The FAST and the curious

Fragment-based lead discovery and optimization via high-throughput X-ray crystallography

BY STEPHEN K. BURLEY

More than a decade ago, most of the large pharmaceutical companies used X-ray crystallography because of its promise to deliver "rationally designed" drugs. Unfortunately, this anticipated revolution in drug discovery failed to materialize. Earlier incarnations of the X-ray method usually proved to be too slow to keep up with the pace of medicinal chemistry. In addition, computational strategies based on the structure of the protein target alone did not permit rational design of small-molecule inhibitors of biochemical and biological function.

Two other would-be revolutionary methods, combinatorial chemistry and high-throughput screening (HTS), came along later, both of which also have failed to live up to expectations. The advent of combinatorial chemistry did allow chemists to make greater quantities of molecules quickly, but these compounds have rarely displayed the druglike properties needed to make it through preclinical testing and into patients. HTS, which was developed to take advantage of the chemical bounty provided by combinatorial approaches, also has failed to deliver useful lead compounds for many targets.

Until the advent of high-brightness synchrotron sources and cryogenic sample-preservation methods in the 1990s, protein structures were attainable only after lengthy and resource-intensive efforts, greatly limiting their contributions to drug discovery. More recently, Vicki Nienaber and colleagues at Abbott Laboratories have demonstrated the feasibility of using X-ray crystallography to screen for drug discovery starting points (*1*). Despite its promise, crystallographic screening has not been adopted widely because the approach depends critically on having high-quality crystals of the target protein early in a drug discovery campaign, producing large numbers of target–small-molecule complex structures in a time frame compatible with the needs of the medicinal chemist, and establishing a reliable, efficient strategy for exploiting structural information to discover and optimize drug candidates. Today, high-throughput X-ray crystallography, advanced computational tools, and parallel organic synthesis have been combined to meet these challenges and provide new opportunities for lead discovery and optimization.

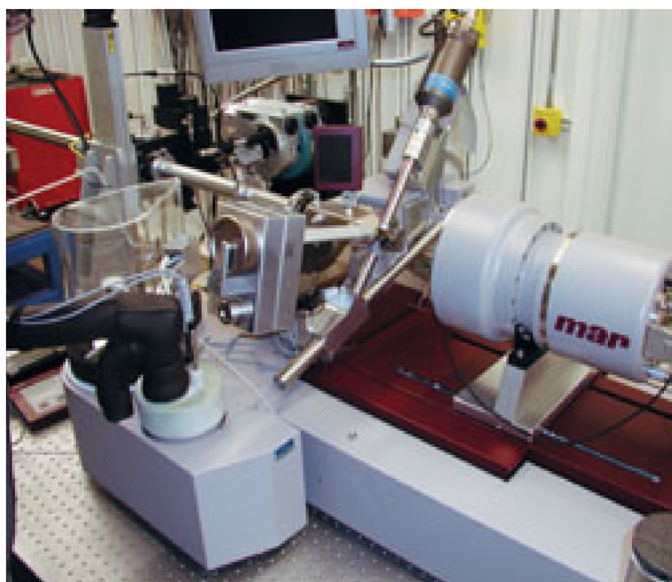

Figure 1. Looking for targets. The Mar Research automated cryogenic sample changer allows unattended diffraction data collection on the SGX beamline at the Advanced Photon Source and high-throughput crystallography in support of FAST. (Courtesy of Structural GenomiX, Inc.)

Fragment hit discovery

In response to the need to reduce discovery timelines and increase the rate at which new chemical entities are brought to market, Structural GenomiX, Inc. (SGX), has developed a novel technology called fragments of active structures (FAST) to serve as an engine for high-quality lead generation (Figure 1). Technical advances have reduced many of the historical barriers, and this new approach exploits the strengths of X-ray crystallographic screening, structure-based lead optimization, computational design, and combinatorial chemistry. For crystallographic screening to deliver maximum value, fragment hits must be suitable for rapid, iterative "hit-to-lead" optimization. The FAST process is based on a library of small, leadlike fragments that are fully enabled for chemical modification.

The SGX fragment-based approach begins with crystallographic screening of a target protein against the FAST library of approximately 1000 small leadlike fragments (MW 150–200 Da). Each fragment has been tailored to include multiple points of diversity, or chemical "handles", which are amenable to rapid parallel synthesis. Fully 50% of the fragments in the library contain one or more bromine atoms, which facilitate both carbon–carbon bond formation and an unambiguous understanding of the precise geometry of fragment binding from the X-ray structures coming from crystallographic screening. The entire library can be screened in 1–2 days using a dedicated synchrotron beamline (SGX-CAT, Advanced Photon Source, *www.aps.anl.gov*).

Crystallographic screening entails soaking preformed protein crystals with mixtures of structurally dissimilar fragments, followed by X-ray diffraction analysis. This approach yields hit rates of 1–5%, which are significantly higher than the 0.010–0.025% typical for traditional HTS. The resulting structures provide "direct looks" at each of the 10–50 leadlike fragments that bind to the crystalline target protein. Bound fragments are evaluated in terms of the binding mode

© 2004 AMERICAN CHEMICAL SOCIETY

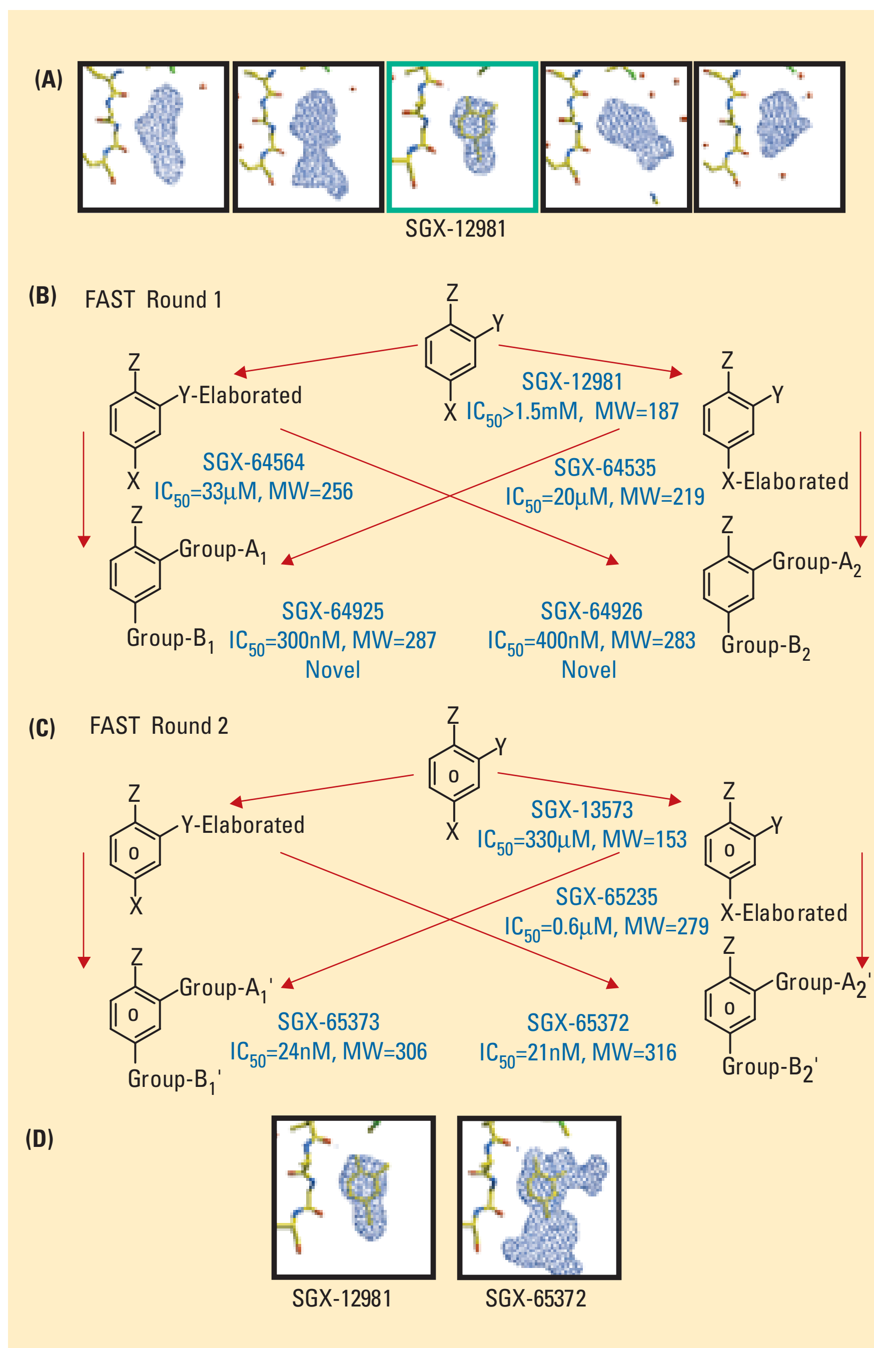

Figure 2. Inflammation inhibitor: FAST lead discovery and optimization with spleen tyrosine kinase (Syk).

(A) Crystallographic screening of Syk yielded five FAST lead discovery hits. A portion of the Syk structure is shown as an atomic stick figure (yellow, carbon; blue, nitrogen; red, oxygen; green, sulfur). The experimental electron density obtained for each FAST hit is shown as a blue mesh outline. The third panel depicts SGX-12981 in yellow (no atom color coding).

(B) Results from the first round of FAST hit optimization applied to SGX-12981. X, Y, and Z denote points of chemical diversity. Synthetic elaborations at points X and Y yielded two compounds with submicromolar affinity (SGX-64925 and SGX-64926).

(C) Results from the second trial of FAST hit optimization applied to SGX-13573, which is closely related in chemical structure to SGX-12981. X, Y, and Z denote points of chemical diversity. Synthetic elaborations at points X and Y yielded two compounds with ~20 nM affinity (SGX-65372 and SGX-65373).

(D) Structure and binding mode of SGX-12981 (left) and SGX-65372 (right). A portion of the Syk structure and SGX-65372 are shown as an atomic stick figure. The experimental electron density obtained for each bound ligand is shown as a blue mesh.

and measured IC_{50} (typically 500 µM–1 mM) to identify the most promising 4–5 candidates for lead optimization.

Elaborating the hits

With detailed insights from the structure of each high-priority fragment bound to the target protein, small one-dimensional libraries are designed by making modifications at each site of chemical diversity. By the application of an advanced computational filter, chemical modifications predicted to yield undesirable steric clashes with the target are excluded, thereby eliminating application of expensive synthetic resources to changes unlikely to increase potency and selectivity. Each one-dimensional library is synthesized and then evaluated using in vitro biochemical assays. Compounds with better IC_{50} values than the starting point or parent fragment (typically 10–100 µM) are then examined by X-ray crystallography to visualize the effect of each chemical modification on binding and to find opportunities for further synthetic chemistry.

The most promising substitutions at each chemical "handle" are combined in small combinatorial (two- or three-dimensional) libraries of modifications at each site of chemical diversity. Each combinatorial library is synthesized and then evaluated using in vitro biochemical assays and X-ray crystallography. At this stage, the best multiply elaborated fragments have IC_{50} values of 10–100 nM and often display considerable selectivity against closely related proteins. Most targets yield at least two structurally distinct lead series that can be evaluated for activity in cells and animal models of disease.

Typical molecular weights of doubly and triply elaborated fragments are <350 Da, which represents an important factor in determining a drug candidate's bioavailability. Similar attention is paid to other determinants of bioavailability, such as solubility, lipophilicity, the number of freely rotatable bonds, and the number of hydrogen-bond donors and acceptors (*2*).

Although the FAST fragment library consists of approximately 1000 compounds, the FAST process's combinatorial nature has unrivaled chemical diversity potential and typically yields novel compounds after only two rounds of chemical synthesis. The number of elaborations possible at each chem-

ical diversity site ranges from 400 to 40,000. Thus, the worst-case scenario of 400 possible modifications at each of two chemical "handles" gives a minimum chemical diversity of 160,000 distinct compounds per fragment, or 160 million distinct compounds from the entire FAST library. Throughout the hit optimization process, fragment modifications generating previously patented compounds are excluded, and computational filters ensure that synthetic chemistry resources are used to explore only novel regions of chemical shape space.

Inflammation inhibitors

FAST has been applied to spleen tyrosine kinase (Syk), an inflammatory disease target that controls degranulation of mast cells in asthma and other inflammatory illnesses. Crystallographic screening with the FAST fragment library yielded five high-priority hits with distinct chemical structures bound in the enzyme's ATP-binding pocket of the enzyme (Figure 2A). One hit, SGX-12981, with IC_{50} > 1.5 mM, is chosen for optimization. Detailed structural analyses revealed two points of chemical diversity available for synthetic elaboration, denoted X and Y. A third site of diversity (Z) in this single-ring heterocycle is an NH_2 group that makes a hydrogen bond with the protein. Following computational filtering of the binding mode and the immediate environments of "handles" X and Y, two small linear libraries were designed. Approximately 20 analogues were synthesized using "handles" X and Y, yielding several compounds with improved binding affinities (IC_{50} = 20–30 µM, Figure 2B). Combining favorable substituents at both positions X and Y yielded doubly elaborated fragments with submicromolar affinities (IC_{50} = 300–400 nM, Figure 2B).

A similarity search identified a homologous heterocycle with an NH_2 group at position Z that was predicted to make a stronger hydrogen bond with the ATP-binding pocket of Syk. X-ray crystallography revealed that this new compound, SGX-13573, bound to Syk in the same manner as SGX-12981, demonstrating the discovery of a more potent fragment with a similar mode of action. Thereafter, a small linear library is designed and synthesized to identify the best substituents at position X on this new fragment. The best singly elaborated fragment gave an IC_{50} value of 600 nM (SGX-65235). Combining optimal substituents at position X with the best substituents at position Y, identified previously with SGX-12981, gave two doubly elaborated compounds, SGX-65372 and SGX-65373, with IC_{50} ~ 20 nM (Figure 2C). These two compounds represent novel kinase inhibitors, and SGX has filed patent applications.

Thus, the SGX-12981 structure bound to Syk (Figure 2D) and subsequent crystallographic results guided the selection of optimal chemical modifications at two sites of chemical diversity, resulting in SGX-65372, a compound with modest molecular weight (~300 Da) and substantially improved potency (Figure 1D, IC_{50} = 21 nM). Cell-based assays have demonstrated SGX-65372 inhibition of Syk target phosphorylation in human B cells and inhibition of rat basophil degranulation. In vivo ADME studies have shown that SGX-65372 is bioavailable and has an acceptable half-life in rats, and in vitro toxicity studies have not detected inhibition of either cytochrome P450s or the hERG potassium channel. Target-profiling studies with SGX-65372 have documented strong target selection for Syk compared to Zap70, the closest neighbor to Syk in the human kinome (*3*), and other tyrosine kinases. Together, these data demonstrate that SGX-65372 represents an advanced lead compound that is suitable for additional medicinal chemistry optimization and further preclinical evaluation.

Changing lead discovery

The FAST process is a highly efficient and adaptable lead discovery strategy for studying large numbers of X-ray structures of lead-like fragments bound to important disease targets. These compounds serve as the building blocks for novel drug candidates by guiding further modifications of initial screening hits. The method exploits high-throughput X-ray crystallography to identify and visualize synthetically enabled fragments productively bound to kinases and other enzymes. Thereafter, the binding mode of each fragment to its target is combined with advanced computational tools to maximally leverage parallel organic synthesis to produce "engineered" leads that are potent, are reasonably selective, and have druglike physicochemical properties. The speed with which a FAST campaign can be accomplished is typically six months, and access to an enormous number of diverse, novel compounds makes it both competitive with and complementary to HTS. FAST may also provide a useful means of approaching targets, such as proteases, for which HTS often fails to yield acceptable medicinal chemistry starting points. This would not be possible without dedicated, full-time access to a high-brightness synchrotron beamline, an optimally designed fragment library, and advanced computational chemistry software and hardware, such as that developed by SGX.

References

(1) Nienaber, V. L.; et al. *Nat. Biotechnol.* **2000,** *18,* 1105–1108.
(2) Wenlock, M. C.; et al. *J. Med. Chem.* **2000**, *46,* 1250–1256.
(3) Manning, G.; et al. *Science* **2002**, *298,* 1912–1934.

Stephen K. Burley is the CSO and senior vice president of research at Structural GenomiX, Inc. (*www.stromix.com*), and adjunct professor at The Rockefeller University (*www.rockefeller.edu*). ■

Comparison of FAST and HTS combined with lead optimization

	FAST	HTS/lead optimization
Screening library size	>160 million compounds	~2 million compounds
Hit rate	1–5%	0.010–0.025%
Lead IC_{50}/lead MW	<100 nM/~350 Da	<100 nM/~500 Da
Novelty potential	High	Limited by HTS library
Time required	3–6 months	8–15 months
Druglike properties	Built in at the outset	Addressed later
Target limitations	Crystallographically enabled targets required (not applicable to G-protein coupled receptors and ion channels at present)	Many targets (e.g., proteases) fail to produce useful starting points for lead optimization