Send Orders for Reprints to reprints@benthamscience.net

Open Innovation Drug Discovery (OIDD): A Potential Path to Novel Therapeutic Chemical Space

Maria Alvim-Gaston*, Timothy Grese, Abdelaziz Mahoui, Alan D. Palkowitz, Marta Pineiro-Nunez, and Ian Watson

Open Innovation Drug Discovery, Discovery Chemistry Research and Technologies; Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana 46285, USA

Abstract: The continued development of computational and synthetic methods has enabled the enumeration or preparation of a nearly endless universe of chemical structures. Nevertheless, the ability of this chemical universe to deliver small molecules that can both modulate biological targets and have drug-like physicochemical properties continues to be a topic of interest to the pharmaceutical industry and academic researchers alike. The chemical space described by public, commercial, in-house and virtual compound collections has been interrogated by multiple approaches including biochemical, cellular and virtual screening, diversity analysis, and *in-silico* profiling. However, current drugs and known chemical probes derived from these efforts are contained within a remarkably small volume of the predicted chemical space. Access to more diverse classes of chemical scaffolds that maintain the properties relevant for drug discovery is certainly needed to meet the increasing demands for pharmaceutical innovation. The Lilly Open Innovation Drug Discovery platform (OIDD) was designed to tackle barriers to innovation through the identification of novel molecules active in relevant disease biology models. In this article we will discuss several computational approaches towards describing novel, biologically active, drug-like chemical space and illustrate how the OIDD program may facilitate access to previously untapped molecules that may aid in the search for innovative pharmaceuticals.

Keywords: Open innovation, chemical space, drug discovery.

1. INTRODUCTION

The innovation gap in drug discovery research is a topic referenced daily in the popular press and in scientific journals [1, 3]. The challenge of identifying new molecular scaffolds for unprecedented biological targets is a key contributor to this gap. Although the set of hypothetical organic chemical structures is essentially limitless (number of possible molecules is estimated to be between 10^{18} and 10^{200}) [4], resource constraints have steered researchers toward techniques that will generate the most compounds in the shortest time, rather than broadly interrogating "chemical space." Chemical space can be defined as the potential universe of chemical structures, wherein the dimensions of that space can be any set of properties chosen in order to differentiate between molecules that are more or less similar. Subsets of this chemical space include "drug-like chemical space", the proportion of chemical space with property constraints appropriate for pharmaceuticals, and "biologically active chemical space", the proportion of chemical space with properties conducive to interacting with biological targets. In fact, drug discovery has substantially broadened our understanding of this chemical space through the exploration of more complex and topologically diverse structures, including natural products, peptides, and fragments.

Today, small molecules can be synthesized in greater numbers and diversity than ever before. The fundamental question of whether there is sufficient overlap between drug-like chemical space and biologically active chemical space to provide sufficient substrate for future drug discovery efforts remains unanswered. Furthermore, the size of a chemical library required to adequately represent this diversity is unknown. Since no single chemical library can practically cover chemical space, it is intriguing to consider how a diverse chemical library assembled from contributing scientists from around the globe, might compare to currently available sources and be utilized to deepen our understanding of biological systems.

To address some of the questions described above, Lilly has established a global network with academic institutions and small biotech companies. Through this network, known as Open Innovation Drug Discovery (OIDD), researchers are invited to submit compounds for evaluation in proprietary biological assays in return for detailed, publication-quality data from those assays. By providing a secure platform protecting intellectual property, the OIDD program lowers the barrier for collaborations between researchers working in different organizations (Fig. **1**). In this paper, we describe the OIDD program and how it provides opportunities for affiliated researchers to uncover possible biological activities of novel chemical structures as well as to test experimental hypotheses. In addition, we will discuss how the unique characteristics of this expanding chemical library may help to enhance efforts to identify new breakthroughs in important areas of unmet medical need.

*Address correspondence to this author at the Eli Lilly & Company, Lilly Corporate Center, Indianapolis, IN 46285 U.S.A.;
Tel: (317) 277 3398; Email: alvimgaston@lilly.com

1873-5294/14 $58.00+.00

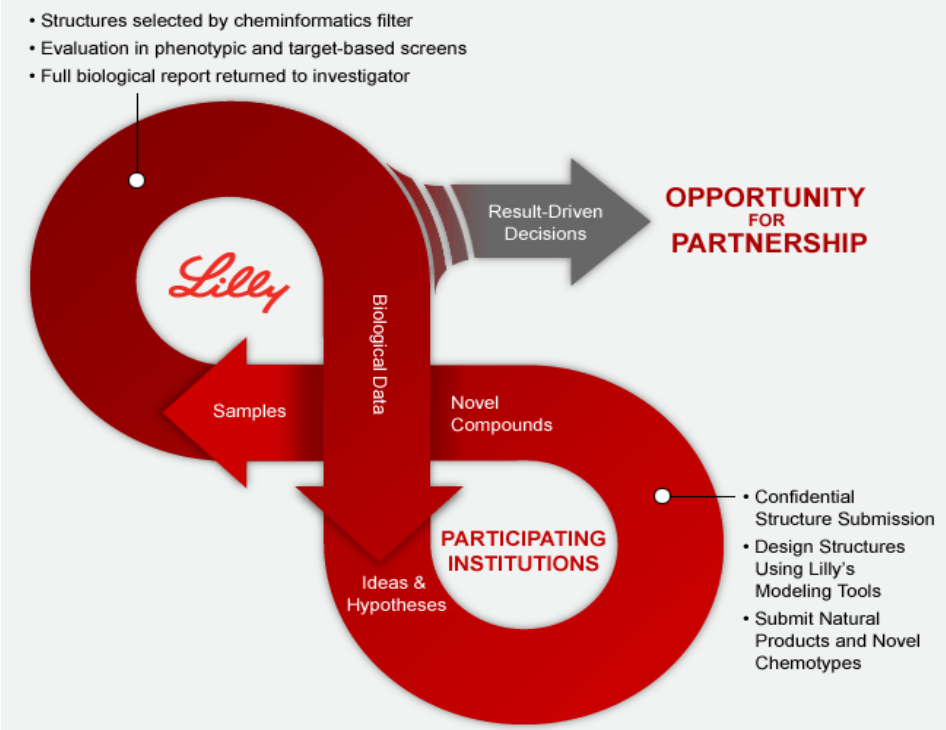

Fig. (1). The Open Innovation Drug Discovery Platform Model.

2. THE OPEN INNOVATION DRUG DISCOVERY PLATFORM AT A GLANCE

The OIDD program was designed to encourage collaboration between Lilly scientists and external researchers and to access the novel chemical diversity present in individual research laboratories throughout the world. The program incorporates both phenotypic and target-based, disease-relevant biological assays and makes these available to researchers for evaluation of their novel compounds. Many scientists have molecules that they would like to explore as potential medicines, but for a variety of reasons they are unable to advance this work. By minimizing the barriers to sharing compounds and scientific data, Lilly hopes to encourage interaction between Lilly scientists and researchers at other institutions.

The OIDD web-based interface serves as the central means of connecting external investigators with Lilly scientists. Via the secure portal, participants upload and submit structures of their compounds to be evaluated for novelty and reasonable drug-like characteristics (*vide infra)*. Should a molecule be selected, investigators submit the physical sample to Lilly for testing. In exchange for the submission, the investigator receives a full report of all the biological data generated. Once testing is completed, Lilly evaluates the data and determines whether to initiate further collaboration discussions.

The experimental centerpiece of the OIDD program is the panel of proprietary *In vitro* biological assays. These assays are organized in project modules that are closely aligned with internal scientific research in areas of long-term strategic interest such as endocrine, cardiovascular, neuroscience, and oncology. Phenotypic modules query complex cellular systems instead of specific targets, thereby interrogating the relevant biological framework without predisposed bias toward mechanism(s). In this case, the modules identify compound actives that may interact with one or more targets or pathways not anticipated by a single mechanism-driven hypothesis. In essence, phenotypic approaches screen multiple mechanisms and targets simultaneously. Furthermore, since the initial readouts from cellular assays are more information-rich, the connection of compound action to disease-relevant phenotypes is established earlier in the drug discovery process. The challenge with this approach lies in the complexity of fully understanding and assessing compound differentiation and elucidating with greater resolution the possible mechanisms of action. Fortunately, this complexity has been substantially reduced with the development of advanced assay technologies and informatics tools that make these challenges tractable for drug discovery.

The target-based modules focus on evaluation of a disease hypothesis through the testing of molecules designed to interact with a specific genomic target believed to be involved in pathogenesis. With the sequencing of the human genome and the development of many high throughput and complimentary drug discovery technologies, target-based drug discovery has been the primary strategy of many pharmaceutical companies during the past 20 years. This approach has been advanced by the development of computational and informatics tools that aid scientists in the design, selection and optimization of molecules for specific enzymes, receptors and other bioactive proteins.

Finally, the OIDD program also includes access to the state-of-the-art TB assay models run by the Infectious Disease Research Institute (IDRI) as part of the Lilly TB Drug Discovery Initiative (tbdrugdiscovery.org). Through this initiative, Lilly has opened access to its drug discovery ex-

pertise and scientific resources, and facilitates interaction between external investigators participating in OIDD and the scientists at IDRI.

Since the inception of the program, about 170,000 virtual structures have been uploaded into the system and about 70,000 have been accepted (Fig. **2**).

To date, approximately 20,000 samples have been received and evaluated in biological assays. As illustrated by the hit rates and promiscuity graphics (Fig. **3**), the OIDD program is indeed uncovering biologically active compounds with specificity across modules, with 82 % of the active compounds showing activity in only a single module. As expected, promiscuity is more apparent in phenotypic versus target-based assays.

3. OIDD CHEMINFORMATICS GUIDELINES

As a matter of practicality, Lilly does not wish to duplicate biological data for compounds already in our collection. In addition, our intent is to focus our biological testing on novel chemical diversity distinct from molecules disclosed in the literature, with reasonable drug-like properties. However, analyses of these parameters for a given compound would normally require disclosure of the chemical structure which may compromise intellectual property rights for an inventor.

In order to avoid this issue, a secure, automated cheminformatics analysis was enabled, whereby structures are converted to a set of bit-strings (molecular fingerprints) on the secure OIDD web application server and stored in the OIDD database. For those computations that need to be performed on Lilly internal cheminformatics servers, fingerprints, and not the structure itself, are transferred.

Within the secure website, submitted molecular structures are processed by the Lilly Medicinal Chemistry Rules (Lilly MedChem Rules) [5]. The Lilly MedChem Rules are comprised of a set of 275 rules, developed over an 18-year period, to identify compounds that may interfere with biological assays, allowing their removal from screening sets. Reasons for rejection include reactivity (e.g., acyl halides), interference with assay measurements (fluorescence, absorbance, quenching), activities that damage proteins (oxidizers, detergents), instability (e.g., latent aldehydes), and lack of druggability (e.g., compounds lacking both oxygen and nitrogen). The structural queries were profiled for frequency of occurrence in drug-like and non-drug-like compound sets and were extensively reviewed by a panel of experts (medicinal chemists). Also, an index of biological promiscuity was developed for profiling the rules and as a filter in its own right [5]. Approximately 600 gene targets with screening data at Lilly were assigned to 17 subfamilies, and the number of subfamilies at which a compound was active was used as a promiscuity index [5]. For certain compounds, promiscuous activity disappeared after sample repurification, indicating interference from concealed contaminants. Because this type of interference is not amenable to substructure search, a "nuisance list" was developed to flag interfering compounds that passed the substructure rules.

Next, and in order to identify and exclude those compounds containing undesirable features, further processing is performed to determine other molecular properties that help identify structures with the highest probability of success. These include MW, cLogP, solubility, and toxicity, among others. In addition to substructure-based filters, the Lilly MedChem Rules discard any molecule with fewer than 7 heavy atoms, or more than 50 heavy atoms. There are also limits on number of rings, number of aromatic rings, ring size and similar features. Unlike other rule sets, a molecule may be discarded due to the combined impact of multiple, undesirable features being present [5].

Finally, the molecular fingerprints are transferred to the internal cheminformatics evaluation servers to enable diversity analyses as compared to the Lilly internal collection, overlap comparisons to publicly available structures, and a check of similarity to controlled substances. An automated process performs these comparisons, destroys any information used to perform the analysis, and returns an overall desirability score. The cheminformatics evaluation is

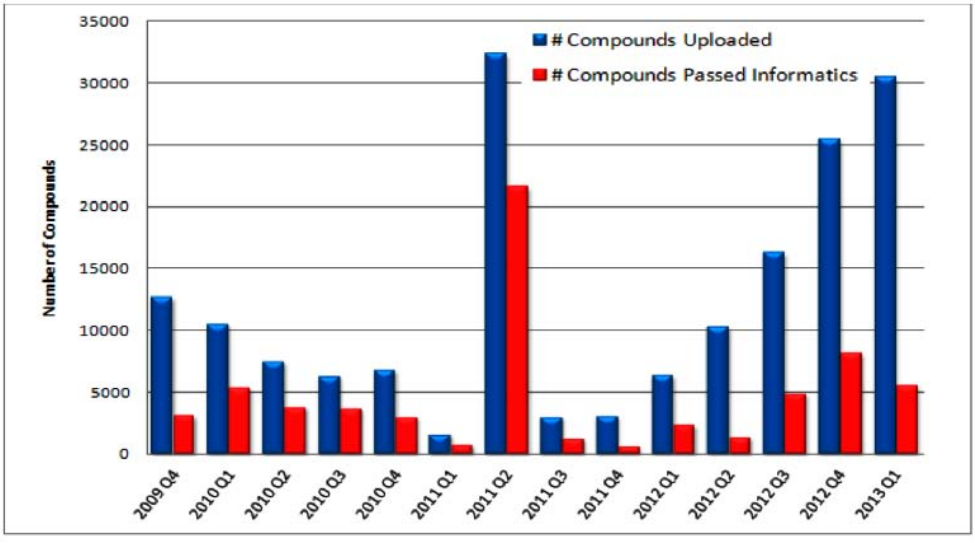

Fig. (2). OIDD Chemoinformatic Filter Acceptance Rate.

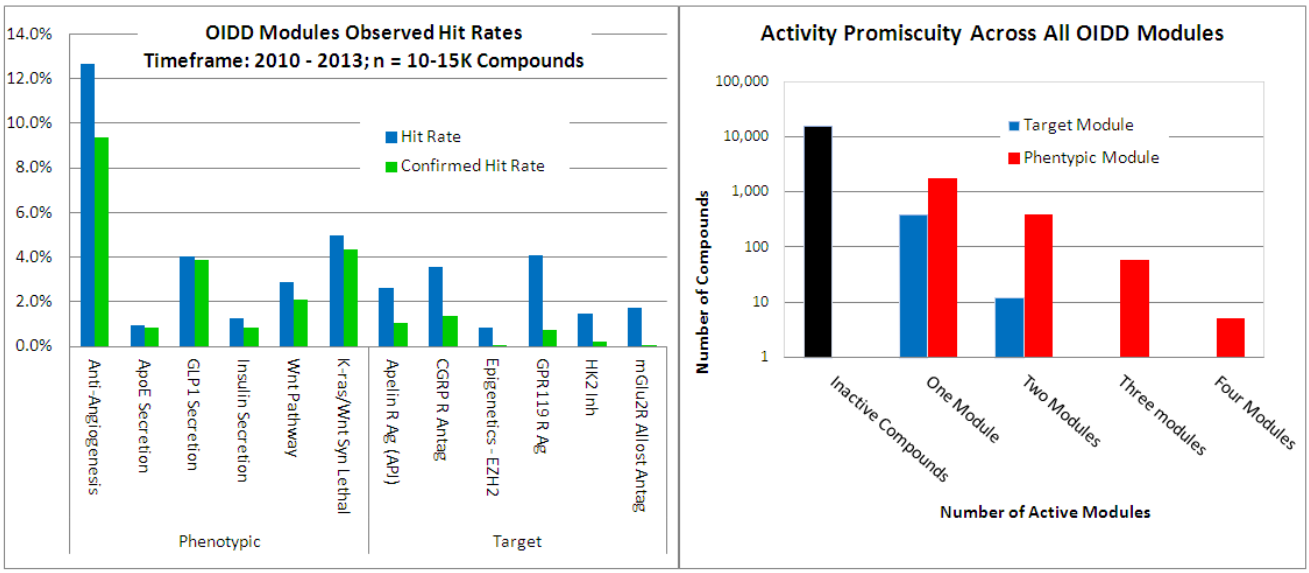

Fig. (3). OIDD Modules Hit Rates & Promiscuity.

accomplished without Lilly personnel being able to view and/or access the submitted structures to retain the confidentiality of the OIDD submitter's compounds (see Fig. **4**).

4. OIDD CHEMINFORMATICS PROCESSING

The OIDD cheminformatics processing is designed to rapidly evaluate large numbers of structurally-blinded molecules. There are three steps performed within the secured area of the OIDD web site: fingerprint generation and Lilly MedChem Rules check, as described in section 3, followed by comparison with previous submissions. The molecular fingerprints are a critical component of the molecular similarity comparisons [6] and model predictions used in the evaluation of the submitted structures. For the molecules that pass the cheminformatic evaluation on the external secure site, these fingerprints are the only information that passes the Lilly firewall to the internal cheminformatics server. It is noteworthy that the fingerprints that cross the firewall contain no identifying information that would enable discovery of a link to an outside entity; additionally the fingerprint is accompanied only by a system-generated random identifier.

Of the many different kinds of molecular fingerprints available, OIDD uses hashed linear path fingerprints, hashed circular fingerprints and encoded whole molecule properties. It is important to note that none of the dictionary-based fingerprints are used in order to increase the encryption of the molecule and limit ability to reconstruct a molecule from this set of fingerprints. The hashed linear path fingerprints are conceptually similar to Daylight3 fingerprints [7], ChemAxon Marvin4 fingerprints [8], a subset of the Tripos5 fingerprints [9], and others. All paths to a given length are generated, and a number in the range $0\text{-}2^{32}$ generated for each path. That number is then hashed to a constant width of 2048 bits. Since there are many more paths than 2048, inevitably there will be hash collisions, where different paths set the same bit. At 2048 bits, on average the collision rate is around 3%. These bits record presence or absence of a feature (or features), and thus are insensitive to repeated features.

The hashed circular fingerprints are conceptually similar to the fingerprints available in Pipeline Pilot6 [10]. For each atom in the molecule, concentric shells are examined. For each shell, a number in the range $0\text{-}2^{32}$ is generated, often including integer overflow. Unlike the linear path fingerprints, these fingerprints include a count of the number of features found. Again, there can be collisions due to accidental corruptions in the computation of atom types and shells, as well as during unsigned integer overflow. To circumvent the issue of repeated features in the linear path fingerprints, these fingerprints are often combined with a small set of whole molecular properties. This augmentation has proven to yield better similarity measures than what would otherwise be obtained by ignoring repeated features. The features used are: number of heavy atoms, size of the largest ring, number of rings, number of ring atoms, aromatic atoms, fused ring atoms, number of heteroatoms, and the number of unsaturated atoms.

The overall similarity measure between two fingerprints is a combination of the differences in these properties and the Tanimoto measure of the fingerprints. The intent is to correct for deficiencies that come from hashed, binary, linear path fingerprints. Over many years of use and tuning, this particular combination has gained widespread acceptance among the Medicinal and Computational Chemistry community, as it represents a good compromise between speed of computation and results. Better similarity measures can be computed using counted features and non linear shapes, but because they are counted, they cannot use the tremendous efficiencies available to binary fingerprints, and so their significantly longer computation times restrict their role to situations where accuracy is more important than speed.

In order to better protect the intellectual property of the OIDD submitters, while deriving maximum information from the limited assay capacity available, molecules that are

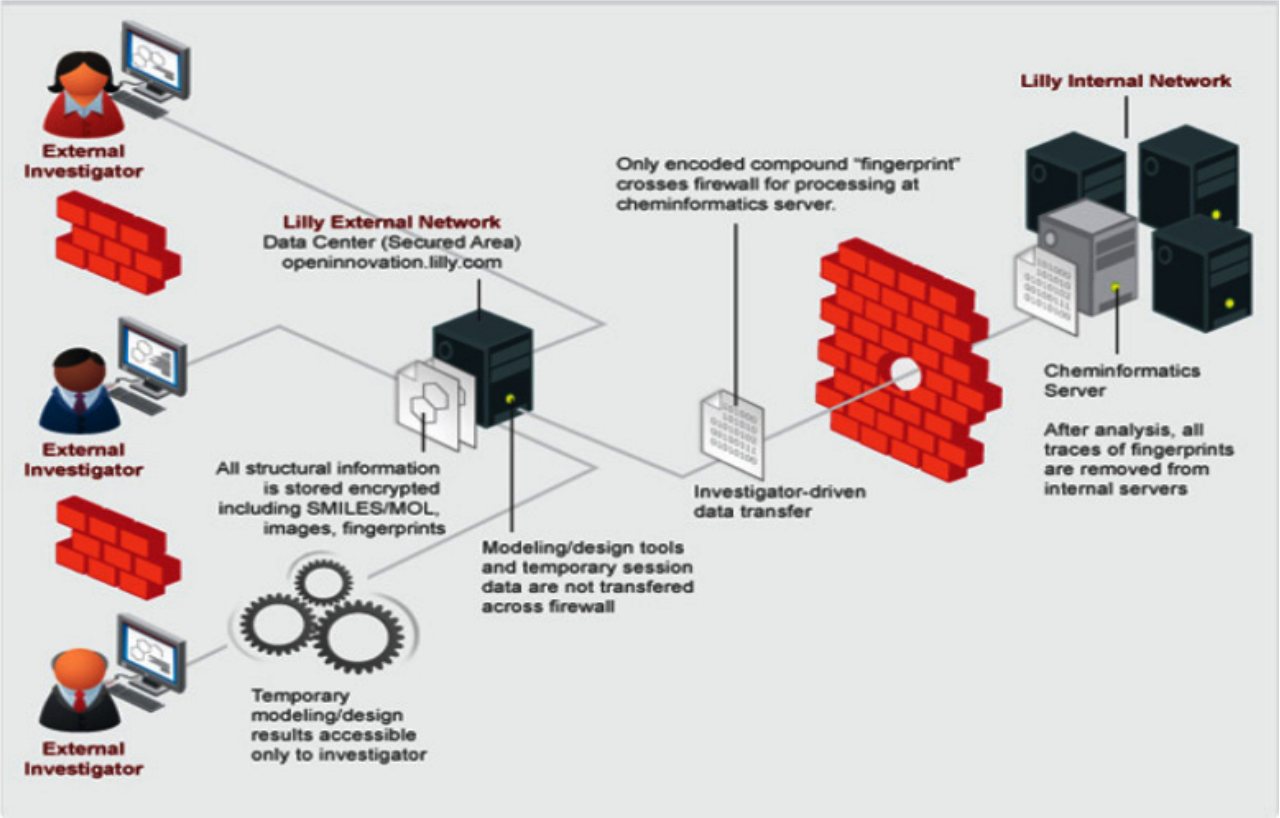

Fig. (4). The Open Innovation Drug Discovery Program Security Provisions.

too similar to those that have previously been submitted by another user are excluded from testing. To accomplish this comparison, the fingerprints of compounds submitted for screening are continually stored and then as part of an evaluation outside the firewall, a Tanimoto similarity comparison of the fingerprints of the new compounds to those previously submitted is carried out. Molecules that are found to have a 90% or higher similarity to a previous submission are discarded for insufficient novelty. Storing these fingerprints outside the firewall, in a secure location, is part of the safety measures that the OIDD program takes to protect the submitted structures.

For molecules passing the cheminformatic evaluation outside the firewall, the corresponding fingerprints pass the Lilly firewall to perform a controlled substance check and comparison with known drugs, the Lilly and PubChem collections. Molecules exhibiting over 90% similarity to a known controlled substance or greater than 85% similar to a known drug are discarded. Kernel-based models are applied as negative selectors, like various models of absorption, distribution, metabolism, and excretion (ADME) properties, solubility predictions, and other relevant physicochemical properties. The object of these computations is to assign a relative desirability score to each fingerprint. A fingerprint's desirability will increase if it is novel compared to our existing collection and if it scores well against one or more predictive models. However desirability will decrease if predictions suggest undesirable ADME or physical properties. A final single numeric, relative desirability score is sent back across the firewall to the external OIDD secure site. That relative desirability score is stored in the database and is used to prioritize molecules to be accepted for biological screening.

This cheminformatics processing minimizes non-drug-like molecules submitted for screening and maximize the chances of screening compounds that will provide the properties profile(s) targeted for development into viable drug candidates.

5. DIVERSITY OF THE OIDD COLLECTION

As described in section 2, the OIDD collection has proved to be a source of biologically active molecules that significantly complements the internal Lilly compound collection. One measure of its diversity is the variety and global nature of compound sources. Since its creation, the OIDD program has steadily grown and currently there are in excess of 700 individual user accounts created by more than 300 institutions distributed across 30 countries around the globe. About half of the participating institutions are located in the United States and about one third in Europe, with the remaining distributed among North America, Africa, Asia, and the Middle East (Fig. **5**).

Because of the selection criteria, it is anticipated that the OIDD collection should evidence significant diversity, compared not only to the Lilly collection, but also to other publically available compound sources. To illustrate this point, a fingerprint similarity comparison of the OIDD collection with the PubChem database, a publically available repository of molecular structures containing about 20 million unique structures, was performed. The results of this comparison showed that the OIDD collection is significantly different

Fig. (5). The OIDD Global Outreach.

from the PubChem collection. The similarity metric used for this comparison was a composite of binary fingerprints and a small set of molecular properties, previously described in the OIDD cheminformatics processing (*vide supra*).

Therefore, for each OIDD molecule the distance from that molecule to each molecule in PubChem was computed in order to determine the shortest distance from the OIDD molecule to any molecule in PubChem. During this exercise, a locally-hosted copy of PubChem was used to ensure that all OIDD fingerprints were never transferred outside the secure space they occupy at Lilly. A histogram of the shortest distances measured in this comparison was then generated, representing number of molecule instances at that particular distance. In this manner, a few molecules within 0.15 distance of a PubChem molecule were observed, but very importantly, a large proportion (71%) of the OIDD molecules are significantly different (distance > 0.1) from anything known in the PubChem database, demonstrating that the OIDD program is indeed accessing very novel molecules (Fig. **6**).

Comparing the chemical space of compound collections is not a trivial task since it is very dependent on the method used and the structural representation of the compounds. Traditionally, compound databases have been compared using physicochemical properties, or fingerprints. These comparisons usually focus on only one or two criteria that do not always necessarily provide a comprehensive assessment of its chemical space [11]. For a more comprehensive analysis of the density coverage of the OIDD collection chemical space and its comparison to the internal Lilly collection and PubChem, projections of the collections on the two principal components (PC) of property space were carried out, using molecular weight, cLogD at pH7.4, aromatic density, fraction of sp3 atoms, and H-bond donor and acceptor as the descriptors. Analysis of the property space by means of PC, illustrates that the OIDD collection has different degrees of overlap with the property space defined by either the Lilly or the PubChem collections, and shows a distinctive distribution in property space in comparison with both collections (Fig. **7**).

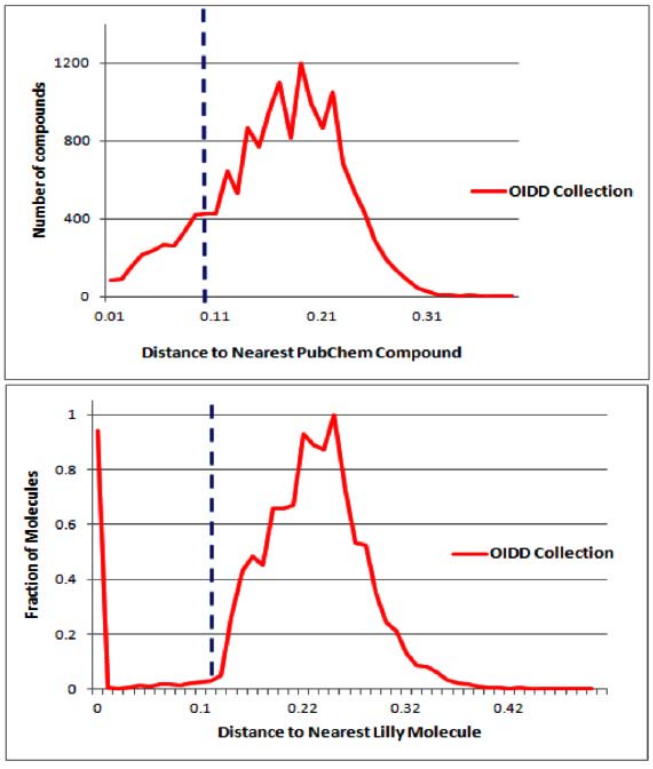

Fig. (6). OIDD Structural Diversity Relative to PubChem & Lilly Collection.

Another method utilized for structural characterization of different compound collections is based on an approach introduced by Sauer and Schwarz utilizing principal moments of inertia (PMI) to describe the overall three-dimensional

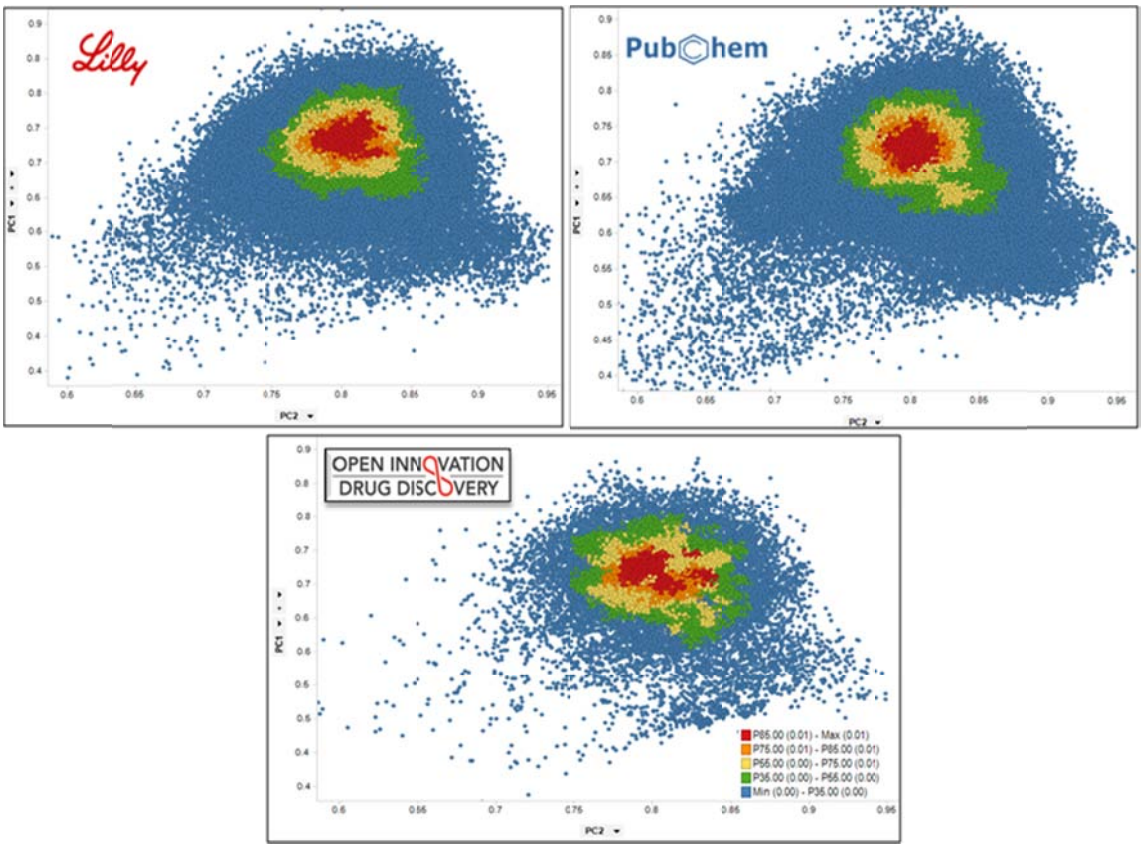

Fig. (7). OIDD Property Space Comparisons with Alternative Diversity Sources.

shape of molecules [12]. More recently, efforts by Clemons and Schreiber have shown this method to be useful for description and comparison of compound collections [13]. A similar analysis performed on the OIDD compound collection shows that the overall shape is heavily "rod-like" in nature with a significant degree of "spherical" character. Overall, the OIDD collection does not have a significant representation of "flat" compounds, correlating with high sp^3 content within the collection. This shape distribution distinguishes the OIDD collection from Lilly and the PubChem compound collection (Fig. **8**).

6. THE OIDD COLLECTION AND DRUG-LIKENESS

In order to be safe and effective, an orally administered drug requires potency and selectivity for its therapeutic target, as well as an acceptable balance of ADME properties. As ADME properties are determined in large measure by the compound's physical properties such as lipophilicity and solubility, investigators make every effort to work at the interface between structure activity relationships (SAR) and structure property relationships (SPR), as they conduct multiple iterations of the "design-test-modify" research cycle. Unfortunately, many times improvement in one property leads to unfavorable changes in another. In terms of compound design, structural modifications that affect molecular size, topology and flexibility are important considerations. Among the key physicochemical assessments (pKa, solubility, permeability, stability and lipophilicity) solubility is recognized as being particularly important. Compounds with insufficient solubility carry a higher risk of failure during discovery and development, since low solubility may compromise other property assays, mask additional adverse properties, influence both pharmacokinetic and pharmacodynamic properties of the compound, and finally affect the ability to develop the compound [14]. Poor aqueous solubility is caused mainly by high lipophilicity and strong intermolecular interactions which make the solubilization of the solid energetically expensive. If solubility is incorrectly estimated, this can lead to erroneous interpretation of results in a number of *in vitro* assays and weaken the interpretation of SAR [15-17].

In order to improve the likelihood of finding orally available small molecule drugs, chemists have long sought to define the boundary conditions for small molecules to be acceptable as potential drugs. Since medicinal chemists can calculate parameters like MW, cLogP, cLogD, rotatable bonds, and topological polar surface area (TPSA), these measures are often substituted at the molecule design stage for more complex properties such as solubility. In 1997 Lipinski formulated the so called "Rule of 5" (RO5) as criteria for oral bioavailability [18], a major breakthrough for the cheminformatics community. The RO5 stated that poor absorption or permeation are more likely when there are more than 5 H-bond donors (HBD); molecular weight (MW) is over 500; cLogP is over 5; and the sum of H-bond acceptors (HBA), N's and O's is over 10. While the rule is useful as a mnemonic device, it is too simplistic to identify all drug-like molecules, as many of today's blockbusters fail Lipinski rules.

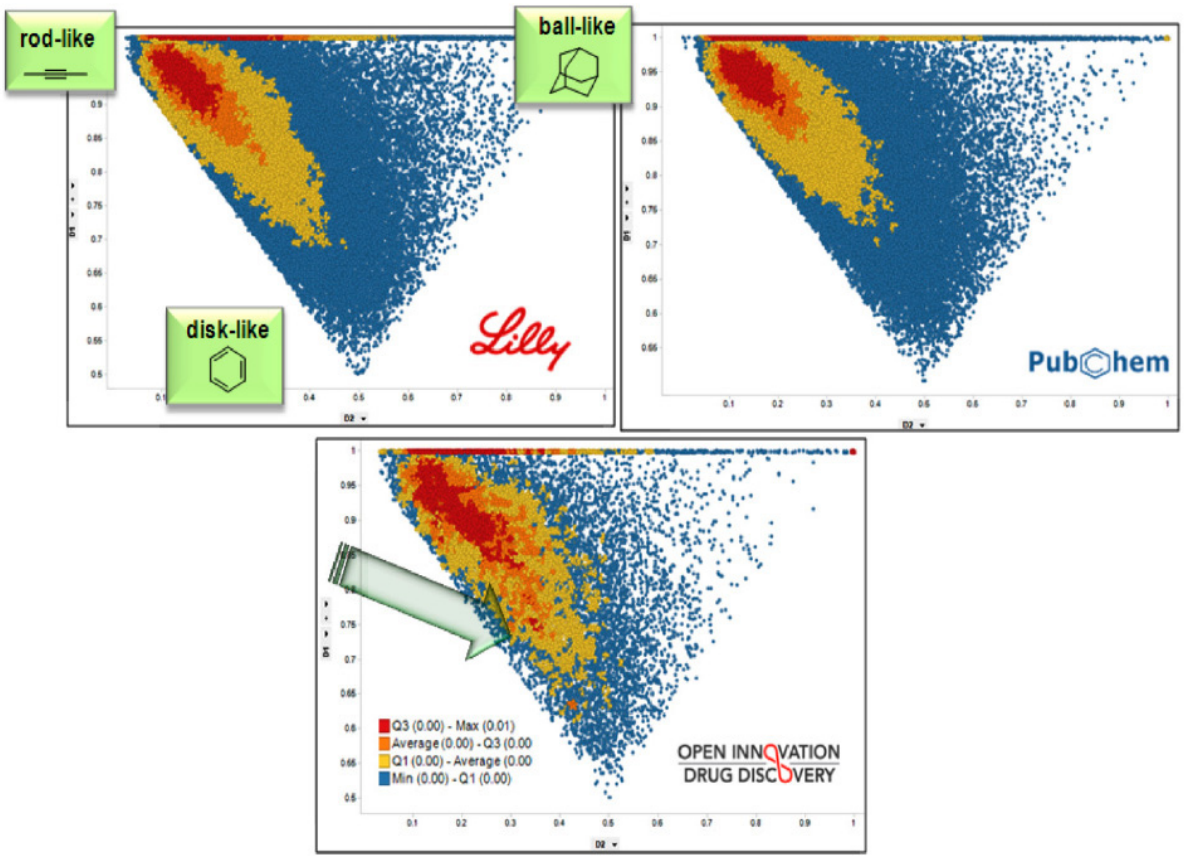

Fig. (8). Principal Moment Inertia (PMI) analysis of the OIDD, Lilly & PubChem collections.

Since Lipinski, various groups continue to refine and publish profiling methods in the attempt to define drug-likeness [19] and lead-likeness [20] and improve predictions to increase the efficiency of drug discovery. The RO5 has been revisited many times with a number of variations, [21-23]. Thus, another modification that defines the concept of lead-likeness is the "Rule of Three" (RO3) [24]. A RO3-compliant compound is defined as one that: has no more than 3 HBD and 3 HBA, MW is less than 300, cLogP is not greater than 3, and has no more than 3 rotatable bonds.

Approaches that avoid the hard cut-offs present in systems similar to RO5 and RO3 and instead allow for acceptable trade-offs to be defined have become more common. These so-called desirability functions map a property value onto a scale between zero and one, representing the desirability of a compound with respect to that particular property. An ideal value will achieve a desirability score of one, while a completely unacceptable value will receive a desirability score of zero. Desirability scores of individual properties can then be easily combined into an general "desirability index" to reflect the overall quality of a compound by adding them together or taking the average or geometric mean. Still, the results may be easily interpreted, since the impact of each individual property to the overall desirability index can be calculated to guide strategies to improve the overall quality.

A group at Pfizer has described an application of desirability functions to prioritize compounds with a greater chance of success against a central nervous system (CNS) target. Its CNS Multi-Parameter Optimization approach (CNS-MPO) employs six calculated physicochemical parameters (MW, cLogP, cLogD, TPSA, HBD and basic pKa) to calculate a desirability index in the range of 0 to 6. This group reported that 74% of a set of marketed drugs for CNS targets achieved a CNS-MPO index of ≥ 4, compared with only 60 % of the Pfizer candidates, a statistically significant difference. At one level, this can be considered as a measure of CNS drug-likeness. Therefore, in order to determine how molecules submitted to OIDD relate to other compound collections using the MPO index, a comparison between the OIDD collection and diverse subsets of the PubChem and Lilly collections was initiated. Within the OIDD collection, 70% achieved a CNS-MPO index of ≥ 4 compared to 64% for PubChem and 68% for the Lilly collection. Furthermore, this trend continues when evaluating the "active" compounds from the OIDD collection, with 66% of all OIDD actives having a MPO index ≥ 4. Clearly, the biological activity observed is not concentrated in the subset of compounds with sub-optimal physicochemical properties.

It has also been reported that a high CNS-MPO index correlates with positive outcomes for several key *In vitro* ADME and toxicity endpoints including permeability, metabolic stability, active transport by P-glycoprotein, cytotoxicity and hERG inhibition [25]. Although this approach was intended to improve the potential to design compounds with good CNS penetration, the underlying principles were fundamental in nature, such that they were extended beyond the CNS drug space. This approach allows for a holistic assessment of drug-like ADME and safety properties that does not rely on a single property, although the selected parameters

contain a significant bias toward lipophilicity as the key element.

As the understanding of the importance of lipophilicity to shape ADME has increased in the past decade and with this relationship firmly established, attention has turned to find other molecular descriptors that can be also correlated to clinical success through their influence on physicochemical properties. Recent studies have provided additional molecular descriptors that complement the well-established physicochemical properties (MW, lipophilicity, and ionization state) by incorporating factors that relate to molecular topology such as aromatic ring count, fraction of sp^3-hybridized carbons (Fsp^3), chiral atom count, aromatic atom count, sp^3 atom count, and fraction of the molecular framework (fMF) [26-31].

Yang and colleagues investigated how various important ADME properties are influenced by two molecular descriptors related to the molecular topology, fMF and Fsp^3 (measure of saturation) [32] and uncorrelated with molecular size and lipophilicity. The results reported not only confirmed that MW, lipophilicity, and ionization state are very important descriptors for ADME predictions, but also demonstrated that ADME properties are substantially influenced by molecular topology. Of particular interest is the finding that both fMF and Fsp^3 influence aqueous solubility (Table **1**). Solubility decreases with increasing fMF and increases with increasing Fsp^3, and this trend is independent of the ionization state of the molecule [33].

Table 1. Summary of the Influence of f_{MF} and Fsp^3 on Investigated ADME Assays [33].

	Fsp^3	f_{MF}
Aqueoussolubility	Fsp^3↑ , solubility↑	f_{MF} ↑ , solubility ↓
Caco-2 permeability	Fsp^3↑ , Caco-2↓	f_{MF} ↑ , Caco-2 ↑
hPPB	Fsp^3 ↑ , fu↑	f_{MF} ↑ , fu ↓
hERG Inhibition	No Influence	f_{MF} ↑ , hERG inhibition ↑
CYP3A4 Inhibition	No Influence	f_{MF} ↑ , CYPD3A4 inhibition ↑

Frequently, more highly complex molecules with respect to saturation have greater three-dimensionality and reduced conformational flexibility. Since protein-binding sites also have a high degree of three-dimensional character, this should increase receptor/ligand complementarily due to the increased opportunity to incorporate out-of-plane substituents and to adjust molecular shape. Additionally, the reduced conformational flexibility that saturation imparts also reduces the number of potential protein interacting partners for a given molecule, and therefore increasing sp^3 character may also result in greater selectivity and fewer off-target effects. Lovering and colleagues have identified a very simple descriptor for saturation which is easily interpretable [28]. As compounds are prepared in the drug discovery setting and transition from discovery through clinical trials to drugs, those that are more highly saturated are more likely to succeed in these transitions. Thus, the average Fsp^3 is reported to be 0.36 for discovery compounds and increased to 0.47 for drugs, and this trend is carried through all of the stages from discovery to drug, where each phase had a higher Fsp^3[28]. If one compares the Fsp^3 values for the Best Selling Drugs set [34], PubChem, Lilly and the OIDD collection, the OIDD molecules compare favorably in this dimension as well (Table **2**).

Table 2. OIDD Calculated Fsp^3 Values Versus Other Datasets.

Data Set	Fsp^3 Average
Lilly collection	0.32
PubChem	0.36
OIDD collection	0.39
Best Selling Drugs	0.41

CONCLUSION AND PERSPECTIVES

The OIDD program has exemplified several new strategies in the way pharmaceutical companies can interact with the external scientific world. It is anticipated that these interactions will produce new science and potentially lower costs and further increase the quality of pharmaceutical innovation. Early results are promising, with a network of hundreds of affiliated academic institutions and small biotech companies submitting thousands of samples for evaluation in the OIDD screening panel, resulting in several collaborations currently ongoing.

The OIDD platform has harvested a unique set of compounds with drug-like characteristics and when compared to both the internal Lilly collection as well as the publically available PubChem dataset, the OIDD collection has shown to be significantly dissimilar and with a different distribution in property space and in shape diversity.

The OIDD business model rests on the ability to automatically evaluate thousands of virtual structures to select those eligible for biological screening, without any human intervention in order to provide complete protection of the submitter's intellectual property. This is achieved via the conversion of chemical structures into molecular fingerprints on the secure OIDD web application before their utilization in chemoinformatics analyses. For those computations that need to be performed on Lilly internal chemoinformatics servers, only these fingerprints, and not the structures themselves, are transferred. As a natural extension of this approach, the OIDD program is reaching out to the global computational chemistry community by introducing a variety of Lilly-developed modeling tools that will provide investigators the option of designing molecules with desirable drug-like properties or suited to a particular biological target. Both SPR and SAR tools will be made available to authorized users to provide participants access to cutting-edge technology via the OIDD website portal.

As the Open Innovation Drug Discovery program continues to evolve, other research modules will likely be added. Scientists from all over the world are taking advantage of this resource to open new opportunities and to test novel therapeutic hypotheses that deepen the understanding of complex biological systems. Through this initiative and others like it, our steadfast goal remains focused on the discovery of novel therapeutics that will improve patients' lives. This is our ultimate measure of success.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the pivotal roles played by Ms. Saba Husain and Mr. Jeff McGowen in the design, implementation and operation of the OIDD program, along with the remaining members of the OIDD support team. We also want to recognize Dr. Jibo Wang and Dr. Christine Humblet for valuable technical discussions regarding computational and cheminformatic methodologies. Finally, we are grateful for the partnership of our colleagues in Discovery IT and Quantitative Biology groups in making this program a tangible success.

REFERENCES

[1] http://www.manufacturing.net/articles/2009/02/the-innovation-gap

[2] Making Academic-Industry Parternerships Work for both parties. *Tuffs CSDD R&D. Management Report*, **2012,** *7*(3), 1-4.

[3] http://www.ngpharma.eu.com/article/A-new-model-for-drug-discovery/

[4] Medina-Franco, J. L. ; Martínez-Mayorga, K.; Giulianotti, M. A.; Houghten, R. A.; Pinilla, C. ; Visualization of the Chemical Space in Drug Discovery Current. *Computer-Aided Drug Design*, **2008**, *4*, 322-333.

[5] Bruns, R. F.; Watson, I. A. ; Rules for Identifying Potentially Reactive or Promiscuous Compounds. *J. Med. Chem*., **2012**, *55*(22), 9763-9772.

[6] Willett, P.; Barnard, J. M.; Downs, G. M.; Chemical Similarity Searching. *J. Chem. Inf. Comp. Sci*., **1998**, *38*(6), 983-996.

[7] Daylight Chemical Information Systems, Inc. www.daylight.com

[8] ChemAxon www.chemaxon.com

[9] Tripos, A Certara™ Company, www.tripos.com

[10] Pipeline Pilot, Accelrys Software Inc.; www.accelrys.com

[11] Dandapani, S.; Marcaurelle, L. A.; Grand Challenge Commentary: Accessing New Chemical Space for "Undruggable" Targets*. Nat. Chem. Biol.,* **2010***, 6,* 861–863.

[12] Sauer, W. H. B.; Schwarz, M. K.; Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J. Chem. Inf. Comput. Sci.,* **2003***, 43*(3), 987–1003.

[13] Wilson, J. A.; Bender, A.; Kaya, T.; Clemon, P. A.; Alpha Shapes Applied to Molecular Shape Characterization Exhibit Novel Properties Compared to Established Shape Descriptors. *J. Chem. Inf. Model.*, **2009***, 49,* 2231–2241.

[14] Alsenz, J.; Kansy, M.; High Throughput Solubility Measurement in Drug Discovery and Development. *Advanced Drug Delivery Reviews,* **2007,** *59*, 546–567.

[15] Bergstrom, K.; Luthman, P.; Accuracy of Calculated pH Dependent Aqueous Drug Solubility. *Eur. J. Pharm. Sci.,* **2004,** *22*(5) 387–398.

[16] Avdeef, A.; pH-Metric Solubility 1:Solubility-pH Profiles from Bjerrum Plots, Gibbs Buffer and pKa in the Solid State. *Pharm. Pharmacol. Commun*. **1998,** *4*, 165–178.

[17] Avdeef, A.; Berger, M.; Brownell, C.; pH-Metric Solubility 2: Correlation Between the Acid-base Titration and the Saturation Shake-flask Solubility-pH Methods. *Pharm. Res*. **2000,** *17*, 85–89.

[18] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. ; Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.,* **1997**, *23*, 3-25.

[19] Lipinski, C. A.; Navigating Chemical Space for Biology and Medicine. *Nature,* **2004**, *432*, 855-861.

[20] Rishton, G. M.; Nonleadlikeness and Leadlikeness in Biochemical Screening. *Drug Discovery Today,* **2003**, *8*, 86-96.

[21] Bhal, S. K.; Kassam, K.; Peirson, I. G.; Pearl, G. M.; The Rule of Five Revisited: Applying LogD in Place of LogP in Drug-Likeness Filters. *Mol. Pharma.,* **2007**, *4*, 556.

[22] Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J.; A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. *J. Combin. Chem*., **1999**, *1*, 55–68.

[23] Veber, D. F.; Johnson, S. R.; Cheng, H.Y, Smith, B.R.; Ward, K.W.; Kopple, K. D.; Molecular Properties that Influence the Oral Bioavailability of Drug Candidate. *J. Med. Chem.,* **2002,** *45*(12), 2615–23.

[24] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H.; A 'Rule of Three' for Fragment-Based Lead Discovery? *Drug Discov. Today*, **2003**, *8*(19), 876–7.

[25] Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A.; Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Drug-like Properties. *ACS Chem. Neurosci.,* **2010,** *1*(6) 435-449.

[26] Ritchie, T. J.; Macdonald, S. J. F.; The Impact of Aromatic Ring Count on Compound Developability- Are Too Many Aromatic Rings a Liability in Drug Design? *Drug Discovery Today,* **2009**, *14*, 1011– 1020.

[27] Ritchie, T. J.; Macdonald, S. J. F.; Young, R. J.; Pickett, S. D.; The Impact of Aromatic Ring Count on Compound Developability: Further Insights by Examining Carbo- and Hetero-aromatic and Aliphatic Ring Types. *Drug Discovery Today,* **2011**, *16*, 164– 171.

[28] Lovering, F.; Bikker, J.; Humblet, C.; Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem*., **2009**, *52*, 6752– 6756.

[29] Yan, A.; Gasteiger, J.; Prediction of Aqueous Solubility of Organic Compounds by Topological Descriptors QSAR. *Comb. Sci.,* **2003**, *22*, 821– 829.

[30] Leeson, P. D.; St-Gallay, S. A.; Wenlock, M. C.; Impact of Ion Class and Time on Oral Drug Molecular Properties. *Med. Chem. Commun.,* **2011**, *2*, 91– 105.

[31] Yang, Y.; Chen, H.; Nilsson, I.; Muresan, S.; Engkvist, O.; Investigation of the Relationship Between Topology and Selectivity for Drug-like Molecules. *J. Med. Chem,.* **2010**, *53*, 7709–7714.

[32] Yang, Y.; Engkvist, O.; Llinàs, A.; Chen, H.; Beyond Size, Ionization State, and Lipophilicity: Influence of Molecular Topology on Absorption, Distribution, Metabolism, Excretion, and Toxicity for Drug-like Compounds. *J. Med. Chem.,* **2012**, *55*(8), 3667–3677.

[33] Andrews, P. R.; Craik, D. J.; Martin, J. L.; Functional Group Contributions to Drug-receptor Interactions. *J. Med. Chem*., **1984,** *27*, 1648–1657.

[34] McGrath, N. A.; Brichacek, M.; Njardarson, J. T.; A Graphical Journey of Innovative Organic Architectures that Have Improved Our Lives. *J. Chem. Ed*., **2010**, *87*, 1348-1349.

Received: July 18, 2013 **Revised: August 14, 2013** **Accepted: August 25, 2013**