RPR's approach to high-speed parallel synthesis for lead generation

Richard F. Labaudiniere

Rhône-Poulenc Rorer (RPR) is developing high-speed parallel synthesis as an essential tool to produce large libraries for lead generation. Libraries of 5000–10,000 members are produced as singles by solution- or solid-phase chemistry for screening in high-throughput assays in order to foster lead generation and accelerate lead optimization. Libraries are designed to optimize chemical diversity by applying a combination of virtual selection methods. The author describes how, in spite of different 2D structural features, library compounds share most of the 'drug-like' properties of compounds from the MDL Drug Data Report database, in terms of pharmacophore coverage and atomic and molecular properties.

As many have described, the medicinal chemistry field of the early 1990s witnessed a paradigm shift from a classical sequential 'hand-crafted' approach to a 'parallel and automated' approach that exploited the power of the emerging combinatorial chemistry and parallel synthesis technologies. This dramatic change in approach was similar to the emergence in the 1980s of automated 'parallel' screening in 96-well microtiter plates.

The transition from the classical sequential approach to the automated parallel approach has accelerated the lead optimization process by rapidly generating structure–activity relationships (SAR) around leads coming from high-throughput screening (HTS). Furthermore, by increasing the speed of lead optimization, the number of concurrent optimization projects and/or the capacity to work in parallel on more chemical series aimed towards the same biological target has also increased.

'Classical' medicinal chemistry is still essential at a more downstream optimization stage, where fine-tuning the pharmacodynamic and pharmacokinetic profiles, the safety margin and the physicochemical properties is required to select the optimal drug candidates for clinical trials. Although, it is clear that the developing field of parallel purification would also have some application at this late stage by allowing relatively large-scale parallel synthesis and purification of analogs for final profiling before selection of the drug candidate.

Another application of combinatorial chemistry/parallel synthesis technologies is the preparation of general libraries for screening. Taking advantage of the substantial increase in screening throughput and the availability of technologies to produce thousands of compounds within a reasonable time frame, medicinal chemists are now able to produce compound libraries of diverse analogs for HTS. This strategy is adopted by most of the biotech companies to generate their collection of screening compounds quickly. It is also an adroit master plan for the major pharmaceutical companies, allowing them to complement their compound collection, accumulated over the years by a classical medicinal chemistry approach, with new and diverse sets of compounds. Other benefits include

Richard F. Labaudiniere, Senior Director Lead Discovery – Rhône-Poulenc Rorer, 500 Arcola Road, SW8, PO Box 5092, Collegeville, PA 19426-0996, USA. tel: +1 610 454 3687, fax: +1 610 454 3311, e-mail: richard.labaudiniere@rp-rorer.com

  1359-6446/98/$19.00. PII: S1359-6446(98)01259-8

Box 1. HSPS – a new tool for drug discovery

In lead generation:
- Adds diversity to company compound collection
- Generates generic and diverse screening sets

In lead optimization:
- Generates SAR rapidly
- Improves potency rapidly
- Improves selectivity rapidly
- Covers a large chemical space rapidly
- Enables broad patents

accelerating the early stage optimization process by generating preliminary SAR at the lead compound selection stage and allowing fast generation of focused libraries.

RPR's approach to high-speed parallel synthesis

In 1994, RPR integrated a combinatorial/parallel synthesis strategy into their drug discovery process. Efforts were essentially concentrated on parallel synthesis, or parallel production of 'singles', rather than production of mixtures by classical combinatorial chemistry. This decision was based on a desire to avoid any deconvolution steps and a strong belief in the rapid evolution of high-speed parallel synthesis technologies.

High-speed parallel synthesis (HSPS) is being used at both the lead generation and lead optimization stages (Box 1). In this review, we will concentrate on the strategy we developed for lead generation.

HSPS and lead generation

The objective is to design and produce, in a high-speed parallel mode, libraries of compounds to be screened in all our primary HTS assays as additional diversity compared to our historical collection and natural extracts. Our clear objective was to increase the proportion of compound for screening coming from our parallel synthesis effort.

In 1995, only 13% of compounds were provided by HSPS, but by 1997 this increased to 39%, and it is forecasted that in 1998, 56% of compounds will come from HSPS (Fig 1.) Certainly, we will increase our chances of finding a hit from chemical libraries, allowing us to generate SAR quickly and hasten the lead optimization process.

Chemical libraries for lead generation

In general, libraries of 5000–10,000 discrete compounds are produced using solution- and solid-phase chemistry,

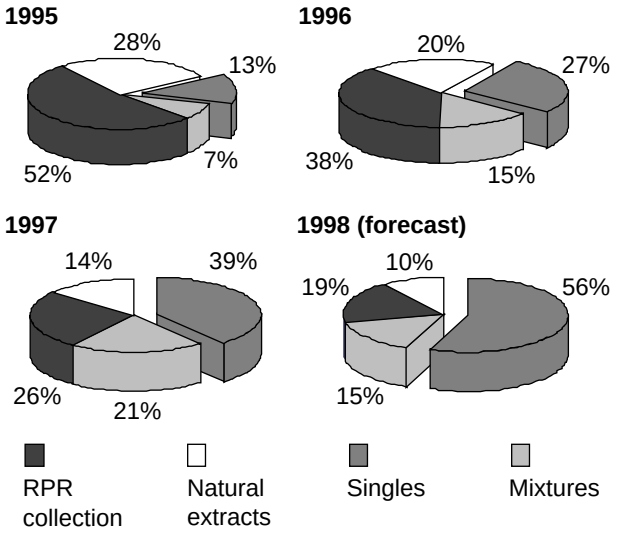


Figure 1. *Ratio of library compounds versus RPR collection/natural extracts. RPR Coll., RPR historical collection subset.*

and these libraries include at least three points of diversity. The trend has been to increase the proportion of compound library produced by solid-phase chemistry from 50% in 1997 to ~80% in 1998. Approximately six months are generally spent on chemistry development to optimize reaction condition and to delineate the scope of each chemical step.

In general, no purification step is performed, but rather, we rely on the systematic work in chemistry development to guarantee the quality of the library produced. Quality control is performed by analysing a representative subset of the library (2–10%) by LC/MS and ELSD/MS. A library is accepted if the majority of analysed compounds are above the 75% purity threshold.

Production is performed on different technology platforms depending on the chemistry involved [Tecan Combitec (Tecan US, Durham, NC, USA)/ACT robots (Advanced ChemTech, Louisville, KY, USA), Charybdis block technology (Charybdis Technologies, Carlsbad, CA, USA), IRORI Accutag (IRORI, La Jolla, CA, USA)]. In fact, most of our solid-phase libraries are now produced using the IRORI RF tag technology. Using microreactors and miniature electronic tags, this technology allows a directed-sorting approach to combinatorial chemistry, which delivers all the advantages of both parallel synthesis and split-and-pool synthesis to produce compounds as singles[1].

Chemical diversity and lead generation libraries

At RPR our objective is to design compound libraries that add diversity to our historical collection, but also provide generic and diverse screening sets towards a generic family of targets [e.g. kinase, proteinase and 7-transmembrane (7-TM) receptors].

Chemical diversity is not only a numbers game. Diversity does reach a plateau when the number of compounds increases, but the goal is to reach this maximum diversity (for a given structure type) with the 'optimum' (i.e. minimum) number of compounds. For example, using a four-component reaction (Ugi chemistry), after three rounds of selection of inputs (amines, acids, aldehydes and isonitriles) to increase the pharmacophore space coverage, we begin to reach a plateau in terms of diversity (Fig. 2). The same result is observed using other methods to measure diversity, such as DiverseSolutions (data not shown).

Because biological targets recognize shape and electronic properties and not 2D structures, we based our strategy for molecular diversity essentially on the 3D pharmacophore method ChemDiverse (Chemical Design, Chipping Norton, UK; Ref. 2 and Davies, E.K. and Briant, C. *Combinatorial Chemistry Library Design using Pharmacophore Diversity* at http://www.awod.com/netsci/Science/Combichem/feature05.html), but customized in-house for our own purposes. With this method, a pharmacophore, which represents the minimum necessary features for interacting with a biological target, is represented by a four-points 3D model defining a chiral polyhedron (Box 2).

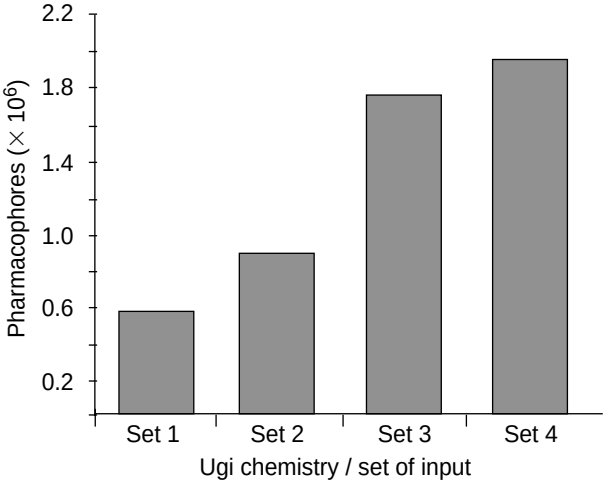
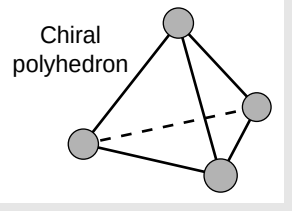


Figure 2. *Pharmacophore coverage with Ugi chemistry by four consecutive libraries.*

Box 2. Four-points 3D pharmacophore method – ChemDiverse

For 3D pharmacophore partitioning:

- Use six different pharmacophore types (features) for four-points pharmacophores:
  - H-bond donors
  - H-bond acceptors
  - Aromatic ring
  - Hydrophobic groups
  - Acid
  - Base
- Use seven or 10 distance ranges per feature–feature distance
- Perform full conformational sampling



For 10 distance ranges and six features, 9.7 million potential four-points 3D pharmacophores are covered.

Six pharmacophore types (features) and 10 distance ranges per feature–feature distance are considered. For each structure, a full conformational sampling is performed and for each conformation, all possible four-points pharmacophore combinations are calculated to define a pharmacophoric key representing all potential pharmacophores covered by this molecule.

The four-points 3D-pharmacophore approach has also been adapted for the design of 'generic' libraries (Mason, J.S. *et al.*, submitted). For that purpose, one pharmacophoric center among the four considered to define a four-points 3D model is fixed – it represents a 'privileged' substructure[3] associated with a family of biological targets (e.g. benzamidine for serine proteases, hydroxamic acid function for zinc metalloproteinases or a diphenylmethane group for 7-TM G-protein-coupled receptors). Maximum diversity is obtained by selecting the building blocks that give the best four-points pharmacophore space coverage. Another strategy is to select inputs for maximizing the coverage of the pharmacophore space defined by all compounds known to interact with the family of targets, for example, using compounds reported in the *MDL Drug Data Report* (MDDR) database.

Most of our compound library for lead generation has been designed using this four-point 3D-pharmacophore

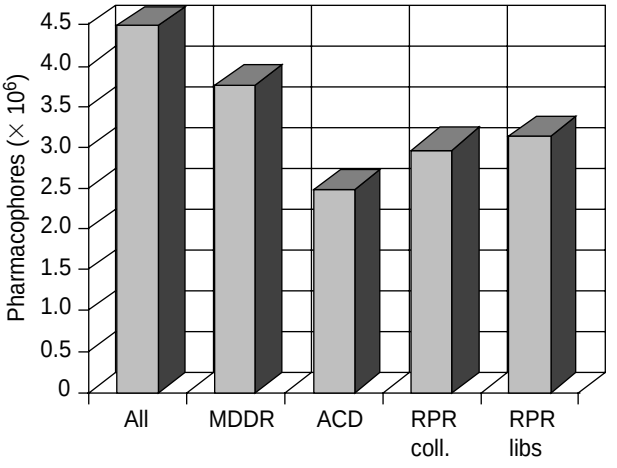


Figure 3. *Total number of pharmacophores in reference sets. ACD,* Available Chemical Directory *(75,000 compounds); MMDR,* MDL Drug Data Report *database (63,000 compounds); RPR coll., RPR historical collection subset (100,000 compounds); RPR libs, RPR compound library subset (100,000 compounds).*

method by selecting diversity inputs to maximize the pharmacophoric space covered by the whole library (molecular weight and cLog*P* distributions are also optimized during the input selection process). Figure 3 shows that the number of pharmacophores covered by the first 100,000 compounds from our libraries is similar to that obtained from our historical compound collection and, importantly, significantly more than the number obtained with the commercially available compound database ACD (*Available Chemical Directory*).

Box 3. Atomic/molecular properties method – DiverseSolutions

Chemistry space defined by:

- Descriptors (typically 4–6):
  - Novel BCUT value descriptors related to 2D, 3D and physicochemical properties
  - Combine relevant atomic properties (e.g. charges, polarizabilities, H-bonding abilities) with connectivity information
- Range of values identified for each descriptor

Chemistry space partitioned into cells (e.g. six bins per descriptor – >47,000 cells)

More recently, we incorporated another method in our arsenal, DiverseSolutions, a cell-based approach, developed by R. Pearlman (University of Texas, Austin, USA), using descriptors called BCUT values that represent a combination of atomic and molecular properties (Box 3; Ref. 4 and Pearlman, R.S., *Novel Software Tools for Addressing Molecular Diversity* at http://www.awod.com/netsci/Science/Combichem/feature08.html). We now routinely use a combination of both methods to select diversity inputs to design diverse lead generation libraries. These two partitioning methods offer the advantage of using high level descriptors relevant to ligand–protein interactions and

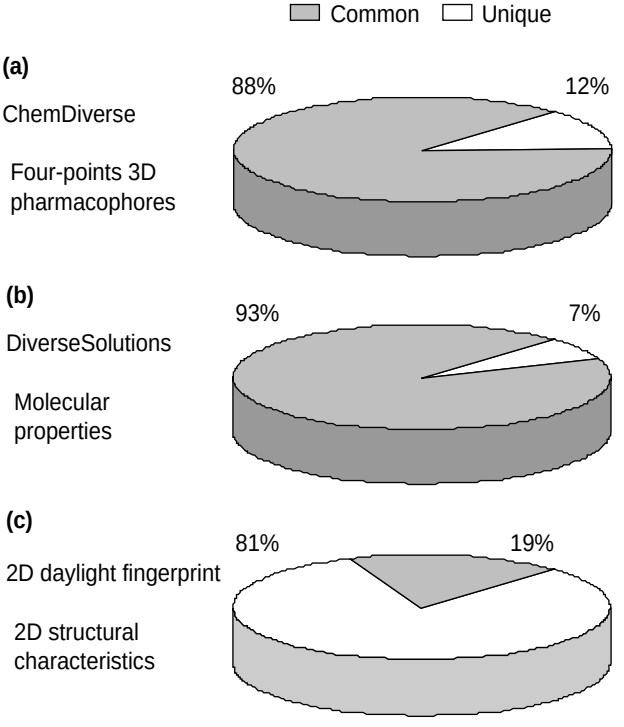


Figure 4. *Comparison of RPR libraries with the* MDL Drug Data Report *(MDDR) 'drug-like' compounds database. (a) Proportion of pharmacophores common to both or unique to RPR libraries based upon ChemDiverse method. (b) Proportion of compounds from RPR libraries distributed in cells occupied by MDDR compounds (common) or in cells occupied by only RPR library compounds (unique) based upon DiverseSolutions method. (c) Proportion of compounds from RPR libraries clustered with MDDR compounds (common) or clustered with only RPR library compounds (unique) based upon 2D daylight fingerprint method.*

the ability to identify diversity voids that can be eventually filled by additional compound libraries[4].

Diversity analysis of lead generation libraries

The first set of 100,000 compounds produced by our HSPS effort was analysed by different diversity methods and compared with the MDDR collection, which is considered as a reference database for drug-like compounds. The diversity methods used were the four-points 3D pharmacophore method, the DiverseSolutions method and the classical 2D daylight fingerprint approach (a clustering analysis method; Weininger, D. and Delany, J. *Daylight Clustering Manual* at http://www.daylight.com/dayhtml/doc/cluster/cluster.toc.html). For each diversity method, properties in common between RPR libraries and MDDR or unique to the RPR were analysed (Fig. 4).

From the analysis, it can be concluded that despite a great structural diversity compared with the MDDR database (Fig. 4c), our libraries possess most of the key features of drug-like molecules. About 90% of our library compounds overlap regions of the pharmacophoric space (Fig. 4a) and the atomic/molecular property space (Fig. 4b) covered by the MDDR 'drug like' compounds.

Conclusion

RPR has established an efficient HSPS Unit for the design and production of lead generation libraries as a principal source of compounds for HTS. Libraries of 5000–10,000 discrete compounds are routinely designed and produced in parallel, which rapidly increases the molecular diversity of our collection, delivers quality leads and generates SAR.

Several leads have already been generated from these libraries for different biological targets, which demonstrates the power of HSPS in the lead generation process. In all the screening assays (biochemical and cell-based) where library compounds have been systematically studied, hits/leads have been identified from different libraries (chemical series) and, in most cases, preliminary SAR were directly generated from the initial library.

REFERENCES

1 Nicolaou, K.C. *et al.* (1995) *Angew. Chem., Int. Ed. Engl.* 34, 2289
2 Davies, E.K., In Chaiken, I.M. and Janda, K.D. (1996) in *Molecular Diversity and Combinatorial Chemistry. Libraries and Drug Discovery*, pp. 309–316, American Chemical Society
3 Evans, B.E. *et al.* (1998) *J. Med. Chem.* 31, 2235–2246
4 Mason, J.S. and Pickett, S.D. (1997) *Perspect. Drug Des. Discovery* 7/8, 85–114