doi:10.2533/chimia.2020.241 Chimia 74 (2020) 241–246 © J-L. Reymond et al.

The Generated Databases (GDBs) as a Source of 3D-shaped Building Blocks for Use in Medicinal Chemistry and Drug Discovery

Kris Meier§, Sven Bühlmann, Josep Arús-Pous, and Jean-Louis Reymond*

§SCS-Metrohm Award for the Best Oral Presentation in Medicinal Chemistry

Abstract: Drug discovery is in constant need of new molecules to develop drugs addressing unmet medical needs. To assess the chemical space available for drug design, our group investigates the generated databases (GDBs) listing all possible organic molecules up to a defined size, the largest of which is GDB-17 featuring 166.4 billion molecules up to 17 non-hydrogen atoms. While known drugs and bioactive compounds are mostly aromatic and planar, the GDBs contain a plethora of non-aromatic 3D-shaped molecules, which are very useful for drug discovery since they generally have more desirable absorption, distribution, metabolism, excretion and toxicity (ADMET) properties. Here we review GDB enumeration methods and the selection and synthesis of GDB molecules as modulators of ion channels. We summarize the constitution of GDB subsets focusing on fragments (FDB17), medicinal chemistry (GDBMedChem) and ChEMBL-like molecules (GDBChEMBL), and the ring system database GDB4c as a rich source of novel 3D-shaped chiral molecules containing quaternary centers, such as the recently reported trinorbornane.

Keywords: Chemical space · Cheminformatics · Drug discovery · Medicinal chemistry · Molecular shape

Kris Meier obtained bachelor and master degrees in chemistry from ETH Zürich. He performed his master's thesis in the group of Prof. François Diederich on novel PKB inhibitors to investigate sulfur–arene interactions. Already during his studies, he was working in Hydrior AG, where he stayed for two more years afterwards developing new surfactants. In 2016 he started his PhD at the University of Bern in the group of Prof. Jean-Louis Reymond, focusing on the synthesis of novel 3D building blocks obtained from the generated database (GDB) project and their usage in medicinal chemistry.

1. Introduction

Discovering new drugs to address unmet medical needs is essential to the progress of modern medicine.[1,2] To assist drug discovery one often uses the concept of the chemical space, which describes all molecules of possible interest as drugs by placing them in a logically organized search space according to their structural and functional similarities.[3,4] Chemical space contains all molecules that are already known, which includes thousands of approved and experimental drugs such as those in DrugBank[5] and millions of molecules with reported biological activities such as those reported in ChEMBL[6] and PubChem,[7] plus a very large number of theoretically possible drug-like compounds, such as the products obtainable by combining known building blocks through known reactions. Millions of these so-called screening compounds are offered by commercial providers and explicitly listed on the ZINC website,[8] while many more are in principle possible as described in public databases such as SCOBIDOO[9] and proprietary projects such as Pfizer's PGVL[10] or Merck's MASSIV chemical space.[11–13]

Beyond these systematic combinations of building blocks, one can reach more deeply into chemical diversity by asking how many of the building blocks might be derived from first principles by assembling atoms through covalent bonds following the rules of organic chemistry. We have taken this approach in our chemical space project by enumerating all possible molecules up to a certain size to form the generated databases GDBs,[14] the largest of which is GDB-17 comprising 166.4 billion molecules up to 17 atoms of C, N, O, S and halogens following simple rules of chemical stability and synthetic feasibility.[15] Herein we review our recent progress in this approach and discuss the chemical diversity of GDBs in the perspective of 3D-shaped molecules and building blocks for drug discovery.

2. Enumeration and Properties of GDB Molecules

The structure of every organic molecule is derived from a parent mathematical graph whose nodes correspond to non-hydrogen atoms and whose edges correspond to covalent bonds. To enumerate all possible molecules, one can simply start with a list of all possible graphs and write out, for each graph, all chemically meaningful combinations of atom and bond types, adding hydrogen atoms to complement valency. One usually writes out each molecule using SMILES (Simplified Molecular Input Line-Entry System), a compact computer-readable format, optionally includ-

*Correspondence: Prof. J.-L. Reymond, E-mail: jean-louis.reymond@dcb.unibe.ch
Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012 Bern, Switzerland

source: https://doi.org/10.48350/148849 | downloaded: 12.10.2024

ing stereochemical information to explicit all possible stereoisomers.[16] SMILES can be converted to molecular drawings using specific programs such as SmilesDrawer.[17]

The enumeration of molecules from graphs is possible thanks to the existence of the program GENG, which produces the starting list of graphs.[18] For example, GENG produces 843,335 connected graphs up to 11 nodes with a maximum node connectivity of four. Applying simple ring strain criteria reduces the list to 15,726 graphs corresponding to chemically possible saturated hydrocarbons. One then converts these saturated hydrocarbons to more molecules by combinatorially enumerating, under consideration of valency rules, all combinations of single, double and triple bonds at the graph edges and carbon, oxygen, nitrogen and fluorine at the graph nodes. This enumeration produces 1.7 billion molecules. One finally applies rules to remove molecules with chemically unstable functional groups (peroxides, gem-diols, enols, bridgehead double bonds, *etc.*), which finally leaves 26.4 million molecules, corresponding to 111 million individual stereoisomers, forming the chemical universe database GDB-11 listing all possible molecules up to 11 non-hydrogen atoms following the chosen selection criteria (Fig. 1a).[19,20]

In the enumeration of GDB-11 all possible Lewis structures are written out and subsequently evaluated. This approach is not feasible for larger molecules due to the exponential increase in the number of molecules as a function of molecule size. By modifying the code such that certain undesirable molecules are not written out in the first place makes it possible to obtain the much larger databases GDB-13, listing 977 million molecules up to 13 atoms of C, N, O, S, and Cl , and GDB-17, listing 166.4 billion molecules up to 17 atoms of C, N, O, S, and halogens and forming the largest explicitly enumerated database of small molecules to date.[15,21] In the case of GDB-17, very drastic selection rules are used to select the graphs above 15 nodes to avoid a combinatorial explosion in the number of enumerated molecules. For example, only one 3- or 4-membered ring is allowed at 16 nodes and no small rings at all at 17 nodes. Furthermore, no non-aromatic carbon–carbon double bonds are allowed at 17 nodes.

A very attractive feature of the GDBs is that they represent an exhaustive coverage of chemical space under a defined set of rules. The database GDB-13 was recently used to evaluate the ability of molecular generative neural networks (MGNNs) to exhaustively cover chemical space.[22] This study showed that an MGNN trained with a random subset of one million molecules from GDB-13 (0.1% of the database) is capable of covering 70% of the entire GDB-13 when generating 2 billion molecules.[23] Creating twice the database size is necessary due to the generation of duplicates. The performance increased to 83% when using randomized SMILES rather than canonical SMILES to represent the molecules.[24] Moreover, only 4.7% of the molecules generated by the MGNN are outside the GDB-13 chemical space (Fig. 1b). This study suggests that the exhaustive exploration of chemical space by enumeration as realized with the GDBs might in future be realized in targeted manner by using MGNNs trained with specific sets of molecules.

Almost all molecules in the GDBs (>99.9%) are outside the known chemical space, simply as a consequence of the very large number of molecules in the databases, which far exceeds the number of molecules of similar size in databases of known molecules such as PubChem. GDB molecules feature a diversity of ring system and functional group combinations spanning a broad range of molecular properties. One can gain an overview of GDB molecules by considering MQN-maps, which are obtained by computing the 42 Molecular Quantum Numbers (MQN, counts of types of atoms, bonds, polar groups and topological features) for each molecule and representing the first two principal components of a principal component analysis (PCA) of the dataset.[25,26] The resulting map can be color-coded as a function of various molecular properties (molecular size, number of rings, *etc.*) and inspected interactively using a dedicated MQN-mapplet application.[27] For example, color-coding GDB-13 by molecular categories illustrates that the database comprises a very broad diversity of heterocycles, while aromatic and heteroaromatic molecules are relatively rare (Fig. 1c).[28]

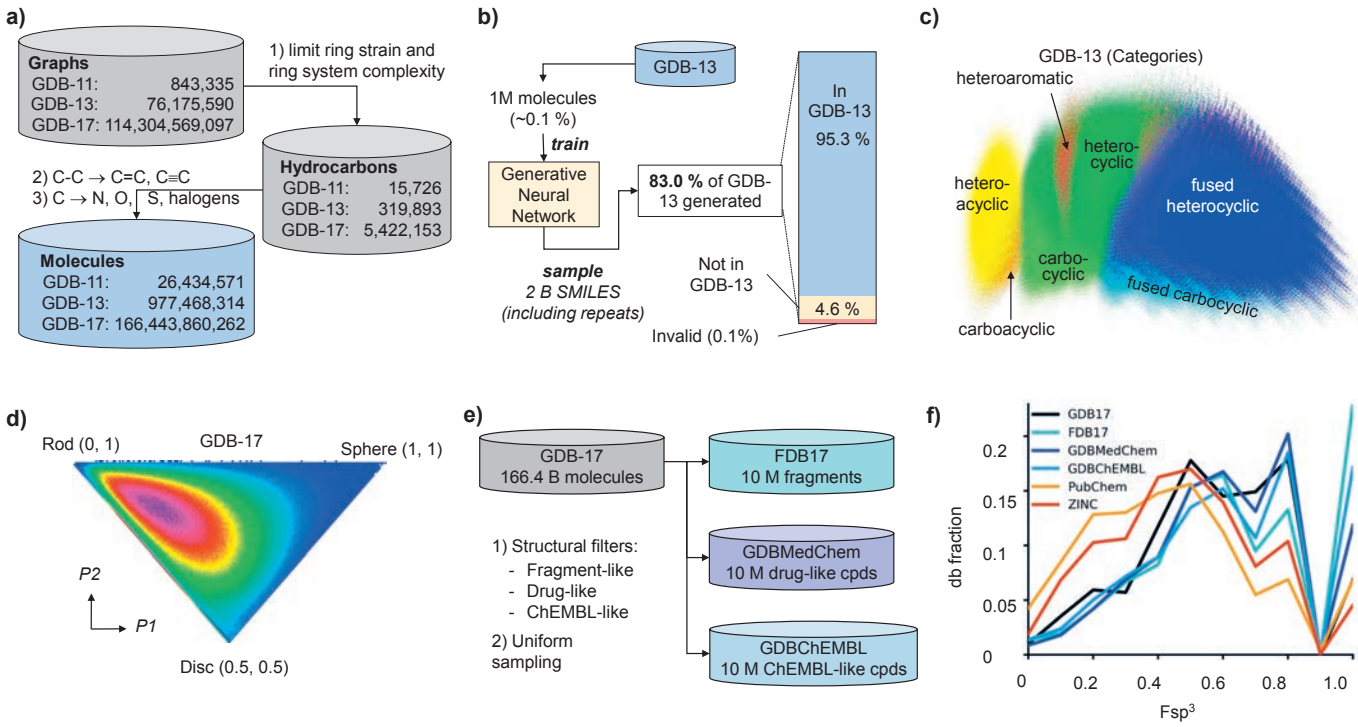

Fig. 1. The generated databases GDB. (a) Principle of GDB enumeration. (b) Generative neural networks can reproduce GDB-13 after learning from only 0.1% of the database. (c) MQN-map of GDB-13 obtained from by principal component analysis, color-coded by the distribution of molecular categories. See ref. [28] for details. (d) Shape analysis of GDB-17 according to principal moments of inertia. (e) Subset assembly of GDB-17. (f) Fsp^3 plot for GDB-17, its three subsets plus PubChem and ZINC.

A striking difference to the known chemical space is that the fraction of aromatic molecules is much lower in the GDBs than in databases of known molecules. Furthermore, compared to known molecules, a much larger percentage of GDB molecules have a high fraction of sp^3-carbon atoms, are rich in stereocenters, and have significant 3D-shape. The prevalence of 3D-shaped molecules in the GDBs is illustrated here for GDB-17 by plotting the molecules in the shape triangle representing the principal moments of inertia allowing to categorize the molecules as rod-like, disk-like or sphere-like (Fig. 1d).[29] 3D-shape in GDB molecules results from the presence of many asymmetric and quaternary carbons as well as saturated and bridged polycyclic systems. 3D-shape is also very typical of natural products and is generally considered a desirable feature in medicinal chemistry as means to escape out of 'flatland',[30] and obtain molecules with higher selectivity and lower toxicity.[31] In addition to their 3D-shape, GDB molecules generally obey Lipinski's rule of 5 for oral bioavailability,[32] and also pass criteria of lead-likeness[33] and fragment-likeness,[34] implying that the majority of GDB molecules are of interest for medicinal chemistry.

3. Selection and Synthesis of Bioactive GDB Molecules

GDB molecules have been to date mostly synthesized in our group for medicinal chemistry projects aiming to identify new modulators of ligand gated ion channels and transporters for which substrates and inhibitors are small molecules in the size range covered by GDB. Molecules were selected from the GDB by a three-step procedure involving: 1) selection of a subset of GDB according to functional group and property choices; 2) virtual screening by docking to the binding pocket of a known protein target; 3) selection from the list of best docking compounds according to synthetic feasibility and novelty. The first proof-of-principle of this approach was demonstrated for inhibitors for the glycine binding site of the NMDA receptor, which uncovered dipeptides **1** and **2** and diketopiperazine **3** as three novel inhibitors of this ion channel (Fig. 2a).[35,36]

The second project was directed at the glutamate transporter GLT-1, also known as EAAT2, which was interesting because known inhibitors such as WAY-855, investigated as possible drugs to treat ischemic stroke injury, are typical GDB-like 3D-shaped molecules.[37] Here a subset of GDB-11 was selected containing saturated primary and secondary amines. For each amine, two carboxyl groups were added at the α- and β- or γ-carbon atoms relative to the amino group to obtain analogs of aspartate or glutamate, which are both recognized by the transporter. Virtual screening by docking to the bacterial glutamate transporter homologue from *Pyrococcus horikoshii* (PDB code: 1XFH) and inspection of the best docking compounds indicated that yet unknown norbornane aspartate analogs such as **4** were among the best performing docking compounds. The *exo*-stereoisomer *rac*-**5**, found to be a weak GLT-1 inhibitor and was prepared starting from cyclopentadiene and dimethyl acetylene dicarboxylate by a Diels-Alder cycloaddition to form the norbornadiene **6**. Michael addition of ammonia and protection of the amine formed a mixture of stereoisomeric bicyclic aspartates from which stereoisomer **7** was separated. Finally, *rac*-**5** was obtained by *xo*-stereoselective hydroformylation of the remaining norbornene double bond to form aldehyde **8** followed by a Wittig olefination. Further optimization led to the benzyl analog *rac*-**10**, obtained by hydrogenation and deprotection of the Wittig product **9**, as the best inhibitor in the series with comparable activity to WAY-855 but better selectivity against EAAC1 (Fig. 2b).[38]

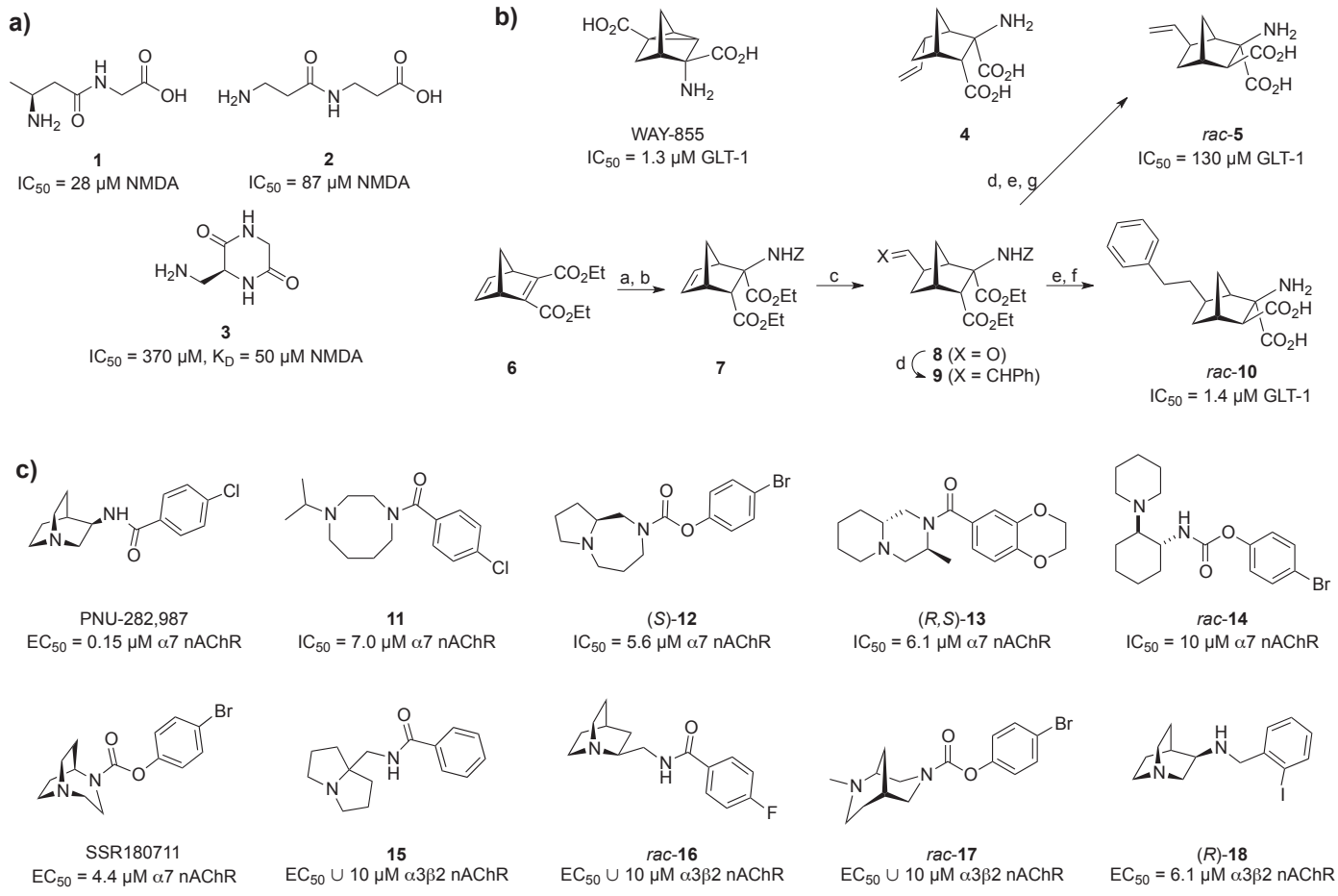

Fig. 2. Medicinal chemistry projects with GDB. (a) NMDA glycine site inhibitors selected by docking of GDB-11 amino acids. (b) Bicyclic aspartate analogs derived from GDB-11 as GLT-1 inhibitors, including synthesis of inhibitor *rac*-**10**. *Conditions:* a) NH_4OAc; b) BnOCOCl, Na_2CO_3; c) CO/H_2, $Rh(CO)_2acac$; d) $H_2C{=}PPh_3$ or $PhHC{=}PPh_3$; e) LiOH; f) H_2, Pd/C; g) AcOH/conc. HCl. (c) α7 nAChR and α3β2 nAChR modulators.

The third project exploited the fact that GDB is extremely rich in 3D-shaped cyclic and polycyclic amines by attempting to identify new analogs of PNU-282,987 or SSR180711, which are both partial agonist of the α7 nicotinic acetyl choline receptor (nAChR) investigated to treat various neurological disorders due to their effect on memory improvement. The aim was to replace the bicyclic diamine core of both compounds with a new diamine from the GDB. To identify suitable diamines, all possible monocyclic and bicyclic aliphatic diamines containing one tertiary and one primary or secondary amine were extracted from GDB-11. The benzamides and carbamates corresponding to the parent drugs were then virtually formed and docked to the acetylcholine binding protein (AChBP) from Lymnaea stagnalis (PDB code: 1UW6) as a model for the α7 nAChR. Twenty-one interesting cyclic diamines were then selected from the best docking compounds, synthesized, acylated with various benzoic acids, and tested on the α7 nAChR, which uncovered several new inhibitors such as **11**, **12** and (*R,S*)-**13** (Fig. 2c). An additional screening against the α3β2 nAChR, a receptor subtype for which no modulators were known, uncovered the first positive allosteric modulators (PAM) of this receptor subtype in form of benzamides **15**, *rac*-**16** and *rac*-**17**.[39] Interestingly in this case a subsequent MQN-similarity search for new analogs of PNU-282,987 in the database ChEMBL followed by optimization led to the identification of *N*-benzylaminoquinuclidine (*R*)-**18** as a much more potent PAM for this receptor, illustrating that many unexplored possibilities also exist within databases of known molecules.[40]

4. Characterization of the GDB Subsets

Due to the combinatorial nature of the enumeration, the most abundant GDB molecules are those with most functional groups and chiral centers, which are also the most challenging to synthesize. To facilitate the selection of realistic synthetic targets, one should select GDB subsets focusing on molecules with reduced complexity. For example, the GDB-13 database containing almost one billion molecules is also available as a subset of 43.7 million molecules from which problematic substructures (N–N and N–O bonds, aldehydes, esters, carbonates, sulfates, epoxides, aziridines, non-aromatic C=C, all C≡C, 3- and 4-membered rings) have been removed.[28] This subset retains sufficient diversity for the search of new drugs, as illustrated by the discovery of new α7 nAChR modulators among nicotine analogs,[41] and can even be further reduced to only 12.9 million molecules by considering only fragment-like molecules. Another GDB-13 subset of 59.5 million molecules features only molecules with fragrance-like properties.[42]

The above subsets reduce functional group density but do not correct for the overabundance of the stereochemically most complex, most functionalized and largest molecules in GDB-13. A better distribution of molecules across size, functional group and stereochemical complexity is achieved in three recent subsets of GDB-17, each containing only 10 million molecules featuring fragments (fragment database FDB17),[43] molecules relevant for medicinal chemistry (GDBMedChem),[44] or molecules similar to bioactive compounds in ChEMBL (GDBChEMBL) (Fig. 1e).[45] In these three cases GDB-17 was first filtered by applying structural filters similar to those discussed above for GDB-13 to form smaller subsets of a few billion molecules. These subsets were then binned according to molecular size, number of stereocenters and number of heteroatoms, and sampled evenly across these parameters to form collections of 10 million molecules.

Despite of the severe limitations on diversity imposed by the selection and sampling procedure, the three GDB-17 subsets retain the distribution of 3D-shaped molecules of the parent database, as illustrated for example by the fraction of sp^3-carbon atoms (Fig. 1f). Furthermore, the subsets retain a very high structural diversity. This can be illustrated by counting the number of different molecular shingles, which are the elements encoded in MHFP6, a structural fingerprint similar but superior to the classical ECFP4 in terms of virtual screening performance.[46–48] Note that the number of different molecular shingles is significantly higher in GDB-17 and the three subsets thereof compared to molecules of similar size in public databases, illustrating the broader structural diversity available in GDB molecules (Table 1). This structural diversity is particularly striking for GDBChEMBL because this subset was created by selecting GDB molecules with a threshold fraction of molecular shingles also occurring in bioactive ChEMBL molecules.

5. The Ring Systems Database GDB4c

Ring systems can be defined as what remains after removing all acyclic atoms and bonds in a molecule and converting all atoms to carbon. As noted already in 2007 with GDB-11, there is striking novelty in the GDB at the level of ring systems. Indeed, not only has the vast majority of molecules never been synthesized, but even the majority of ring systems are not exemplified in any known molecule. The database GDB4c provides a very extensive insight into the potential for novel rings systems by listing 916,130 ring systems up to four rings that are possible under a set of constraints (max. ring size 14, max. 2 rings larger than 6, max. one ring larger than 7, no atoms shared by two 3- or 4-membered rings for tetracyclic ring systems) including optional aromatization of 5- and 6-membered rings whenever possible.[49]

When considering the 79,502 ring systems that can be identified in known molecules (ZINC, PubChem, ChEMBL and Reaxys combined) without constraints for ring size or number of rings, collected in a reference database (RDB), one finds that only 12,536 of these comply with the constraints of GDB4c, which implies that 98.6% of the rings systems in GDB4c are novel. Note that the occurrence of ring systems in known molecules follows a power law, with benzene alone accounting for 50% of occurrences and the ten most abundant ring systems accounting for nearly 90% of all occurrences, which shows that known molecules are particularly poor in ring system diversity.

While aromatic rings are very abundant in RDB (46.9%), illustrating that planarity dominates in known molecules, 79.5% of the ring systems in GDB4c do not contain any aromatic rings and are therefore 3D-shaped. Note that while ring systems with a six-membered ring as largest ring are most abundant in RDB, the majority of GDB4c ring systems are macrocycles (ring size ≥ 8). The ring systems database is also available as GDB4c3D containing the 3D-structure of the 6,555,929 stereoisomers derived from GDB4c. In GDB4c3D ring systems are labelled with a number, called p-value that reflects the overall ring strain as mea-

Table 1. Shingle count comparison of GDB-17 subsets vs. 10 M collections of ZINC and PubChem compounds up to 17 heavy atoms.

Database	Total shingles	Shingles unique to db
GDB-17[a]	33,036,456	22,433,445
FDB17[b]	29,135,003	20,732,026
GDBMedChem[c]	24,280,303	15,558,424
GDBChEMBL[d]	22,463,337	14,952,479
ZINC[e]	2,455,487	863,571
PubChem[f]	12,750,754	9,437,808

[a]GDB-17: 10 M uniform sample; [b]FDB17: 10.1 M; [c]GDBMedChem: 9.9 M; [d]GDB-ChEMBL: 9.8 M; [e]ZINC: 10 M uniform sample, HAC ≤ 17; [f]PubChem: 10 M uniform sample, HAC ≤ 17.

sured by the deviation of bond angles from their optimum value. Interestingly, the distribution of p-values in GDB4c3D is comparable to that among known ring systems (RDB3D) and among molecules from the Cambridge Structural Database (CSD), which shows that GDB4c molecules do not have a particular ring strain compared to known molecules (Fig. 3a).

Much of the novelty of GDB4c lies in intrinsically chiral, 3D-shaped ring systems with unusual topologies often including quaternary centers (Fig. 3b). A striking example is trinorbornane **19**, which was noted as an original but yet unknown ring system in GDB-11 in 2007,[20] and has been synthesized by the group of Marcel Mayor in Basel in 2017 using an eight-step sequence. Unsaturated ketone **20** was converted into cyclopentadiene, which reacted in an intramolecular Diels-Alder reaction to form the central norbornane and one of the external norbornanes (**21**) simultaneously. Radical ring closure yielded the third norbornane unit of trinorbornane **19** (Fig. 3c).[50] Similarly 3D-shaped, chiral ring systems containing a quaternary center also occur in natural products, as exemplified here with the tricyclic ring system **22** found in the recently discovered Crokonoid A (**23**).[51] Further 3D-shaped chiral ring systems that are typical of GDB4c but are already known include twistane **24** obtained by intramolecular α-alkylation of a [2.2.2]bicyclooctanone **25** and deoxygenation of **26**,[52] and tris-homocubane reported as the trioxa tris-homocubane **27** resulting from epoxidation of barrelene **28** followed by acid-catalyzed cyclization of the symmetrical triepoxide stereoisomer **29**.[53]

6. Conclusion and Outlook

The GDBs offer an overwhelming diversity of new molecules. In previous projects synthetic targets were selected to achieve specific bioactivities by choosing molecules according to their similarities to known bioactive compounds and to their predicted binding affinities to a target protein. In all cases synthetic considerations limited choices to the small fraction of GDB molecules that were not only novel but also simple to synthesize. These synthetic considerations led to the idea of defining smaller subsets of the GDBs to facilitate the identification of interesting yet feasible molecules. Similarity search portals are freely accessible at *http://gdb.unibe.ch* to search the GDBs by similarity to any molecule of choice. GDB subsets can also be explored in undirected manner using interactive layouts such as Faerun and TMAP accessible at the same website.[54,55] We are currently synthesizing GDB molecules identified using these search tools as novel building blocks for medicinal chemistry to be later functionalized to achieve specific target activities.

Acknowledgements

This work was supported financially by the Swiss National Science Foundation

Received: February 3, 2020

[1] K. H. Bleicher, H. J. Bohm, K. Muller, A. I. Alanine, *Nat. Rev. Drug Discov.* **2003**, *2*, 369.
[2] D. H. Drewry, R. Macarron, *Curr. Opin. Chem. Biol.* **2010**, *14*, 289, DOI: 10.1016/j.cbpa.2010.03.024.
[3] J. L. Medina-Franco, K. Martinez-Mayorga, M. A. Giulianotti, R. A. Houghten, C. Pinilla, *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 322, DOI: 10.2174/157340908786786010.
[4] M. Awale, R. Visini, D. Probst, J. Arus-Pous, J. L. Reymond, *Chimia* **2017**, *71*, 661, DOI: 10.2533/chimia.2017.661.
[5] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z. T. Dame, B. Han, Y. Zhou, D. S. Wishart, *Nucleic Acids Res.* **2014**, *42*, D1091, DOI: 10.1093/nar/gkt1068.
[6] A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit, A. R. Leach, *Nucleic Acids Res.* **2017**, *45*, D945, DOI: 10.1093/nar/gkw1074.
[7] A. Gindulyte, B. A. Shoemaker, B. Yu, J. He, J. Zhang, J. Chen, L. Zaslavsky, P. A. Thiessen, Q. Li, S. He, S. Kim, T. Cheng, E. E. Bolton, *Nucleic Acids Res.* **2018**, *47*, D1102, DOI: 10.1093/nar/gky1033.
[8] T. Sterling, J. J. Irwin, *J. Chem. Inf. Model.* **2015**, *55*, 2324, DOI: 10.1021/acs.jcim.5b00559.
[9] F. Chevillard, P. Kolb, *J. Chem. Inf. Model.* **2015**, *55*, 1824, DOI: 10.1021/acs.jcim.5b00203.
[10] Q. Hu, Z. Peng, J. Kostrowicki, A. Kuki, *Methods Mol. Biol.* **2011**, *685*, 253, DOI: 10.1007/978-1-60761-931-4_13.

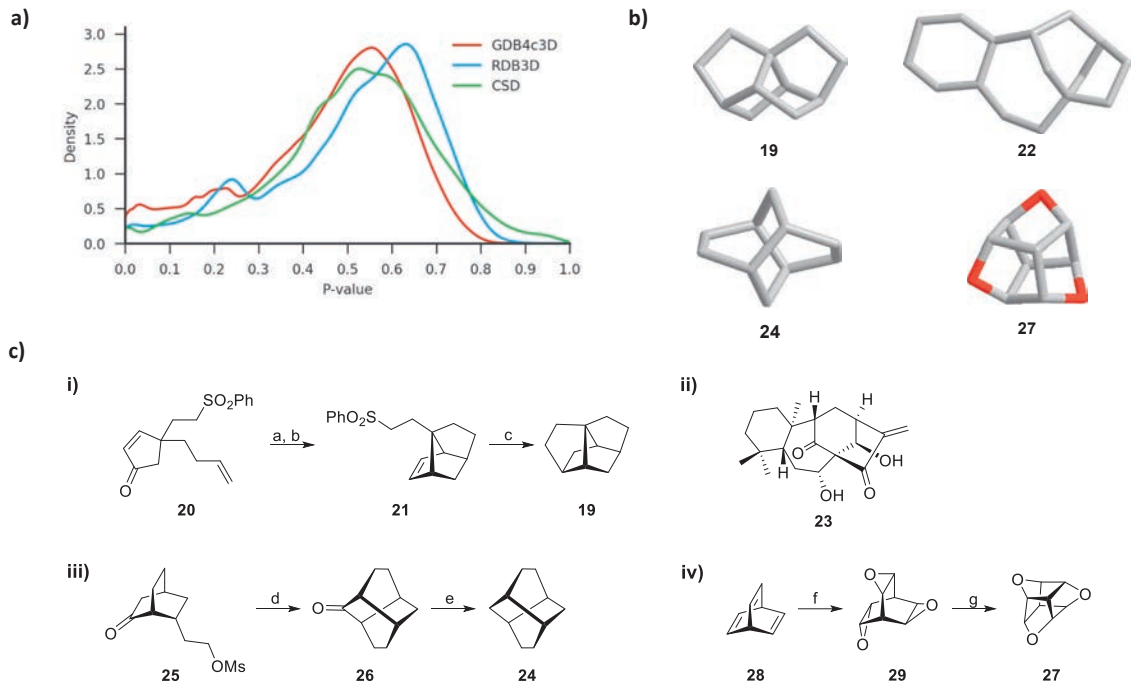

Fig. 3. Comparison of GDB4c and known molecules. (a) Kernel density estimation of the p-value distribution in each database. (b) Examples of 3D-shaped, intrinsically chiral ring systems from GDB4c occurring in known molecules. (c) i) Trinorbornane (**19**) synthesis. ii) Structure of Crokonoid A (**23**). iii) Twistane (**24**) synthesis. iv) Trioxa tris-homocubane (**27**) synthesis. *Conditions:* a) $NaBH_4$, $CeCl_3$; b) cat. *p*-TsOH; c) SmI_2; d) NaH; e) H_2NNH_2, KOH; f) Oxone, $NaHCO_3$; g) $BF_3{\cdot}Et_2O$.

[11] W. P. Walters, *J. Med. Chem.* **2019**, *62*, 1116, DOI: 10.1021/acs.jmedchem.8b01048.
[12] N. van Hilten, F. Chevillard, P. Kolb, *J. Chem. Inf. Model.* **2019**, *59*, 644, DOI: 10.1021/acs.jcim.8b00737.
[13] T. Hoffmann, M. Gastreich, *Drug Discov. Today* **2019**, *24*, 1148, DOI: 10.1016/j.drudis.2019.02.013.
[14] J. L. Reymond, L. Ruddigkeit, L. C. Blum, R. Van Deursen, *WIREs comput. Mol. Sci.* **2012**, *2*, 713.
[15] L. Ruddigkeit, R. van Deursen, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **2012**, *52*, 2864, DOI: 10.1021/ci300415d.
[16] D. Weininger, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31, DOI: 10.1021/ci00057a005.
[17] D. Probst, J.-L. Reymond, *J. Chem. Inf. Model*. **2018**, *58*, 1, DOI: 10.1021/acs.jcim.7b00425.
[18] B. D. McKay, *Congressus Numerantium* **1981**, *30*, 45.
[19] T. Fink, H. Bruggesser, J. L. Reymond, *Angew. Chem., Int. Ed.* **2005**, *44*, 1504.
[20] T. Fink, J. L. Reymond, *J. Chem. Inf. Model.* **2007**, *47*, 342.
[21] L. C. Blum, J. L. Reymond, *J. Am. Chem. Soc.* **2009**, *131*, 8732.
[22] J. Arús-Pous, M. Awale, D. Probst, J.-L. Reymond, *Chimia* **2019**, *73*, 1018, DOI: 10.2533/chimia.2019.1018.
[23] J. Arús-Pous, T. Blaschke, S. Ulander, J.-L. Reymond, H. Chen, O. Engkvist, *J. Cheminf.* **2019**, *11*, 20, DOI: 10.1186/s13321-019-0341-z.
[24] J. Arús-Pous, S. V. Johansson, O. Prykhodko, E. J. Bjerrum, C. Tyrchan, J.-L. Reymond, H. Chen, O. Engkvist, *J. Cheminf.* **2019**, *11*, 71, DOI: 10.1186/s13321-019-0393-0.
[25] K. T. Nguyen, L. C. Blum, R. van Deursen, J.-L. Reymond, *ChemMedChem* **2009**, *4*, 1803, DOI: 10.1002/cmdc.200900317.
[26] R. van Deursen, L. C. Blum, J. L. Reymond, *J. Chem. Inf. Model.* **2010**, *50*, 1924.
[27] M. Awale, R. van Deursen, J. L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 509, DOI: 10.1021/ci300513m.
[28] L. C. Blum, R. van Deursen, J. L. Reymond, *J. Comput.-Aided Mol. Des.* **2011**, *25*, 637.
[29] W. H. Sauer, M. K. Schwarz, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 987, DOI: 10.1021/ci025599w.
[30] F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **2009**, *52*, 6752, DOI: 10.1021/jm901241e.
[31] T. J. Ritchie, S. J. Macdonald, R. J. Young, S. D. Pickett, *Drug Discov. Today* **2011**, *16*, 164, DOI: 10.1016/j.drudis.2010.11.014.
[32] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **1997**, *46*, 3, DOI: 10.1016/s0169-409x(00)00129-0.
[33] S. J. Teague, A. M. Davis, P. D. Leeson, T. Oprea, *Angew. Chem., Int. Ed.* **1999**, *38*, 3743.
[34] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discov. Today* **2003**, *8*, 876, DOI: 10.1016/s1359-6446(03)02831-9.
[35] K. T. Nguyen, S. Syed, S. Urwyler, S. Bertrand, D. Bertrand, J.-L. Reymond, *ChemMedChem* **2008**, *3*, 1520, DOI: 10.1002/cmdc.200800198.
[36] K. T. Nguyen, E. Luethi, S. Syed, S. Urwyler, S. Bertrand, D. Bertrand, J. L. Reymond, *Bioorg. Med. Chem. Lett*. **2009**, *19*, 3832.
[37] J. Dunlop, S. Eliasof, G. Stack, H. B. McIlvain, A. Greenfield, D. Kowal, R. Petroski, T. Carrick, *Br. J. Pharmacol.* **2003**, *140*, 839, DOI: 10.1038/sj.bjp.0705509.
[38] E. Luethi, K. T. Nguyen, M. Bürzle, L. C. Blum, Y. Suzuki, M. Hediger, J.-L. Reymond, *J. Med. Chem.* **2010**, *53*, 7236, DOI: 10.1021/jm100959g.
[39] L. Bréthous, N. Garcia-Delgado, J. Schwartz, S. Bertrand, D. Bertrand, J.-L. Reymond, *J. Med. Chem*. **2012**, *55*, 4605, DOI: 10.1021/jm300030r.
[40] J. J. Burgi, M. Awale, S. D. Boss, T. Schaer, F. Marger, J. M. Viveros-Paredes, S. Bertrand, J. Gertsch, D. Bertrand, J. L. Reymond, *ACS Chem. Neurosci.* **2014**, *5*, 346, DOI: 10.1021/cn4002297.
[41] L. C. Blum, R. van Deursen, S. Bertrand, M. Mayer, J. J. Bürgi, D. Bertrand, J.-L. Reymond, *J. Chem. Inf. Model.* **2011**, *51*, 3105, DOI: 10.1021/ci200410u.
[42] L. Ruddigkeit, M. Awale, J. L. Reymond, *J. Cheminf.* **2014**, *6*, 27, DOI: 10.1186/1758-2946-6-27.
[43] R. Visini, M. Awale, J. L. Reymond, *J. Chem. Inf. Model.* **2017**, *57*, 700, DOI: 10.1021/acs.jcim.7b00020.
[44] M. Awale, F. Sirockin, N. Stiefl, J.-L. Reymond, *Mol. Inf.* **2019**, *38*, 1900031, DOI: 10.1002/minf.201900031.
[45] S. Bühlmann, J. L. Reymond, *Front. Chem*. **2020**, DOI: 10.3389/fchem.2020.00046.
[46] D. Probst, J.-L. Reymond, *J. Cheminf.* **2018**, *10*, 66, DOI: 10.1186/s13321-018-0321-8.
[47] H. L. Morgan, *J. Chem. Doc.* **1965**, *5*, 107, DOI: 10.1021/c160017a018.
[48] S. Riniker, G. A. Landrum, *J. Cheminf.* **2013**, *5*, 26, DOI: 10.1186/1758-2946-5-26.
[49] R. Visini, J. Arús-Pous, M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.* **2017**, *57*, 2707, DOI: 10.1021/acs.jcim.7b00457.
[50] a) L. D. Bizzini, T. Müntener, D. Häussinger, M. Neuburger, M. Mayor, *Chem. Commun.* **2017**, *53*, 11399, DOI: 10.1039/C7CC06273G; b) L. D. Bizzini, T. Bürgi, M. Mayor, *Helv. Chim. Acta* **2020**, *103*, e2000019, DOI: 10.1002/hlca.202000019.
[51] Y.-Y. Fan, S.-Q. Shi, G.-Z. Deng, H.-C. Liu, C.-H. Xu, J. Ding, G.-W. Wang, J.-M. Yue, *Org. Lett.* **2020**, DOI: 10.1021/acs.orglett.9b04484.
[52] H. W. Whitlock, *J. Am. Chem. Soc.* **1962**, *84*, 3412, DOI: 10.1021/ja00876a047.
[53] S. I. Kozhushkov, T. Preuß, D. S. Yufit, J. A. K. Howard, K. Meindl, S. Rühl, C. Yamamoto, Y. Okamoto, P. R. Schreiner, B. C. Rinderspacher, A. de Meijere, *Eur. J. Org. Chem.* **2006**, *2006*, 2590, DOI: 10.1002/ejoc.200600019.
[54] D. Probst, J.-L. Reymond, *Bioinformatics* **2018**, *34*, 1433, DOI: 10.1093/bioinformatics/btx760.
[55] D. Probst, J.-L. Reymond, *J. Cheminf.* **2020**, *12*, 12, DOI: 10.1186/s13321-020-0416-x.

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License CC BY_NC 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (*http://chimia.ch/component/sppagebuilder/?view=page&id=12*).

The definitive version of this article is the electronic one that can be found at doi:10.2533/chimia.2020.241