

# Designed chemical libraries for hit/lead optimisation

The integration of high-throughput synthetic chemistry with sophisticated library design technology offers advantages in terms of speed and generality.

Dr Tony Cooper and Dr Kathe Andrews-Cramer, Tripos

*At first, screening was considered 'low tech' and had much less aesthetic scientific appeal than the medicinal chemistry approach ...*

Two quite different approaches have both made very successful contributions to drug discovery. On the one hand, **medicinal chemistry** - involving the low throughput, manual synthesis of carefully chosen (designed) compounds, which were fully characterised and tested in a range of related *in vitro* and *in vivo* assays - was applied successfully for the discovery of drugs such as cimetidine, ranitidine and salbutamol. In marked contrast, **screening** (or 'random screening' as it was called initially) involved testing collections of chemical samples in simple biological assays - generally *in vitro*. At first, screening was considered 'low tech' and had much less aesthetic scientific appeal than the medicinal chemistry approach, but it proved to be very effective in drug discovery. The approach was initially applied to crude natural product extracts prepared from plant material or microbial broths, and was responsible for the discovery of many drugs in current use such as beta lactam, aminocyclitol and macrolide antibiotics, anti-cancer agents such as taxol, and the immunosuppressant, cyclosporin.

A large number of positively screened compounds went straight into the clinic, with no significant chemical modification at all. This was partly because the complex chemical structures presented serious challenges to the synthetic chemist, requiring considerable time and effort to overcome. Secondly, many of these natural products offered the prospect of high therapeutic

gain in areas such as cancer (taxol) and transplants (cyclosporin). Screening has continued to grow in importance as a drug discovery strategy as a greater range of *in vitro* assays have become available and the pace of drug discovery has increased. Samples from collections of synthetic compounds also started to be used in screening, while throughput increased dramatically as automated systems were introduced.

When combinatorial chemistry technology started to develop about ten years ago, it seemed that huge libraries containing perhaps millions of compounds could be subjected to high-throughput screening (HTS) and that, as a result, new drugs would be discovered - perhaps even directly from screening. However, this has clearly not happened. Instead, we have a situation in which many drug discovery research groups have generated large numbers of hits in many different biological screening assays. The major problem now facing pharmaceutical discovery groups is to prioritise HTS hits for further study, and make chemical sense of the unprecedented rapid accumulation of screening data.

While large synthetic libraries may appear attractive from the drug discovery perspective, there are several factors which detract from their value and suggest that large screening libraries may not be the most efficient use of resources:

- Many of the largest libraries are based on chemistry that is not 'drug-like', such as peptides or oligonucleotides.

While large synthetic libraries may appear attractive from the drug discovery perspective, there are several factors which detract from their value ...

- Analytical data is limited because of the large number of compounds and the small mass of each sample. Loop injection mass spectrometry is the most frequently used method; however, this can give optimistic results due to its high sensitivity, and provides no indication of yield or purity. Hence, the number of compounds actually present in large libraries may be much smaller than the theoretical number.
- Perhaps most importantly, without sophisticated design input, large libraries may contain a significant proportion of very similar compounds that tend to display similar biological properties. This reduces the *effective* size and value of the library since 'chemical space' is not explored in an efficient manner.

### New drug discovery strategy

At Tripos, we have developed a drug discovery process for hit/lead optimisation which embodies the advantages of both the designed medicinal chemistry approach and high-throughput synthesis/screening strategies (Figure 1). A key component of this process is ChemSpace™ - a proprietary *in silico* technology which enables the construction and rapid searching of vast virtual libraries of three-dimensional chemical structures to identify those molecules which are similar to given hit/lead compounds. In this way, we design synthetically accessible libraries of compounds to facilitate the efficient exploration of the chemical space around a given biologically active molecular structure. The key steps in this process are outlined below:

- The starting point is the two-dimensional

chemical structure of a hit or lead compound, or series of biologically active analogues. These may originate from random screening, competitors' patents, known biologically active compounds, known drugs, studies of the molecular structure of the target and so on. ChemSpace can then search a general 'research virtual library' containing greater than  $10^{13}$  structures to explore easily accessible chemistries, supply scaffold ideas and guide the construction of 'custom virtual libraries.'

- Once desired scaffolds and chemistries have been chosen, ChemSpace is programmed with the synthetic chemistry routes - including reactions and reagents - which will be used to make the desired compounds. If necessary, this route may contain multiple steps, and solid and/or solution phase techniques, and may include stereo centres or involve stereospecific transformations. The virtual libraries (VLs) reside in a large relational database that references the structures of all the compounds that could possibly be made using all different chemical intermediates used in the chosen synthetic scheme and that are available - both in-house and commercially. The sizes of the virtual libraries vary. For example, with just two sites of diversity on the lead compound, more than 100M structures would normally be generated. The VL includes 3-D structures that are built within ChemSpace using CONCORD, and side-chains are aligned using a rule-based protocol. A typical VL can be built with a few hours of human input and, at most, two days of computational time on an R10K CPU.

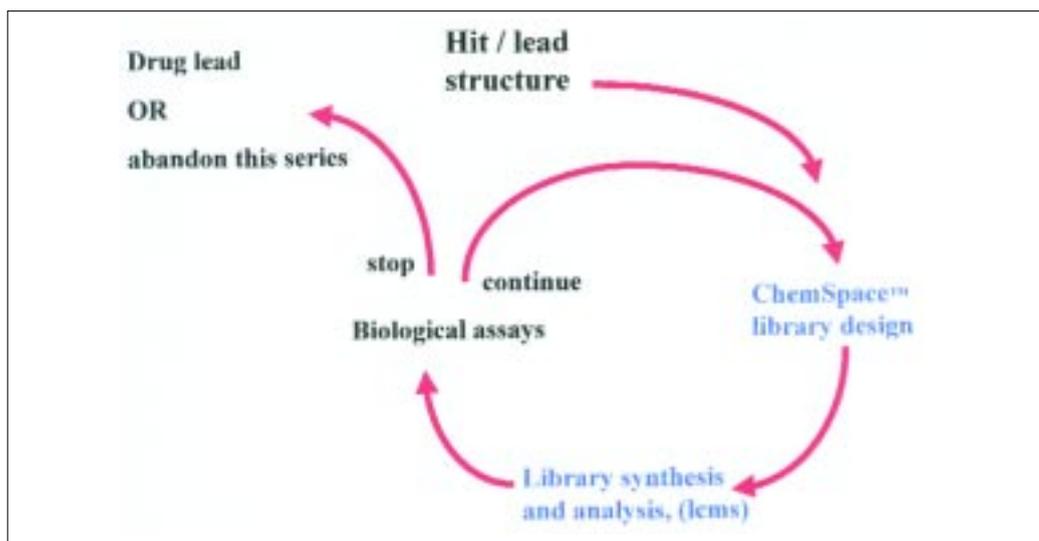


Figure 1. The iterative drug discovery process.

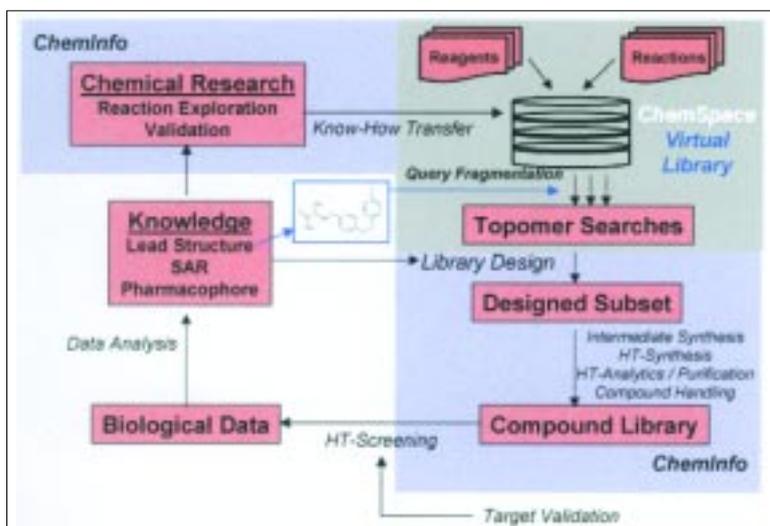


Figure 2. The interface between compound library design (ChemSpace™) and library synthesis informatics (ChemInfo).

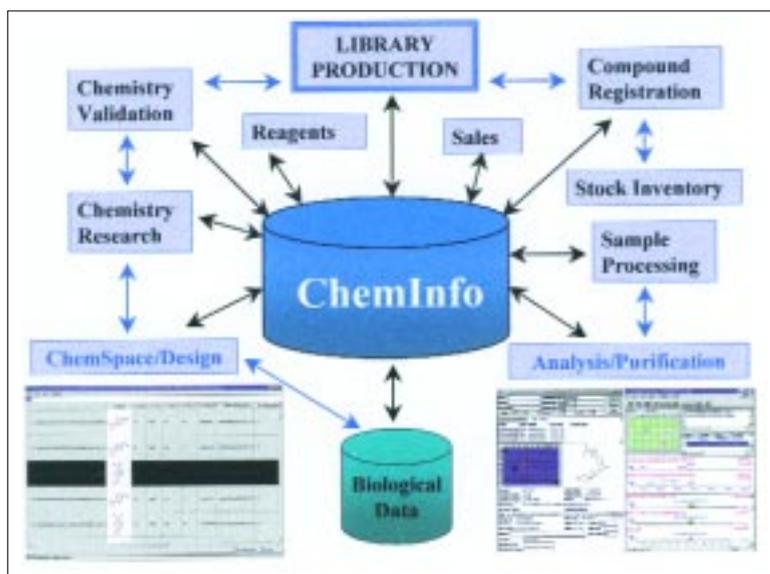


Figure 3. Areas of the Tripos process requiring data exchange through ChemInfo and two screen snapshots.

- The major value of ChemSpace is to search the VL for compounds which have a similar 3-D shape to the original hit/lead structures. The Tripos-proprietary topomeric shape descriptor is used for this purpose. Validation studies of this descriptor have been published, showing that it exhibits 'neighbourhood behaviour' (1a and b). Practical use of the technology, along with the published validation experiments, have shown us that when searching our virtual libraries, we ideally specify a shape similarity within 120 topomer shape units to the

hit/lead structures (where 60 units is equivalent to replacing a hydrogen atom with a methyl group on a typical drug molecule). Many leads yield search hit-lists containing thousands of combinatorial structures at stringent search radii of 80 or 90 units. More exploratory searching can be done to identify different chemistries and new structural domains by expanding the radius to 150 units. Search times are extremely fast ( $2 \times 10^{13}$  structures per hour), and other criteria such as 'topomer features' (pharmacophoric interactions) and 'SAR tuning' can also be applied.

The ChemSpace technology was applied in a successful collaboration between Tripos and Bristol Myers Squibb, with the results published in a joint paper (2). A recent collaboration with Arena Pharmaceuticals (San Diego, CA) resulted in a preclinical candidate less than nine months following target identification by using ChemSpace combined with medicinal chemistry to move rapidly through the 'make and test' iterations shown in Figure 1 (3). The ChemSpace technology is now heavily used at both the US and UK Tripos sites to accelerate drug discovery projects with partners.

### Design must be integrated with synthesis

Our designed combinatorial libraries range in size from a few hundred to a few thousand structures. They are constructed to facilitate the exploration of SAR around a hit/lead compound in an efficient manner - that is, maximum information from a given number of compounds synthesised. ChemSpace design can also be valuable when patenting a series of active compounds by highlighting those compounds around a lead which are likely to show similar biological activity (and hence should be synthesised and tested) by virtue of shape similarity. In this way, we try to minimise the opportunities for competitors to file subsequent patents.

ChemSpace forms a key part of the iterative discovery process, as illustrated in Figure 1, but it must also be closely integrated with the other elements of the cycle for the whole system to operate efficiently and generate significant advantage over more random strategies. Thus, we have interfaced our design technology with a high-throughput chemical synthesis process that is monitored with a cheminformatics system called ChemInfo. The interaction between these two systems is illustrated in Figure 2. ChemInfo is an in-house system built by Tripos software engineers in St Louis (MO, USA), to support the chemical library synthesis group at Tripos Receptor Research Ltd in the UK. ChemInfo facilitates the smooth flow of information between many areas of our drug discovery process, as illustrated in Figure 3. A

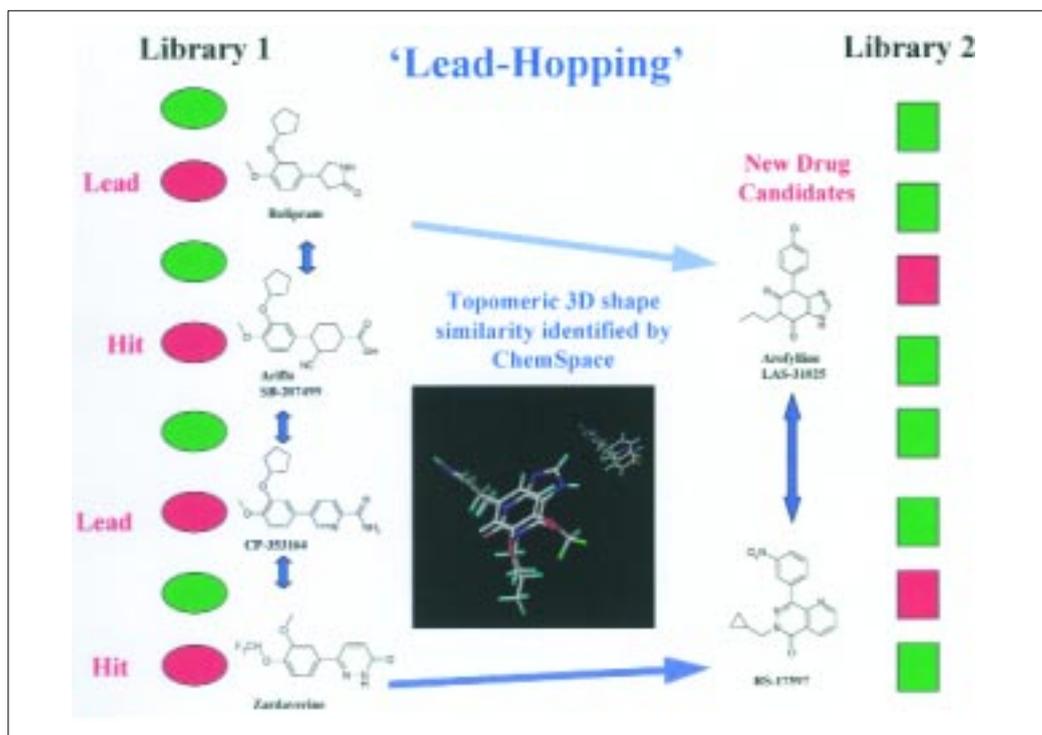


Figure 4. 'Lead hopping' - the use of ChemSpace™ to identify topomeric shape similarity in non-analogues.

*Our designed combinatorial libraries range in size from a few hundred to a few thousand structures*

screen-shot of the "chemists' interface" to ChemSpace is shown, with each row representing a combinatorial reaction. Following further analysis, the design is completed and moved into the realm of ChemInfo, where all aspects of synthesis and analysis are tracked. For hit/lead optimisation libraries, it is important that every compound is checked for structural assignment and a minimum specified level of purity in order to avoid false negatives and ensure maximum accuracy of the SAR data that is generated in each iteration of the process. We submit every library sample to LC-MS analysis and Figure 3 shows a screen snapshot of Micromass LC-MS data linked to compound synthesis information within ChemInfo. NMR is used in the research phase of library synthesis, and for specific compounds in the final library.

#### 'Lead hopping'

The use of ChemSpace as part of our process for the synthesis of designed chemical libraries for hit/lead optimisation is well established in our laboratories. This usually involves intra-library 'lead hops' - where hit structures are designed as combinatorial analogues of a lead molecule. However 'lead hopping' appears most powerful when applied across chemistries and libraries, to identify novel classes of drugs and cover new and broad areas of patent space.

To assess the general applicability of lead hopping with ChemSpace, lead structures were taken from 34 recent medicinal chemistry publications and used as queries for topomeric shape searching of a research virtual library containing  $2.6 \times 10^{13}$  virtual structures, representing seven simple combinatorial chemistries (4). Eighty-five per cent of the leads yielded more than 500 'hit structures' at a relatively close search radius of 120 topomeric shape units or less, with many hit-lists containing thousands of suggested combinatorial structures, readily synthesised using commercial reagents. Furthermore, the use of a single, defined conformation to assess shape similarity was validated using FlexS™, an independent and rigorous computational method for structural overlays of flexible molecules.

In a retrospective validation of topomeric shape similarity as a predictor of biological activity, a series of 26 phosphodiesterase 4 (PDE4) inhibitors (all in clinical trials) were subjected to pair-wise topomer shape comparisons (5). These compounds inhibit the conversion of cAMP to 5'-AMP and appear promising in a number of therapeutic areas. By definition, all compounds within a combinatorial library are analogues. This can be visualised in Figure 4, where the rolipram analogues in Library 1 can be recognised by the presence of a dialkoxy-phenyl moiety. Topomeric shape differences between molecules in Library 1 were

typically in the range of 80 to 120 shape units, indicating that these potent PDE4 inhibitors are close topomeric neighbours. Years of research went into the development of these analogues by elegant medicinal chemistry methodology, yielding a large class of preclinical candidates for anti-inflammatory and anti-asthmatic therapies. A technology such as ChemSpace could accelerate the development of these discoveries by offering medicinal chemists 'all possibilities' - allowing them to use their intuition and expertise in the selection of the 'right' compounds to make and test. The rolipram-like PDE4 inhibitors are good examples of intra-library lead hops.

Side-effects and the desire for specificity and applicability to new therapeutic areas drove discovery towards a second class of PDE4 inhibitors, exemplified in Library 2 by the xanthine analogue, arofylline. Topomeric shape comparisons of the structures within Library 2 showed striking shape similarity, with differences typically 80 to 90 shape units. More interestingly, when Library 1 was compared pairwise with members of Library 2, several close topomers were identified, including zardaverine and RS-17597 - shown in fragmented form in the inset and representing an inter-library lead hop. Upon further examination, it was found that many of these structures could have been synthesised using commercial reagents in a high-throughput combinatorial fashion. Thus, ChemSpace could have suggested many of these structures, as well as novel ones.

This example illustrates the potential of 'lead hopping' because ChemSpace is able to identify significant shape similarity between the compounds in these different libraries via their 3-D structures. It has become clear that this technology offers the possibility for fast identification of shape-similar compounds that are not analogues of given hit/lead structures. This is important for a number of reasons. It can enable the discovery of novel compounds that are significantly more active, with reduced side effects and increased specificity; also, new biologically active compounds may be found which lie outside existing patents. Furthermore, it is now becoming possible to identify active compounds that have better pharmacological/ADME profiles than the original hit. It may also be possible to avoid toxicity problems encountered with an initial hit. For example, the H2 receptor antagonist, burimamide, showed carcinogenic activity in animal tests and was discarded, but the topomerically shape-similar molecule, cimetidine (Tagamet), replaced it and went on to become a safe and effective drug.

### Acknowledgement

We would like to thank all our colleagues, particularly Peter Hecht, Dick Cramer and Bernd Wendt.

*Dr Tony Cooper qualified in chemistry at Imperial College, London. After working for Glaxo Research and then Maybridge Chemical Company, he formed Receptor Research Ltd in 1990. This company was acquired by Tripos in November 1997, and Dr Cooper is now Chief Scientific Officer of Tripos Receptor Research Ltd at their laboratories in Bude, Cornwall.*

*Dr Kathe Andrews-Cramer received her PhD in Biochemistry at the University of Illinois, Urbana, in 1988, followed by post-doctoral research on neuroreceptors at Washington University, St Louis, until 1991. After several years teaching and managing the undergraduate genetics and molecular biology laboratories at Washington University, she entered the world of computational chemistry at Tripos Inc. She is currently a Research Scientist at Tripos, where her R&D efforts are focused on ChemSpace™ and small molecule design projects.*

*... this technology offers the possibility for fast identification of shape-similar compounds that are not analogues of given hit/lead structures*

### References

- 1a). Patterson DE, Cramer RD, Ferguson AM, *et al.* (1996). Neighborhood behavior: a useful concept for validation of molecular diversity descriptors. *J Med Chem*, 39, 3049-30. 1.
- 1b). Cramer RD, Poss MA, Hermsmeier MA, *et al.* (1999). Prospective identification of biologically active structures by topomer shape similarity searching. *J Med Chem*, 42, 3919-3933.
2. Hodgkin E and Andrews-Cramer K (2000). Compound collections get focused. *Modern Drug Discovery*, Jan/Feb 2000, 55-60.
3. Andrews KM and Cramer RD (2000). Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries. *J Med Chem*, In press.
4. Andrews KM and Cramer RD. Topomer shape and topomer feature matching in groups of known drugs. Manuscript in preparation.