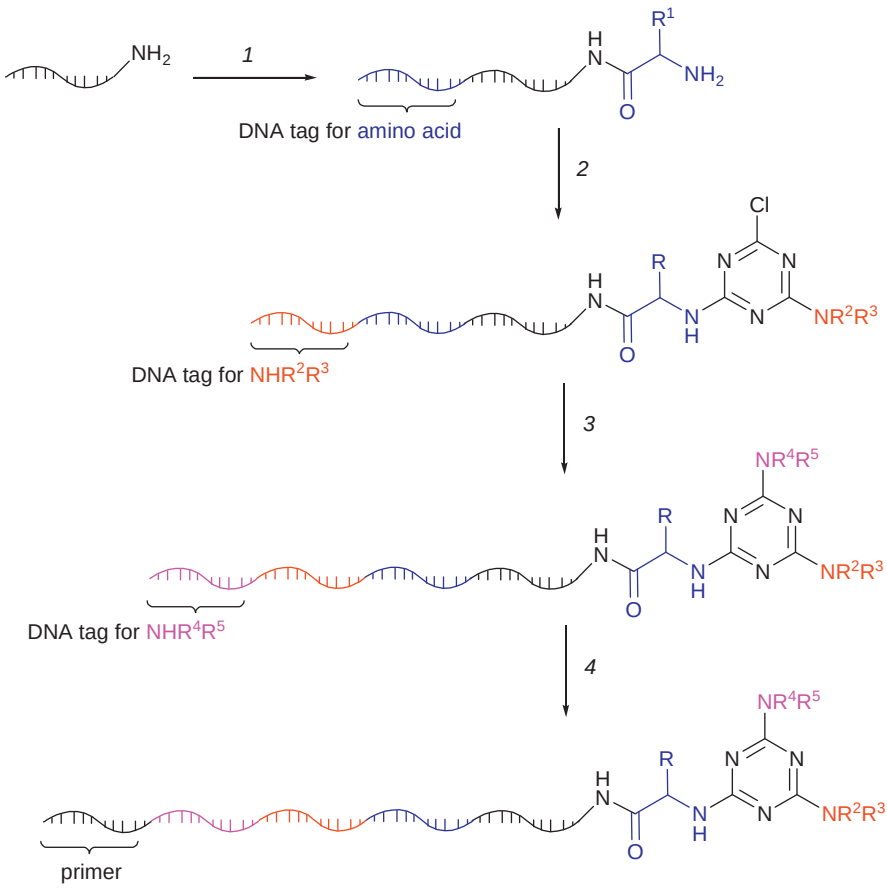
1.2.4 Emerging Approaches for Ligand Discovery Using Large Small Molecule Libraries

Combinatorial biochemical techniques,[41] such as RNA display, phage display and aptamer SELEX, can allow large (*e.g.* 10^8–10^{13}) libraries of macromolecules to be screened on the basis of binding to a target protein. These techniques enable the discovery of novel macromolecules that bind to, and hence modulate the function of, the target protein. In contrast, most screens of small molecule function are restricted to much smaller libraries, essentially because individual molecules usually need to be separately purified before they are (usually individually) assayed: even in large pharmaceutical companies, high-throughput screens are restricted to smaller (generally $<10^6$) compound libraries.

Extremely powerful methods are now emerging to increase the efficiency of screens for small molecule binders.[42–44] These methods enable the synthesis of extremely large small molecule libraries in a manner that is broadly analogous to combinatorial biochemical approaches. In some cases, an evolutionary approach is possible,[42,44] with increasing active compounds being selectively fed into subsequent rounds of chemical synthesis. Large libraries of constrained

peptides have been encoded by phages, and have been screened to yield an inhibitor that was specific to human plasma kallikrein.[42] A broader range of chemistry, which nonetheless still needs to be compatible with aqueous conditions (and DNA), is possible using DNA-encoded[43] or DNA-programmed[44] organic synthesis.

A split-pool approach has been used to prepare a library of 800 000 000 DNA-encoded small molecules (see Scheme 1.8 for an example for a synthesis of a DNA-encoded library).[43] After each diversification step, the small molecules were derivatised with a DNA tag to encode the substituent that had been added. At the end of the synthesis, the small molecules – each of which was



Scheme 1.8 Synthesis of a DNA-encoded small-molecule library. The DNA tag is not drawn to scale. *Reagents and conditions*: (1) (a) attach DNA tag; (b) acylation with Fmoc-protected amino acid; (c) purify; (d) deprotect; (2) (a) attach DNA tag; (b) cyanuric chloride; (c) substitute with R^2R^3NH; (3) (a) attach DNA tag; (b) substitute with R^4R^5NH; (c) purify; (4) ligate primer.

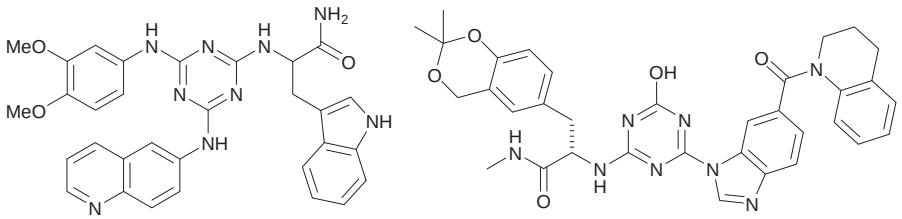


Figure 1.2 Structures of some inhibitors of Aurora A kinase (**22**) and p38 MAP kinase (**23**) discovered using a DNA-encoded small-molecule library.

derivatised with a unique DNA tag – were pooled. The library was then interrogated by affinity selection (*i.e.* ability to bind the target protein). The identity of the binding molecules was deconvoluted by high-throughput sequencing of their DNA tags, to yield extremely rich structure-activity relationships for the active compounds. The general approach has been used to identify inhibitors of two enzymes: Aurora A kinase (*e.g.* **22**) and p38 MAP kinase (*e.g.* **23**) (Figure 1.2).

DNA can also be used to *programme* a series of organic transformations (rather than simply to record the history of a synthesis).[44] In this format, all of the transformations may be performed in a single pot, with the synthesis of each individual molecule being determined by base-pairing between the DNA-tagged substrate and its complementary DNA-tagged reagent.

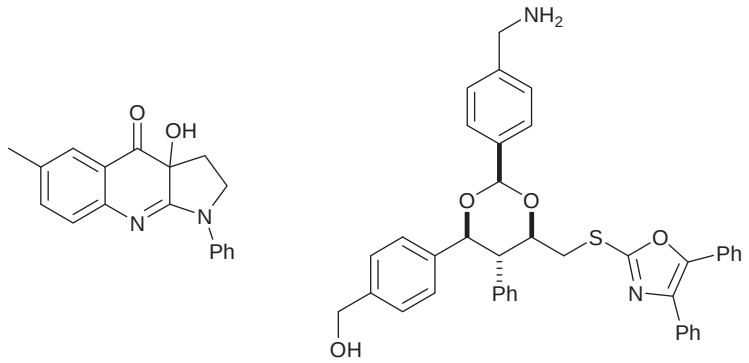
DNA-programmed synthesis lends itself to an evolutionary approach: the tags for small molecule binders may be amplified after each round of affinity selection, and fed into a subsequent generation of synthesis. The approach has been used to discover, from a library of 100 000 000 compounds, ligands for the *N*-terminal SH3 domain of the proto-oncogene Crk.[44a]

1.3 Discovery of Small Molecule Probes

In this section, the discovery of small molecules for the interrogation of biological mechanisms will be described. An appropriate small molecule tool may be discovered either on the basis of modulation of the activity of a specific protein (reverse chemical genetics, Section 1.3.1) or its phenotypic effect (forward chemical genetics, Section 1.3.2). Emphasis will be placed on the approaches that may be used to discover useful small molecule tools, and the insights into biological mechanisms that may be accrued.

1.3.1 Discovery of Small Molecule Probes Using a Reverse Chemical Genetic Approach

The reverse chemical genetic approach begins the discovery of a small molecule modulator of a specific target protein. The small molecule is usually identified



blebbistatin uretupamine

Figure 1.3 Structures of blebbistatin and uretupamine.

using an *in vitro* assay of the activity of a purified protein. However, it is also possible to identify ligands that modulate the cellular function of a specific protein.

Blebbistatin (Figure 1.3) was discovered using a high-throughput screen for small molecules that inhibit non-muscle myosin II.[8b] Blebbistatin was then used to study myosin II-dependent cell processes. The compound blocked cell blebbing rapidly (within $\sim$2 min) and reversibly, and disrupted directed cell migration and cytokinesis in vertebrate cells. Blebbistatin was also used in combination with other ligands to demonstrate that exit from the cytokinetic phase of the cell cycle is dependent on ubiquitin-mediated protein degradation.

Small molecule microarrays[45,46] can allow the discovery of small molecule binders to a specific protein: the methodology is accessible to many academic groups, and can allow high-throughput screening of large compound libraries. A small molecule microarray was used to discover uretupamine (Figure 1.3) which binds to Ure2p, the central repressor of genes involved in nitrogen metabolism.[27a,b] A 3780-member library of small molecules was used to create a small molecule microarray which was then challenged with fluorescently tagged Ure2p. Transcription profiling was used to determine the effects of uretupamine on the regulation of genes in yeast. Uretupamine was found to modulate only a subset of the functions of Ure2p: the regulation of the glucose-sensitive subset of the downstream genes. The example illustrates that reverse chemical genetics can be a more refined tool for interrogating cellular protein function than reverse genetics since the effects of uretupamine were more specific than those resulting from deletion of the *URE2* gene.

Small molecule microarrays were also used to discover robotnikinin (Section 1.4), a ligand that binds to the extracellular Sonic Hedgehog (Shh) protein.[47] Approximately 10 000 diversity-oriented synthesis compounds and natural products were immobilised on a single microscope slide. The small molecule microarray was challenged with the 20 kDa N-terminal fragment (ShhN) of Shh: some structurally related macrocycles were identified as hits and, after

optimisation, robotnikinin emerged. Robotnikinin has been used to interrogate the Sonic Hedgehog signalling pathway (see Section 1.4). Small molecule microarrays have also been used to discovery ligands for human immunoglobulin G (IgG),[46] the transcription factor Hap3p,[28] and calmodulin.[45]

Valuable small molecule tools may also be obtained by structure-guided ligand design. The interaction between the lysine 382-acetylated p53 and the bromodomain of the coactivator CBP is essential for p53-induced transcription of the cell cycle inhibitor p21 in response to DNA damage. A structure-guided approach was used to discover small molecule inhibitors of the CBP/p53 interaction.[48] Thus, the structure of the bromodomain of CBP (in complex with a peptide mimic of lysine 382-acetylated p53) enabled the design of a focused small molecule library to target the CBP/p53 interaction: NMR spectroscopy was then used to identify specific small molecules that bound to the bromodomain of CBP. The active molecules were shown to inhibit CBP/p53 association: the small molecules led to destabilisation of p53, and, thus, deactivation of its transcriptional activity in response to DNA damage. In this investigation, the chemical genetic approach enabled the functional consequence of a protein-protein interaction to be determined in a cellular context.

1.3.1.1 General Reverse Chemical Genetic Approaches for Targeting Specific Members of Families of Macromolecules

A disadvantage of the chemical genetic approach is that it usually lacks generality: in general, for each macromolecule to be targeted, the discovery of a tailored ligand is required. However, more general approaches have been developed to target protein kinases (Section 1.3.1.1.1) and duplex DNA sequences (Section 1.3.1.1.2).

1.3.1.1.1 Inhibitor-sensitive Protein Kinase Mutants. Shokat has pioneered a remarkable approach that allows specific protein kinases to be targeted with small molecule ligands.[49] The design of highly specific ATP-competitive inhibitors of protein kinases is hampered by the large number of proteins with ATP binding sites (including $\sim$600 protein kinases). The key to Shokat's approach lies in engineering a protein kinase that is uniquely sensitive to a "bulky" kinase inhibitor. The strategy of the approach is illustrated in Figure 1.4. The structure of the protein kinase Hck in complex with the ligand PP1 is shown in Panel B. By replacing the *p*-tolyl substituent of PP1 with a bulkier substituent (such as a 1-naphthyl or a 1-naphthylmethyl substituent), the ligand is no longer able to inhibit Hck (or indeed any other protein kinase) because of an unfavourable interaction with the gatekeeper residue (T338 in Hck). However, mutating the gatekeeper residue to a smaller residue allows the bulkier ligand to be (uniquely) accommodated by the mutant protein kinase.

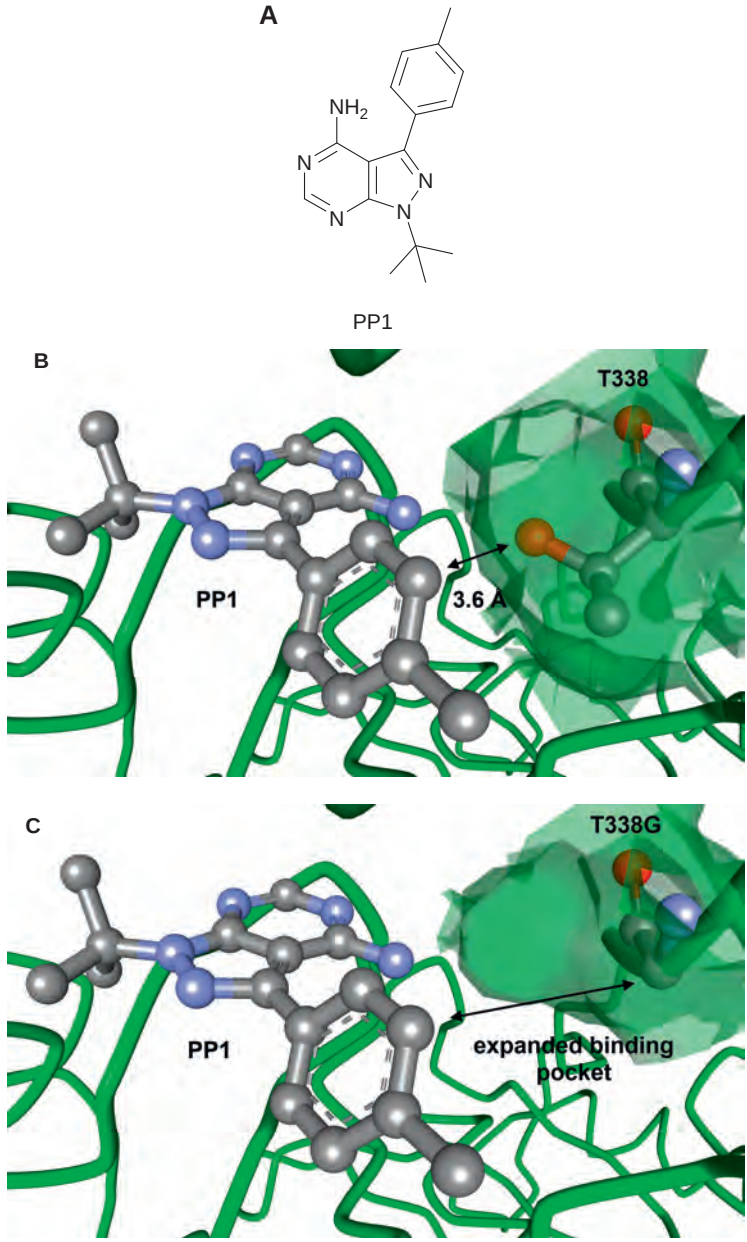


Figure 1.4 Development of inhibitor-sensitive protein kinases illustrated using the protein kinase Hck. Panel A. Chemical structure of the ligand PP1. Panel B. Crystal structure of PP1 in complex with the protein kinase Hck. The *p*-tolyl substituent of PP1 is in close proximity to the side chain of the gatekeeper residue T338. Panel C: Model of the structure of the Hck · PP1 complex in which the side chain of the gatekeeper residue 338 has been removed *in silico*. The expanded binding pocket in the T338G mutant is able to accommodate bulkier analogues of PP1.

The approach has been demonstrated to be rather general, and has been used to interrogate many biological mechanisms that are regulated by protein kinases.[50] An early investigation focused on the role of the cyclic-dependent kinase Cdc28 (CDK1) in the regulation of the cell cycle in yeast. Specific inhibition of the inhibitor-sensitive mutant of Cdc28 caused a pre-mitotic cell-cycle arrest. Remarkably, the observation was in stark contrast to the G1 arrest that is typically observed in temperature-sensitive *cdc28* mutants.[51] Thus, in this case, the insights gleaned from a conditional (temperature-sensitive) genetic approach were different to those observed using a conditional chemical genetic approach.

1.3.1.1.2 Chemical Tools for Investigating DNA Function. The structural regularity of duplex DNA has presented a special opportunity to identify small molecules that target a specific DNA sequence. Dervan has developed rules that allow a a specific polyamide to be designed to bind to any pre-determined DNA sequence.[52] The ligands were inspired by the natural product, distamycin, that binds to 9 base-pairs of AT-rich DNA. The ligands recognise specific DNA sequences through recognition of the "edges" of intact Watson–Crick base pairs in the minor groove of DNA (Figure 1.5).[53]

The ability to design polyamide ligands to target any sequence of duplex DNA is exceptionally valuable for probing the cellular role of specific DNA sequences. A polyamide was designed to target the binding site (5′-AGTACT-3′) of the TFIIIA transcription factor (Figure 1.6).[54] The polyamide competed effectively with the TFIIIA transcription factor for its target DNA sequence: the ligand thus selectively down-regulated, both *in vitro* and *in vivo*, the transcription of a gene (the 5S RNA gene) controlled by TFIIIA (relative to a control gene).

1.3.2 The Forward Chemical Genetic Approach

The forward chemical genetic approach begins the discovery of a small molecule modulator using a phenotypic assay. In this section, emphasis will be placed on the challenges associated with identifying the target of the small molecule modulator, and the insights into biological mechanisms that may be accrued (Section 1.3.2.2). Some of the challenges (and opportunities) associated with identifying the target of a small molecule modulator are described in Section 1.3.2.1.

1.3.2.1 The Challenge of Target Identification

A major challenge associated with the forward chemical genetic approach is the identification of the target of the small molecule probe.[11] Affinity chromatography is a common approach to target identification. This approach requires the attachment of the small molecule probe to a solid support: the site of the attachment of the linker must be chosen carefully to avoid the ablation of the probe's biological activity. Trapoxin, is a cyclic tetrapeptide whose