

February 13, 2008 | Bio-IT World > The Search for Unusual Suspects

The Search for Unusual Suspects

Scaffold hopping expands the range of core molecular shapes for lead generation.

By Vicki Glaser

Feb. 1, 2008 | In the search for novel intellectual property (IP) that can maximize product life, medicinal chemists are embracing fragment-based design strategies. These can overcome some of the limitations of high-throughput screening (HTS) and “leverage information in the literature,” says Kenneth Foreman, a computational chemist at OSI Pharmaceuticals.

If a compound has a desired activity, ideally with in vivo proof of concept, and “if we can take that compound and make something that does essentially the same thing that we own, we can leapfrog into a competitive position,” says Foreman.

Although medicinal chemists have utilized such strategies for years, decisions were often based on assumptions about the core structure. Today, the literature reveals a trend toward smaller, more focused libraries, and fragments rather than full-fledged compounds as the starting point for lead generation. Increasingly there is a focus on the design of ligands that have a desired biological activity, using an approach called scaffold hopping — or, alternatively, ligand or core hopping. Ironically, by starting small, researchers have been able to expand the searchable chemical space and are delivering a broader variety of novel chemical structures that can serve as the basis for large-scale in silico screens for both lead discovery and lead optimization.

Supporting the predictions of core structure is one way in which current computational tools for scaffold hopping can have an impact — using the scientific literature to give medicinal chemists a head start. “The tools are helpful by being unbiased toward which fragments are important and which are not,” says Foreman. “The more ligand information you have, the more likely that you will be able to pick out the key pharmacophoric elements.”

Additionally, “the lack of bias in these computational tools can aid in converting [HTS] hits that have certain undesirable core features into new, IP-able cores,” says Foreman.

Software tools can help identify the “key generic interacting elements—the rings, donors, and acceptors”—and use this information to match compounds from either an internal database or an external vendor collection to the reference compound, he explains.

What has increased the feasibility of scaffold hopping in a realistic timeframe is the combination of ever-increasing computer processing speed and a more mature knowledge base regarding structure- and ligand-based design.

“Running these approaches in parallel—for example, running a shape-matching, ligand-based approach at the same time as a docking, structure-based approach—and looking for the commonalities,” can identify fragments and compounds that meet multiple criteria, says Mehran Jalaie, a computational chemist at Pfizer. The combination of 2-D strategies and shape-based, 3-D approaches to explore the same chemical space allows chemists to pick up molecules or fragments that look fundamentally different but share pharmacophoric features.

Furthermore, today’s faster computer speeds are enabling parallel searches based on shape-matching features and electrostatic properties. “This is allowing us to look at molecules in a generic way and identify new molecules that fit the desired paradigm,” says Jalaie. As processing speeds increase, the software tools will soon be able to search a larger chemical space resulting in an almost interactive process.

Varied Architecture

The subject of lead generation and structure-based small molecule drug design traditionally conjures images of a lock-and-key configuration, with the goal of generating a physical or in silico model of a drug target and a custom-designed molecule that can bind within its active site. However, over the years, computer-assisted modeling strategies driven by docking strategies have met with limited success in boosting drug discovery productivity.

By the 1990s, when combinatorial library synthesis surged to the forefront, the emphasis shifted from target structure-based design to ligand design and the production of large compound libraries intended to probe a more diverse sample of chemical space. This approach also proved less successful than anticipated, but along the way medicinal chemists started thinking about molecules as fragments — as a collection of sub-structures.

The term “scaffold hopping” was coined by former Hoffmann-La Roche researcher Gisbert Schneider. “It defines the techniques used to identify isofunctional molecules — molecules that have the same bioactivity but different architecture — in other words, different chemotypes,” says Schneider, the Beilstein Endowed Chair for Cheminformatics at Johann Wolfgang Goethe-University in Frankfurt, Germany.

As an example, consider β -lactam, a patented scaffold found at the core of many current antibiotics. Scaffold hopping offers a route to avoid IP conflicts as well as perhaps the side effects associated with β -lactams to find bioisosteric replacements — compounds that mimic the primary activity of β -lactams but have a different chemical structure. The goal is describing molecules “not on an atom-and-bond level, but on a more abstract, conceptual level that ignores chemical structure,” says Schneider.

Conventional medicinal chemistry suffers from a relatively narrow focus on a comfort zone of familiar synthetic chemistry — narrow at least in comparison to the whole of chemical space. A few years ago, an analysis in the *Journal of Medicinal Chemistry* of the molecular frameworks of known drugs revealed limited diversity in terms of compound shape. Half of the 5,120 drug compounds evaluated shared only 32 different frameworks. Schneider and Kristina Grabowski, writing last year in *Current Chemical Biology*, advocate exploring unique molecular frameworks that “complement ‘drug space’” based on their computer-based analysis and comparison of the architecture of drugs and natural products. Their analysis identified more than 1,000 scaffolds not found in a survey of drug compound libraries.

One challenge in scaffold hopping is that chemists are unlikely to venture too far outside the chemistry space they know best. Products such as Tripos’s AllChem and BioSolveIT’s FeatureTrees help remove that obstacle.

BioSolveIT developed the FTrees tool for “fuzzy similarity searching” to facilitate virtual HTS. Feature Tree, its underlying topological descriptor, captures connectivity and physico-chemical properties of functional groups. An FTrees alignment defines the optimal similarity of two descriptors, enabling SAR detection. According to the company, FTrees can search a catalog of 60,000 compounds in 15 seconds on a standard PC. For de novo design via fragment space searching, the software can process 1018 compounds in about five minutes. Pfizer has employed this technology to screen its

virtual combichem collection of about 3 billion compounds using FTrees Fragment Space (FTreesFS

) to perform similarity searches.

Courting Chemists

Some chemists may be reluctant to embrace core hopping due to the nature of its output. Scaffold searching can yield molecules that look quite different from those chemists typically synthesize, raising questions about the wisdom of investing time into producing unfamiliar compounds. This reluctance may be reasonable if based on concerns of synthetic feasibility, as scaffold hopping and in silico de novo design strategies typically do not take synthetic tractability into account. Schneider's group alleviates this concern by constructing new molecules based on building blocks derived from existing drugs or natural products. "We produce chimeras of these fragments on the computer, and because these scaffolds are drug-like they are better accepted by the chemists," he says.

"There is an inherent conflict between what IT says to make versus the cost and time" a medicinal chemist must invest to produce a compound, says Tripos CSO Richard Cramer. Tripos intends for AllChem, a product under development, to improve productivity at this juncture. It incorporates a chemistry engine that generates the most accessible and pharmaceutically acceptable scaffolds and R-groups based on established sets of feasible reactions and easily obtainable building blocks. "The resulting database enables researchers to search a chemical space of at least 1020 structures, searching for novelty while at the same time staying grounded in synthetic reality and providing chemists with ready-to-use, computed synthesis routes," says Cramer.

Patent space is becoming increasingly crowded, notes Christian Lemmen, CEO of BioSolveIT. "If a lead dies, for whatever reason, then [often] a whole series dies altogether, which generates the need for sufficient diversity in the pipeline," says Lemmen.

Activity Focus

Unlike traditional medicinal chemistry, which tends to generate analogues of active compounds, core hopping searches for novel scaffolds while preserving the activity of a potential lead. It utilizes computational tools that focus on properties relevant for binding rather than on chemical structure.

Fragment-based design strategies can be employed to enable a scaffold hop. One of the main challenges for in silico ligand design is "the accurate prediction of binding affinities and the ability to detect low-affinity binders, which is often the case for fragments," says Lemmen. BioSolveIT's ReCore is a ligand-based design tool that "abstracts from the ligand structure by taking only the vectors connecting the core piece to its R-groups," explains Lemmen.

As its name infers, ReCore targets the core of a molecule for removal, searches 3D-fragment libraries for a suitable replacement, and generates a new scaffold, maintaining the surrounding components to create a chemically distinct query compound. ReCore was developed by Patrick Maass, at the Center for Bioinformatics, University of Hamburg, in collaboration with Hoffmann-La Roche. The developers recently described its use for scaffold hopping based on small-molecule crystal structure conformations (J. Chem. Inf. Model. 2007; doi: 10.1021/ci060094h).

Tripos's topomer technology began in the company's ChemSpace virtual library design software. Cramer describes a topomer as a molecular fragment that is aligned according to a set of rules. Topomer searching treats molecules as a collection of fragments. Developed by Cramer and previously used as an in-house product development tool, Tripos is now releasing two products — Topomer Search and Topomer CoMFA.

Topomer Search aids users in searching corporate databases and scaffold hopping to find compounds with a similar shape as an identified lead. Cramer describes in-house studies comparing topomer searching with docking strategies that showed ligand similarity searching to be more effective for predicting desired bioactivity and of value for generating patentable chemical entities. While compounds identified by lead hopping from off-the-shelf compounds may look similar to a target protein, they "are structurally dissimilar enough to look different to the patent office and also differ based on 2D fingerprints," says Cramer.

Topomer CoMFA, which is in beta testing, takes the topomer concept further and can help distinguish between hits of similar shape by developing a 3D quantitative structure-activity relationship (QSAR). When used for lead optimization, the technology exploits the corporate database as a source of molecular fragments or scaffolds for use in virtual screening to rank compounds based on their predicted potency.

Molecular Alignments

Although their names may not be entirely descriptive, ROCS, EON, and BROOD represent a family of software tools from OpenEye Scientific Software designed for lead hopping. ROCS, the company's flagship shape-matching product performs shape and chemical functionality comparisons of a database of compounds to a query molecule. This approach has proven successful in large part because similarity in three dimensions is not necessarily related to similarity in two dimensions.

"People tend to look at molecules as 2D structures, whether on paper or on a computer screen," says Paul Hawkins, senior applications scientist at OpenEye. "And they tend to tinker with the molecule at the 2D level, making small, incremental changes." The result is typically a set of analogues of an existing compound that explore a limited region of chemical space.

"To find truly different molecules that have similar activity, or lead-hops, requires thinking about similarity in 3D," Hawkins says. "Similarity searching based on shape and electrostatic properties goes beyond atomic compositions and how those atoms are joined together," explains Hawkins. "Molecules with very similar shapes might look very different when depicted in two dimensions."

EON and BROOD are evolutions of the company's 3D property matching technology. Based on a shape alignment of two molecules, EON calculates and compares their electrostatic potentials, while BROOD uses the same principles to compare molecular fragments rather than whole molecules. EON identifies molecules predicted to have similar activity, and BROOD returns a set of suggested R-group replacements and information on possible starting points for exploring chemical space. Future versions will build in estimations of properties such as solubility and other ADME qualities for rank ordering.

"The ultimate arbiter of whether a compound 'works' should be how well it matches the 3D structure of the protein's active site," emphasizes Hawkins, keeping in mind that the structure of neither the target nor the ligand is static. Both have intrinsic flexibility and it is not possible from a 2D model to predict how the protein might adapt to accommodate the shape of a ligand.

"Structure-based approaches have traditionally been viewed as superior in many applications in drug discovery," observes Hawkins. However, a weakness of traditional structure-based methods and docking engines is the approximations that must be made to get to a solution in a reasonable timeframe; "these approximations compromise the quality of the output," he says. "Since we can treat an active ligand in a rigorous way in terms of its shape and electrostatic properties, we can do at least as good a job as with protein-ligand docking for identifying interesting compounds for screening."

Filling the Toolbox

Several companies have developed computational tools that can facilitate lead hopping and devised innovative ways of generating new scaffolds and lead compounds. SimBioSys designed eHiTS LASSO (electronic High Throughput Screening Ligand Activity by Surface Similarity Order) as a tool for virtual ligand screening. It generates an interacting surface point type molecular descriptor from the 3D structure of a ligand that is conformation independent and, together with a neural network machine learning technique, screens molecular databases at a rate of 1 million structures in less than 1 minute.

Chemical Computing Group's MOE pharmacophore modeling methodology comprises tools for scaffold replacement that identify substitution points on a scaffold molecule and the location of potential R-group substituents, as well as tools for pharmacophore elucidation and searching, and the Pharmacophore Consensus application that suggests pharmacophore queries based on a set of aligned active compounds.

and flexible docking. The Catalyst software suite can be used for 3D pharmacophore modeling, pharmacophore-based alignment of molecules, and generation of pharmacophore hypotheses based on SAR data.

Recently, Schneider and colleagues described an approach that combines a 3D alignment-free pharmacophore descriptor with artificial neural networks (ANNs). The ANNs were trained for receptor selectivity using known antagonists with $IC_{50} < 1 \mu\text{m}$ as reference structures. Self-organizing maps (SOM) helped select structurally diverse compounds for bioactivity testing via virtual screening. Several hits were identified, the most potent with functional IC_{50} values ranging from 9 to 21 μm . Various chemotypes were identified that did not overlap with the scaffolds of reference compounds based on atom types, bond order, carbon scaffold, and/or ring size.

Defining a Bioactivity Fingerprint

The StARLITe (structure activity relationships from the literature) database of bioactive molecules developed by BioFocus DPI contains medicinal chemistry information abstracted from the scientific literature. It includes some 400,000 compound records representing approximately 48,000 chemical series and 3,440 distinct molecular targets, and covering more than 1.4 million assay data points. Experimental data linked to biological activities include sequence data mapped to assay results and links to synthetic chemical routes and assay protocols.

Based on either sequence data or an activity pattern, the user can search active ligands for privileged scaffolds or search targets for closely related receptors. The software ranks the scaffolds based on fragment specificity (compared to all known scaffolds in the StARLITe database) and fragment elegance (synthetic ease). It identifies a subset of scaffolds that are synthetically accessible for use in activity-focused library design.

StARLITe's compound-activity mapping capability can be used to profile hits from HTS to identify alternative activities and to identify scaffolds that share similar patterns of bioactivity. The software queries bioequivalent scaffolds and determines a similarity score, which can then be used to develop a bioactivity fingerprint.

Edith Chan, a research fellow at BioFocus, extracts compounds from the literature, breaks them down into scaffolds and R-groups, and uses the scaffolds to search StARLITe and generate a focused library for a protein and its homologues based on the target's amino acid sequence and SAR data linked to individual compounds and scaffolds. She describes this approach as "knowledge-based hopping," in contrast to de novo ligand design strategies.

Where can scaffold hopping lead? If done properly, and with a little bit of luck, to previously unexplored chemical space, new and perhaps "funny looking" molecular structures, and a series of bioactive compounds that could bring new life to drug targets in pharmaceutical pipelines for which no good lead candidates have been found.

For reprints and/or copyright permission, please contact RMS, 1808 Colonial Village Lane, Lancaster, PA;

(717) 399-1900 ext 100 or via email to bio-itworld@reprintbuyer.com.