

Cheminformatic Analysis of Natural Products and their Chemical Space

Stefan Wetzel^a, Ansgar Schuffenhauer^b, Silvio Roggo^b, Peter Ertl^b, and Herbert Waldmann^{*a}

Abstract: Cheminformatic methods allow the detailed characterization of particular and characteristic properties of natural products (NPs) and comparison with related characteristics of drugs and other compounds. An overview of the most important properties of natural products and analogues and their difference with respect to drugs and synthetic compounds is presented. Moreover, different approaches to charting the chemical space populated by natural products are reviewed and their underlying principles are delineated. Some insights about NP chemical space are described together with possible applications of methods charting chemical space. Strengths and weaknesses of the different approaches will be discussed with respect to possible applications in compound collection design.

Keywords: Chemical space · ChemGPS · Cheminformatics · Natural products · SCONP

1. Introduction

Natural products (NPs) have been selected during evolution to bind to various proteins during their life-cycle, *e.g.* in biosynthesis and degradation and while exerting their mode of action. Thus NP structures are good starting points for the discovery and development of protein ligands and, therefore, ultimately for drug discovery. Indeed many current drugs originate from NPs although their structures were chemically modified in many cases.^[1] After a recession of NPs in the pharmaceutical industry in the 1990s recent years have witnessed a comeback of NPs in drug discovery. NPs were often viewed as being ‘too complex’ and ‘sufficiently examined’ but once again are recognized as a valuable source of interesting and diverse structures.^[2,3] NPs in gener-

al constitute biologically validated starting points for library design and many of their core structures have been recognized to be privileged structures.^[4,5] Known examples for privileged structures in NPs that have successfully been applied in library design are the benzodiazepine^[4] and the indole scaffolds.^[6,7]

With the upcoming new focus on NPs in drug discovery interest in their chemical and structural properties and in ways to explore them in drug discovery have increased as well. Modern cheminformatics methods allow the rapid prediction of a plethora of chemical and physico-chemical properties of available NP structures.^[8] Nowadays, many of these properties are routinely used in library design and optimization. The most prominent and widely used example may be Lipinski’s ‘Rule-of-Five’. The Rule-of-Five denotes a set of property rules describing orally bioavailable drug space. They were empirically derived from known orally available drugs.^[9] Chemical properties constitute one way to describe the part of chemical space occupied by a compound set; structural features is another. Several approaches have been taken to map and navigate chemical space reflecting chemical properties and structure.

2. Natural Product Properties

The systematic evaluation of NP properties yields a better understanding of what distinguishes NPs from drugs or combinatorial chemistry compounds. This knowledge

may be applied in the synthesis of compounds with NP-like properties. There are several publications describing the results of statistical analyses of NP properties. In 1999 Henkel *et al.* published a first comparison of molecular properties of natural products, synthetic compounds and drugs.^[10] Two years later, Lee and Schneider^[11] published a paper evaluating trade drugs and natural products also addressing the general notion that natural products may not be very drug-like. In 2003 Feher and Schmidt^[12] published one of the most comprehensive comparisons of NPs, drugs and combinatorial chemistry compounds and compared these three groups using 40 different properties. Ertl and Schuffenhauer analyzed properties and structural features of more than 130,000 NP molecules and identified substructures typical for particular classes of source organisms.^[13] One of the most recent analyses of natural product properties and scaffolds was conducted by Grabowski and Schneider in 2007.^[14] In the following paragraphs the results of these publications will be summarized.

Henkel *et al.* analyzed two natural product databases, the Dictionary of Natural Products (DNP)^[15] and the Bioactive Natural Product Database.^[16] Synthetic compounds were taken from the Chemicals Dictionary (ACD)^[17] and the Bayer AG in-house collection. The analysis revealed that molecular weight distribution was comparable for drugs and NPs but synthetic compounds were found to have a lower molecular weight. The average NP molecule was found to contain three stereogenic

*Correspondence: Prof. Dr. H. Waldmann^a

Tel.: + 49 231 133 2400

Fax: + 49 231 133 2499

E-Mail: herbert.waldmann@mpi-dortmund.mpg.de

^aMax-Planck Institut für Molekulare Physiologie

Abteilung IV – Chemische Biologie

Otto-Hahn-Strasse 11

D-44229 Dortmund

and Universität Dortmund, Chemische Biologie

Dortmund

^bNovartis Institutes for BioMedical Research

CH-4002 Basel

centers, three times as many as the average drug molecule. The authors also reported that NPs contain more oxygen but significantly less nitrogen compared to the other compound classes. Additionally, Henkel *et al.* performed an analysis of pharmacophoric groups, *i.e.* structural units prone to interact with biological macromolecules. These include different functional groups like acids or alcohols and their isosteres as well as structural motifs, *e.g.* aryl or alkyl moieties. As expected, NPs incorporate more oxygen-containing groups while drugs and synthetic compounds are abundant with nitrogen-containing ones. Since the statistics on heteroatoms as well as on pharmacophoric groups were reported as fraction of molecules incorporating them, no average number per molecule can be given. In general, differences in the pharmacophoric groups of NPs and drugs tend to be smaller than between NPs and synthetic compounds. Particularly interesting may be the largest differences which can be found for arenes. NPs contain half as many aryl groups as synthetics. Alcohols are three times more frequent in NPs compared to synthetic compounds. Henkel *et al.* also did a detailed analysis of the NP compound classes represented in the DNP. The result showed a clear predominance of alkaloids and terpenes over most other natural product classes in the DNP. Although this may have changed by now it should be kept in mind that the results of every analysis inherently reflect the dataset which was analyzed.

Lee and Schneider's paper in 2001 compared a set of parameters mainly related to the Rule-of-Five for trade drugs and natural products.^[11] In this survey, heteroatom frequencies were given as average atoms per molecule. The natural product molecules contained an average of 1.4 nitrogen atoms per molecule, about one less than the drug set while the average number of oxygen atoms per molecule was reported to be about four in both sets. The calculated logP values indicate that natural products are more lipophilic than drug molecules (2.9 vs 2.1). One of the most surprising findings, however, was the low rate of Rule-of-Five violations. Only 10% of the NPs violated the rules although these rules have been derived exclusively from orally available drugs. A comparable rate of violation was found in the drug set suggesting that, at least with respect to the Rule-of-Five, small natural product molecules were found to be more drug-like than they were thought to be.

A more comprehensive study on the properties of natural products in comparison to those of drugs and libraries resulting from combinatorial chemistry was conducted by Feher and Schmidt in 2003.^[12] They analyzed the drugs from the Dictionary of Drugs and all other structures from the cata-

logues of compound library vendors. Altogether, the authors calculated more than 40 molecular properties for all test sets. They found the average number of stereogenic centers to be 6.2 for NPs and 2.3 for drugs. Notably, these averages are twice as high as those found by Henkel *et al.* four years earlier while the ratio of 3:1 remained the same. The NPs contained on average two more oxygen atoms than the drugs which is in agreement with Henkel *et al.* although not with the results of Lee and Schneider. Half as many nitrogen atoms were found in NPs compared to drugs and even much less sulphur and halogens. NPs contain about two rings more than drugs while at the same time their degree of ring fusion is twice as high. NPs were found to be generally more unsaturated, however they contain considerably less aromatic rings. The number of rotatable bonds found in NPs is two less than in drugs. Together with the high degree of unsaturation and ring fusion, this indicates that NPs are on average more rigid than drugs. Feher and Schmidt also distinguished between genuine natural products and 'seminaturals', *i.e.* derivatives. They found the seminatural compounds to be more drug-like in their properties especially with respect to the number of stereogenic centers, aromaticity, ring fusion and heteroatom distribution.

Grabowski and Schneider^[14] compared the properties and scaffolds of drugs, pure natural products (PNP), NP derivatives and, particularly interesting, of marine natural products in their 2007 publication. While most of their findings are comparable to those of Feher and Schmidt, Grabowski and Schneider found that 10% of the drug molecules but 18% of the pure natural products and 30% of the marine natural products analyzed by them showed at least two violations of Lipinski's Rule-of-Five. The violation rate for drugs matches the value published earlier by Lee and Schneider, but for the pure natural products the figure is twice as high. The higher rate found for marine natural products may be due to the higher molecular weight and the increased number of acceptors per molecule. Notably, the marine natural products analyzed contain fewer rings and fewer atoms in aromatic rings but more rotatable bonds than PNPs, which indicates higher flexibility.

The analysis of Ertl and Schuffenhauer^[13] on the largest set of NP structures studied so far (more than 130,000 molecules) confirmed the results obtained by previous studies for smaller sets. Additionally, the authors identified the scaffolds and substituents which are typical for NPs and which may serve as inspiration for the design of NP-like combinatorial libraries. Study of NPs produced by different classes of organisms (bacteria, fungi, plants, and

animals) could identify some substructural features typical for these groups.

From the data described one can conclude that NPs differ from drugs in several properties although the majority of them do not violate Lipinski's Rule-of-Five. Properties which distinguish NPs from drugs are heteroatom distribution, number of rings, fraction of aromatic rings and degree of ring fusion. Certainly some of these differences may reflect empirical knowledge and synthetic methodology applied in drug discovery. But others may be explained by the limitations of the biosynthesis of NPs, *e.g.* the availability of certain elements. The very low nitrogen content in NPs from plants as shown by Henkel *et al.*^[10] for example, may partly be due to nitrogen being a growth-limiting factor for many plants. In contrast, the higher halogen content found in NPs originating from algae may reflect the higher availability of halogens in the marine environment. Some comparisons of NP properties to those of drugs may be partly biased because drugs are almost exclusively small molecules while some NP classes also incorporate much larger structures, *e.g.* macrocycles with different properties.^[18] In general, one should keep in mind that natural products are often viewed as a particular but more or less homogenous and related class of compounds like, for example, drugs or synthetic compounds. But in fact they consist of many diverse subclasses. The properties and scaffold architecture of NPs vary between these subclasses and depend on the organism of origin, the biosphere this organism lives in and the role of the NPs in nature, *e.g.* as metabolite, poison, neurotransmitter *etc.* Our knowledge about these subclasses also varies considerably. While, for example, plant NPs have been studied extensively for decades, exploration of marine NPs has been less frequent and less comprehensive. Average NP properties give some insight into their distinguished features and highlight some prominent differences to other classes of compounds, however, they are less well suited to describe and explore the diversity of NP chemical space.

2.1. Navigating Natural Product Chemical Space

NP chemical space encompasses the regions of chemical space which are occupied by natural products. Visualizations of NP chemical space facilitate analysis and exploration of the chemical diversity in Nature. The knowledge gleaned from such an endeavour certainly is interesting in its own right but can also serve as a source of inspiration for the design of new compound libraries. Several approaches to charting chemical space have been taken so far exploring different aspects of chemical space. Some visualize the chemical property space

based on the properties of the compounds populating it, and others are based on chemical structure, *i.e.* they explore chemical structure space. Both approaches employ different methodology and thus yield different but complementary results.

Feher and Schmidt depicted the chemical space of drugs, compounds originating from combinatorial chemistry and natural products based on their properties in a 2D scatter plot diagram using principal component analysis (PCA). PCA is a mathematical method which reduces an n -dimensional vector space to a space of smaller dimension, mostly two or three, while keeping the characteristics that contribute most to the variance of the data. The new base vectors are formed by linear combinations of the n base vectors of the previous vector space and each axis consists of the sum of fractions of the different properties initially calculated, the so-called 'loading'.

In order to obtain a more recent insight into differences between NPs and other molecules, we performed an analysis of representative sets of natural products, bioactive molecules and synthetic organic compounds using substructural features instead of properties. The natural products were taken from the Dictionary of Natural Products (DNP).^[19] An *in silico* deglycosylation using the Molinspiration toolkit^[20] yielded 113,664 unique aglycons from which 15,000 were selected to form the representative set. The 15,000 bioactive molecules were representatively selected from the World Drug Index^[21] and the MDDR database^[22] (together about 120,000 structures) and the same number of organic molecules was selected from a large internal database of commercially available compounds. As a descriptor, two-atom fragments were used which were present in at least 0.3% of the molecules. This yielded a matrix with 110 columns for the corresponding fragment frequencies. After normalization, a PCA was performed to arrive at three principal components which were then displayed in the 3D scatter plot diagram shown in Fig. 1.

The observation that combinatorial chemistry space is much smaller than natural product or drug space and more sharply defined matches Feher and Schmidt's observation. However, in their dataset there was almost no difference between drug and natural product space. In the diagram shown in Fig. 1 the two sets only share a small overlap and can otherwise be distinguished. This difference may be explained by the fact that Feher and Schmidt used simple calculated properties to characterize molecules, while in the analysis described above a large set of substructure features was used.

One obvious limitation of the PCA approach is its dependency on the dataset and

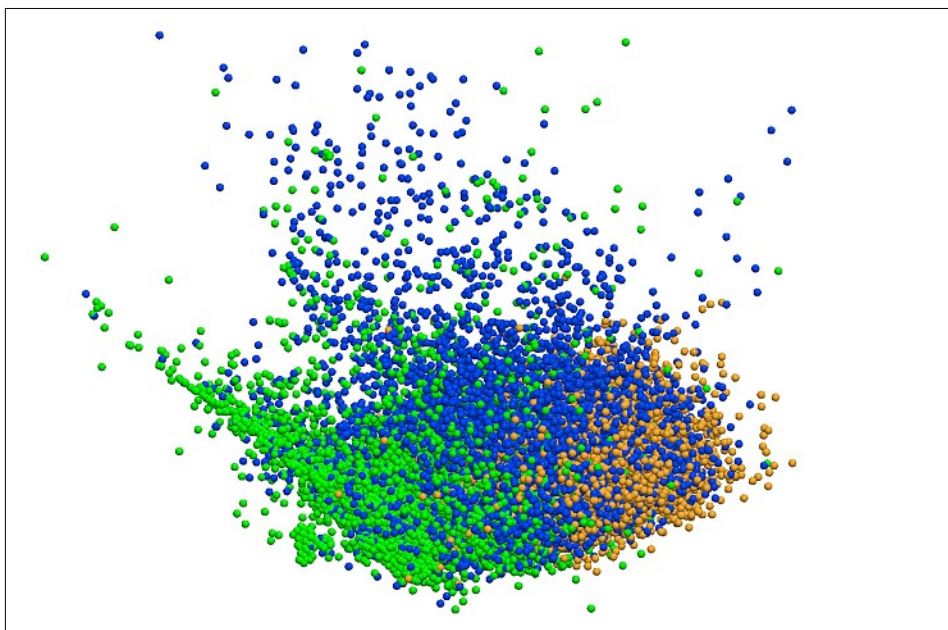


Fig. 1. Scatter plot of the combined chemical space of natural products (green), bioactive molecules (blue) and synthetic organic molecules (orange). Natural products and synthetic compounds occupy different parts of chemical space, and drugs are somewhere inbetween. Moreover, the space occupied by drugs and natural products is significantly larger and more diffuse than the space characteristic for the synthetic compounds.

the descriptors used. This is due to the PCA method which employs the characteristics contributing most to the variance of the dataset as the first principal component, those contributing second most as the second component *etc.*

The ChemGPS approach published by Oprea and Gottfries in 2001^[23] also used PCA and more than 60 descriptors. However, the authors proposed a method to overcome the limited comparability by applying a dedicated reference set of compounds to determinate the loading of the PCA. These reference compounds were chosen to have extreme properties which place them at the outer rims of the chemical space of interest. The positions of the compounds mapped can then be interpolated from the references. The overall approach resembles very much the Navstar global positioning system (GPS) where satellites in geo-stationary orbits (far from earth) are used as references for triangulation of a position on earth, hence the name 'ChemGPS'. Moreover, the loading of the three principal components was optimized in a way that each of the three axes can be translated into interpretable descriptors, *e.g.* size, hydrophobicity and flexibility. This is a considerable advantage compared to standard PCA since many of their principal components do not directly relate to interpretable chemistry. The maps of chemical space generated by this method resemble the 3D scatter plot in Fig. 1.

ChemGPS was applied to natural products by Larsson *et al.*^[24] using natural cyclooxygenase (COX) inhibitors as an example. The authors found that ChemGPS was able to discriminate clusters of com-

pounds with different activity, *e.g.* COX2 protein inhibition, COX2 mRNA inhibition and COX1 enzyme inhibition. They were also able to identify some properties which may be important with respect to these individual activities. However, some outliers were identified whose properties had to be extrapolated rather than interpolated from the reference set. The authors concluded that for charting natural product space a new training set with more NP-like properties would fine tune the approach towards ChemGPS-NP.^[25]

In conclusion, PCA-based approaches and, in particular, ChemGPS are well suited for the comparison of compound sets with respect to their properties or structural features captured by the descriptors used. The ability to rapidly process and display large numbers of compounds makes it valuable for the comparison of large libraries or databases. However, comparing the chemical space depicted by different PCA models is difficult, even more so if different sets of descriptors have been used. Additionally, conversion of positions in chemical space into chemistry and chemical structure is often difficult so that their use in synthesis planning or library design is limited.

Recently, Waldmann and co-workers developed a different approach to navigating chemical space which is based on chemical structure rather than on calculated properties.^[26] Their structural classification of natural products (SCONP) is based on a hierarchical classification of scaffolds which consist of rings systems and their aliphatic linkers but not other chains.^[27] Dissecting these scaffolds into smaller structures ac-

According to a set of rules yields a series of one-to-many parent-child relationships with the smaller scaffold being the parent of the larger one. The rules used to guide this iterative dissection process reflect synthetic and medicinal knowledge. Each parent-child assignment was guided by the following rules until only one possible parent scaffold was left:

- i) The parent scaffold has to be a substructure of the child scaffold.
- ii) The parent scaffold has to have fewer rings than the child scaffold.
- iii) Breaking of ring bonds is forbidden.
- iv) The parent with the highest number of heteroatoms is chosen.
- v) The largest parent scaffold is selected.
- vi) The more frequent parent scaffold, *i.e.* the one representing more NPs, is selected.

The relationships can be displayed in a tree-like diagram which is depicted in Fig. 2.

The scaffold tree diagram shown in Fig. 2 provides an overview over natural product structure space contained in the Dictionary of Natural Products (DNP). It depicts the more frequent scaffolds, *i.e.* those representing at least 0.2% of the dataset, and their relationships in a genealogy-like fashion. At the given threshold the carbocyclic section contains most scaffolds, followed by the O-heterocycles and the N-heterocycles. While in the carbocyclic section many scaffolds with three and four rings can be found, the O-heterocyclic section contains mainly one- and two-ring scaffolds. Among the N-heterocycles only some branches extend beyond the first ring. Koch *et al.*^[26] found the scaffolds with three rings to be most abundant across the whole dataset with the two- and four-ring scaffolds being within one standard deviation. Together, these three scaffold types account for more than 50% of all scaffolds found within the DNP. An analysis of the molecular volume distribution of the NPs incorporating a two-, three- or four-ring scaffold showed that it ranges from 100 to 500 Å³ with a maximum at 250 Å³. This is found to be comparable to molecular volumes of the drugs in the World Drug Index and well in the range of the size of protein cavities determined by Klebe and coworkers.^[28] The scaffold tree shown in Fig. 2 is exclusively built from scaffolds which are represented in NPs themselves. This leads to 'holes' in branches where the intermediate scaffolds do not occur in NPs present in the DNP. These missing links, however, may provide promising opportunities for synthetic molecules with interesting biological properties.

One should keep in mind that due to historical reasons, there is a bias for plant ingredients in the DNP although that might slowly change with new entries originating from other species. The scaffold hierarchy

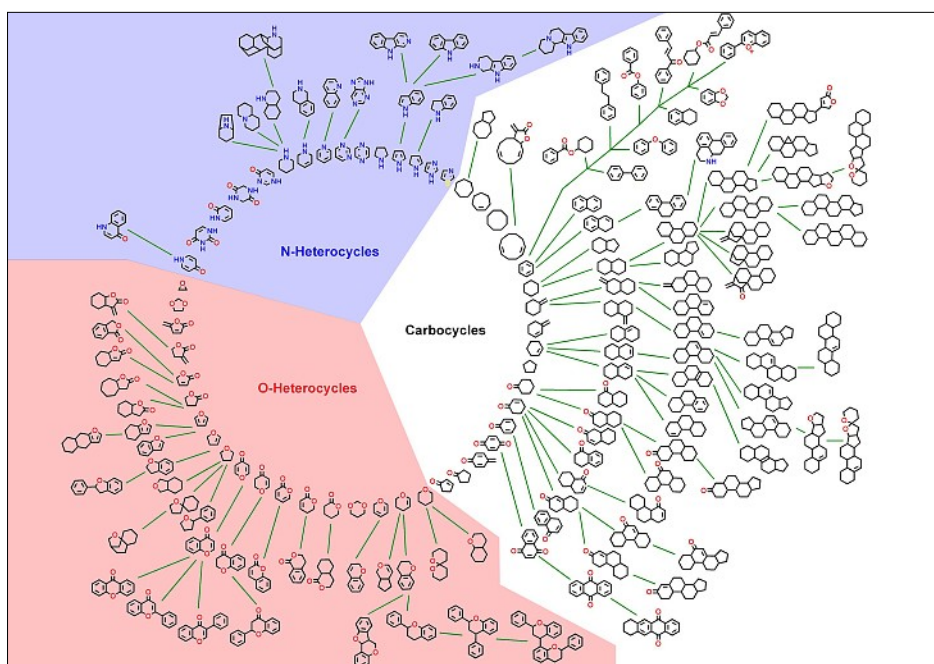


Fig. 2. Scaffold tree generated from the Dictionary of Natural Products 14.2. For clarity, only scaffolds are shown which are present in at least 0.2% of molecules in the dataset.

created from one molecule depends on the rest of the data set because only scaffolds present in at least one molecule are allowed. While one might argue that thus only 'true' natural product scaffolds enter the scaffold tree, comparisons with trees from different data sets are very difficult if not impossible.

A new rule set was recently introduced by Schuffenhauer *et al.*^[29] This set includes modifications to make the dissection of

scaffolds independent of the dataset, *i.e.* the parent generated from one scaffold will always be the same irrespective of the dataset used. This change in methodology may infer scaffold constructs which are not present in molecules in the dataset but provides the links otherwise missing, *e.g.* in the natural product scaffold tree. Thus, comparisons between datasets become possible, an apparent application of such a tool. The

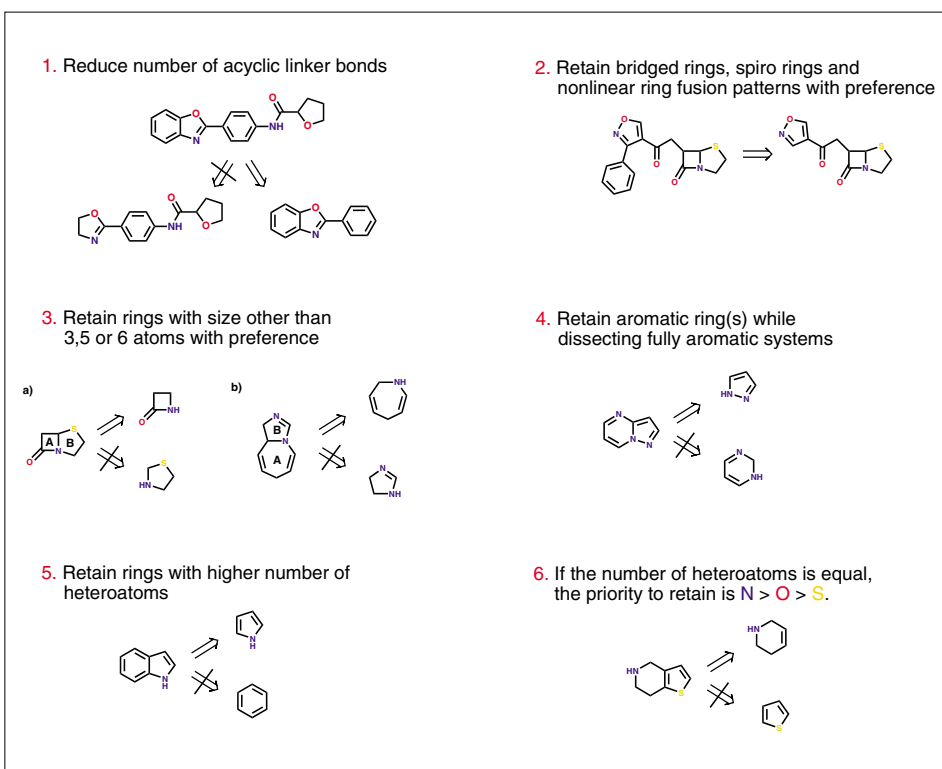


Fig. 3. Six of the thirteen rules guiding the parent-child assignment in the newly developed scaffold tree generating algorithm

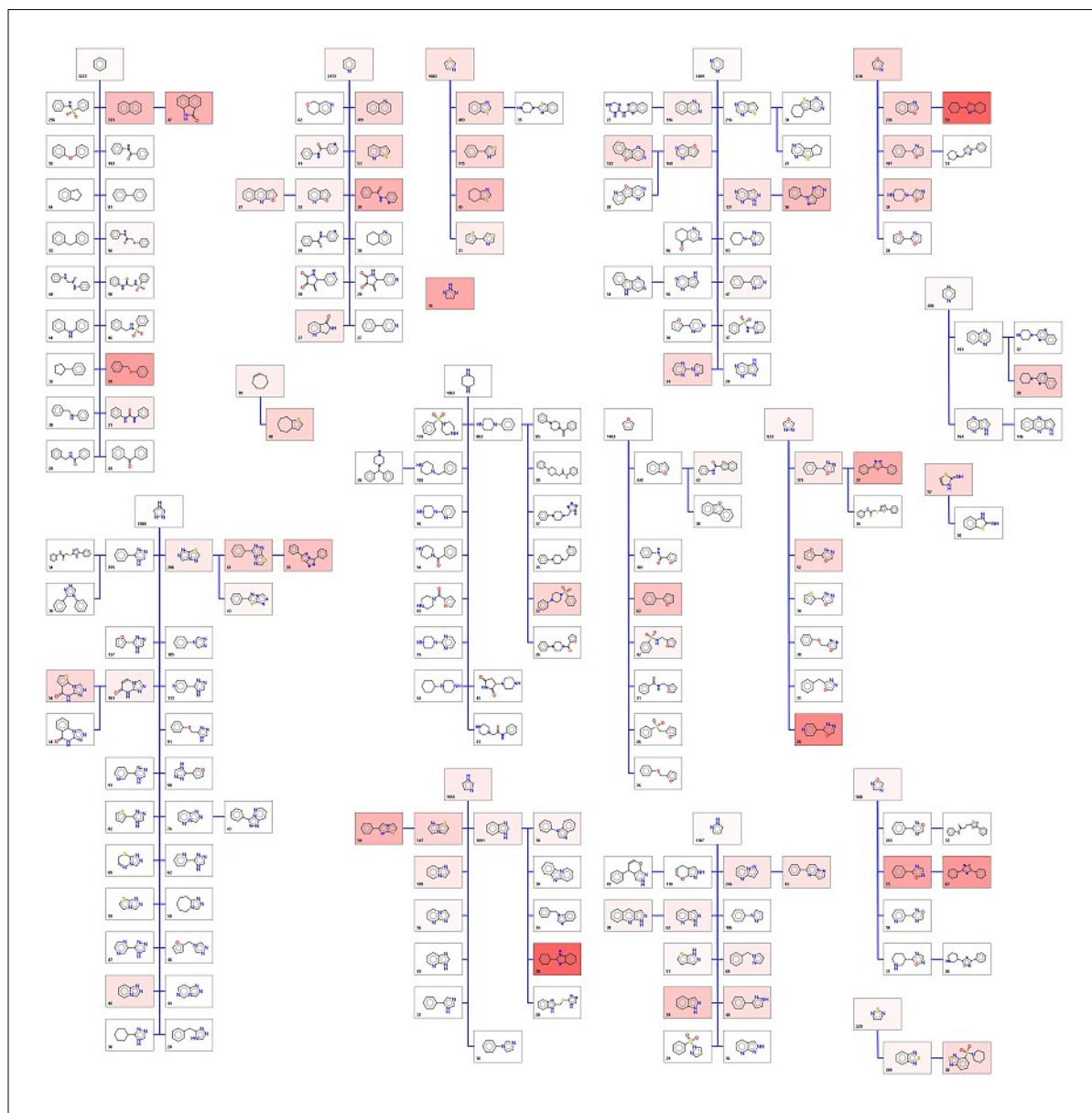


Fig. 4. Scaffold tree based on the pyruvate kinase assay dataset from PubChem containing 602 active compounds and 50000 inactive molecules. Only scaffolds are shown that represent at least 0.02% of the molecules in the dataset and for which at least 5% of the molecules were active on pyruvate kinase.

new rules themselves are based mostly on medicinal and synthetic chemistry knowledge and are more detailed than the rules used in the initial classification by Waldmann *et al.* They include, among others, rules which for example keep macrocycles intact, retain unusual structural motifs like spiro- or bridged compounds or conserve aromaticity. Six of the 13 rules in the new rule set are shown in Fig. 3.

This rule set was first applied in the analysis of the pyruvate kinase screen dataset^[30] available in PubChem.^[31] The resulting scaffold tree is shown in Fig. 4. Color intensity reflects the fraction of active molecules represented by the individual scaffold.

Scaffolds representing a high fraction of bioactive molecules can easily be identified from Fig. 3. Interestingly, in some cases also the next smaller scaffold contains a similar fraction of active molecules. This gives a quick overview over the performance of a library and allows rapid identification of starting points for closer examination and

data mining, *e.g.* structure–activity relationship (SAR) analyses. Schuffenhauer *et al.* also gave a first example for the use of the scaffold tree to find the common core for SAR analysis as implemented for example in Pipeline Pilot. The use of chemical structure, which is the universal language of chemists, renders the results from the analysis directly and intuitively accessible to chemists.

3. Conclusions

While many natural products adhere to the Rule-of-Five, they nonetheless differ from other compound sets, especially drugs, in many other properties, *e.g.* heteroatom composition, number of rings or stereogenic centers. In general, NPs are found to contain more oxygen and fewer nitrogen atoms than drugs but are more lipophilic. They contain on average more stereogenic centers and their degree of unsaturation is higher than in drugs and they incorporate

less aromatic rings. Although NPs in general contain more rings than drugs, most of them are non-aromatic and part of single fused ring system. Since natural products consist of several subclasses with diverse properties, their properties depend strongly on the chosen data set. The databases are subject to change over time and thus the properties regarded as NP-like may change as well. Notably, many seminatural compounds, mainly natural product derived analogues, were found to be much closer to drugs than to pure natural products in many of their properties.

Several approaches towards charting chemical space in general and natural product space in particular have been presented. ChemGPS and the SCONP scaffold tree differ substantially in their methodology and results, but can both be applied to a variety of problems and may often give complementary results. ChemGPS is favorable for the assessment and comparison of large datasets according to their physico-chemi-

cal properties and is amenable to very large numbers of compounds. Some chemical parameters describing the molecules of interest can also be derived. In contrast, the scaffold tree may be of better use in library design. SCONP delineated the natural product scaffold space of the most frequent scaffolds. It can easily be seen that carbocycles are more abundant than O-heterocycles and N-heterocycles. Among the carbocycles also more complex scaffolds consisting of three and four rings can be identified while in the heterocyclic compounds these seem to be less frequent. Some insight can be gained about the frequency of individual scaffolds in NP chemical space which may be interesting for library design. The structural basis establishes an immediate link to chemistry and makes the approach intuitively accessible to chemists. The pyruvate kinase tree example demonstrates the value of the scaffold tree as a data mining tool linking information to structures. The natural product scaffold tree not only allows exploration of Nature's diversity but also indicates opportunities for synthetic organic chemistry by identifying holes in the branches which Nature has probably not filled.

The question remains how to transform the information gleaned from the exploration of chemical space into synthetic chemistry and, last but not least, compounds. One possible approach to utilize the knowledge extracted from chemical space analysis has recently been presented by Waldmann and co-workers.^[32] Utilizing the SCONP approach, the authors explored the biologically relevant and prevalidated structural space populated by nature so far. Scaffolds identified from the SCONP scaffold tree shown in Fig. 2 were used as templates for the design of NP-derived compound collections. Four new classes of phosphatase inhibitors could be discovered from these compound collections by biochemical screening. Waldmann and co-workers created the term 'biology-inspired synthesis' (BIOS) for this approach, *i.e.* synthetic efforts of molecules enriched in biochemical and biological activity by exploration of the chemical space relevant to nature.

Together with other tools available for the design of natural product based libraries^[33] and modern synthetic organic chemistry methods the exploration of larger parts of Nature's chemical space seems possible at reasonable effort.

Acknowledgements

SW is grateful for a Novartis graduate scholarship.

Received: April 26, 2007

- [1] D. J. Newman, G. M. Cragg, K. M. Snader, *J. Nat. Prod.* **2003**, *66*, 1022.
- [2] J.-Y. Ortholand, A. Ganesan, *Curr. Opin. Chem. Biol.* **2004**, *8*, 271.
- [3] I. Paterson, A. Anderson Edward, *Science* **2005**, *310*, 451.
- [4] B. E. Evans, K. E. Rittle, M. G. Bock, R. M. DiPardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson *et al.*, *J. Med. Chem.* **1988**, *31*, 2235.
- [5] G. Mueller, *Drug Discovery Today* **2003**, *8*, 681.
- [6] C. Rosenbaum, P. Baumhof, R. Mazitschek, O. Muller, A. Giannis, H. Waldmann, *Angew. Chem., Int. Ed.* **2004**, *43*, 224.
- [7] C. Rosenbaum, S. Roehrs, O. Mueller, H. Waldmann, *J. Med. Chem.* **2005**, *48*, 1179.
- [8] H. van de Waterbeemd, E. Gifford, *Nat. Rev. Drug Discovery* **2003**, *2*, 192.
- [9] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **2001**, *46*, 3.
- [10] T. Henkel, R. M. Brunne, H. Muller, F. Reichel, *Angew. Chem., Int. Ed.* **1999**, *38*, 643.
- [11] M.-L. Lee, G. Schneider, *J. Comb. Chem.* **2001**, *3*, 284.
- [12] M. Feher, J. M. Schmidt, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218.
- [13] P. Ertl, A. Schuffenhauer, in 'Natural Products as Drugs', Eds. F. Petersen, R. Amstutz, Birkhaeuser Verlag, Basel, **2007**.
- [14] K. Grabowski, G. Schneider, *Curr. Chem. Biol.* **2007**, *1*, 115.
- [15] 'CRC Dictionary of Natural Products', CRC Press, 78318 structural entries, June **1996**, <http://www.crcpress.com/>.
- [16] 'Bioactive Natural Product Database', Szenzor Management Consulting Company, Budapest, Hungary, Berdy, 29432 entries of natural products with described biological activity, status: July **1996**.
- [17] 'ACD: Available Chemicals Directory', Version 93.2, Molecular Design Ltd. Information Systems Inc., San Leandro, CA, USA, 182822 entries.
- [18] L. A. Wessjohann, E. Ruijter, D. Garcia-Rivera, W. Brandt, *Mol. Diversity* **2005**, *9*, 171.
- [19] 'CRC Dictionary of Natural Products', v15.1, CRC Press, **2006**, <http://www.crcpress.com/>.
- [20] Molinspiration Cheminformatics, <http://www.molinspiration.com>.
- [21] 'WDI, Derwent World Drug Index', <http://www.derwent.com/products/lr/wdi/>.
- [22] 'MDL Drug Data Report', <http://www.prous.com/product/electron/mddr.html>.
- [23] T. I. Oprea, J. Gottfries, *J. Comb. Chem.* **2001**, *3*, 157.
- [24] J. Larsson, J. Gottfries, L. Bohlin, A. Backlund, *J. Nat. Prod.* **2005**, *68*, 985.
- [25] J. Larsson, J. Gottfries, S. Muresan, A. Backlund, *J. Nat. Prod.*, ACS ASAP.
- [26] M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17272.
- [27] G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887.
- [28] S. Schmitt, D. Kuhn, G. Klebe, *J. Mol. Biol.* **2002**, *323*, 387.
- [29] A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. A. Koch, H. Waldmann, *J. Chem. Inf. Model.* **2007**, *47*, 47.
- [30] J. Inglese, D. S. Auld, A. Jadhav, R. L. Johnson, A. Simeonov, A. Yasgar, W. Zheng, C. P. Austin, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11473.
- [31] Pyruvate kinase assay in PubChem, <http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=361>.
- [32] A. Noeren-Mueller, I. Reis-Correa, Jr., H. Prinz, C. Rosenbaum, K. Saxena, H. J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10606.
- [33] L. O. Haustedt, C. Mang, K. Siems, H. Schiewe, *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 445.