Journal of Computer-Aided Molecular Design, **16:** 299–300, 2002.
KLUWER/ESCOM

Recent advances in molecular diversity

Chris L. Waller
Pfizer Global Research & Development, Ann Arbor Laboratories, 2800 Plymouth Road 016/330W/BIF/6, Ann Arbor, MI 48105, USA

Over the past decade, the size of the typical mid- to large pharmaceutical company compound collection has increased dramatically. Collections of several millions of unique compounds plated and available for screening are not uncommon in the industry. Along with the size of the collections, the very nature of the collections themselves has changed. Small collections typically of the order of 100,000s of discrete compounds synthesized as small series in traditional medicinal chemistry-lead optimization campaigns have been dramatically 'enhanced' by the addition of 'combinatorial libraries' with 1000s of representatives in each set being registered.

The impact of these screening library enhancement efforts has been less than dramatic. In retrospect, the early efforts to enhance our screening collections with large libraries of large combinatorial compounds with relatively conservative structural modifications and questionable analytical profiles should have seemed like a bad idea. Previous and on-going attempts to 'multiplex' or create intentional mixtures of the above mentioned compounds in an effort to increase biological screening throughput, decrease screening cost, and conserve chemical matter has further enhanced the complexity of our tasks as modern-day pharmaceutical researchers.

Several scientific and technical advances have come to the 'rescue'. In recent years, spiking of the collections by biasing the similarity of the selected compounds toward those of compounds previously demonstrated to possess pharmaceutical activities has been the vogue. The opening chapter by Mary Bradley presented in this review provides a concise and comprehensive overview of the content, similarity, and diversity of many of the commercially available databases.

More recently, analyses on the commercially-available chemical databases, those published by Chris Lipinski seminal among these, have stimulated the recent trend to supplement corporate collections with pharmaceutically acceptable (read intestinally absorbable) compounds by computationally filtering large actual and virtual collections. The so-called 'Rule of Five', consisting of simple guidelines around desirable molecular properties, has become the dogma of the library design and compound acquisition efforts of most pharmaceutical companies. Paul Charifson and Pat Walters provide some insight into the rationale behind this trend in their chapter.

The emphasis on enhancing our corporate collections with compounds displaying properties of known drugs has led some researchers to suggest that we may be biasing our collections to be too drug-like, not hit- or lead-like. As any researcher involved in a lead optimization effort can attest to, compounds tend to grow in size and complexity during this process. Rarely do we trim any foliage from the tree. Expanding on this simple observation, Tudor Oprea provides evidence to support the notion that perhaps we should focus on additional or alternative properties in our efforts to enhance our compound collections.

Diversity analysis techniques, as well as molecular diversity itself, take many forms. While it is well outside the scope of this effort to provide a comprehensive review, I was interested in providing some overview of the types of diversity analysis techniques currently in use. Dimitris Agrafiotis provides an update on techniques that attempt to simultaneously optimize numerous molecular properties associated with

Correspondence e-mail: chris.waller@pfizer.com

molecular diversity. At the other end of the spectrum, Alexander Golbraikh and Alexander Tropsha report on some diversity analysis techniques implemented in structure–activity studies, specifically with respect to the composition and selection of training and test sets for quantitative structure–activity relationship model-building exercises.

Technology and computer hardware, and lack thereof, have until only recently precluded us from dealing with collections surpassing 10^6 in size. These barriers are being shattered on a daily basis. However, these real limitations have spawned numerous creative ways to assess diversity. Valerie Gillet provides a comparison study of reagent- versus product-based diversity analysis techniques.

The historical compound collection, as a source for leads to be optimized for biological activity, selectivity, and absorption, distribution, metabolism, and excretion (ADME) in follow-up medicinal chemistry campaigns, contains many compounds not suitable for use as pharmaceuticals. The vast majority of candidate compounds fail due to poor ADME properties or overt toxicity. While the 'Rule of Five' may be a surrogate source of information for some of the ADME-relevant properties, toxicity is much more difficult to predict using simple 'rules of thumb'. Many more direct approaches have been developed to assist us in our mission to enhance our screening collections with compounds possessing more favorable ADME and toxicity profiles. Sean Ekins, Bruno Boulanger, Peter Swaan, and Maggie Hupcey provide a comprehensive overview of the field of ADME/toxicity research as it relates to molecular diversity. Scott Boyer and Ismael Zamora, in their contribution, further extend this topic in a discussion on new techniques in predictive metabolism.

There are numerous examples of marketed pharmaceuticals that were derived from humble natural product extracts. While the advent and widespread adoption of combinatorial chemistry as a more directed means to enhance the diversity of the corporate screening collection has prompted the fall in popularity of natural products research in pharmaceutical companies, concepts from natural products research are incorporated into our compound collection diversity enhancement design strategies. Douglas Horton, Gregory Bourne, and Mark Smythe provide a nice example of this in their contribution on the design and synthesis of privileged structures derived from natural products. This concept is further illustrated in the final article by Jürgen Bajorath in which the design of hybrid structures, or compounds embodying both synthetic and natural product characteristics, is presented.