

Part I: Concept and Theory

1

The Concept of Fragment-based Drug Discovery

Daniel A. Erlanson and Wolfgang Jahnke

1.1

Introduction

Fragment-based drug discovery builds drugs from small molecular pieces. It combines the empiricism of random screening with the rationality of structure-based design. Though the concept was articulated decades ago, the approach has become practical only recently.

Historically, most drugs have been discovered by one of two methods. The first of these was famously summarized by Nobel Laureate Sir James Black, who noted that the best way to find a new drug is to start with an existing one. Indeed, any successful drug spawns a surge of similar molecules, as illustrated by the number of chemically similar COX-2 inhibitors or HIV protease inhibitors on the market and in development. Though often disparaged as “me-too” or “patent-busting”, such efforts are productive. The first drug to market is rarely the best; one need only consider the state of HIV medication now compared to a decade ago to appreciate this fact. Even the search for new drugs often begins with known starting points in the form of natural ligands such as substrates, co-factors or inhibitors.

For diseases and targets where no drug or other starting point exists, the second major route of drug discovery, random screening, is essential. This approach to drug discovery is perhaps the oldest and most venerable but requires serendipity. Indeed, it was a serendipitous observation of bacterial killing by fungus that led Alexander Fleming to the discovery of the natural product penicillin. Many highly successful drugs, from cyclosporine to paclitaxel, have been discovered by screening collections of compounds. With each medicinal chemistry program, more chemical compounds and their analogs are added to corporate screening libraries.

The invention of combinatorial chemistry in the late 1980s and early 1990s vastly expanded the number of compounds in chemical collections, just as the development of sophisticated automation equipment and miniaturization of biological assays led to the advent of high-throughput screening, or HTS. Today, most major pharmaceutical companies and many biotechnology companies have in-house collections of hundreds of thousands or even millions of molecules.

In parallel to HTS, more rational routes for drug discovery have been sought. Structure-based drug design attempts to design inhibitors *in silico* on the basis of the three-dimensional structure of the target protein.

Among the latest developments in drug discovery is a concept called fragment-based drug design, or fragment-based screening (FBS). In contrast to conventional HTS, where fully built, “drug-sized” chemical compounds are screened for activity, FBS identifies very small chemical structures (“fragments”) that may only exhibit weak binding affinity. Follow-up strategies are then applied to increase affinity by elaborating these minimal binding elements. Fragment-based drug design thus attempts to build a ligand piece-by-piece, in a modular fashion. Structural information plays a central role in most follow-up strategies. Therefore, fragment-based drug design can be viewed as the synthesis of random screening and structure-based design.

1.2

Starting Small: Key Features of Fragment-based Ligand Design

Fragment-based screening promises to have a great impact on drug discovery because of several advantages, which are summarized in the following sections.

1.2.1

FBS Samples Higher Chemical Diversity

Typical chemical libraries used for HTS contain 10^5 to 10^6 individual compounds. Though a million-compound library sounds vast, it covers only a very small portion of “drug space”, the theoretical set of possible small, drug-like molecules. In fact, a widely quoted estimate (actually a back-of-the-envelope calculation in a footnote in a review of structure-based drug design) places this number at 10^{63} molecules [1], a number beyond the comprehension of anyone except perhaps astrophysicists. A recent estimate of the total number of molecules available for screening in all the commercial and academic institutions on the Earth is around 100 million, or 10^8 , so even a planet-wide screening effort would not even scratch the surface of diversity space [2]. This will never change in any meaningful way. To understand why, imagine assembling a library of 10^{63} molecules. Even if miniaturization advances to the point where we need only 1 pmol of each molecule (about 0.5 ng for a 500-Da molecule), this would still require gathering 5×10^{47} tons of material, roughly 26 orders of magnitude larger than the mass of our planet. Clearly, libraries screened in HTS will always explore only a tiny fraction of drug space.

The explored fraction of diversity space swells when working with smaller molecules (“fragments”), because there are fewer possible small molecules than possible large molecules. If we screen small molecular fragments, rather than drug-sized molecules, we can cover exponentially larger swaths of diversity space with much smaller collections of molecules. To illustrate, imagine two sets of compounds, each

consisting of 1000 fragments. If we were to exhaustively make all binary combinations with a single asymmetric linker, this would yield $(1000 \text{ molecules}) \times (1000 \text{ molecules}) = 1\,000\,000$ molecules to synthesize and screen, a daunting task. In contrast, if we could identify the five best fragments in each set and only combine and screen those, we would only need to synthesize and test $[(1000 \text{ molecules}) + (1000 \text{ molecules})] + [(5 \text{ molecules}) \times (5 \text{ molecules})] = 2025$ molecules. This number is clearly much more manageable, and still covers the same chemical diversity space.

A first-principles computational analysis suggests that there are roughly 13.9×10^6 stable, synthetically feasible small molecules with a molecular weight less than or equal to 160 Da (44×10^6 once stereoisomers are considered, although the approach excludes compounds containing three- and four-membered rings and elements other than carbon, hydrogen, oxygen, nitrogen, and halogens) [3]. This is still a large number, but it is at least a comprehensible number, especially compared with 10^{63} . It shows that, with fragment-based screening, a higher (although still very small) proportion of diverse drug space can be covered. From a technical standpoint as well, focusing on these smaller fragments could simplify many aspects of the drug discovery process, from compound acquisition and synthesis through data management.

1.2.2

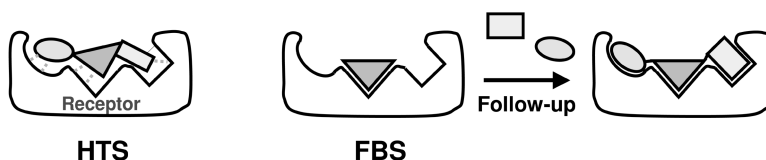
FBS Leads to Higher Hit Rates

Imagine a small fragment with high but imperfect complementarity to a target protein. Now imagine adding a methyl group at exactly the right spot to increase complementarity even further: rendering the fragment more complex in the right manner leads to slightly increased affinity to the target protein. But imagine adding the methyl group at any other spot, so that it protrudes from this fragment towards the receptor such that the modified fragment can no longer bind to the target: rendering the fragment more complex in the “wrong” manner ablates affinity for the receptor. Notably, there are many more ways to increase complexity in the “wrong” manner, and doing so often leads to a decrease of binding affinity by several orders of magnitude, whereas in the lucky case of increasing complexity in the “right” manner, binding is generally only enhanced by one or two orders of magnitude. This simple example makes sense intuitively, and a more rigorous theoretical analysis comes to the same conclusion: as molecules become more complex, additional chemical groups are much more likely to ablate binding than to enhance it [4]. The probability of binding (the “hit rate” in screening) thus decreases with increasing ligand complexity. Libraries containing smaller compounds (“fragments”) are expected to exhibit higher hit rates, although the resulting affinities are generally weak and so require sensitive detection methods.

1.2.3

FBS Leads to Higher Ligand Efficiency

Screening drug-sized molecules is thought to favor ligands with several sub-optimal binding interactions, rather than those with a few optimal interactions. This is schematically shown in Fig. 1.1: the drug-sized molecule on the left side is identified by HTS since it binds to the receptor. However, none of the binding interactions are optimal, since establishing one optimal interaction would disrupt another interaction. All binding interactions are thus compromised and do not retain the full strength they would have without the molecular strain.

**Fig. 1.1**

Potential drawback of HTS (left), and principle and advantages of FBS (right): In HTS, fully assembled, “drug-sized” ligands are identified, but with multiple compromised, non-optimal binding interactions. In FBS, ligands for individual subpockets are identified separately, and show few but good binding interactions. Follow-up strategies such as fragment elaboration or linking are used to increase ligand affinity.

Relative to their molecular size, fragments can thus show more favorable binding energies than drug-sized molecules. The binding energy, normalized by the number of heavy atoms in the ligand, is referred to by the term ligand efficiency [5]. Smaller fragments can have higher ligand efficiency, leading to smaller drugs with better chances for favorable pharmacokinetics [6, 7]. This concept is also being applied to conventional HTS with the advent of “lead-like”, instead of “drug-like,” compound libraries [8].

1.3**Historical Development**

The basic concept of fragment-based drug discovery was developed about 25 years ago by William Jencks, who wrote in 1981 that the affinities of whole molecules could be understood as a function of the affinities of separate parts:

“It can be useful to describe the Gibbs free energy changes for the binding to a protein of a molecule, A–B, and of its component parts, A and B, in terms of the “intrinsic binding energies” of A and B (ΔG_A^i and ΔG_B^i) and a “connection Gibbs energy” (ΔG^s) that is derived largely from changes in translational and rotational entropy [9].”

This paper received considerable attention and spawned academic interest in ligand–receptor interactions. Nakamura and Abeles studied the inhibition of HMG-CoA reductase by small molecule inhibitors, and found that those inhibitors could be understood as a linkage of two fragments, each binding to distinct sites on the enzyme [10]. The paper also provided a theoretical framework for understanding a very early study in which biotin was deconstructed into component fragments, which were found to bind weakly to streptavidin [11].

Despite these developments, Jencks' formulation did not immediately have an impact on drug discovery. The practical implementation of the theoretical promise required overcoming two difficult barriers: finding fragments and linking them.

Finding weakly binding fragments is inherently difficult because the binding interactions are easily disrupted. Moreover, there are hidden hazards in looking for weak binders: apparent hits could be “false positives”. For example, compounds forming aggregates at low to mid-micromolar concentrations can inhibit biochemical functional assays without specifically interacting with the target [12–14].

But even if a true weak hit was identified, what could be done with it? Jencks provided an elegant theoretical framework for combining two weakly binding fragments into a single molecule, but enormous practical difficulties remained: First, one had to find a suitable fragment. One then had to find a second fragment; and that second fragment had to bind in close proximity to the first. Finally, one had to figure out how to link these two fragments while not distorting the binding mode of either. It was no wonder that the field remained largely theoretical and computational for well over a decade.

All this changed in 1996, when researchers at Abbott published the first practical demonstration of fragment-based drug discovery, called SAR-by-NMR [15]. In this approach, Shuker, Hajduk, Meadows, and Fesik used nuclear magnetic resonance (NMR) as a robust binding assay with sensitivity for weak interactions to identify fragments and to determine their binding sites, which revealed how to link the fragments. A flood of papers followed, initially from Abbott but soon from other research groups as well. Today, well over a dozen companies, from small biotechs to multinational pharmaceutical companies, as well as a large number of academic laboratories, are pursuing some form of fragment-based drug discovery. Many of the laboratories that have been leading the conceptual, theoretical and experimental development of FBS are represented in the following chapters. Below, we provide an overview of what to expect.

1.4

Scope and Overview of This Book

The text can be roughly divided into three sections: background and computational approaches are covered in Chapters 1–7, experimental methods and applications are covered in Chapters 8–14, and the last two Chapters, 15 and 16, describe related and emerging fields in chemistry that have the potential to inform and transform fragment-based drug discovery.

An intellectual sister to fragment-based drug discovery is the concept of multivalency, which is covered in a comprehensive review by Krishnamurthy, Estroff, and Whitesides in Chapter 2. This chapter focuses on the application of multivalency to the design of high-avidity ligands, and thus sets the conceptual framework for fragment-based drug design.

A special case of multivalency, hetero-oligovalency, is often encountered in fragment-based drug design when two ligands that bind to adjacent binding pockets are linked to form a high-affinity ligand. Murray and Verdonk discuss the entropic effects associated with this process in theory and experiment in Chapter 3. Basic concepts such as ligand efficiency and ligand hot spots are introduced in this chapter as well.

The identification and characterization of a protein binding site is key for ligand design. One way to achieve this is by mapping organic solvent binding sites in a protein, as in the multiple solvent crystal structure method, described in chapter 4 by Ringe & Mattos.

The quality of the fragment library is a crucial success factor for fragment-based drug design. The design of fragment libraries is the topic of Chapters 5 and 6. Oprea and Blaney outline the concepts of chemical space, lead-likeness, and fragment-like leads, using both *de novo* calculations and data mining in Chapter 5. This chapter illustrates the difficulty in trying to sample even a relatively small section of “fragment space”, and provides a number of specific examples to direct researchers toward the most fruitful regions.

In Chapter 6, Vieth and Siegel “dissect” existing drugs into their component fragments and demonstrate that there are considerable differences between oral and injectable drugs. This clearly has implications for the choice of fragments in a screening collection.

In the years after Jencks’ formulation of fragment-based drug discovery, before experimental methods became sufficiently sensitive to discover fragments, computational approaches were the dominant activity. In Chapter 7, Stultz and Karpus discuss the multi-copy simultaneous search (MCSS) program, one of the earliest and most powerful approaches to *in silico* fragment-based drug discovery, and its use for ligand design.

Chapter 8, by Sem, begins the “applied” section of the book and covers NMR-based approaches to fragment assembly. This chapter is a comprehensive review of the subject and covers the theory, various approaches, and specific examples.

The following chapter, by Hajduk, Huth, and Sun, discusses the original “SAR-by-NMR” approach and summarizes successes with this method. It also considers the success and requirements of fragment linking versus fragment elaboration from both a theoretical and an experimental vantage point. The authors draw important conclusions regarding the limits of these approaches as well as the sizes of libraries that should be assembled to maximize the likelihood of success in fragment elaboration approaches.

Chapters 10 and 11 focus on X-ray crystallography applications for fragment-based drug design. Davies, van Montfort, Williams, and Jhoti describe the process established at Astex by using NMR and X-ray for fragment screening, and X-ray crystallography as the basis for fragment optimization. Blaney, Nienaber, and Bur-

ley outline the crystallography-driven fragment-based ligand design at SGX and illustrate it with case studies.

Chapters 8–11 represent predominantly either NMR or X-ray crystallography approaches, while Chapter 12, by Abad-Zapatero, Stamper, and Stoll, describes the synergies that can result by marrying these techniques. The authors describe in two case studies how the combined use of these powerful biophysical techniques can rapidly advance medicinal chemistry programs.

Chapter 13 covers two somewhat unusual topics: use of mass spectrometry (MS) to identify fragments, and fragment-based discovery methods applied to an RNA target. Although the structural resolution of MS is necessarily less than that of either NMR or X-ray, Griffey and Swayze demonstrate that the technique can be powerfully applied to a challenging drug target.

A further use of MS, Tethering, is discussed in Chapter 14. This technique differs from other approaches in that it uses a transient covalent bond between the fragment and the target protein. The technology can be used to both identify fragments as well as to link two fragments. Erlanson, Ballinger, and Wells review the theory and practice of this method of fragment-based drug discovery.

Finally, the last two chapters touch on two fields that are themselves areas of vibrant research and that also overlap with fragment-based drug discovery. Chapter 15, by Röper and Kolb, introduces the powerful technique of Click chemistry, while Chapter 16, by Hochgürtel and Lehn, discusses dynamic combinatorial chemistry. Both of these approaches have been successfully applied to fragment-based drug discovery, albeit in a few limited studies. It is therefore fitting that we should end this volume here, at the intersection of emerging fields, where the opportunities are great, if only dimly perceived.

References

- 1 Bohacek, R.S., McMartin, C., Guida, W.C. **1996**, The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* 16, 3–50.
- 2 Hann, M.M., Oprea, T.I. **2004**, Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* 8, 255–263.
- 3 Fink, T., Bruggesser, H., Reymond, J.L. **2005**, Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew. Chem. Int. Ed. Engl.* 44, 1504–1508.
- 4 Hann, M.M., Leach, A.R., Harper, G. **2001**, Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864.
- 5 Hopkins, A.L., Groom, C.R., Alex, A. **2004**, Ligand efficiency: a useful metric for lead selection. *Drug. Discov. Today* 9, 430–431.
- 6 Kuntz, I.D., Chen, K., Sharp, K.A., Kollman, P.A. **1999**, The maximal affinity of ligands. *Proc. Natl Acad. Sci. USA* 96, 9997–10002.
- 7 Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. **1997**, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25.
- 8 Teague, S.J., Davis, A.M., Leeson, P.D., Oprea, T. **1999**, The design of leadlike combinatorial libraries. *Angew. Chem. Int. Ed. Engl.* 38, 3743–3748.

- 9 Jencks, W.P. 1981, On the attribution and additivity of binding energies. *Proc. Natl Acad. Sci. USA* 78, 4046–4050.
- 10 Nakamura, C.E., Abeles, R.H. 1985, Mode of interaction of beta-hydroxy-beta-methylglutaryl coenzyme A reductase with strong binding inhibitors: compactin and related compounds. *Biochemistry* 24, 1364–1376.
- 11 Green, N.M. 1975, Avidin. *Adv. Protein Chem.* 29, 85–133.
- 12 McGovern, S.L., Caselli, E., Grigorieff, N., Shoichet, B.K. 2002, A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* 45, 1712–1722.
- 13 McGovern, S.L., Helfand, B.T., Feng, B., Shoichet, B.K. 2003, A specific mechanism of nonspecific inhibition. *J. Med. Chem.* 46, 4265–4272.
- 14 Ryan, A.J., Gray, N.M., Lowe, P.N., Chung, C.W. 2003, Effect of detergent on “promiscuous” inhibitors. *J. Med. Chem.* 46, 3448–3451.
- 15 Shuker, S.B., Hajduk, P.J., Meadows, R.P., Fesik, S.W. 1996, Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531–1534.