ACS Chemical Neuroscience

Review

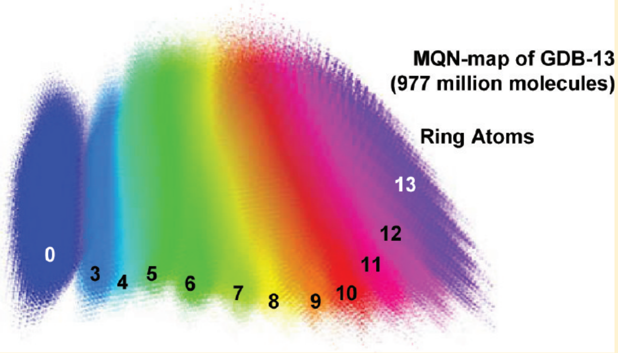
pubs.acs.org/chemneuro

Exploring Chemical Space for Drug Discovery Using the Chemical Universe Database

Jean-Louis Reymond* and Mahendra Awale

Department of Chemistry and Biochemistry, University of Berne, Freiestrasse 3, 3012 Berne, Switzerland

ABSTRACT: Herein we review our recent efforts in searching for bioactive ligands by enumeration and virtual screening of the unknown chemical space of small molecules. Enumeration from first principles shows that almost all small molecules (>99.9%) have never been synthesized and are still available to be prepared and tested. We discuss open access sources of molecules, the classification and representation of chemical space using molecular quantum numbers (MQN), its exhaustive enumeration in form of the chemical universe generated databases (GDB), and examples of using these databases for prospective drug discovery. MQN-searchable GDB, PubChem, and DrugBank are freely accessible at www.gdb.unibe.ch.



KEYWORDS: *chemical space, databases, virtual screening, docking, acetylcholine, glutamate*

Small molecule drugs exert their action by binding to specific molecular constituents of the cell such as to modulate biochemical processes in a disease modifying manner. The magnitude and specificity of binding depends on the complementarity between the drug molecule and its target in terms of shape, polarity, and chemical functionality. The success of small molecule drugs stems from the facts that (a) their size matches that of most biologically relevant binding sites; (b) the structural and functional diversity available in small molecules is sufficient to achieve strong and specific binding to most of these binding sites; and (c) the pharmacokinetics of small molecule drugs can be optimized while retaining target binding to enable efficacy and safety in vivo.[1]

Genome sequencing, proteomics, structural biology, and yeast two-hybrid screens have documented the extremely large number and diversity of potential drug targets and their interactions.[2] On the other hand, chemists have learned to deliver potent and selective ligands on demand by combining molecular design and synthesis methods with bioactivity assays and activity optimization protocols.[3] Although the experimental evidence of polypharmacology shows that many drug molecules are not selective and hit multiple targets,[4] the driving hypothesis of medicinal chemistry remains that a specific small molecule ligand can be found for any binding site.[5]

The hypothesis above assumes that the number and diversity of drug-sized molecules is sufficient to address all the different binding sites in biology. Hence, the following fundamental chemical question arises: how many molecules are in principle possible? This question was formulated in the early days of organic chemistry as soon as it was realized that organic structures can be described as graphs,[6] and over the years has remained a playground of theoretical chemistry.[7] The systematic enumeration of molecules found applications in the 1960s in the area of computer aided structure elucidation (CASE),[8] and since the 1990s to address molecular diversity in the context of combinatorial chemistry for drug discovery.[9] Considerations on the number of possible molecules has led to the concept of the "chemical space" to describe the ensemble of all organic molecules to be considered when searching for new drugs.[10] Herein we discuss the exploitation of this concept for drug discovery, with focus on the exhaustive enumeration of the small molecule chemical space realized in our group in form of the chemical universe database GDB and its utilization for ligand discovery in the area of neurotransmitter receptors and transporters.

1. THE KNOWN CHEMICAL SPACE

Whereas the theoretically possible chemical space is very large (see below), one may at first consider the known chemical space, that is, the ensemble of all organic molecules reported thus far. Thanks to several open access initiatives, this chemical space is currently accessible to the public (Table 1). Most of the databases listed in Table 1 can be searched for single compounds or their analogues by structure, name, or bioactivity.

The number of known molecules is impressive and interesting; however, this number alone does not provide any information on what these molecules are. The concept of "chemical space" suggests a representation in form of a geographical map to illustrate the distribution of molecules and their properties. To obtain such a map, one first creates a

Received: April 18, 2012
Accepted: April 25, 2012
Published: April 25, 2012

Table 1. The Known Chemical Space[a]

database	description	size[a]	Web address	ref
PubChem	known molecules from various public sources	32.5 M	http://pubchem.ncbi.nlm.nih.gov	11
Chemspider	online resource from the Royal Society of Chemistry	26.0 M	http://www.chemspider.com/	12
ZINC	commercially available small molecules	21.0 M	http://zinc.docking.org	13
NCI Open	anticancer and AIDS compounds with screening data	0.25 M	http://cactus.nci.nih.gov/ncidb2.1	14
ChemDB	commercially available small molecules	4.1 M	http://cdb.ics.uci.edu	15
BindingDB	bioactive molecules with binding affinity data	0.36 M	http://www.bindingdb.org	16
ChemBank	small molecules annotated with screening data	1.2 M	http://chembank.broadinstitute.org/	17
ChEMBL	small molecules annotated with experimental data	1.1 M	https://www.ebi.ac.uk/chembldb	18
CTD	comparative toxicogenomics database	0.17 M	http://ctdbase.org	19
HMDB	human metabolome database	0.0085 M	http://www.hmdb.ca	20
SMPDB	small molecule pathway database	0.001 M	http://www.smpdb.ca	21
DrugBank	experimental and approved small molecule drugs	0.0065 M	http://www.drugbank.ca	22

[a] Open access collections as of April 2012. Corporate collections and nonopen access sources are not listed.

property space by assigning dimensions to series of molecular descriptors. Each molecule is placed in this multidimensional property space using the descriptor values as positional coordinates, as first introduced by Pearlman and Smith.[10b] One then uses principal component analysis (PCA) to extract the most relevant dimensions (in form the principal components, PC), and represents projections of the chemical space in a PC-plane, usually the (PC1,PC2) plane.[23] Alternatively, one can also classify the descriptor value vectors using self-organizing maps, which consist of two-dimensional grids of nodes grouping the most similar vectors, and hence the most similar molecules, in nearby nodes.[24]

Thousands of different molecular descriptors exist, and any combination of these descriptors may be selected to produce a property space formally possessing tens to hundreds of different dimensions, from which a chemical space map can be derived.[23] We recently proposed a set of 42 integer value descriptors called molecular quantum numbers (MQNs).[25] MQNs count elementary features of molecules including atom and bond types, polar groups, and topological features, which are all easily identified in a structural formula by anyone with basic knowledge of organic chemistry (Table 2). The MQN system defines a simple and universal chemical space to classify organic molecules, in analogy to the periodic system classifying the elements according to their atomic and principal quantum numbers.[26] For most databases, a relevant fraction (>70%) of the variance of MQN-space is covered within the first two or three PCs, implying that maps derived from projections in the PC-planes provide a relevant overview of their chemical space.

The MQN-map of PubChem in form of the (PC2,PC3) plane provides a representative example (Figure 1).[27] In this map, molecules of increasing size are distributed concentrically around the center where the smallest molecules are located, as illustrated by color-coding with the number of non-hydrogen atoms (HAC, heavy atom count, Figure 1A). The horizontal (PC2) axis represents molecular rigidity, with acyclic, flexible molecules at left, and cyclic, rigid molecules at right, as illustrated by color-coding by the fraction of cyclic atoms in the molecules (Figure 1B). The vertical axis (PC3) represents polarity, as illustrated by color-coding by the fraction of hydrogen bond acceptor atoms in the molecule (Figure 1C). Molecules of different classes occupy distinct regions of this map (Figure 1D). For example, acyclic branched alkanes, which were enumerated by Cayley, the inventor of graph theory, as the first attempts to consider chemical space, form a thin stripe extending to the southwest of the map.[6] Peptides, which are also acyclic but more polar, stretch out directly west. The increasingly more cyclic and polar oligosaccharides and oligonucleotides populate the northwest and northeast portion of the map, while polycyclic hydrocarbons such as diamondoids[28] and graphenes[29] stretch out directly at east corresponding to entirely cyclic molecules. Groups of related bioactive compounds often cluster together on MQN maps. For example, a group of the 2445 ligands active on nicotinic acetylcholine receptors (nAChR) reported in ChEMBL are concentrated on the center right portion of the map, which corresponds to cyclic aromatic and heteroaromatic molecules of up to 30 heavy atoms.

The idea behind any representation of chemical space is to be able to use the positional information within this space to search for bioactive molecules, thus performing virtual screening to select compounds for in vitro testing. In that respect, the relevance of any chemical space must be judged by its ability to group compounds with similar bioactivity together.[10,23] This is, for example, the case for the above-mentioned MQN-space, as can be exemplified by the efficient recovery of groups of bioactive compounds such as those in the DUD (Database of Useful Decoys)[30] using MQN-distances as selection criteria.[27] Many chemical spaces constructed from descriptors of chemical structure, including also binary substructure or pharmacophore fingerprint spaces, perform well for virtual screening, whereby a variety of similarity measures can be used as distance measures.[31] However, the main limitation of such similarity searches is that nearest neighbor relationships often indicate compounds that are structurally similar and therefore rather unsurprising from the point of view of analoguing. In that respect, it should be noted that MQN-similarity does not select for substructure similarity and can reveal nontrivial lead-hopping relationships between actives.[27]

2. THE UNKNOWN CHEMICAL SPACE

Virtual screening is mostly used to select compounds from existing collections such as to focus the time and resources dedicated to experimental testing on the most promising molecules. Naturally, the approach can be extended to also save the time and resources dedicated to organic synthesis. This implies to perform virtual screening on virtual rather than actual molecules, hence the idea to explore the yet unknown chemical space. This concept forms the basis for the field of de novo drug design, which attempts to design bioactive compounds in silico prior to their synthesis.[32]

Table 2. The 42 Molecular Quantum Numbers (MQNs)

atom counts (12)	
c	carbon
f	fluorine
cl	chlorine
br	bromine
i	iodine
s	sulfur
p	phosphorus
an	acyclic nitrogen
cn	cyclic nitrogen
ao	acyclic oxygen
co	cyclic oxygen
hac	heavy atom count
polarity counts (6)[a]	
hbam	H-bond acceptor sites
hba	H-bond acceptor atoms
hbdm	H-bond donor sites
hbd	H-bond donor atoms
neg	negative charges
pos	positive charges
bond counts (7)	
asb	acyclic single bonds
adb	acyclic double bonds
atb	acyclic triple bonds
csb	cyclic single bonds
cdb	cyclic double bonds
ctb	cyclic triple bonds
rbc	rotatable bond count
topology counts (17)[b]	
asv	acyclic monovalent nodes
adv	acyclic divalent nodes
atv	acyclic trivalent nodes
aqv	acyclic tetravalent nodes
cdv	cyclic divalent nodes
ctv	cyclic trivalent nodes
cqv	cyclic tetravalent nodes
r3	3-membered rings
r4	4-membered rings
r5	5-membered rings
r6	6-membered rings
r7	7-membered rings
r8	8-membered rings
r9	9-membered rings
rg10	≥10 membered rings
afr[c]	atoms shared by fused rings
bfr[c]	bonds shared by fused rings

[a] Polarity counts consider the ionization state predicted for the physiological pH = 7.4. hbam counts lone pairs on H-bond acceptor atoms, and hbdm counts H-atoms on H-bond donating atoms. [b] All topology counts refer to the smallest set of smallest rings. [c] afr and bfr count atoms repectively bonds shared by at least two rings. afr and bfr enhance the differentiation of polycyclic systems with nonplanar shapes such as norbornanes, as discussed in ref 27a.

The playground for de novo drug design concerns all organic molecules that are of potential interest as drugs, usually the molecules following Lipinski's "rule of five" (Ro5).[33] The Ro5 states that a molecule displays favorable pharmacokinetic properties in terms of absorption and distribution if at least two of the following four criteria are met: (1) MW ≤ 500 Da (not too large), (2) logP ≤ 5 (not too lipophilic), (3) HBA ≤ 10, and (4) HBD ≤ 5 (not too hydrogen-bonding). Whereas the known chemical space including public databases and corporate collections probably contains on the order 100 million molecules, it has been estimated that the Lipinski virtual chemical space might contain as many as 10^{60} compounds when considering only basic structural rules,[10a,34] or a more modest 10^{20}–10^{24} molecules if combination of known fragments are considered.[35]

These size estimates suggest that this entire chemical space is far too large for an exhaustive enumeration, even using today's computers. One is therefore left with a partial, targeted enumeration as the only option to produce molecules for virtual screening. Virtual libraries were first designed for combinatorial chemistry by combining fragments using established synthetic routes.[36] The program BREED, which systematically generates combinations from a list of fragments, is a good current application of this principle.[37] A group at Pfizer have used such a combinatorial strategy to enumerate what seems to be the largest virtual library reported so far. The Pfizer Global Virtual Library (PGVL) lists approximately10^{12} virtual molecules that can be potentially synthesized from validated reaction protocols.[38] Alternatively, one can couple compound enumeration with virtual screening in form of genetic algorithms that perform cycles of molecule generation and fitness selection. This approach restricts enumeration to compounds with the highest probability of a given bioactivity, and forms the bulk of de novo drug design methods to date.[32]

While the targeted enumeration approaches in de novo drug design offer a practical method to find new ligands, they do not address the initial fundamental question of describing the entire chemical space. This question may not be tractable for the far too vast Lipinski space, yet exhaustive enumeration offers an opportunity to characterize the chemical space of very small organic molecules, a question which has recently gained particular relevance in the context of fragment-based drug discovery.[39] Our group has reported the first exhaustive enumeration of chemical space for fragment-sized organic molecules, which produced an impressive number of molecules up to 11 atoms (generated database up to 11 atoms: GDB-11, with C, N, O, F, 26.4 million cmpds with 153 ± 7 Da)[40] and 13 atoms (generated database up to 13 atoms: GDB-13, with C, N, O, Cl and S, 977 million compds 180 ± 8 Da).[41] These databases list molecules as SMILES strings,[42] which represent the structural formula. Conversion to 3-dimensional stereoisomers and conformers can be performed using a stereoisomer generator such as CORINA.[43]

An overview of GDB-13 is provided by MQN-maps, which show that the database spans from acyclic to polycyclic molecules (Figure 2A) with varying numbers of H-bond acceptor atoms (Figure 2B), and mostly consists of heterocyclic and fused heterocyclic compounds (Figure 2C). In the case of GDB-11, we have shown that the vast majority of GDB-molecules larger than 10 atoms are chiral.[40b] Almost all GDB-molecules follow the Ro5[33] as well as lead-likeness[44] and fragment-likeness[45] rules, because these rules restrain molecular size.

A striking feature of the GDB databases is that the number of molecules is very large compared to known molecules and grows exponentially with the number of atoms in the molecules (Figure 2D). Due to the sheer number of molecules in GDB-13, the vast majority of them (>99.9%) has never been synthesized, and this would be even more true for an exhaustive enumeration with more atoms. However, the currently available

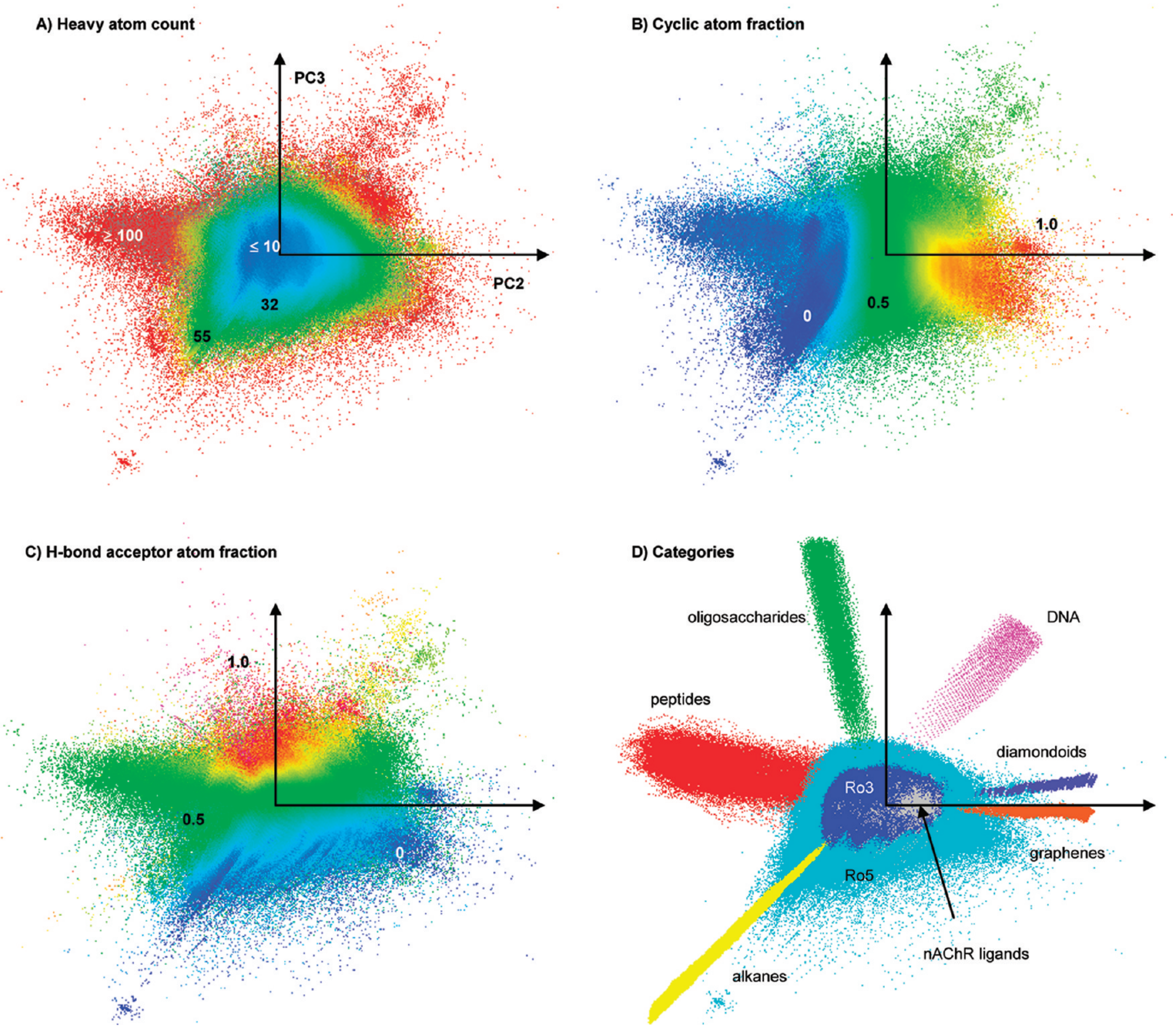


Figure 1. Color-coded MQN-map of the PubChem chemical space (19.2 million structures) as the (PC2,PC3)-plane, marked as the horizontal and vertical axes starting from the (0,0) coordinate where hydrogen is located. The values corresponding to each color are indicated on the maps. PC2, PC3 refer to the 2nd and 3rd principal component, respectively, in the PCA of the MQN data for PubChem. (A) Average number of non-hydrogen atoms per molecule. (B) Average fraction of cyclic atoms per molecule. (C) Average fraction of H-bond acceptor atoms per molecule. (D) Compound categories including computationally enumerated molecules (up to hac = 500) for each category. Ro3 are Congreve's "rule of 3" molecules, and Ro5 are Lipinski's "rule of 5" molecules.

computing power and data storage capacity will probably limit exhaustive enumeration to fragment-sized molecules only (HAC ≤ 20). The enumeration of larger molecules has been approached by Oprea et al. focusing on scaffold topologies.[46] This description does not explicitly enumerate molecules but describes structural types in broad terms, and, for example, shows that only a small subset of the possible scaffold topologies occur in known molecules, suggesting avenues for innovation.

Although the enumeration of GDB considers only functional groups and ring systems that are chemically stable and in principle synthetically accessible,[40,41] many of the GDB molecules appear to be far too challenging for synthesis. To simplify the exploitation of GDB toward the potentially least problematic and synthetically more tractable molecules, we have generated subsets of the database excluding substructures and functional groups that are problematic from the point of view of medicinal and synthetic chemistry.[47] For instance many GDB-molecules contain nonaromatic carbon−carbon double or triple bonds (63% of GDB-13), small rings (3- and 4-membered rings, 54% of GDB-13), nonaromatic N−N− and N−O bonds from oximes and hydrazones (35% of GDB-13), or metabolically unstable groups (e.g., aldehydes, epoxides, aziridines, esters, carbonates, sulfates, 29% of GDB-13). Eliminating molecules featuring any of these substructures leaves a restricted subset of 43.7 million molecules, which is 20-fold smaller than the entire database, yet still exceeds the number of molecules up to 13 atoms in PubChem by 2 orders of magnitude (Figure 2D). A freely accessible MQN-searchable version of GDB-13 is available at www.gdb.unibe.ch in which eight different such restriction criteria can be applied at will, which defines 256 different subsets of various sizes.[47] It should

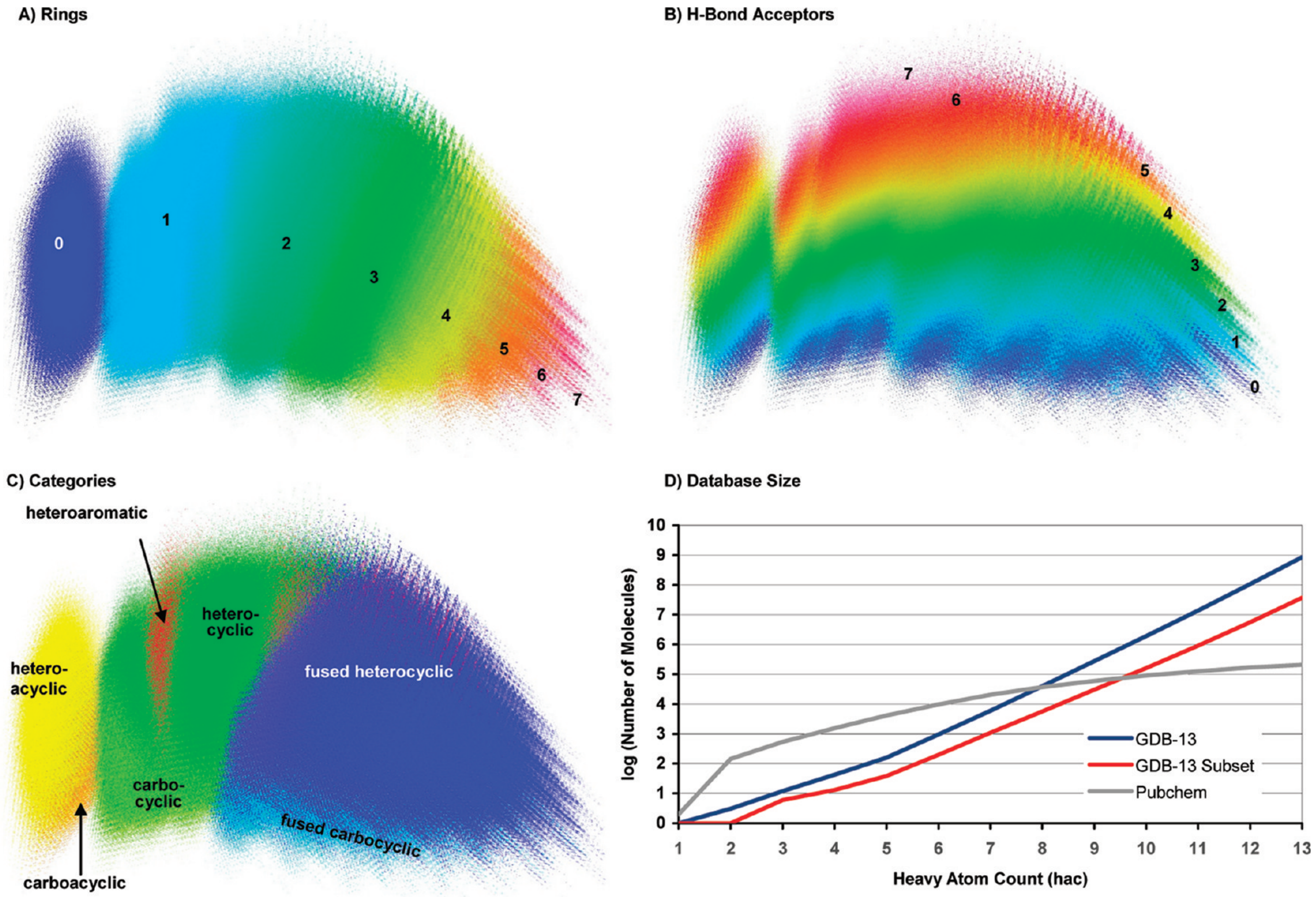
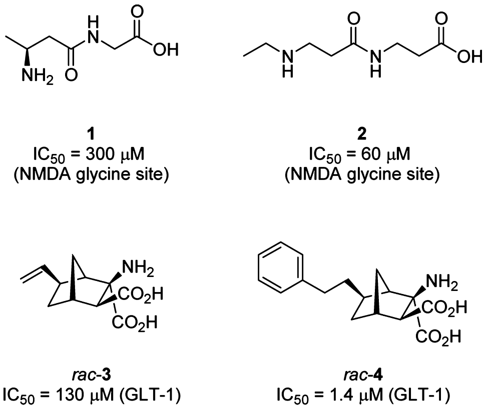


Figure 2. Overview of the chemical universe database GDB-13 containing 977 million structures up to 13 atoms of C, N, O, Cl, S. (A−C) Color-coded MQN-map of the (PC1,PC2) plane. PC1 (horizontal dimension) and PC2 (vertical dimension) refer to the 1st respectively 2nd principal component in the PCA of the MQN data for GDB-13. (D) Size of the GDB database, its 43.7 M subset, and PubChem as a function of molecular size.

be noted that defining these restricted subsets does not solve the synthetic challenge to actually prepare any of the GDB-13 molecules. Their synthesis has to be considered on a case by case basis and in the vast majority of cases represents a nontrivial task even for molecules from the subsets.

3. LIGAND DISCOVERY FROM GDB

To translate our exploration of the virtual chemical space into real molecules, we have focused on drug discovery projects for neurotransmitter receptors and transporters, because these targets often require small molecule ligands such as those enumerated in GDB. Our first proof of concept was based on GDB-11 and dedicated to the glycine site of the NMDA receptor, for which a high-resolution crystal structure with bound glycine was available.[48] A fragment-based Bayesian classifier, which determines a bioactivity probability score for any compound from the product of the relative frequency of occurrence of all its substructures in known active versus inactive compounds,[49] was used to select 15 000 virtual analogues of known NMDA-receptor ligands from GDB-11. These ligands were converted to 70 000 stereoisomers using CORINA[43] and docked using Autodock 3.0.5.[50] The top 1% scoring ligands, which contained several known NMDA-glycine site ligands, were inspected, and 23 compounds were selected for synthesis and testing. An interesting series of dipeptides such as **1** was identified and optimized to yield dipeptide **2** as a micromolar ligand (Figure 3). In a second study, GDB-11 was used to enumerate 250 000 possible analogues of aspartate and glutamate, both of which are substrates of the glutamate transporter GLT-1.[51] A similar docking approach followed by synthesis and testing led to the discovery of a low micromolar



1
IC$_{50}$ = 300 μM
(NMDA glycine site)

2
IC$_{50}$ = 60 μM
(NMDA glycine site)

rac-**3**
IC$_{50}$ = 130 μM (GLT-1)

rac-**4**
IC$_{50}$ = 1.4 μM (GLT-1)

Figure 3. NMDA glycine site and GLT-1 inhibitors identified from GDB-11 by virtual screening, synthesis and testing. Activities were determined by electrophysiology (NMDA glycine site) or by radioactive ligand uptake inhibition (GLT-1) for the human receptors expressed in *Xenopus* oocytes.

inhibitor of this transporter in form of a norbornane-aspartate derivative *rac*-**3** and its optimized analogue *rac*-**4**.

A related strategy was applied to search for nicotinic acetylcholine receptor (nAChR) ligands in GDB in form of analogues of the known α7 nAChR partial agonists PNU-282,987 (**5**)[52] and SSR180711 (**6**) (Figure 4).[53] A limited set of aromatic acyl groups appearing in known active analogues of these ligands[54] was combined with a large diversity of diamines extracted from GDB-11.[55] Connecting all aliphatic diamines containing a tertiary and a primary or secondary amine separated by a two-carbon spacer with five selected acyl groups yielded a total of 1.2 million virtual analogues of **5** and **6**, from which a random selection of 70 000 ligands (6.2% of the library) was subjected to virtual screening by docking using both Autodock 3.0.5[50] and Glide.[56] Docking was performed on the crystal structure of the *Lymnaea signalis* acetylcholine binding protein with bound nicotine (AChBP, PDB ID: 1UW6),[57] a homologue of the human α7 nAChR useful for structure-based drug discovery.[58]

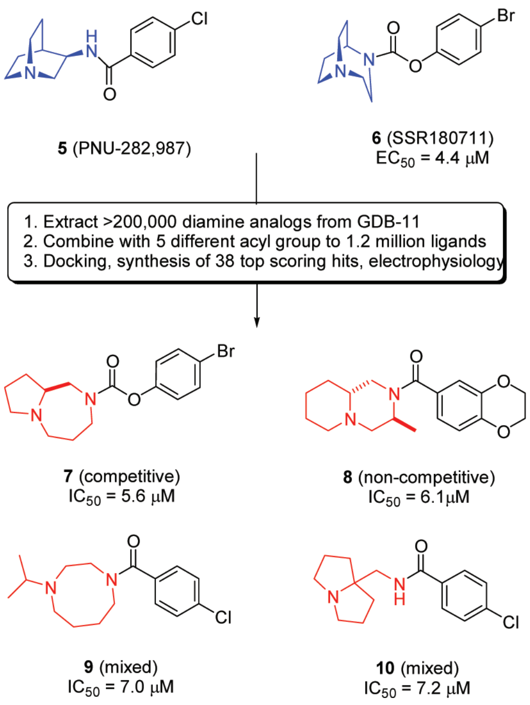


Figure 4. Discovery of α7 nAChR inhibitors by fragment-based diversification of known ligands using GDB-11.

Easily accessible diamines were selected from the 1000 top scoring molecules from each docking method, choosing diamines which were unknown or at least not previously described in the α7 nAChR literature. The diamines were synthesized and acylated with various acyl groups, eventually yielding a total of 38 ligands which were evaluated for modulation of the human α7 nAChR by electrophysiology. Although no agonistic effects were observed, several of the ligands displayed significant inhibition of the receptor. A detailed characterization of four inhibitors showed that at least one of them, compound **7**, acted as a competitive antagonist of acetylcholine, presumably by direct binding to the nicotinic site as suggested by docking. The other three ligands **8**−**10** showed mixed or noncompetitive inhibition, suggesting additional interactions with the receptor, such as direct blockade of the ion channel (Figure 4).

The above studies relied on structure-based drug discovery using docking to select ligands for synthesis and testing. However the method is limited to scoring at most a few million potential ligands, which is clearly too low to tackle very large databases such as GDB. Our next study was dedicated to testing a ligand-based virtual screening approach that would be compatible with GDB-13 and its almost one billion structures. In particular, we were interested to see if the concept of MQN-space discussed above (section 1) could be used for virtual screening in a prospective study. A preliminary study showed that GDB-13 molecules that were nearest neighbors of known drugs in MQN-space, using the city-block distance CBD_{MQN} (the sum of the absolute differences between value pairs across all 42 MQNs) as distance measure, showed strong shape similarity to the drug as measured by the shape-similarity score ROCS (Rapid Overlay of Chemical Structures),[59] suggesting that the MQN-distance measure might select for bioactive analogues of known drugs.[47]

As an application example, we selected to search for new nicotinic ligands by MQN-similarity to nicotine (**11**), a natural product with 12 atoms well within the chemical space of GDB-13 (Figure 5).[60] The fact that known nicotinic ligands up to 13 atoms from ChEMBL were much closer to nicotine in MQN-space (average CBD_{MQN} =22.8 ± 12.5) compared to GDB-13 molecules (average CBD_{MQN} = 38.8 ± 11.1) suggested that the selection procedure should indeed work. Among 31 504 MQN-space nearest neighbors of nicotine selected from the functional group filtered GDB-13 subset of 43.7 million structures discussed above (section 3), 48 were indeed already listed as nicotinic ligands in ChEMBL. Another 692 compounds were listed in ZINC, from which 61 compounds were acquired from commercial sources for experimental evaluation by electrophysiology. While the positive control neonicotine (**12**) gave the known agonistic effect, 11 compounds of the 60 other test compounds (18%) effected 60% inhibition of the ACh evoked current or more. Closer characterization of three of them showed that these acted as micromolar inhibitor with both competitive or noncompetitive mode of action.

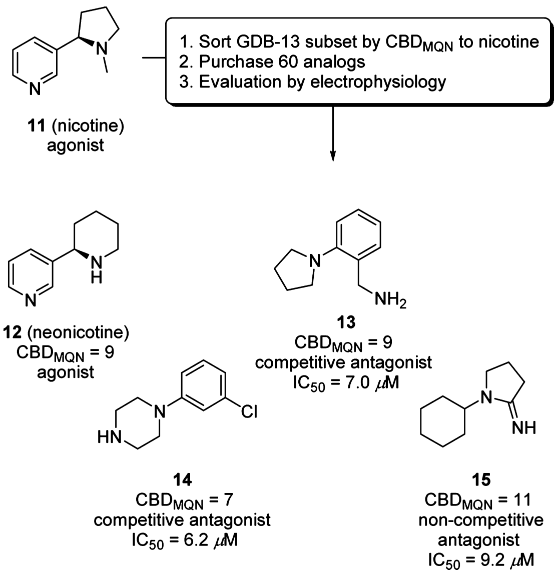


Figure 5. Discovery of α7 nAChR inhibitors by nearest neighbor searching in the MQN-space of GDB-13.

Although **13−15** did not act as agonists like nicotine and neonicotine do, a subsequent evaluation by docking showed that these ligands were essentially indistinguishable from nicotine or neonicotine in terms of docking pose or docking energy. This suggested that a more sophisticated virtual screening procedure based on docking rather than MQN similarity might have selected the very same compounds for testing.

4. CONCLUSION AND OUTLOOK

Small molecule drugs are essential to the success of modern medicine. Considerations on the size of chemical space indicate that the vast majority of possible molecules are still unknown and yet to be synthesized and tested, even at the level of relatively small, fragment sized molecules such as those in the chemical universe databases GDB-11 and GDB-13. The key challenge in exploiting this vast resource lies in the throughput and predictive value of virtual screening. Indeed, while it is not difficult to identify thousands of high-scoring molecules from chemical space, the success rates of virtual screening predictions rarely exceed 1−5% upon in vitro testing. Such success rates are spectacular compared to random screening, but are too low for committing large synthetic resources for preparing the modeled compounds. In the first three examples discussed above, synthetic resources were engaged to follow rather conservative library designs, a strategy which certainly contributed to success independent of the structure-based scoring schemes followed.

In the last example where analogues of nicotine were identified in GDB-13 by proximity in MQN-space by contrast, the virtual screening hits were much more diverse. In this case, we chose the faster "purchase and test" approach to reach a proof-of-concept of the approach. The key aspect of this experiment concerned the speed of the virtual screening method. Indeed searching by MQN-similarity is remarkably fast. A freely accessible web-based application is available from our webpage www.gdb.unibe.ch to search GDB-13 or any of its subsets for MQN-nearest neighbors of any molecule.[61] A typical similarity search such as that used to retrieve MQN-nearest neighbors of nicotine requires only a few seconds of computing time. Engaging significant synthetic resources to prepare selected GDB-13 virtual screening hits will however require additional scoring such as shape matching or docking to prioritize hits.

The chemical space exploration strategies discussed here should be generally applicable to various targets including many CNS targets where small molecule drugs perform best. Experiments along these lines are currently in progress in our laboratory, including the development of improved compound enumeration, classification and virtual screening schemes, and the implementation of chemical synthesis to evaluate the methods by prospective drug discovery.

AUTHOR INFORMATION

Corresponding Author

*Fax: + 41 31 631 80 57. E-mail: jean-louis.reymond@ioc.unibe.ch.

Author Contributions

J.L.R. designed the research and wrote the paper. M.A. prepared tables and figures.

Funding

This work was supported financially by the University of Berne, the Swiss National Science Foundation, and the NCCR TransCure.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) (a) van der Horst, E., Peironcely, J. E., van Westen, G. J., van den Hoven, O. O., Galloway, W. R., Spring, D. R., Wegner, J. K., van Vlijmen, H. W., Ijzerman, A. P., Overington, J. P., and Bender, A. (2011) Chemogenomics approaches for receptor deorphanization and extensions of the chemogenomics concept to phenotypic space. *Curr. Top. Med. Chem. 11*, 1964−1977. (b) Bon, R. S., and Waldmann, H. (2010) Bioactivity-guided navigation of chemical space. *Acc. Chem. Res. 43*, 1103−1114.

(2) (a) Harrison, P. M., Kumar, A., Lang, N., Snyder, M., and Gerstein, M. (2002) A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res. 30*, 1083−1090. (b) Montelione, G. T., and Szyperski, T. (2010) Advances in protein NMR provided by the NIGMS Protein Structure Initiative: impact on drug discovery. *Curr. Opin. Drug Discovery Dev. 13*, 335−349. (c) Bruckner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci. 10*, 2763−2788.

(3) Bleicher, K. H., Bohm, H. J., Muller, K., and Alanine, A. I. (2003) Hit and lead generation: Beyond high-throughput screening. *Nat. Rev. Drug Discovery 2*, 369−378.

(4) (a) Garcia-Serna, R., and Mestres, J. (2010) Anticipating drug side effects by comparative pharmacology. *Expert Opin. Drug Metab. Toxicol. 6*, 1253−1263. (b) Brown, J. B., and Okuno, Y. (2012) Systems biology and systems chemistry: new directions for drug discovery. *Chem. Biol. 19*, 23−28.

(5) (a) Schuffenhauer, A., Brown, N., Selzer, P., Ertl, P., and Jacoby, E. (2005) Relationships between Molecular Complexity, Biological Activity, and Structural Diversity. *J. Chem. Inf. Model. 46*, 525−535. (b) Klebe, G. (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today 11*, 580−594.

(6) Cayley, E. (1875) Ueber die analytischen Figuren, welche in der Mathematik Bäume genannt werden und ihre Anwendung auf die Theorie chemischer Verbindungen. *Chem. Ber. 8*, 1056−1059.

(7) (a) Brinkmann, G., Caporossi, G., and Hansen, P. (2003) A survey and new results on computer enumeration of polyhex and fusene hydrocarbons. *J. Chem. Inf. Comput. Sci. 43*, 842−851. (b) Dias, J. R. (2010) The polyhex/polypent topological paradigm: regularities in the isomer numbers and topological properties of select subclasses of benzenoid hydrocarbons and related systems. *Chem. Soc. Rev. 39*, 1913−1924.

(8) (a) Lederberg, J., Sutherland, G. L., Buchanan, B. G., Feigenbaum, E. A., Robertson, A. V., Duffield, A. M., and Djerassi, C. (1969) Applications of artificial intelligence for chemical inference. I. Number of possible organic compounds. Acyclic structures containing carbon, hydrogen, oxygen, and nitrogen. *J. Am. Chem. Soc. 91*, 2973−2976. (b) Steinbeck, C. (2004) Recent developments in automated structure elucidation of natural products. *Nat. Prod. Rep. 21*, 512−518.

(9) Renner, S., Popov, M., Schuffenhauer, A., Roth, H. J., Breitenstein, W., Marzinzik, A., Lewis, I., Krastel, P., Nigsch, F., Jenkins, J., and Jacoby, E. (2011) Recent trends and observations in the design of high-quality screening collections. *Future Med. Chem. 3*, 751−766.

(10) (a) Bohacek, R. S., McMartin, C., and Guida, W. C. (1996) The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev. 16*, 3−50. (b) Pearlman, R. S., and Smith, K. M. (1998) Novel software tools for chemical diversity. *Perspect. Drug Discovery Des. 9−11*, 339−353. (c) Dobson, C. M. (2004) Chemical space and biology. *Nature 432*, 824−828. (d) Reymond, J.-L., Ruddigkeit, L., Blum, L. C., and van Deursen, R. (2012) The

enumeration of chemical space. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, DOI: 10.1002/wcms.1104.

(11) (a) Bolton, E. E., Wang, Y., Thiessen, P. A., and Bryant, S. H. (2008) Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Rep. Comput. Chem. 4*, 217−241. (b) Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., and Bryant, S. H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res. 37*, W623−W633.

(12) Williams, A. J. (2008) Public chemical compound databases. *Curr. Opin. Drug Discovery Dev. 11*, 393−404.

(13) Irwin, J. J., and Shoichet, B. K. (2005) ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model. 45*, 177−182.

(14) Voigt, J. H., Bienfait, B., Wang, S., and Nicklaus, M. C. (2001) Comparison of the NCI open database with seven large chemical structural databases. *J. Chem. Inf. Comput. Sci. 41*, 702−712.

(15) Chen, J., Swamidass, S. J., Dou, Y., Bruand, J., and Baldi, P. (2005) ChemDB: a public database of small molecules and related chemoinformatics resources. *Bioinformatics 21*, 4133−4139.

(16) (a) Chen, X., Liu, M., and Gilson, M. K. (2001) BindingDB: a web-accessible molecular recognition database. *Comb. Chem. High Throughput Screening 4*, 719−725. (b) Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K. (2007) BindingDB: a web-accessible database of experimentally determined protein−ligand binding affinities. *Nucleic Acids Res. 35*, D198−D201.

(17) Seiler, K. P., George, G. A., Happ, M. P., Bodycombe, N. E., Carrinski, H. A., Norton, S., Brudz, S., Sullivan, J. P., Muhlich, J., Serrano, M., Ferraiolo, P., Tolliday, N. J., Schreiber, S. L., and Clemons, P. A. (2008) ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res. 36*, D351−D359.

(18) (a) Warr, W. A. (2009) ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). *J. Comput.-Aided Mol. Des. 23*, 195−198. (b) Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overington, J. P. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res. 40*, D1100−D1107.

(19) Davis, A. P., King, B. L., Mockus, S., Murphy, C. G., Saraceni-Richards, C., Rosenstein, M., Wiegers, T., and Mattingly, C. J. (2011) The Comparative Toxicogenomics Database: update 2011. *Nucleic Acids Res. 39*, D1067−D1072.

(20) Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J., and Forsythe, I. (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res. 37*, D603−D610.

(21) Frolkis, A., Knox, C., Lim, E., Jewison, T., Law, V., Hau, D. D., Liu, P., Gautam, B., Ly, S., Guo, A. C., Xia, J., Liang, Y., Shrivastava, S., and Wishart, D. S. (2010) SMPDB: The Small Molecule Pathway Database. *Nucleic Acids Res. 38*, D480−D487.

(22) Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eisner, R., Guo, A. C., and Wishart, D. S. (2011) DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs. *Nucleic Acids Res. 39*, D1035−D1041.

(23) (a) Oprea, T. I., and Gottfries, J. (2001) Chemography: The art of navigating in chemical space. *J. Comb. Chem. 3*, 157−166. (b) Medina-Franco, J. L., Maggiora, G. M., Giulianotti, M. A., Pinilla, C., and Houghten, R. A. (2007) A similarity-based data-fusion approach to the visual characterization and comparison of compound databases. *Chem. Biol. Drug Des. 70*, 393−412. (c) Rosen, J., Gottfries, J., Muresan, S., Backlund, A., and Oprea, T. I. (2009) Novel chemical space exploration via natural products. *J. Med. Chem. 52*, 1953−1962. (d) Geppert, H., Vogt, M., and Bajorath, J. (2010) Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J. Chem. Inf. Model. 50*, 205−216.

(24) (a) Bauknecht, H., Zell, A., Bayer, H., Levi, P., Wagener, M., Sadowski, J., and Gasteiger, J. (1996) Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: Dopamine and benzodiazepine agonists. *J. Chem. Inf. Comput. Sci. 36*, 1205−1213. (b) Schmuker, M., and Schneider, G. (2007) Processing and classification of chemical data inspired by insect olfaction. *Proc. Natl. Acad. Sci. U.S.A. 104*, 20285−20289. (c) Schneider, G., Hartenfeller, M., Reutlinger, M., Tanrikulu, Y., Proschak, E., and Schneider, P. (2009) Voyages to the (un)known: adaptive design of bioactive compounds. *Trends Biotechnol. 27*, 18−26.

(25) Nguyen, K. T., Blum, L. C., van Deursen, R., and Reymond, J. L. (2009) Classification of organic molecules by molecular quantum numbers. *ChemMedChem 4*, 1803−1805.

(26) Wang, S. G., and Schwarz, W. H. (2009) Icon of chemistry: the periodic system of chemical elements in the new century. *Angew. Chem., Int. Ed. 48*, 3404−3415.

(27) (a) van Deursen, R., Blum, L. C., and Reymond, J. L. (2010) A searchable map of PubChem. *J. Chem. Inf. Model. 50*, 1924−1934. (b) van Deursen, R., Blum, L. C., and Reymond, J. L. (2011) Visualisation of the chemical space of fragments, lead-like and drug-like molecules in PubChem. *J. Comput.-Aided Mol. Des. 25*, 649−662. (c) Awale, M., and Reymond, J.-L. (2012) Cluster analysis of the DrugBank chemical space using molecular quantum numbers. *Bioorg. Med. Chem.*, DOI: 10.1016/j.bmc.2012.03.017.

(28) (a) Dahl, J. E., Liu, S. G., and Carlson, R. M. (2003) Isolation and structure of higher diamondoids, nanometer-sized diamond molecules. *Science 299*, 96−99. (b) Schwertfeger, H., Fokin, A. A., and Schreiner, P. R. (2008) Diamonds are a chemist's best friend: diamondoid chemistry beyond adamantane. *Angew. Chem., Int. Ed. 47*, 1022−1036.

(29) Allen, M. J., Tung, V. C., and Kaner, R. B. (2009) Honeycomb Carbon: A Review of Graphene. *Chem. Rev. 110*, 132−145.

(30) Huang, N., Shoichet, B. K., and Irwin, J. J. (2006) Benchmarking sets for molecular docking. *J. Med. Chem. 49*, 6789−6801.

(31) (a) Willett, P., Barnard, J. M., and Downs, G. M. (1998) Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci. 38*, 983−996. (b) Khalifa, A. A., Haranczyk, M., and Holliday, J. (2009) Comparison of Nonbinary Similarity Coefficients for Similarity Searching, Clustering and Compound Selection. *J. Chem. Inf. Model. 49*, 1193−1201.

(32) (a) Schneider, G., and Fechner, U. (2005) Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discovery 4*, 649−663. (b) Reymond, J. L., Van Deursen, R., Blum, L. C., and Ruddigkeit, L. (2010) Chemical space as a source for new drugs. *MedChemComm 1*, 30−38.

(33) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev. 23*, 3−25.

(34) Kirkpatrick, P., and Ellis, C. (2004) Chemical space. *Nature 432*, 823.

(35) Ertl, P. (2003) Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J. Chem. Inf. Comput. Sci. 43*, 374−380.

(36) (a) Danziger, D. J., and Dean, P. M. (1989) Automated Site-Directed Drug Design: A General Algorithm for Knowledge Acquisition about Hydrogen-Bonding Regions at Protein Surfaces. *Proc. R. Soc. London, Ser. B 236*, 101−113. (b) Lewell, X. Q., Judd, D. B., Watson, S. P., and Hann, M. M. (1998) RECAP-retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci. 38*, 511−522. (c) Leach, A. R., and Hann, M. M. (2000) The in silico world of virtual libraries. *Drug Discovery Today 5*, 326−336. (d) Patel, H., Bodkin, M. J., Chen, B., and Gillet, V. J. (2009) Knowledge-based approach to de

novo design using reaction vectors. *J. Chem. Inf. Model. 49*, 1163−1184.

(37) Pierce, A. C., Rao, G., and Bemis, G. W. (2004) BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. *J. Med. Chem. 47*, 2768−2775.

(38) Boehm, M., Wu, T.-Y., Claussen, H., and Lemmen, C. (2008) Similarity Searching and Scaffold Hopping in Synthetically Accessible Combinatorial Chemistry Spaces. *J. Med. Chem. 51*, 2468−2480.

(39) (a) Foloppe, N. (2011) The benefits of constructing leads from fragment hits. *Future Med. Chem. 3*, 1111−1115. (b) Leach, A. R., and Hann, M. M. (2011) Molecular complexity and fragment-based drug discovery: ten years on. *Curr. Opin. Chem. Biol. 15*, 489−496.

(40) (a) Fink, T., Bruggesser, H., and Reymond, J. L. (2005) Virtual exploration of the small-molecule chemical universe below 160 Da. *Angew. Chem., Int. Ed. 44*, 1504−1508. (b) Fink, T., and Reymond, J. L. (2007) Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *J. Chem. Inf. Model. 47*, 342−353.

(41) Blum, L. C., and Reymond, J. L. (2009) 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc. 131*, 8732−8733.

(42) Weininger, D. (1988) Smiles, a Chemical Language and Information-System 0.1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci. 28*, 31−36.

(43) Sadowski, J., and Gasteiger, J. (1993) From Atoms and Bonds to 3-Dimensional Atomic Coordinates - Automatic Model Builders. *Chem. Rev. 93*, 2567−2581.

(44) Teague, S. J., Davis, A. M., Leeson, P. D., and Oprea, T. (1999) The Design of Leadlike Combinatorial Libraries. *Angew. Chem., Int. Ed. 38*, 3743−3748.

(45) Congreve, M., Carr, R., Murray, C., and Jhoti, H. (2003) A rule of three for fragment-based lead discovery? *Drug Discovery Today 8*, 876−877.

(46) (a) Pollock, S. N., Coutsias, E. A., Wester, M. J., and Oprea, T. I. (2008) Scaffold topologies. 1. Exhaustive enumeration up to eight rings. *J. Chem. Inf. Model. 48*, 1304−1310. (b) Wester, M. J., Pollock, S. N., Coutsias, E. A., Allu, T. K., Muresan, S., and Oprea, T. I. (2008) Scaffold topologies. 2. Analysis of chemical databases. *J. Chem. Inf. Model. 48*, 1311−1324.

(47) Blum, L. C., van Deursen, R., and Reymond, J. L. (2011) Visualisation and subsets of the chemical universe database GDB-13 for virtual screening. *J. Comput.-Aided Mol. Des. 25*, 637−647.

(48) (a) Nguyen, K. T., Syed, S., Urwyler, S., Bertrand, S., Bertrand, D., and Reymond, J. L. (2008) Discovery of NMDA glycine site inhibitors from the chemical universe database GDB. *ChemMedChem 3*, 1520−1524. (b) Nguyen, K. T., Luethi, E., Syed, S., Urwyler, S., Bertrand, S., Bertrand, D., and Reymond, J. L. (2009) 3-(Aminomethyl)piperazine-2,5-dione as a novel NMDA glycine site inhibitor from the chemical universe database GDB. *Bioorg. Med. Chem. Lett. 19*, 3832−3835.

(49) Bender, A. (2011) Bayesian methods in virtual screening and chemical biology. *Methods Mol. Biol. 672*, 175−196.

(50) Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem. 19*, 1639−1662.

(51) Luethi, E., Nguyen, K. T., Burzle, M., Blum, L. C., Suzuki, Y., Hediger, M., and Reymond, J. L. (2010) Identification of selective norbornane-type aspartate analogue inhibitors of the glutamate transporter 1 (GLT-1) from the chemical universe generated database (GDB). *J. Med. Chem. 53*, 7236−7250.

(52) Bodnar, A. L., Cortes-Burgos, L. A., Cook, K. K., Dinh, D. M., Groppi, V. E., Hajos, M., Higdon, N. R., Hoffmann, W. E., Hurst, R. S., Myers, J. K., Rogers, B. N., Wall, T. M., Wolfe, M. L., and Wong, E. (2005) Discovery and structure-activity relationship of quinuclidine benzamides as agonists of alpha7 nicotinic acetylcholine receptors. *J. Med. Chem. 48*, 905−908.

(53) Biton, B., Bergis, O. E., Galli, F., Nedelec, A., Lochead, A. W., Jegham, S., Godet, D., Lanneau, C., Santamaria, R., Chesney, F., Leonardon, J., Granger, P., Debono, M. W., Bohme, G. A., Sgard, F., Besnard, F., Graham, D., Coste, A., Oblin, A., Curet, O., Vige, X., Voltz, C., Rouquier, L., Souilhac, J., Santucci, V., Gueudet, C., Francon, D., Steinberg, R., Griebel, G., Oury-Donat, F., George, P., Avenet, P., and Scatton, B. (2007) SSR180711, a novel selective alpha7 nicotinic receptor partial agonist: (1) binding and functional profile. *Neuropsychopharmacology 32*, 1−16.

(54) Walker, D. P., Wishka, D. G., Piotrowski, D. W., Jia, S., Reitz, S. C., Yates, K. M., Myers, J. K., Vetman, T. N., Margolis, B. J., Jacobsen, E. J., Acker, B. A., Groppi, V. E., Wolfe, M. L., Thornburgh, B. A., Tinholt, P. M., Cortes-Burgos, L. A., Walters, R. R., Hester, M. R., Seest, E. P., Dolak, L. A., Han, F., Olson, B. A., Fitzgerald, L., Staton, B. A., Raub, T. J., Hajos, M., Hoffmann, W. E., Li, K. S., Higdon, N. R., Wall, T. M., Hurst, R. S., Wong, E. H., and Rogers, B. N. (2006) Design, synthesis, structure-activity relationship, and in vivo activity of azabicyclic aryl amides as alpha7 nicotinic acetylcholine receptor agonists. *Bioorg. Med. Chem. 14*, 8219−8248.

(55) Garcia-Delgado, N., Bertrand, S., Nguyen, K. T., van Deursen, R., Bertrand, D., and Reymond, J.-L. (2010) Exploring a7-Nicotinic Receptor Ligand Diversity by Scaffold Enumeration from the Chemical Universe Database GDB. *ACS Med. Chem. Lett. 1*, 422−426.

(56) Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem. 47*, 1739−1749.

(57) Celie, P. H., van Rossum-Fikkert, S. E., van Dijk, W. J., Brejc, K., Smit, A. B., and Sixma, T. K. (2004) Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron 41*, 907−914.

(58) (a) Ulens, C., Akdemir, A., Jongejan, A., van Elk, R., Bertrand, S., Perrakis, A., Leurs, R., Smit, A. B., Sixma, T. K., Bertrand, D., and de Esch, I. J. (2009) Use of acetylcholine binding protein in the search for novel alpha7 nicotinic receptor ligands. In silico docking, pharmacological screening, and X-ray analysis. *J. Med. Chem. 52*, 2372−2383. (b) Reymond, J. L., van Deursen, R., and Bertrand, D. (2011) What we have learned from crystal structures of proteins to receptor function. *Biochem. Pharmacol. 82*, 1521−1527.

(59) Hawkins, P. C., Skillman, A. G., and Nicholls, A. (2007) Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem. 50*, 74−82.

(60) Blum, L. C., van Deursen, R., Bertrand, S., Mayer, M., Burgi, J. J., Bertrand, D., and Reymond, J. L. (2011) Discovery of α7-Nicotinic Receptor Ligands by Virtual Screening of the Chemical Universe Database GDB-13. *J. Chem. Inf. Model. 51*, 3105−3113.

(61) Reymond, J. L., Blum, L. C., and Van Deursen, R. (2011) Exploring the Chemical Space of Known and Unknown Organic Small Molecules at www.gdb.unibe.ch. *Chimia 65*, 863−867.