ELSEVIER

Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry perspective

Hongyu Zhao

R4MC, AP-10, Metabolic Disease Research, Global Pharmaceutical Research and Development, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064-6098, USA

Hit selection and lead generation are crucial for the success of the resource-demanding lead-optimization phase in drug discovery, and represent a major research area of medicinal chemistry today. Ligand-binding efficiency, ligand complexity, ligand–target profile complementarity and chemical tractability are important parameters in hit selection. As synthesis and assay throughput improve, a large number of analogs based on the same scaffold can be rapidly synthesized and tested. Consequently, more chemistry resources could be devoted to scaffold modifications to expand the candidate pool in lead generation. Most recently discovered druggable targets are promiscuous toward lipophilic ligands, and the hydrophobic portions of hit compounds should be preferentially modified in analog and scaffold design.

Introduction

The role of the medicinal chemist changes as the drug discovery paradigm shifts [1]. Under the current paradigm, most drug discovery programs start with an HTS campaign, followed by an uncertain lead-generation and -optimization process [2]. As a result, today's medicinal chemists work in a much larger research area than their counterparts of a generation ago. Considerable amounts of time and effort are spent in generating compounds of a quality that is suitable for lead optimization. At this stage, most programs are primarily driven by *in vitro* assays and a large number compounds can be tested weekly (i.e. up to several hundred compounds per week). Advances in the synthesis, characterization and purification of drug-sized molecules, along with the greater and growing availability of commercial building blocks, enable many analogs based on the same scaffold to be synthesized quickly. Consequently, more chemistry resources could be devoted to scaffold modifications to expand the candidate pool in lead generation. Sometimes, lack of high-quality screening hits renders this effort necessary. The importance of selecting hit compounds with an appropriate scaffold and scaffold hopping in lead generation are discussed here. In this article, lead generation refers to the identification of potent and selective compounds with adequate pharmacokinetic (PK) properties. Lead optimization is reserved for the process of optimizing the efficacy and safety profiles to develop a clinical candidate.

Corresponding author: Zhao, H. (Hongyu.zhao@abbott.com)

Hit selection: focus on the scaffold

Hits are identified by high-throughput or virtual screening of corporate compound collections, or sometimes from published patents. Depending on the nature of the target, assay method and quality and quantity of the compounds screened, various numbers of hits can be identified in an HTS campaign. Compounds containing prohibited structural motifs, such as reactive functional groups, are removed from the list without further testing [3]. Confirmation of structure, purity, potency, selectivity, mode-of-action and evaluation of patentability, chemical tractability and PK liability are required before further action is taken [3–5]. After these exercises, there might still be more hits than the chemistry team can fully explore. At this stage, the aim is to select the hit compounds with the best developmental potential. Namely, to identify those that captured most features of the potential drug.

Depending on the target, some of these features are easier to engineer into a series than others. Basically, medicinal chemists are good at adding and replacing peripheral functional groups on a scaffold to improve the properties of a hit compound. By contrast, the molecular scaffold is more complicated to manipulate because

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.12.003

this process is more labor intensive and considerable loss of activity is often associated with structural changes at the scaffold level. Therefore, selecting the compounds with the optimal scaffold is vital to the success of the medicinal chemistry effort in the later stages. Ligand efficiency (LE), ligand complexity, ligand–target profile complementarity and chemical tractability are important parameters in hit selection.

Several methods have been proposed to measure ligand-binding efficiency [6–8], all of which focus on evaluating how well the functional groups or heavy atoms are used in binding. For hit selection, some chemists argue that more-efficient and less-complex ligands have a better chance of being developed into a drug because the molecular weight increases that are likely to be encountered during lead optimization are better tolerated by such ligands [9,10]. Although high ligand efficiency is an advantage, the nature of the target, other features of the compounds and the current medicinal chemistry paradigm also need to be considered [11,12]. In fact, a survey of recent lead-generation and -optimization efforts on new targets revealed that, most often, the ligand efficiency increased at this stage [13], suggesting that ligand efficiency should not be the only parameter considered in hit selection [12].

Another element of ligand quality is the structural complexity of the ligand, especially the features that are associated with its scaffold. It has been shown that 50% of known drugs share only 32 molecular frameworks [14], implying an optimal level of skeletal complexity is an intrinsic character of druglikeness. Recent analysis revealed that a certain level of molecular complexity is necessary to achieve the desired biological activity [15]. The preferred ligands for different targets can have dramatically different molecular complexity, from Vancomycin at one end of the extreme to Aspirin at the other. Determination of the complexity requirements of the preferred ligand scaffold for the target is important in hit selection because molecular complexity is difficult to manipulate productively at the scaffold level. An historical problem in drug discovery is finding peptidomimetics with simpler structure but similar activity. It is also difficult to improve molecular properties by increasing the complexity of a molecule because more-complex molecules are less likely to be similarly complementary to the protein-binding pocket [16], and more medicinal chemistry resources are needed. A structural analysis of hit compounds might reveal ligand preference for the target. Also, experimental or computational analysis of the binding pocket [17], or inspection of the structural characteristics of the endogenous ligands of the protein or potent ligands of homologous proteins also provides useful information about the nature of the target. If a certain level of molecular complexity is required for a ligand to bind adequately to a particular target, this feature should be present in the hit compound. For two compounds with equal ligand efficiency but different complexity, the more complex compound might have several peripheral functional groups that could interact unfavorably with the protein. These defects can be corrected readily by medicinal chemists. For example, as shown in Figure 1, compound **1**, based on an effective scaffold, was inactive at agonizing estrogen receptor α (ERα) [18]. However, replacement of the methoxy group in **1** with a hydroxyl group resulted in a highly potent ERα agonist (Compound **2**, EC_{50} 4 nM).

By contrast, the less complex compound might achieve its potency through optimal peripheral functional groups on a less-efficient scaffold. This defect is harder to fix, although techniques are emerging to address this problem (e.g. see scaffold hopping in the subsequent section on lead generation). To this end, the development of two series of protein tyrosine phosphatase 1B (PTPT1B) inhibitors is shown in Figure 2. The SAR study on both series was done by the same group of medicinal chemists with the same level of support, and the compounds were tested in the same assay. The ligand efficiency (defined by $LE = \Delta G/N_{non\text{-}hydrogen\ atoms}$) for compound **3** (LE = 0.22) was lower that that of **5** (LE = 0.27). However, compound **5** had fewer features than **3** and lacked chemically modifiable positions on the key binding motif (i.e. the aryl carboxylic acid portion). After roughly equal optimization efforts, low nanomolar ligands (e.g. inhibitor **4**, K_i = 0.018 μM; Figure 2) were found based on **3** [19]. However, the most potent compound (i.e. inhibitor **6**) from the isoxazole series only demonstrated a K_i of 0.92 μM [20]. Replacement of the allyl ether linker in **6** with larger amide-containing linkers only resulted in weaker inhibitors.

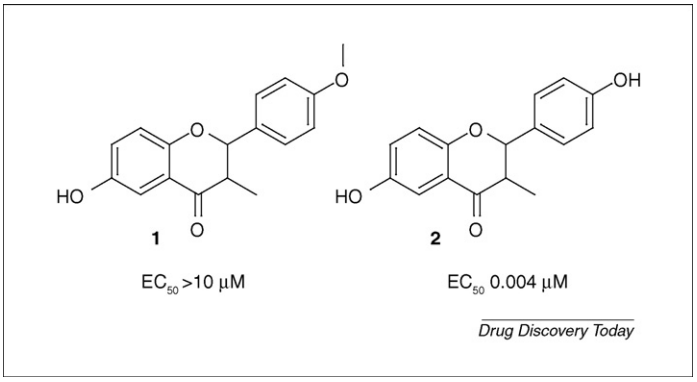

FIGURE 1
An effective scaffold with an intolerable peripheral substitution.

The molecular complexity could have a similar impact on the optimization of other molecular properties besides potency such as selectivity and PK, and should be an important factor to consider in hit selection.

Frequently, the hit list contains lipophilic and hydrophilic compounds with otherwise comparable qualities. Hydrophobic interactions are considered easier to identify than polar interactions, but this observation was based on the experience with conventional drug targets. For some targets like phosphatases, the opposite seems true. It is usually accepted that a balanced hydrophilicity–hydrophobicity profile is necessary for good PK properties [21]. The nature of the binding pocket of the target has a considerable impact on hit selection. If the target has a tendency to bind to lipophilic molecules (e.g. G-protein-coupled receptors; GPCRs), the hits that provide more polar interactions might have an advantage because the hydrophobic interactions could be more-readily identified later on. By contrast, if the target preferentially binds polar molecules (e.g. phosphatases), following the same logic, the hit compounds that provide more hydrophobic interactions might be better choices. An isothermal titration calorimetric (ITC) determination of enthalpic and entropic contribution to ligand-binding free energy would provide a more quantitative measure of how well the desired features have been captured by a particular compound [22]. Enthalpy-driven compound selection would enable easier optimization for lipophilic

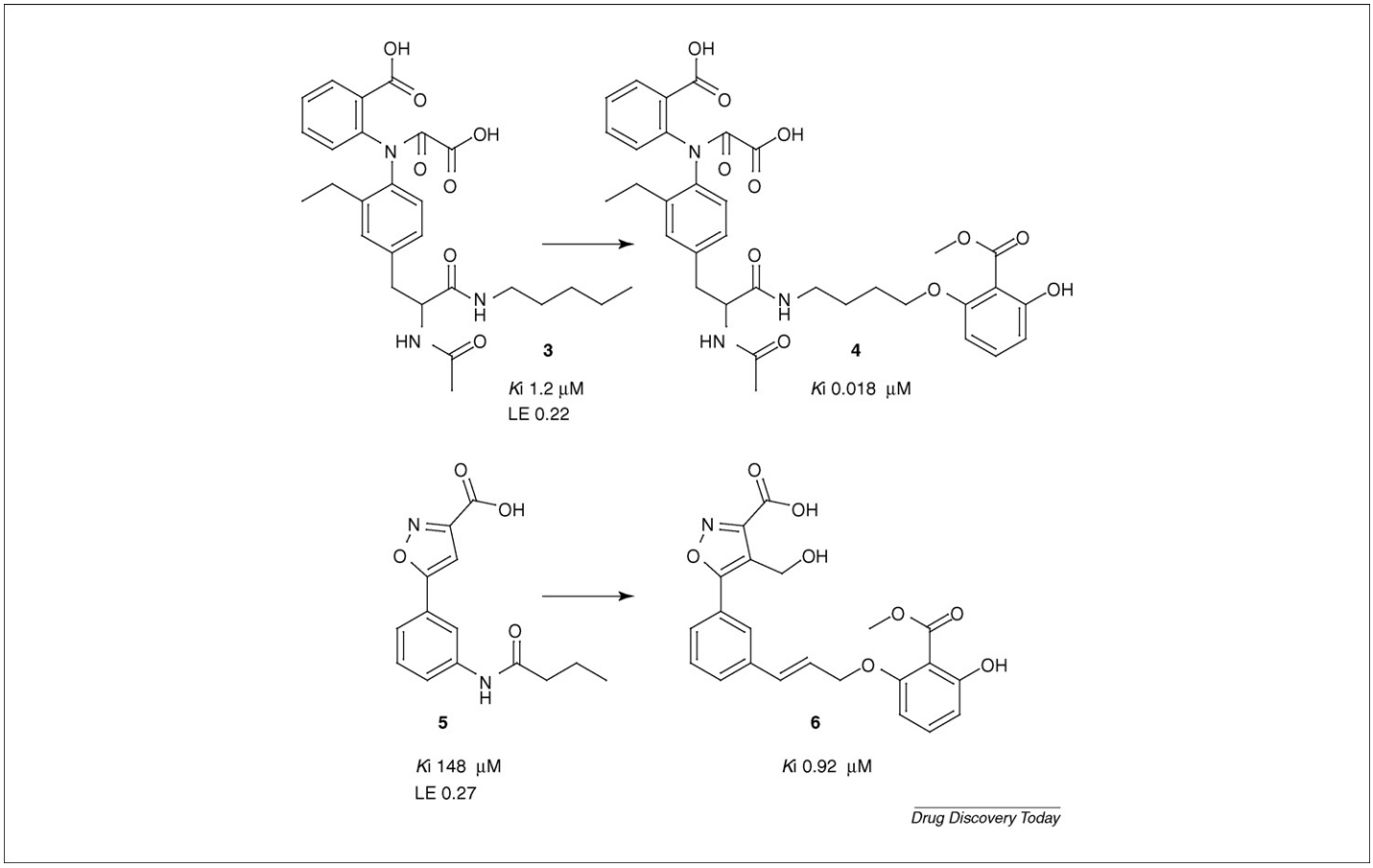

FIGURE 2

Two hits (compounds 3 and 5) with similar ligand efficieny (LE) but different complexity resulted in protein tyrosine phosphatase 1B (PTP1B) inhibitors with different potencies (compounds 4 and 6).

binding pockets, and entropy-driven compound selection would enable easier optimization for hydrophilic binding pockets.

Molecules that are easier to synthesize and more amenable for SAR development should be given higher priority in hit selection. The more positions on a scaffold that can be modified using robust chemistry, the more diverse the analogs that can be accessible for the collection of SAR information. Chemical tractability is apparently a relative term depending on the synthetic insights and skills available. Recently, a strategy coined 'diverted total synthesis of small-molecule natural products' was put forward for lead optimization [23]. This approach starts with complex natural product leads and has played a key role in developing epothilone analogs, the most advanced of which is in Phase II clinical trials for cancer therapy [24].

Lead generation: identification of an effective scaffold

At this stage, the goal is to identify multiple diverse analogs that possess adequate potency, selectivity and favorable PK properties, to ensure a desirable level of drug exposure in targeted tissues. This will set the stage for the more resource-demanding lead-optimization phase of efficacy and safety evaluation in different animal models. Lead optimization is a complex multiple-property optimization process and only high-quality leads could realistically be developed into drugs [25]. The key at the lead-generation stage is to identify an appropriate scaffold capable of providing sufficient numbers of high-quality analogs. Many excellent discussions on lead-generation strategies have been reported in the literature in recent years [2–5,11–13]. I will focus on one specific aspect of this process – scaffold design.

Once a hit is selected, its developmental potential is evaluated against the qualities required by the potential drug. Apparently, both the quality of the hit compounds and the druggability of the target [26,27] play parts in this assessment. Druggability can be considered as the level of promiscuity of the protein and which part of the chemical space it is promiscuous in [28]. If the target is promiscuous in the available or tangible druglike chemical space, it is easy to find high-quality ligands. For example, non-nucleoside reverse-transcriptase has >30 chemically distinct classes of inhibitors [29] – and it is not trivial to find interesting compounds that do not bind cytochrome p450 enzymes. If the hit compounds already possess the most desired qualities, lead generation would be straightforward and traditional analog synthesis might be a good strategy. However, the screening of old compound collections or combinatorial libraries against novel targets sometimes yields poor-quality hits. Even the carefully selected ones cannot always be transformed into useful leads by analog synthesis. Poor-quality compounds occur in different forms. For example, a scaffold could have only one position that gives productive SAR (e.g. the other positions might be chemically inaccessible, intolerable to chemical modifications or provide flat SAR); or all analogs in the series could suffer a major setback (e.g. poor PK, low selectivity, potency plateau, patent issues, and so on). If a scaffold cannot

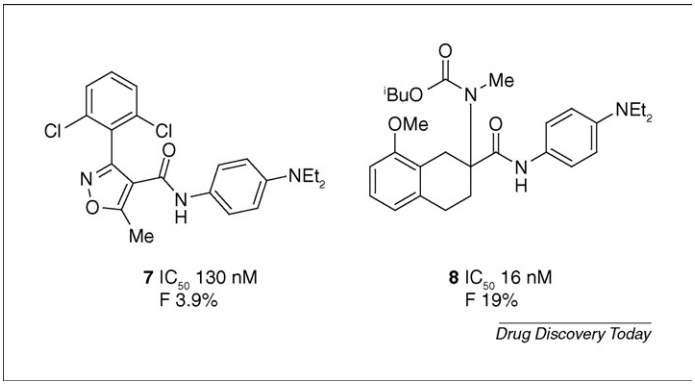

FIGURE 3

Scaffold modification of growth hormone secretagogue receptor (GHS-R) antagonist 7 led to an improved antagonist 8.

provide multiple structurally orthogonal analogs with adequate potency, selectivity and PK properties, the success rate of the following lead-optimization phase is likely to be low and the scaffold needs to be modified.

Scaffold hopping is a computational technique that identifies compounds containing a topologically different scaffold from the parent compound, but with similar or improved activity and other properties from a given database [30–32]. This technique has gained increasing popularity in recent years, indicating a shift from traditional analog design to scaffold design in the lead-generation phase. Medicinal chemists can also design and synthesize scaffolds *de novo*, which has been practiced throughout drug discovery history. This approach requires minimal computational resources and is not limited by compound availability, as long as they can be synthesized within a reasonable time frame, and is of practical value to medicinal chemists.

Several basic guidelines, all of which might not be suitable for all cases, can be useful for practicing this technique. First, an adequate magnitude of change should be estimated. Usually, less-efficient and overly simple or complex compounds need more-drastic changes. The more drastic the changes involved, the more scaffolds are required to synthesize, validate or invalidate these chemistry-demanding efforts. A pharmacophore for the hit series needs to be hypothesized, which can be accomplished after some initial SAR studies. An evaluation of the promiscuity of the target is necessary to determine the likelihood of success, as well as which part of the pharmacophore is more tolerable to chemical modifications. For example, if the target protein is more promiscuous toward lipophilic ligands, the lipophilic portion of the hit should be modified more dramatically. Apparently, more-promiscuous

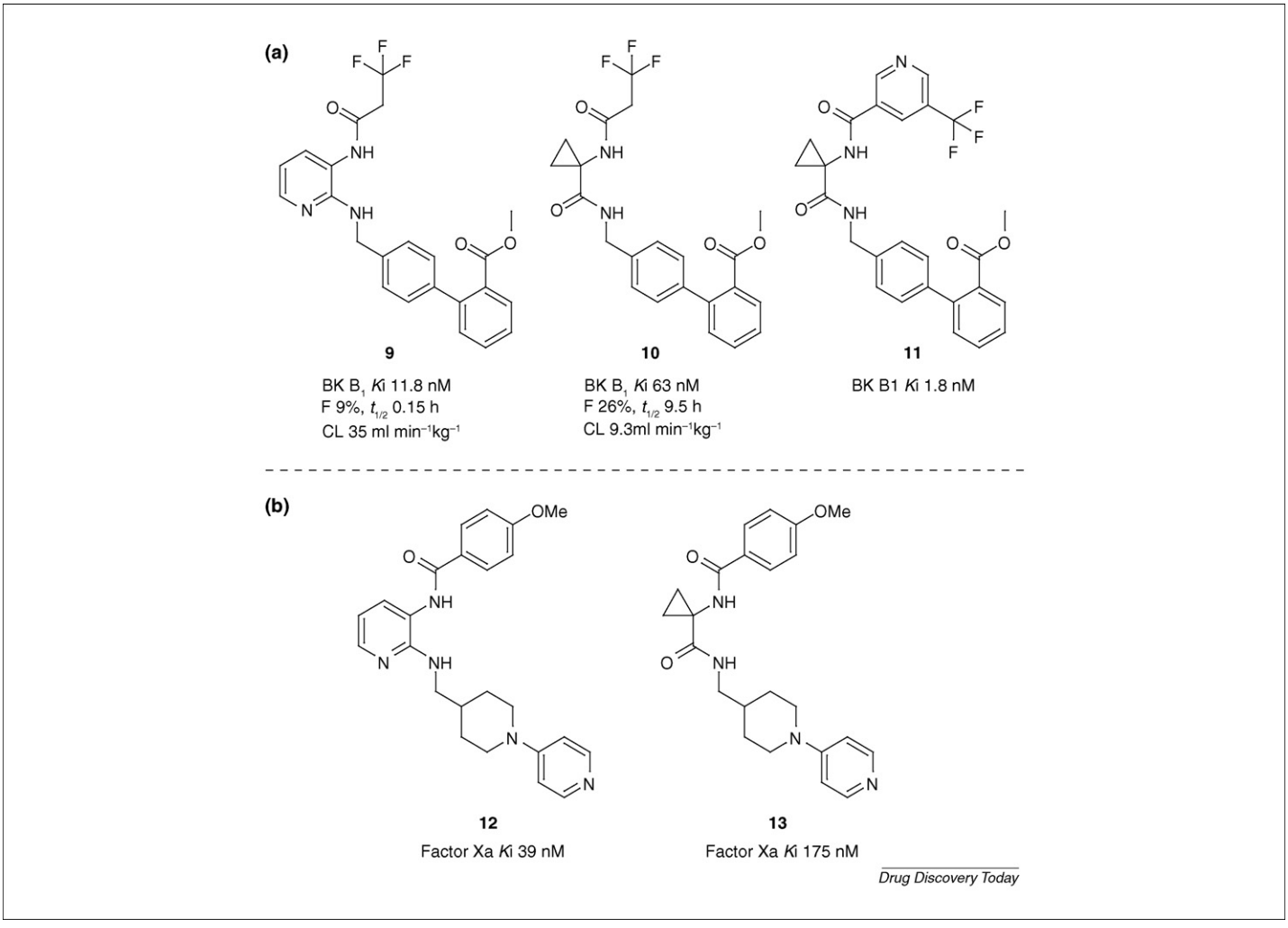

FIGURE 4

Cyclopropylamino acid amide as a pharmacophoric replacement of 2,3-diaminopyridine. (a) Bradykinin B_1 antagonists. **(b)** Factor Xa inhibitors.

proteins will accommodate more-diverse sets of scaffolds. It would be advantageous for designed scaffolds to be in clusters so that a larger chemical space can be evaluated. Also, the designed scaffolds should be suitable for SAR development if they prove successful. Because the topology of the pharmacophore is being modified, a considerable loss of biological activity can be expected for most of the new scaffolds. To ensure a reasonable success rate, the other properties [33] of the pharmacophore should be kept as constant as possible. Attention should be paid to conserving carbon atom hybridization. Synthetic and medicinal chemistry insights are crucial in this iterative process of conception and rejection of new templates based on their similarity to the pharmacophore, synthetic accessibility and the size of chemical space they cover.

The discovery of tetralin carboxamide growth hormone secretagogue receptor (GHS-R) antagonists from our [34] laboratory could be an example to illustrate some of these points (Figure 3). The screening hit compound **7** was made from a simple amide-coupling reaction of commercially available materials. The patent position of this series of compounds was questionable and they suffered from poor PK profiles with all analogs having <5% oral bioavailability in rats. The GHS-R is a GPCR, which are well-known for their promiscuity for lipophilic ligands. The poor quality of the hit compound and the promiscuity of the target justified a scaffold modification. The hydrophobic portion in **7**, the phenylisoxazole ring, presented a good motif for topology change. SAR suggested the dihedral angle between the phenyl ring and the isoxazole ring could be further optimized. A tetralin carboxamide scaffold was designed and the follow-up SAR studies identified a series of new GHS-R antagonists with a much improved patent position and better PK profiles, as exemplified by compound **8**.

Recently, Wood *et al.* [35] reported an interesting scaffold modification where a 2,3-diaminopyridine motif was successfully replaced with a cyclopropylamino acid amide group (Figure 4) in a series of bradykinin B_1 receptor antagonists. The electron-rich diaminopyridine ring in compound **9** was metabolically unstable and the reactive metabolites could complicate further development of analogs of **9**. Thus, compound **10** was designed and was only approximately fivefold less active than the parent compound **9**. Furthermore, a more active analog **11** was identified based on the new scaffold (Figure 4). Compound **10** also demonstrated improved PK properties over **9** (Figure 4). This imaginative operation was carefully designed. The oxygen atom in the carbonyl group in **10** resembled the pyridine nitrogen atom in **9**. The carbonyl group in **10** also served to retain the electronic and stereoelectronic characters of the

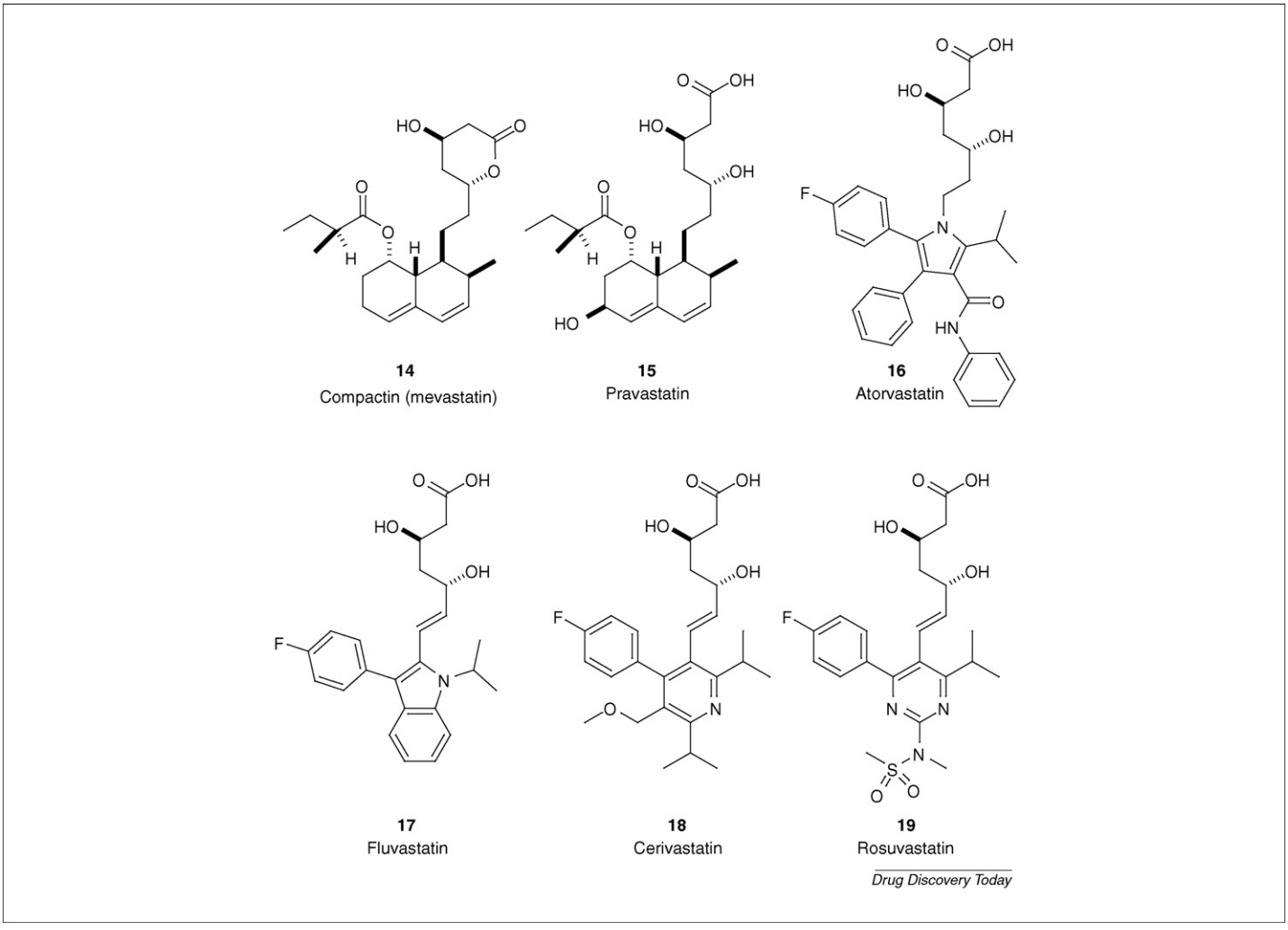

FIGURE 5

Identification of synthetic statins by scaffold hopping.

benzylic amine. The ethylene group in the cyclopropane ring was installed to control the local conformation. Interestingly, the larger ring systems (e.g. 4–6-membered rings) led to less-active analogs because the partially sp^2-hybridized carbon atoms in the cyclopropane ring more closely resembled the sp^2-hybridized carbons in the pyridine ring in several fundamental aspects. The favorable hyperconjugative overlap between the carbonyl π bond and the cyclopropyl C–C bond orbitals, and the strain of the cycloproane ring, forced **10** and **11** to adopt a conformation almost identical to that of **9**. Sterically, the cyclopropane ring is also closest to the planar pyridine ring. Because of these fundamental similarities between the two scaffolds, Wood *et al.* suggested the operation (i.e. **9** to **10**) could be of broad use. Indeed, the same operation was performed on a series of factor Xa inhibitors and similar results were observed (Figure 4). Inhibitor **13** was only fivefold less active than the parent compound **12**.

A classic example of scaffold hopping is the discovery of several synthetic 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) inhibitors – statins – based on the natural product compactin (**14**; Figure 5). It was determined that compounds **14**,**16**–**19** all bind to HMGR in a similar fashion [36]. The conformational flexibility of the enzyme enabled the formation of a promiscuous hydrophobic pocket near the active site, to which the hydrophobic portion of the structurally diverse statins binds. ITC measurements [37] revealed that the binding for **15**–**18** was entropy-driven: entropy contributed >50% to binding free energy at 25 °C. This suggested the hydrophobic groups provided important interactions with the enzyme for these statins. A unique polar interaction between the sulfone group in rosuvastatin (**19**) and Arg568 of HMGR rendered the binding to be enthalpy-driven: enthalpy contributed 76% to the binding free energy at 25 °C. The discovery of the synthetic statins highlighted the importance of exploiting promiscuous binding pockets via scaffold hopping to develop differentiating products. Lipitor®, the calcium salt of atorvastatin (**16**), went on to become the all time best seller despite its late entrance to the market.

Although scaffold modification is often useful or even crucial for successful lead generation, in many cases high-quality leads can be identified without resorting to this risky approach. One needs to inspect the quality of the hit compounds and the nature of the target protein so the necessity and success rate of this endeavor can be evaluated.

Concluding remarks

The practice of medicinal chemistry has changed dramatically in the past 25 years or so, mostly because the starting point of the drug discovery process shifted toward lower-quality screening hits. Selection of compounds with the best developmental potential is important for the success of the more resource-intensive lead-optimization phase. Strategies to facilitate this process have been formulated and new knowledge and expertise have been generated in the past decade. Today, medicinal chemists spend much of their time in transforming screening hits to lead-quality compounds. The throughput of synthesis and biological assays enables a large number analogs to be synthesized and tested quickly and more chemistry effort could be devoted to scaffold modifications. The identification of lead compounds containing an optimal scaffold, through either direct selection from screening hits or scaffold hopping, is highly consequential to the success of lead optimization and should be a main focus of the lead-generation effort.

Acknowledgements

I would like to thank Dr Cele Abad-Zapatero for insightful discussions in preparing this manuscript.

References

1 Lombardino, J.G. and Lowe, J.A. (2004) The role of the medicinal chemist in drug discovery — then and now. *Nat. Rev. Drug Discov.* 3, 853–862
2 Bleicher, K.H. *et al.* (2003) Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discov.* 2, 369–378
3 Davis, A.M. *et al.* (2005) Components of successful lead generation. *Curr. Top. Med. Chem.* 5, 121–139
4 Alanine, A. *et al.* (2003) Lead generation - enhancing the success of drug discovery by investing in the hit to lead process. *Comb. Chem. High Throughput Screen.* 6, 51–66
5 Steinmeyer, A. (2006) The hit-to-lead process at Schering AG: strategic aspects. *ChemMedChem* 1, 31–36
6 Andrews, P.R. *et al.* (1984) Functional group contributions to drug-receptor interactions. *J. Med. Chem.* 27, 1648–1657
7 Hopkins, A.L. *et al.* (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431
8 Abad-Zapatero, C. and Metz, J.T. (2005) Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* 10, 464–469
9 Hann, M.M. and Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* 8, 255–263
10 Teague, S.J. *et al.* (1999) The design of leadlike combinatorial libraries. *Angew. Chem. Int. Ed. Engl.* 38, 3743–3748
11 Wunberg, T. *et al.* (2006) Improving the hit-to-lead process: data-driven assessment of drug-like and lead-like screening hits. *Drug Discov. Today* 11, 175–180
12 Schnecke, V. and Bostroem, J. (2006) Computational chemistry-driven decision making in lead generation. *Drug Discov. Today* 11, 43–50
13 Keseru, G.M. and Makara, G.M. (2006) Hit discovery and hit-to-lead approaches. *Drug Discov. Today* 11, 741–748
14 Bemis, G.W. and Murcko, M.A. (1996) The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* 39, 2887–2893
15 Schuffenhauer, A. *et al.* (2006) Relationships between molecular complexity, biological Activity, and structural diversity. *J. Chem. Inf. Model.* 46, 525–535
16 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864
17 Vajda, S. and Guarnieri, F. (2006) Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr. Opin. Drug Discov. Devel.* 8, 354–362
18 Mestres, J. and Veeneman, G. (2003) Identification of ''latent hits'' in compound screening collections. *J. Med. Chem.* 46, 3441–3444
19 Liu, G. *et al.* (2003) Discovery and structure-activity relationship of oxalylarylaminobenzoic acids as inhibitors of protein tyrosine phosphatase 1B. *J. Med. Chem.* 46, 2093–2103
20 Zhao, H. *et al.* (2004) Isoxazole carboxylic acids as protein tyrosine phosphatase 1B (PTP1B) inhibitors. *Bioorg. Med. Chem. Lett.* 14, 5543–5546
21 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25
22 Campoy, A.V. and Freire, E. (2005) ITC in the post-genomic era...? Priceless. *Biophys. Chem.* 115, 115–124
23 Wilson, R.M. and Danishefsky, S.J. (2006) Small molecule natural products in the discovery of therapeutic agents: the synthesis connection. *J. Org. Chem.* 71, 8329–8351
24 Rivkin, A. *et al.* (2005) On the remarkable antitumor properties of fludelone: how we got there. *Angew. Chem. Int. Ed. Engl.* 44, 2838–2850
25 Proudfoot, J.R. (2002) Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg. Med. Chem. Lett.* 12, 1647–1650
26 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730

27 Hajduk, P.J. *et al.* (2005) Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* 48, 2518–2525
28 Hopkins, A.L. *et al.* (2006) Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol* 16, 127–136
29 De Clercq, E. (1998) The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral Res.* 38, 153–159
30 Boehm, H.J. *et al.* (2004) Scaffold hopping. *Drug Discov. Today: Technol.* 1, 217–224
31 Lloyd, D.G. *et al.* (2004) Scaffold popping in *de novo* design. ligand generation in the absence of receptor information. *J. Med. Chem.* 47, 493–496
32 Schneider, G. *et al.* (1999) ''Scaffold-hopping'' by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed. Engl.* 38, 2894–2896
33 Kier, L.B. and Hall, L.H. (2004) Bioisosterism: quantitation of structure and property effects. *Chem. Biodivers.* 1, 138–151
34 Zhao, H. *et al.* (2004) Discovery of tetralin carboxamide growth hormone secretagogue receptor antagonists via scaffold manipulation. *J. Med. Chem.* 47, 6655–6657
35 Wood, M.R. *et al.* (2006) Cyclopropylamino acid amide as a pharmacophoric replacement for 2,3-diaminopyridine. application to the design of novel bradykinin B1 receptor antagonists. *J. Med. Chem.* 49, 1231–1234
36 Istvan, E.S. and Deisenhofer, J. (2001) Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* 292, 1160–1164
37 Carbonell, T. and Freire, E. (2005) Binding thermodynamics of statins to HMG-CoA reductase. *Biochemistry* 44, 11741–11748