Enamine real database: making chemical diversity real

ABSTRACT

The Enamine REAL DataBase (RDB) covers rigorously validated chemical space of over 29,000,000 virtual HTS compounds, over 10,000,000 of which comply to drug likeness Rule-of-5 standards. The high efficiency of our RDB methodology is based on 30 optimized reactions, 54 optimized chemical procedures applied to 18,000 proprietary in house and 9,000 purchased building blocks and our efficient algorithms for calculating the synthetic feasibility of all virtual structures. Optimized schemes for RDB production allows the synthesis of 20,000 compounds a month with an average feasibility rate of 65 %.

Modern high throughput screening (HTS) methods allow fast and effective evaluation of biological activity of large collections of compounds (1). Successful screening of new therapeutic targets depends on the structural and functional diversity of the compounds screened. Usually researchers try to cover as large chemical space as possible during the initial stage of HTS. The result of these initial screens will determine subsequent steps of drug discovery and development. Therefore, suppliers of HTS compounds face two major challenges:

1) timely and cost effective production of large and diverse libraries of compounds in commercially acceptable periods of time;
2) continuous expansion and optimization of chemical space of synthetically accessible compounds with potential biological activity.

Preparation of large numbers of HTS compounds has been approached via parallel synthesis and other combinatorial chemistry methods (2). Biological screening of multi-component mixtures resulting from one-pot combinatorial procedures is simple and efficient on libraries generation stage, but requires much effort for isolation and identification of active compounds from the mixture. The identification becomes straightforward when combinatorial libraries of individual compounds are screened. The most significant drawback of combinatorial libraries of individual HTS compounds is the low synthetic feasibility caused by limited capacities of computational methods for the estimation of success rate. In 2001 Enamine Ltd (3) initiated the development of an efficient production and management scheme to produce large chemically diverse collections of individual compounds. These efforts resulted in the REAL DataBase (RDB) which is a thoroughly validated library of virtual screening compounds whose high synthetic feasibility enables us to consider them as a part of our stock. Currently the Real Data Base is a major source of Enamine screening compounds with production rate of about 20,000 samples a month; more than 500,000 compounds (ca 50% of entire Enamine stock) were generated using the RDB approach. Thirty one-step reactions and fifty four chemical procedures were optimized at the Enamine R&D department in order to increase yields and simplify isolation and purification of compounds in RDB (Table 1) (4-7). The optimized processes comprise alkylation, acylation, arylation, condensation and heterocyclization reactions that combine 20,378 building blocks to generate diverse sets of compounds for HTS. The current collection of building blocks used for generation of RDB consists of 4867 amines; 290 alcohols, 497 phenols, 784 CH-acidic compounds, 251 NH-acidic compounds, 381 sulfamides, 1510 thiols, 5309 carboxylic acids, 195 sulfochlorides, 2996 alkylating agents, 376 arylation agents and 2992 active intermediates of other classes. Selection of building blocks for RDB is based on systematic scaffold analysis and starts from thorough investigation of medical and chemical literature by the team of experienced synthetic chemists and chemoinformatics personnel.

Reaction type	Synthetic procedures	Validated set of compounds	Drug-like compounds
Acylation	16	9,390,016	3,818,068
Alkylation	8	11,543,743	4,384,755
Arylation	4	631,614	211,355
Aminomethylation	3	389,452	103,493
Epoxide ring opening	2	127,789	53,332
Reductive amination	2	132,633	50,092
Heterocyclizations	21	10,845411	3,846,248

Table 1. Representative reactions and their substrates used for generating RDB sets

Combination of building blocks that can react with each other results in 101,136,755 virtual structures. Removal of identical structures generated via different synthetic routes yields 65,786,004 unique structures. Chemical feasibility and current availability of building blocks results in 29,215,900 chemically accessible virtual compounds designated as our Primary REAL DataBase. Identification and removal of synthetically unfeasible virtual structures is the most complicated stage of generating RDB. Estimation of accessibility of a particular structure is approached through comparative analysis of synthetic routes to its synthesis and selecting the optimal one(s). Every route is characterized by a combination of initial reagents, and every reagent has its own set of scores attributed to every available reaction and representing reactivity, steric hindrance, solubility and other parameters. In order to evaluate the chemical accessibility of a RDB compound via a particular chemical route, a set of scores for every reagent is added up and compared to a set of scores for another reagent(s) and resulting scores are compared to statistically established and empirically verified thresholds. A structure is regarded accessible if the threshold is exceeded. A choice of the most appropriate procedure for synthesis of every synthetically accessible structure also depends on initial data from successful syntheses actually carried out with corresponding reagents. Systematic analysis of all the above mentioned factors ensures the high reliability of RDB.

Application of Lipinsky Rule of 5 (8) (Ro5) as well as toxicity and ADME (9) filters to the primary REAL DataBase affords virtual drug-like sets. Drug-likeness filters are used to create the Silver drug-like set (10,686,375 structures) and is based on Lipinsky rules in which the number of H-bond acceptors is allowed to be under 12. The distributions of physical properties for this set are shown in figure 1 and can also be characterized by the following mean values: MW=410 Da, LogP=3.1, H-Bond acceptors=6.9, H-bond donors=0.9. The majority of drug-like structures in the Silver collection are derived from two-component reactions, introducing only small common units. This results in good diversity of the Silver set (0.79, calculated by Ched™ software) which is comparable to the diversity of Enamine stock. The application of additional ADME-T filters to the Primary REAL DataBase results in about 4,000,000 virtual structures of our Golden REAL DataBase drug-like collection.

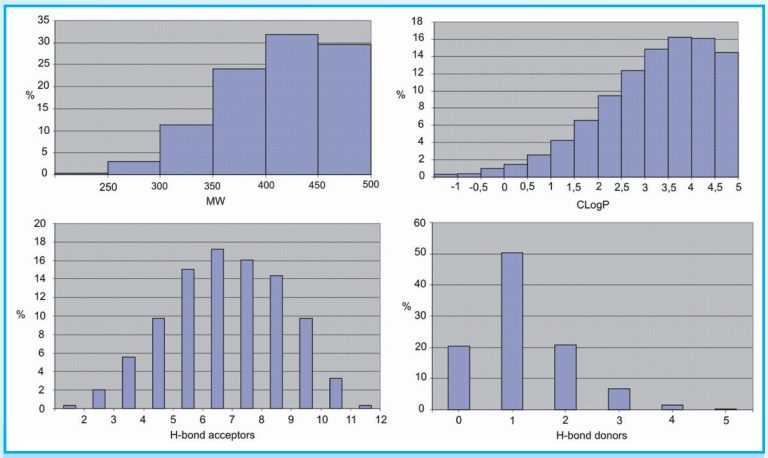

Figure 1. Distributions of physical properties of the Silver drug-like RDB collection

The high rate of RDB compounds production is achieved by division of labour at all stages of the production scheme e.g. weighing and dissolving reagents, conducting the reactions and purification of target compounds (10). Every produced screening compound is individually isolated and characterized. In house optimized synthetic protocols allow us to isolate most compounds through simple precipitation filtering procedures that ensure 90+% purity. In more complicated cases highly efficient polymeric scavenging systems are employed to remove excessive reagents from reaction mixtures. High-performance rotary evaporating units, preparative HPLC and TLC techniques are used for isolation of soluble compounds and oils. The identity and purity of all compounds are controlled by LCMS and ^{1}H NMR spectroscopy. In the case of complex compounds 2D NMR studies and single crystal X-ray analysis of representative structures are carried out for unambiguous determination of structure and stereochemistry. Enamine's in house logistic system and proprietary software allow us to track and manage all aspects of the process including synthesis conditions with initial reagents, compilation of analytical data and other information useful in determining conditions for optimal yield. Over the last three years RDB has become a commercial source of original screening compounds for pharmaceutical and biotech companies worldwide. Enamine offers a flexible system for ordering large sets (more than 10,000) of RDB compounds. In the first stage, the client provides Enamine with their prefered drug-likeness criteria which is used in filtering the RDB collection to generate a customized set of structures fitting the customer's needs. Using this set, the client selects target structures assuming their 0.6 synthetic feasibility rate. Table 2 summarizes results of three RDB orders over the last years. Synthetic feasibility for RDB compounds constantly increases as additional information on reagents and their behaviour in utilized reactions is gained and processed. The synthesis of RDB compounds can be initiated with the full or partial financial commitment of the client. Synthesis can be done with a partial front payment or on a "no risk" basis with compounds paid for in full upon delivery. The price of RDB compounds depends on the size of the library as well as the mass of sample required and is usually negotiated up front and the subject of an agreement between the client and Enamine. Samples can be supplied on exclusive basis (subject to additional agreement).

In conclusion, RDB is a unique service in which Enamine can generate structurally and functionally diverse drug-like libraries with high likelihood of success within an acceptable time frame. These libraries can be customized to the specific needs of the client by using selection criteria provided by the client. With over 10 years of synthetic experience supported by efficient combinatorial procedures and division of labour as well as fully computerized processing ensures Enamine can provide highly diverse, customized libraries to clients in a timely fashion. All of this can be done with no up front risk to the client.

	Compounds Selected	Compounds synthesized, %	Lead time, days
SET A	8,432	67.0	250
SET B	40,828	48.9	365
SET C	76,711	70.5	225

REFERENCES AND NOTES

1. G.L. Patrick An Introduction to Medicinal chemistry. 3-d Ed. Oxford University Press. 2005, P 299-325.
2. Combinatorial chemistry. From theory to application. Ed. W. Bannwarth, B. Hinzen, 2 -nd edition, 2006, Wiley VCH, Weinheim. 672 P.
3. www.enamine.net
4. S. V Ryabukhin, A. S. Plaskon, D. M. Volochnyuk, A. A. Tolmachev Synthesis, 2007, 20, 3163-3170
5. S. V. Ryabukhin, A. S. Plaskon, E. N. Ostapchuk, D. M. Volochnyuk, O. V. Shishkin, A. N. Shivanyuk, A. A. Tolmachev, Org. Lett. 2007; 9, 4215-4218.
6. S. V. Ryabukhin, A. S. Plaskon, D. M. Volochnyuk, S. E. Pipko, A. N. Shivanyuk, A. A. Tolmachev, J. Comb. Chem.; 2007, 9, 1073-1078.
7. A. V. Bogolubsky, S. V. Ryabukhin, S. V. Stetsenko, A. A. Chupryna, D. M. Volochnyuk, A. A. Tolmachev, J. Comb. Chem.; (Article); 2007; 9(4); 661-667.
8. C.A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney Adv. Drug Deliv. Reviews, 1997, 23, 3-10.
9. M. M. Olah, C. G. Bologa, T. I. Oprea Cur. Drug Discovery Techn., 2004, 1, 211-220
10. A. A. Tolmachev Patent of Ukraine, 11687, 2006.

ALEXANDER N. SHIVANYUK[1],
SERGEY V. RYABUKHIN[1],
ANDREY V. BOGOLYUBSKY[2],
DMYTRO M. MYKYTENKO[2],
ALEXANDER A. CHUPRYNA[2],
WILLIAM HEILMAN[2],
ALEXANDER N. KOSTYUK[2],
ANDREY A. TOLMACHEV[1]

1. Department of Chemistry
Kyiv Taras Shevchenko University
Volodymyrska Street 64, Kyiv 01033, Ukraine

2. Enamine Ltd.
Alexandra Matrosova Street 23A, Kyiv 01103, Ukraine