doi:10.2533/chimia.2011.863 Chimia 65 (2011) 863–867

Exploring the Chemical Space of Known and Unknown Organic Small Molecules at *www.gdb.unibe.ch*

Jean-Louis Reymond*[a], Lorenz C. Blum[a], and Ruud van Deursen[b]

Abstract: Organic small molecules are of particular interest for medicinal chemistry since they comprise many biologically active compounds which are potential drugs. To understand this vast chemical space, we are enumerating all possible organic molecules to create the chemical universe database GDB, which currently comprises 977 million molecules up to 13 atoms of C, N, O, Cl and S. Furthermore, we have established a simple classification method for organic molecules in form of the MQN (molecular quantum numbers) system, which is an equivalent of the periodic system of the elements. Despite its simplicity the 42 dimensional MQN system is surprisingly relevant with respect to bioactivity, as evidenced by the fact that groups of biosimilar compounds form close groups in MQN space. The MQN space of the known organic molecules in PubChem and of the unknown molecules in the Chemical Universe Database GDB-13 can be searched interactively using browser tools freely accessible at *www.gdb.unibe.ch*.

Keywords: Cheminformatics · Computer-aided drug design · Chemical space · Databases · Virtual screening

Introduction

Organic molecules with up to approximately 40 non-hydrogen atoms, the so-called small molecules, are of particular interest for medicinal chemistry since they comprise many biologically active compounds which are potential drugs.[1,2] In the history of chemistry the discovery of bioactive molecules originally relied on the isolation of the active principles from biological extracts. The molecular collections were later expanded by synthetic organic chemistry, which developed methods to prepare the natural products themselves as well as their analogs and entirely novel compounds. The pace of organic synthesis accelerated significantly in the 1990s with the invention of combinatorial and automated parallel synthesis.[3] Currently it is estimated that over 100 million small organic molecules may have been prepared, many of which reside in corporate screening collections. Nevertheless, in the last few years several large databases of small organic molecules have become freely accessible online.[4] For instance PubChem, a publicly available database that archives the molecular structures and bioassay data within the National Institute of Health (NIH) Roadmap for Medical Research Initiative, currently lists over 30 million compounds.[5] Furthermore, many custom synthesis companies offer large collections of drug-like small molecules for purchase and screening. These rapidly increasing collections are conveniently collected in the public access database ZINC, which currently lists approximately 13 million compounds.[6] Additionally, a range of somewhat smaller, more specialized databases are available such as the NCI open database,[7] ChemDB,[8] BindingDB,[9] ChemBank,[10] Chembl,[11] HMBD,[12] SMPBD,[13] and DrugBank (Table 1).[14]

This plethora of molecules has created a major problem: how should one analyze such immensely large numbers of compounds in a logical manner and gain a reasonable overview of the chemical diversity at hand?[15,16] Such an overview can be an enormous advantage to perform *de novo* drug design.[17–19] In this article we present our approach to the problem. First, we have developed a systematic molecule enumeration method which has allowed us to create lists of all possible organic molecules up to a certain size in the form of the Chemical Universe Databases GDB-11 (up to 11 atoms of C, N, O, F)[20,21] and GDB-13 (up to 13 atoms of C, N, O, Cl, S).[22,23] These databases provide a vast resource to inspire compound design and synthesis. For example we have used these databases to discover several new ligands for neurotransmitter receptors[24–26] and transporters.[27] Second, we have established a simple classification method in form of the MQN (molecular quantum numbers) system, which is the equivalent of the periodic system of the elements for organic molecules.[28] Despite its simplicity the MQN system is surprisingly relevant to bioactivity, as evidenced by the fact that groups of biosimilar compounds form close groups in MQN space. The MQN space of the known organic molecules in PubChem and of the unknown molecules in the Chemical Universe Database GDB-13 can be searched interactively using browser tools freely accessible at *www.gdb.unibe.ch*.

The Chemical Universe Database

Organic molecules consist of covalently bonded atoms of carbon, hydrogen, oxygen, nitrogen, sulfur, halogens, and a few other elements. The number of molecules allowed by valency rules and consisting of stable geometries and functional groups must be a finite number for a given number of atoms. The generated databases GDB enumerate these combinations systematically. The enumeration starts with a unique collection of mathematical graphs[29] which are converted to saturated hydrocarbons

*Correspondence: Prof. J.-L. Reymond[a]
E-mail: jean-louis.reymond@ioc.unibe.ch
[a]Department of Chemistry and Biochemistry
NCCR TransCure and NCCR Chemical Biology
University of Bern
Freiestrasse 3, CH-3012 Bern
[b]Biomolecular Screening Facility
NCCR Chemical Biology
Global Health Institute
School of Life Sciences
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne

Table 1. Open access databases of organic molecules.

Database	Description	Size	webaddress
PubChem	Known molecules with experimental data	30.3 M	*http://pubchem.ncbi.nlm.nih.gov*
ZINC	Commercially available small molecules	13.0 M	*http://zinc.docking.org*
NCI Open Database	Anticancer and AIDS compounds with screening data	0.25 M	*http://cactus.nci.nih.gov/ncidb2.1*
ChemDB	Commercially available small molecules	4.1 M	*http://cdb.ics.uci.edu*
BindingDB	Bioactive small molecules with measured binding affinities to proteins	0.32 M	*http://www.bindingdb.org*
ChemBank	Annotated database of small molecules and screening data	1.2 M	*http://chembank.broadinstitute.org*
Chembl	Bioactive drug-like small molecules annotated with experimental data	1.1 M	*https://www.ebi.ac.uk/chembldb*
HMDB	Human metabolome database	0.0079 M	*http://www.hmdb.ca*
SMPDB	Small molecule pathway database	0.001 M	*http://www.smpdb.ca*
DrugBank	Experimental and approved small molecule drugs	0.0068 M	*http://www.drugbank.ca*
GDB-11	All possible small molecules up to 11 atoms of C, N, O, F	26.4 M	*http://www.gdb.unibe.ch*
GDB-13	All possible small molecules up to 13 atoms of C, N, O, S, Cl	980 M	*http://www.gdb.unibe.ch*

by considering nodes as carbon atoms and edges as carbon–carbon bonds. Hydrogen atoms are not considered in the enumeration and are added to complement valency. Hydrocarbons with acceptable geometries are selected and expanded to skeletons (unsaturated hydrocarbons) by converting single bonds to double and triple bonds. Acceptable skeletons are again selected by applying valency rules and geometrical strain criteria, and molecules are enumerated by substituting oxygen, nitrogen, sulfur and halogen atoms for carbon atoms systematically following valency rules. A set of functional group selection rules is finally applied to retain chemically stable molecules only.

We have reported these enumerations up to 11 atoms of C, N, O, F and up to 13 atoms of C, N, O, Cl, S following similar but slightly different selection rules, producing the databases GDB-11 with 26.4 million structures respectively GDB-13 with 977 million structures.[20–22] These databases and selected subsets are distributed as text files enumerating the SMILES[30] through our website *www.gdb.unibe.ch*, and have been downloaded by over 500 different users worldwide. The enumeration of GDB-17 has been completed (over 150 billion structures) and will be reported in the near future. For larger sizes of molecules, we have reported the 'chemical space travel' algorithm which automatically generates intermediate structures between a starting molecule A and a target molecule B by traveling from A to B *via* a sequence of structural mutations.[31] This approach is by far not exhaustive but offers a practical approach for exploring the chemical space at molecular sizes up to 50 atoms which are not accessible by exhaustive enumeration.

The value of GDB as a resource for inspiring new molecules can be illustrated by analyzing polycyclic hydrocarbons. For instance in the context of GDB-11 we enumerated 124 polycyclic hydrocarbons up to 11 carbons atoms without 3- or 4-membered rings, or acyclic bonds. Three of these were unknown either as the compounds themselves or as substructures including heteroatom and unsaturated analogs. These comprised in particular the chiral, *C*2-symmetrical, norbornane analog **1**, which contains three different norbornanes in its structure (Fig. 1). This compound is not strained, however it has been never been prepared in over 100 years of norbornane chemistry. The same applies for the spiro-tricyclic hydrocarbon **2**, which is also chiral and contains two symmetry-related bicyclo[3.2.1.]octanes. On the other hand the related chiral, *C*3-symmetrical, tris-homocubane has been reported as the trioxa analog **3**, which contains three symmetry-related 7-oxanorbornanes in its structure. Compound **3** was prepared by de Meijere *et al.* in a high-yielding acid-catalyzed cyclization step from a tri-epoxide precursor.[32]

GDB provides a useful resource for drug discovery. In our first application of GDB, we used a Bayesian classifier to select 15,000 analogs of drugs known to interact with the NMDA receptor glycine site in GDB-11. The 2D-structures were then converted to 70,000 stereoisomeric 3D-structures using CORINA[33] and screened by high-throughput docking to the glycine binding site of the NMDA receptor. A small set of 23 compounds with particularly high docking scores and suitable predicted binding poses, which were mostly amino acids, was eventually prepared and tested by radioligand displacement assay and electrophysiology. Screening revealed an interesting series of dipeptides as a new class of NMDA glycine site inhibitors. In particular the N-ethyl β-alanine dipeptide **4** and the diketopiperazine **5** were identified after ligand optimization.[24,25]

GDB was also used as a resource to enumerate polycyclic analogs of aspartate

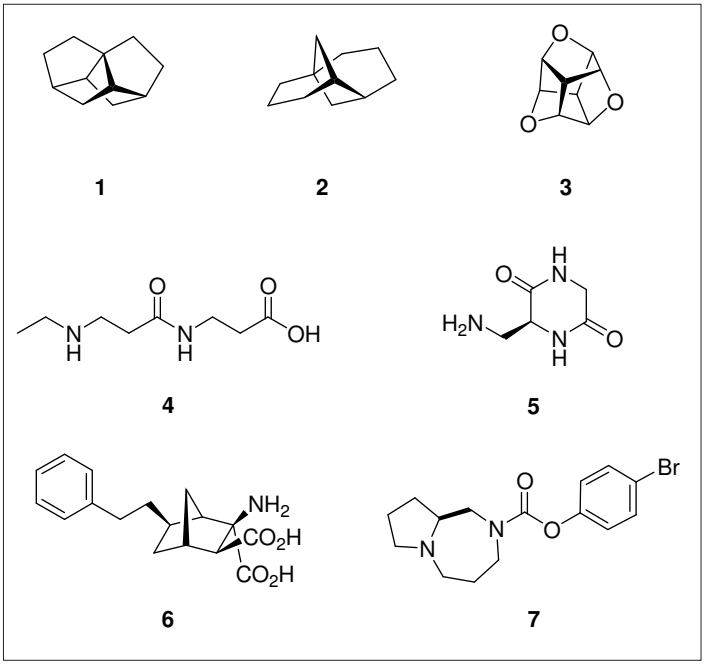


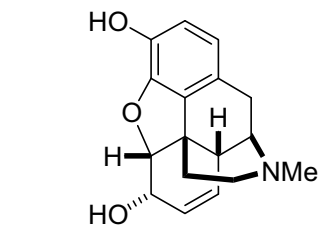
Fig. 1. Examples of molecules derived from GDB.

and glutamate in a substructure approach, which through the application of high-throughput docking and ligand optimization led to the discovery of a new class of norbornane-based aspartate analogs such as **6** as inhibitors of the glutamate transporter GLT-1.[27] In a related strategy, GDB was used to enumerate diamine scaffold analogs of the nicotinic acetyl choline receptor (α7 nAChR) ligands PNU-282,987[34] and SSR180711.[35] High-throughput docking to the nicotine binding site of the acetyl choline binding protein (1uw6.pdb) was used to select the most promising diamine scaffolds. Thirty-eight new ligands were prepared and tested for modulation of the acetylcholine response of recombinant human α7 nAChR by electrophysiology in *Xenopus* oocytes. This allowed the identification of several new ligands for the receptor, such as carbamate **7** which acts as a micromolar competitive inhibitor.[26]

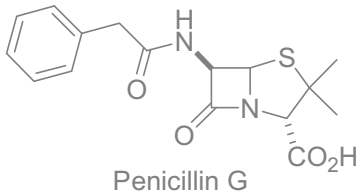
The MQN System of Organic Molecules

In the periodic system, elements are classified according to their atomic and principal quantum numbers, which creates a two-dimensional map grouping isotopes of the same element in the same position.[36] This map provides a useful and chemically meaningful overview of the elements and was instrumental in integrating the chemical knowledge acquired during the 19th century, and in guiding the discovery of the still missing elements. In this section we discuss the MQN system, which provides a related classification system for organic molecules based on molecular quantum numbers (MQNs).[28] MQNs are defined as a set of 42 integer value descriptors of molecular structure counting atom types, bond types, polar groups and topological features.

The classification of organic molecules historically focused on systematic naming and assigning reference indexes, which was essential for archiving experimental data. However the approach did not provide a useful overview of known and unknown molecules in terms of their diversity and structural, physico-chemical and biological properties. Through the advent of cheminformatics the concept of chemical space has provided a new paradigm to classify and understand molecular diversity.[1,37,38] Chemical spaces, which are typically multidimensional, can be constructed by assigning dimensions to descriptors or combinations of descriptors. Descriptors are values calculated from a structural formula which can be as simple as the molecular weight (MW) or the number of non-hydrogen atoms (hac: 'heavy atom count'), or as complex as the calculated water-octanol partition coefficient (clogP) or the topological polar surface area (TPSA).



Morphine



Penicillin G

Bonds		
Acyclic single bonds	**3**	8
Acyclic double bonds	**0**	3
Acyclic triple bonds	**0**	0
Cyclic single bonds	**18**	11
Cyclic double bonds	**4**	3
Cyclic triple bonds	**0**	0
Rotatable bonds	**0**	4
Polar groups		
H-Bond donor atoms	**3**	1
H-Bond donor sites	**3**	1
H-Bond acceptor atoms	**3**	4
H-Bond acceptor sites	**3**	7
Positive charges	**1**	0
Negative charges	**0**	1

Atoms		
Carbon	**17**	16
Fluorine	**0**	0
Chlorine	**0**	0
Bromine	**0**	0
Iodine	**0**	0
Sulphur	**0**	1
Phosphor	**0**	0
Acyclic nitrogen	**0**	1
Cyclic nitrogen	**1**	1
Acyclic oxygen	**2**	4
Cyclic oxygen	**1**	0
Heavy atom count	**21**	23
Topology		
Acyclic monovalent nodes	**3**	6
Acyclic divalent nodes	**0**	2
Acyclic trivalent nodes	**0**	2
Acyclic tetravalent nodes	**0**	0
Cyclic divalent nodes	**8**	6
Cyclic trivalent nodes	**9**	6
Cyclic tetravalent nodes	**1**	1
3-Membered rings	**0**	0
4-Membered rings	**0**	1
5-Membered rings	**1**	1
6-Membered rings	**4**	1
7-Membered rings	**0**	0
8-Membered rings	**0**	0
9-Membered rings	**0**	0
≥ 10 membered rings	**0**	0
Atoms shared by fused rings	**7**	2
Bonds shared by fused rings	**6**	1

Fig. 2. Molecular Quantum Numbers (MQN) and their values for morphine and penicillin G.

In a property space any compound is located at the coordinates corresponding to the values of the selected descriptors for this compound. The MQN-system consists of a 42-dimensional property space in which each of the 42 dimensions corresponds to a different descriptor of the molecular structure. We have selected 42 MQNs as descriptors such that they describe simple but relevant features that can be counted in the structural formula by anyone with basic training in organic chemistry (Fig. 2). Because the descriptors have only integer values, MQN space is not continuous and consists of individual MQN bins corresponding to different combinations of MQN values. Molecules occupying the same MQN bins are MQN isomers, which may be compared to isotopes in the periodic system of the elements.

For most databases principal component analysis (PCA) of the MQN data shows that 60–80% of the variability is covered in the first three principal components (PC). One can therefore gain a graphical overview of these databases in form of two-dimensional projections of the MQN space, which form MQN maps. Fig. 3 shows such an MQN map for the case of PubChem combined with a few well-known molecule categories.[39] PC2 (horizontal axis) represents molecular rigidity and PC3 (vertical axis) represents polarity, such that the most polar molecules appear far north. PC1 is perpendicular to the plane and codes for molecular size, however compounds with increasing size also extend radially from the center in the (PC2,PC3)-projection.

Browsing MQN Space

The similarity of compounds in a given property space can be quantified by a distance measure in this space. Through a careful selection of descriptors one can obtain property spaces where compounds sharing similar physicochemical properties are close to one another. In our MQN space compounds with similar physicochemical, and most interestingly with similar biological activities, are found in close proximity.[39,40] This can be demonstrated by searching for compounds sharing the same bioactivity as a selected reference

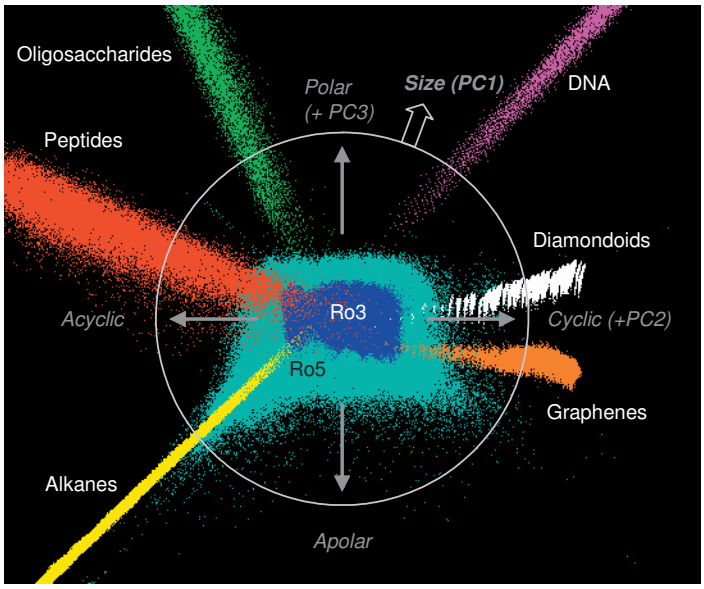


Fig. 3. Two-dimensional projection of the 42-dimensional MQN-space of PubChem by principal component analysis. The (PC2,PC3) plane is shown. PC1 corresponds to molecular size and extends radially from the center. Ro3 (blue area) are small organic molecules in PubChem following Congreve's 'rule of three'. Ro5 (cyan area) are small organic molecules in PubChem following Lipinski's 'rule of five'. The other categories are virtual collections of the corresponding compounds.

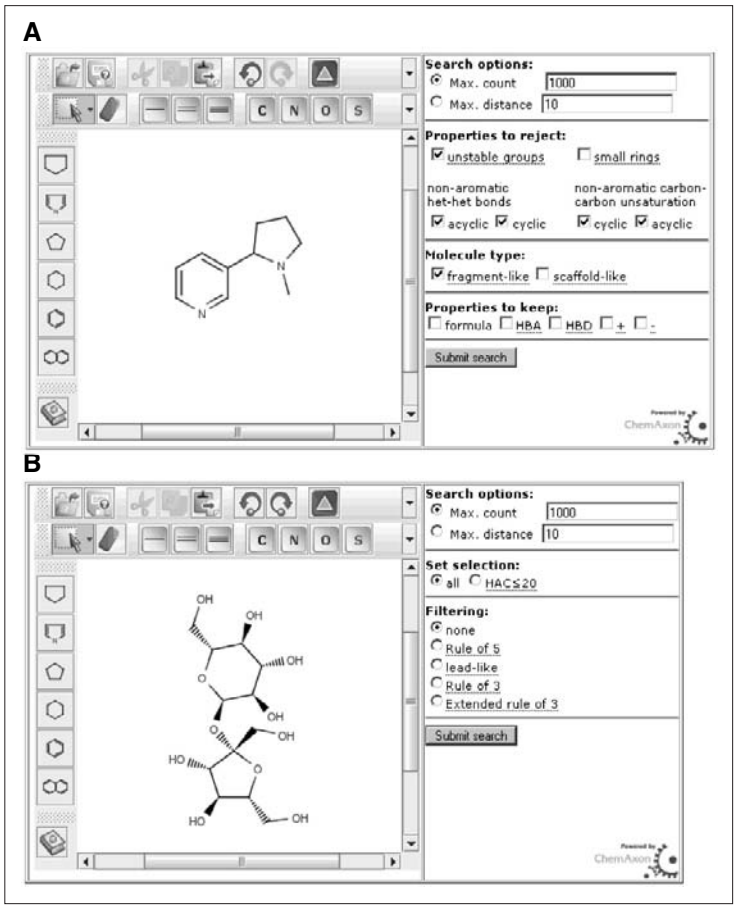


Fig. 4. Webbrowser windows accessible at *www.gdb.unibe.ch* to search for nearest CBD_{MQN} neighbors. **A.** MQN-browser for GDB-13, allows searching of 977 million structures in GDB-13, shown with nicotine as example query. **B.** MQN-browser for PubChem, allowing search of 20 million organic molecules in PubChem, shown with sucrose as example query. Note that MQNs do not encode stereochemistry.

compound using a distance measure to that reference compound in MQN-space as sorting function.[41,42] Sorting by MQN similarity performs similarly well to sorting based on substructure fingerprints[43] for many groups of known bioactives such as the DUD dataset.[44] Interestingly MQN similarity searches group together compounds with similar bioactivities but entirely different substructures, thus providing a useful tool for identifying lead-hopping relationships.[45]

The databases PubChem and GDB-13 can be explored in MQN-space at *www.gdb.unibe.ch* using interactive browsers. By opening the browsers, a window appears for drawing the structural formula of a query molecule for which MQN-analogs are desired (Fig. 4). The browser returns the most MQN-similar compounds by using the city-block distance in MQN-space (CBD_{MQN}) as distance measure. CBD_{MQN} is simply the sum of the absolute differences between MQN-values. The searches are usually complete in less than 30 seconds and provide a list of molecules from the database with the indication of the CBD_{MQN}-distance from the query molecule at which they were found. The list of molecules can be stored as a text file containing the SMILES. Search options are available to search only certain subsets of the databases. In the case of PubChem, the search can be restricted 'rule of 5',[46] 'lead-likeness'[47] and 'rule of 3'[48] criteria.[40] For GDB-13, options are available to exclude certain structural features which tend to be frequent in the enumeration but are not very attractive for synthesis or for medicinal chemistry. These include metabolically unstable functional groups such as aldehydes, epoxides, or esters, as well as structures such as small rings, non-aromatic C–C double bonds or non-aromatic N–N or N–O bonds from oximes and hydrazones.[23] One can also restrict the search to fragment-like and scaffold-like (no acyclic carbon atoms) structures. These restrictions help focus MQN-similarity searches to the most relevant subsets in the perspective of a possible synthesis. We expect that the browsers will be broadly useful for presorting these large databases in view of selecting compounds for advanced virtual screening (*e.g.* docking) and for synthesis.

Conclusion and Outlook

The immensity of chemical space is a challenging problem for chemistry. Describing and visualizing the known and unknown chemical space is an enabling technology to understand molecular diversity and design new molecules with desired properties. Our approach follows systematic enumeration of the GDB databases and the classification of organic molecules with the MQN system. Current research in our group also concerns the exploitation of molecular diversity to design biologically active compounds, in the framework of collaborative projects such as the NCCR Chemical Biology and the NCCR TransCure.

Acknowledgement

This work was supported financially by the University of Berne and the Swiss National

Science Foundation. ChemAxon is gratefully acknowledged for their support and the free-web license of JChem.

Received: August 26, 2011

[1] C. M. Dobson, *Nature* **2004**, *432*, 824.
[2] K. H. Bleicher, H. J. Bohm, K. Muller, A. I. Alanine, *Nat. Rev. Drug Discov.* **2003**, *2*, 369.
[3] P. J. Edwards, *Curr. Opin. Drug Discov. Devel*. **2009**, *12*, 899.
[4] A. J. Williams, *Curr. Opin. Drug Discov. Devel*. **2008**, *11*, 393.
[5] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, S. H. Bryant, *Nucleic Acids Res.* **2009**, *37*, W623.
[6] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177.
[7] J. H. Voigt, B. Bienfait, S. Wang, M. C. Nicklaus, *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 702.
[8] J. Chen, S. J. Swamidass, Y. Dou, J. Bruand, P. Baldi, *Bioinformatics* **2005**, *21*, 4133.
[9] X. Chen, M. Liu, M. K. Gilson, *Comb. Chem. High Throughput Screen*. **2001**, *4*, 719.
[10] K. P. Seiler, G. A. George, M. P. Happ, N. E. Bodycombe, H. A. Carrinski, S. Norton, S. Brudz, J. P. Sullivan, J. Muhlich, M. Serrano, P. Ferraiolo, N. J. Tolliday, S. L. Schreiber, P. A. Clemons, *Nucleic Acids Res*. **2008**, *36*, D351.
[11] J. Overington, *J. Comput.-Aided Mol. Des.* **2009**, *23*, 195.
[12] D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H. J. Vogel, I. Forsythe, *Nucleic Acids Res.* **2009**, *37*, D603.
[13] A. Frolkis, C. Knox, E. Lim, T. Jewison, V. Law, D. D. Hau, P. Liu, B. Gautam, S. Ly, A. C. Guo, J. Xia, Y. Liang, S. Shrivastava, D. S. Wishart, *Nucleic Acids Res.* **2010**, *38*, D480.
[14] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, D. S. Wishart, *Nucleic Acids Res.* **2011**, *39*, D1035.
[15] J. L. Reymond, R. Van Deursen, L. C. Blum, L. Ruddigkeit, *Med. Chem. Commun.* **2010**, *1*, 30.
[16] L. B. Akella, D. DeCaprio, *Curr. Opin. Chem. Biol.* **2010**, *14*, 325.
[17] G. Schneider, U. Fechner, *Nat. Rev. Drug Discov.* **2005**, *4*, 649.
[18] G. Klebe, *Drug Discov. Today* **2006**, *11*, 580.
[19] P. Kolb, R. S. Ferreira, J. J. Irwin, B. K. Shoichet, *Curr. Opin. Biotechnol.* **2009**, *20*, 429.
[20] T. Fink, H. Bruggesser, J. L. Reymond, *Angew. Chem. Int. Ed.* **2005**, *44*, 1504.
[21] T. Fink, J. L. Reymond, *J. Chem. Inf. Model.* **2007**, *47*, 342.
[22] L. C. Blum, J. L. Reymond, *J. Am. Chem. Soc.* **2009**, *131*, 8732.
[23] L. C. Blum, R. van Deursen, J. L. Reymond, *J. Comput. Aided Mol. Des.* **2011**, *25*, 637.
[24] K. T. Nguyen, S. Syed, S. Urwyler, S. Bertrand, D. Bertrand, J. L. Reymond, *ChemMedChem* **2008**, *3*, 1520.
[25] K. T. Nguyen, E. Luethi, S. Syed, S. Urwyler, S. Bertrand, D. Bertrand, J. L. Reymond, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3832.
[26] N. Garcia-Delgado, S. Bertrand, K. T. Nguyen, R. van Deursen, D. Bertrand, J.-L. Reymond, *ACS Med. Chem. Lett.* **2010**, *1*, 422.
[27] E. Luethi, K. T. Nguyen, M. Burzle, L. C. Blum, Y. Suzuki, M. Hediger, J. L. Reymond, *J. Med. Chem*. **2010**, *53*, 7236.
[28] K. T. Nguyen, L. C. Blum, R. van Deursen, J. L. Reymond, *ChemMedChem* **2009**, *4*, 1803.
[29] B. D. McKay, *Congressus Numerantium* **1981**, *30*, 45.
[30] D. Weininger, *J. Chem. Inf. Comp. Sci*. **1988**, *28*, 31.
[31] R. van Deursen, J. L. Reymond, *ChemMedChem* **2007**, *2*, 636.
[32] C. Weitemeyer, A. De Meijere, *Angew. Chem. Int. Ed.* **1976**, *15*, 686.
[33] J. Sadowski, J. Gasteiger, *Chem. Rev*. **1993**, *93*, 2567.
[34] A. L. Bodnar, L. A. Cortes-Burgos, K. K. Cook, D. M. Dinh, V. E. Groppi, M. Hajos, N. R. Higdon, W. E. Hoffmann, R. S. Hurst, J. K. Myers, B. N. Rogers, T. M. Wall, M. L. Wolfe, E. Wong, *J. Med. Chem.* **2005**, *48*, 905.
[35] B. Biton, O. E. Bergis, F. Galli, A. Nedelec, A. W. Lochead, S. Jegham, D. Godet, C. Lanneau, R. Santamaria, F. Chesney, J. Leonardon, P. Granger, M. W. Debono, G. A. Bohme, F. Sgard, F. Besnard, D. Graham, A. Coste, A. Oblin, O. Curet, X. Vige, C. Voltz, L. Rouquier, J. Souilhac, V. Santucci, C. Gueudet, D. Francon, R. Steinberg, G. Griebel, F. Oury-Donat, P. George, P. Avenet, B. Scatton, *Neuropsychopharmacology* **2007**, *32*, 1.
[36] S. G. Wang, W. H. Schwarz, *Angew. Chem. Int. Ed*. **2009**, *48*, 3404.
[37] R. S. Bohacek, C. McMartin, W. C. Guida, *Med. Res. Rev*. **1996**, *16*, 3.
[38] R. S. Pearlman, K. M. Smith, *Perspec. Drug Discov. Des.* **1998**, *9*, 339.
[39] R. van Deursen, L. C. Blum, J. L. Reymond, *J. Chem. Inf. Model.* **2010**, *50*, 1924.
[40] R. van Deursen, L. C. Blum, J. L. Reymond, *J. Comput. Aided Mol. Des.* **2011**, *25*, 649.
[41] P. Willett, J. M. Barnard, G. M. Downs, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983.
[42] A. A. Khalifa, M. Haranczyk, J. Holliday, *J. Chem. Inf. Model*. **2009**, *49*, 1193.
[43] P. Willett, *Drug Discov. Today* **2006**, *11*, 1046.
[44] N. Huang, B. K. Shoichet, J. J. Irwin, *J. Med. Chem*. **2006**, *49*, 6789.
[45] G. Schneider, W. Neidhart, T. Giller, G. Schmid, *Angew. Chem. Int. Ed*. **1999**, *38*, 2894.
[46] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Del. Rev.* **1997**, *23*, 3.
[47] S. J. Teague, A. M. Davis, P. D. Leeson and T. Oprea, *Angew. Chem. Int. Ed*. **1999**, *38*, 3743.
[48] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discov. Today* **2003**, *8*, 876.