Natural Product-likeness Score and Its Application for Prioritization of Compound Libraries

Peter Ertl,* Silvio Roggo, and Ansgar Schuffenhauer

Novartis Institutes for BioMedical Research, CH-4002 Basel, Switzerland

Received August 3, 2007

Natural products (NPs) have been optimized in a very long natural selection process for optimal interactions with biological macromolecules. NPs are therefore an excellent source of validated substructures for the design of novel bioactive molecules. Various cheminformatics techniques can provide useful help in analyzing NPs, and the results of such studies may be used with advantage in the drug discovery process. In the present study we describe a method to calculate the natural product-likeness score—a Bayesian measure which allows for the determination of how molecules are similar to the structural space covered by natural products. This score is shown to efficiently separate NPs from synthetic molecules in a cross-validation experiment. Possible applications of the NP-likeness score are discussed and illustrated on several examples including virtual screening, prioritization of compound libraries toward NP-likeness, and design of building blocks for the synthesis of NP-like libraries.

INTRODUCTION

Natural products (NPs) are chemical entities produced by living organisms. Of special interest for drug discovery is the class of NPs defined as secondary metabolites, i.e., metabolites which are not directly necessary for host survival. They are typically produced by organisms such as bacteria, plants, or various marine invertebrates and are usually used as "chemical warfare" to protect parent organisms from predators or, on the other hand, used as a means of attack. To efficiently fulfill this role the NPs have been optimized in a very long natural selection process for optimal interactions with biological macromolecules. NPs are therefore an excellent source of validated substructures for the design of new drugs.[1] Indeed, many drugs in the current pharmacopeias are NPs, and many others are of NP origin.[2] In the pharmaceutical industry we can witness presently a real explosion of interest in NPs.[3] After several years of deceleration, caused by various reasons, most notably exaggerated expectations in the novel drug discovery technologies, the NPs are again the center of attention of the pharmaceutical industry as a promising and reliable source of new bioactive molecules. Several startups focusing entirely on NP-based drug discovery have emerged,[4] and traditional pharmaceutical companies are increasing their investments in the natural product-based drug discovery.

Structures of NPs have become also a new and welcome source of inspiration for the design of combinatorial libraries. It is a well-known fact that the first generation of combinatorial libraries, containing mostly large, hydrophobic molecules with many rotatable bonds, was rather a disappointment concerning their biological activity. But these negative results also had a positive effect. Chemists learned that not only the amount of molecules synthesized is important but also their properties. This led to the re-evaluation of combichem design strategies and the introduction of a concept of diversity oriented synthesis (DOS)[5,6] focused on the replacement of "classical" flat aromatic heterocyclic chemistry by small molecules with high skeletal- and stereodiversity covering broad areas of structural space. These are synthesized using highly branched synthesis pathways in analogy to the highly branched biosynthesis pathways of natural products.[7] NPs with their high diversity are indeed very well suited as a source of bioactive substructures for the design of new types of combinatorial libraries.[8,9] In addition to diversity, however, NPs also have another advantage. They contain numerous bioactive substructures validated by nature's long evolution. A biology-oriented synthesis (BIOS) based on this fact has been introduced recently.[10] BIOS builds on a diversity created by nature and aims at its local extension in areas of proven biological relevance using as a starting point simplified core structures of NPs.

In order to introduce the NP-like features into the design of novel libraries the properties which are typical for NPs need to be known. Several studies focused therefore on the analysis of NPs from the cheminformatics point of view. Henkel et al.[11] was probably the first to analyze differences in molecular properties and structural features between NPs and synthetic molecules and found distinct differences (such as the number of bridgehead atoms or the frequencies of various functional groups). Stahura et al.[12] identified a set of descriptors which were able to distinguish NPs from synthetic molecules based on their Shannon entropy. Schneider with collaborators[13,14] analyzed a set of NPs to identify whether they contain novel scaffold architectures for potential use in combinatorial chemistry. Several such scaffolds have been identified which were not present in marketed drugs. Feher and Schmidt[15] compared the distribution of various molecular properties among NPs, drugs, and molecules originating from combinatorial chemistry, identifying the number of chiral centers, the presence of aromatic rings, the

* Corresponding author phone: +41 61 3240685; fax: +41 61 3243357; e-mail: peter.ertl@novartis.com.

10.1021/ci700286x CCC: $40.75
Published on Web 11/23/2007

degree of saturation, and the number of various heteroatoms as being the most important. Koch et al.[16] analyzed a large database of NPs to study NP scaffolds by arranging them in the form of a tree and used this information to navigate within the scaffold universe to identify its interesting regions. Analysis of property distribution of more than 130 000 NP structures as well as identification of substructures typical for particular classes of source organisms have been reported by Ertl and Schuffenhauer.[17] And finally Wetzel et al.[18] discussed recently various approaches to analyze and chart the chemical space covered by NPs.

Most authors agree that although the NPs differ also in their global physicochemical properties (such as logP, polar surface area, etc.) from synthetic molecules, the major differences between these two classes of molecules are in their structural characteristics, such as the number of aromatic rings, stereocenters, and distribution of nitrogen and oxygen atoms.

In the present study we perform a more detailed chem-informatics analysis of structural features of a large collection of NPs, focusing on identification of those substructures which distinguish NPs from common synthetic molecules. Based on this analysis a score has been developed which may be used to assess the natural product-likeness of individual molecules and whole compound libraries.

DEVELOPMENT OF NATURAL PRODUCT-LIKENESS SCORE

With respect to the importance of NPs in the drug discovery process discussed in the previous section, it would be advantageous to have the possibility of comparing the characteristics of studied molecules with those of NPs. A similar measure, called drug-likeness,[19,20] is used routinely to assess the similarity of screened molecules to the known drugs. Although there is no formal definition of drug-likeness (different drug-likeness scores use a broad range of approaches to calculate it), this characteristic is generally understood as a measure of how close is the molecule under study to the area of chemical universe occupied by the common drugs. Recently also a method to calculate metabolite-likeness has been introduced[21] to characterize the chemical space covered by compounds involved in metabolic reactions and also a method to calculate peptide-likeness as a measure of "peptide flavor"[22] which may be used by peptidomimetic bioisosteric design. Analogically to these descriptors, we define natural product-likeness simply as a measure of similarity to the NP molecules.

Preparation of the Data. The largest commercially available database of NP structures is the CRC Dictionary of Natural Products (DNP).[23] We used this database as a source of reference NP molecules. Before actual substructure analysis the molecules have been standardized by normalizing charges and by removing small disconnected fragments (counterions, etc.). Structures having less than 6 atoms or containing metals have also been removed. In the next step all molecules have been deglycosylated (i.e., all sugar substituents have been removed). The main role of sugar moieties in NPs is to affect pharmacokinetic properties of parent structures and make them more soluble.[24] In many cases sugar units do not affect the biological activity of aglycon directly, although several notable exceptions to this general rule exist. The presence of various sugar units is therefore the most typical structural characteristic of NP molecules. And because we did not want this feature to surpass other more interesting structural elements of NPs, particularly the structural characteristics of central scaffolds, the sugar units have been removed before the actual substructure analysis. The deglycosylation step preceding the actual substructure processing parallels the strategy from our previous study of NP scaffolds.[16] For the removal of sugar units we used a recursive deglycosylation procedure written in Java. In this procedure sugar rings at the periphery of the molecule have been identified and removed including also the attached nonring substituents. The procedure was repeated until no such sugar rings could be identified. In this step 1 to over 80 sugar units were removed from 21 670 molecules. The NP database after deglycosylation contained 115 590 unique aglycons.

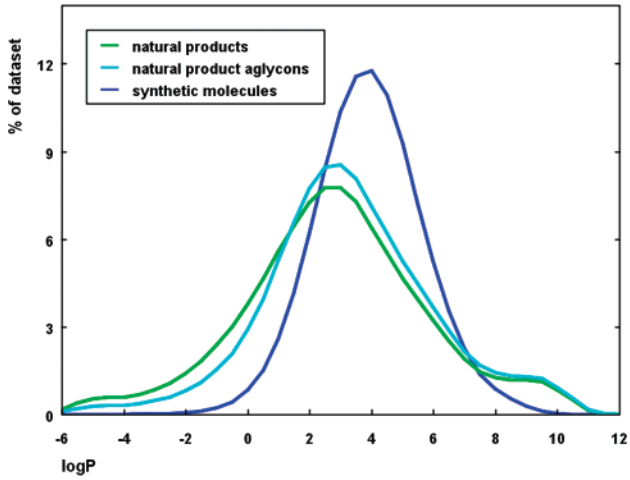

Figure 1. Distribution of calculated logP for natural products and synthetic organic molecules.

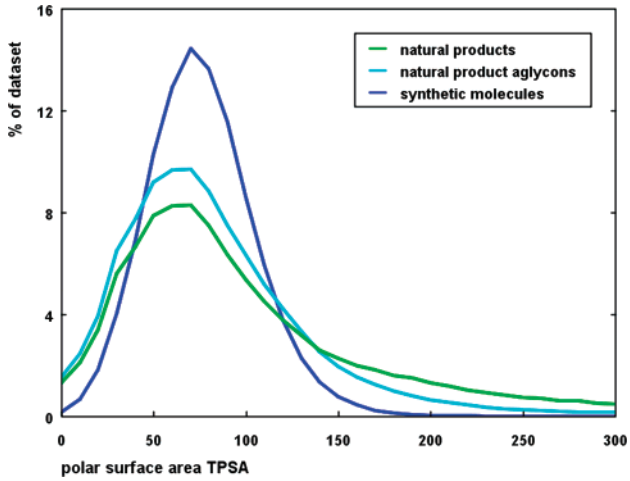

Figure 2. Distribution of polar surface area for natural products and synthetic organic molecules.

The characteristics of NPs have been compared with those of synthetic molecules (SMs). For this purpose we selected 290 000 structures from the in-house collection of commercially available synthetic compounds by representative selection. These molecules represented in our comparative analysis the currently available "synthetic organic chemistry" space.

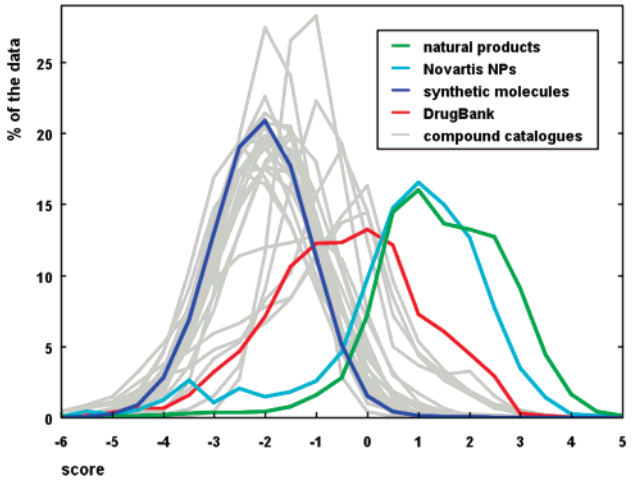

Figure 3. Distribution of the NP-likeness score for various molecular collections.

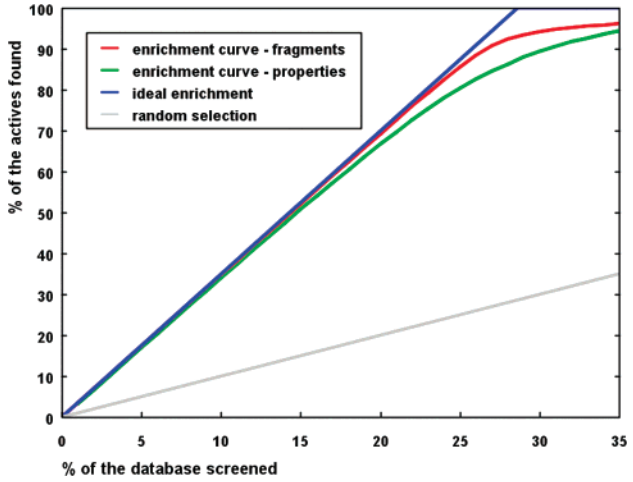

Figure 4. Enrichment plot obtained as an average of 5 cross-validation runs: red − enrichment when using atom-centered HOSE fragments and green − enrichment when using calculated properties and simple substructure features.

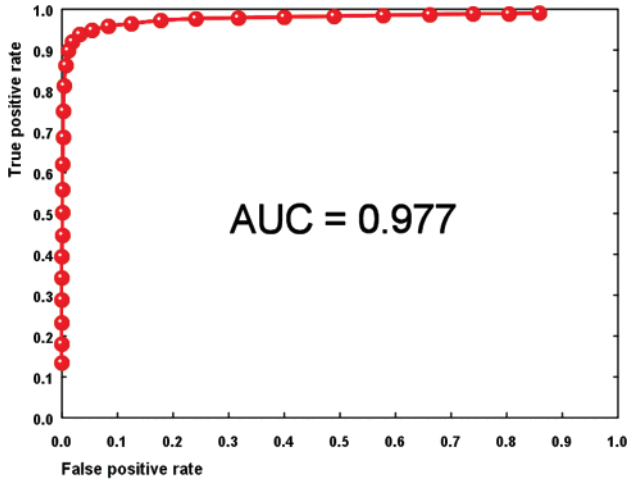

Figure 5. ROC curve (for atom-centered fragments) obtained as an average of 5 cross-validation runs. The area under the curve is 0.977.

The cheminformatics analysis and molecular processing including molecule cleaning, normalization, calculation of various molecular properties, and substructure analysis was performed using the PipelinePilot[25] and Molinspiration[26] software. Additionally, several specialized modules (for example recursive deglycosylation procedure or custom fragmentation) have been written in-house in Java.

Development of the NP Score. As discussed already in the Introduction, calculated global molecular properties differ between NPs and SMs. As an example distribution of calculated octanol−water partition coefficient (logP) and topological polar surface area (TPSA)[27] for NPs, deglycosylated NPs, and SMs are shown in Figures 1 and 2. One can see that NPs are generally more hydrophilic than SMs, while the TPSA has a similar mean for the both sets but a broader distribution for NPs. For structure-based characteristics such as the number of aromatic atoms, the number of stereocenters, or the number of oxygen and nitrogen atoms in the molecule the differences are even more pronounced.[17]

For the development of the NP-likeness score, however, we decided to use more complex structural features. One can expect better separation between NPs and SMs by using more specific substructures than relatively simple molecular properties, and, what is even more important, the knowledge about substructure features which are typical for NPs resulting from this analysis may be used directly in the design of novel NP-like molecules.

To characterize molecule structural features we used atom centered fragments introduced by Bremser in 1978 as HOSE codes to estimate molecule spectra.[28] Under various names (for example, atom environments, extended atoms, circular substructures, or atom-centered fragments) this type of substructure descriptors has been shown to be very useful also in other areas of cheminformatics, including similarity searching, estimation of molecular properties, or development of models for bioactivity prediction.[29−32] In addition to atom centered fragments with two levels of neighbors we used in this analysis also pairs of these fragments including information about the number of bonds (up to 6) separating them.

Once a set of fragments for NPs and SMs is generated, one has to use an appropriate measure to compare distribution of fragments between these two sets. Willett et al. compared various fragment weighting schemes for substructure analysis.[33] We used the score defined by eq 1, because it provided the best results for the calculation of substituent drug-likeness in our earlier study.[34]

$$f_i = \log(\mathrm{nact}_i/\mathrm{ninact}_i * \mathrm{ninact}_{\mathrm{total}}/\mathrm{nact}_{\mathrm{total}}) \qquad (1)$$

In the equation nact_i is the number of NPs which contain fragment i, ninact_i is the number of SMs which contain fragment i, $\mathrm{nact}_{\mathrm{total}}$ is the total number of NPs, and $\mathrm{ninact}_{\mathrm{total}}$ is the total number of SMs in the training set. The NP-likeness of the whole molecule is then simply calculated as a sum of contributions of fragments f_i in the molecule, normalized relative to the molecule size. In principle, this is an application of a naïve Bayesian statistics, because contributions of fragments are considered to be independent of each other. The calculated score is typically in the range from −5 to 5. The higher the score is, the higher the probability is that the molecule is a NP. Distribution of the score for the training sets of NPs and SMs as well as some other data sets (see below) is shown in Figure 3.

Besides the naïve Bayesian classifier a broad range of techniques is available to separate two classes of objects.[35] Popular in cheminformatics are, for example, support vector

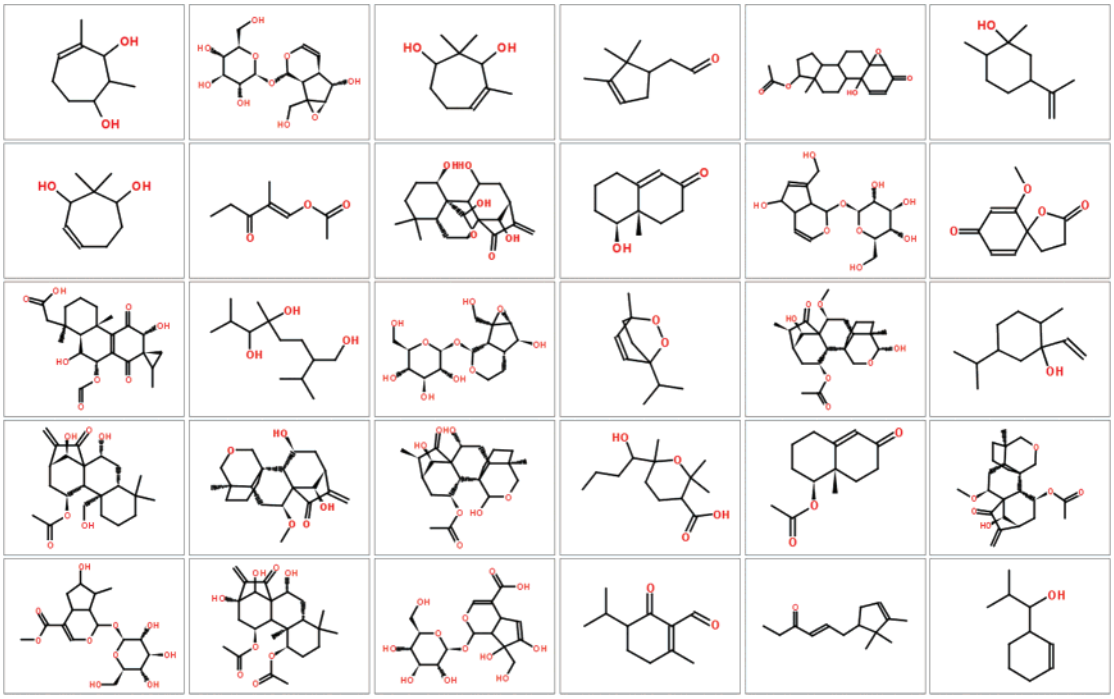

Figure 6. Example of structures from the MDPI collection[38] with high calculated NP-likeness.

machines, neural networks, decision trees, or various clustering techniques. The major advantages of the approach we used is that the method is not parametric, and, therefore, it is not sensitive to overfitting as most other machine-learning approaches are. Additionally, the Bayesian classifier can directly identify particular substructure features responsible for NP-likeness.

Validation of the Score. Before used in actual applications, the new NP-likeness score has to be validated. We performed two types of validation experiments.

In the classical cross-validation study the data were randomly divided into two halves—training and test sets, respectively. The training set was used for development of a classification model, and then the performance of the model was evaluated by calculating scores for the molecules in the test set and comparing them with the actual molecule class (NP or SM). The resulted enrichment plot obtained as an average of five cross-validation runs is shown in Figure 4. For comparison we generated also a model by using calculated molecular properties and simple structure characteristics. We used those properties which have been shown in previous studies to differ significantly between NPs and SMs—logP, PSA, total number of non-hydrogen atoms, number of oxygens and nitrogens, number of aromatic atoms, number of potential stereocenters, and number of rotatable bonds. The standard PipelinePilot Bayesian module was applied to do the classification. The cross-validation enrichment using this "simple properties" model is also shown in Figure 4 and exhibits only slightly worse performance than classification by HOSE fragments. A second statistical measure we used to characterize the quality of our NP-likeness model was a receiver operating characteristics (ROC) curve. This curve is shown in Figure 5. The area under the ROC curve (AOC) is 0.977. This number is the probability that when an active and an inactive molecule are selected randomly, the active molecule will have a higher score than the inactive one. Both graphs document excellent predictivity of the NP-likeness model based on HOSE fragments. The enrichment in cross-validation mode shown in Figure 4 is for the first 20% of data practically identical with the ideal enrichment curve.

In the second validation experiment we selected those NP structures from the in-house Novartis collection which were not present in the Dictionary of Natural Products and calculated the NP-likeness for these molecules. Distribution of the resulting score is shown in Figure 3. This was a more stringent test, because the Novartis NP collection contains also the number of novel structural classes which are not present in the DNP. Despite this, the method correctly identified 93.9% of the structures as NPs by using the optimal cutoff suggested by the ROC curve.

POSSIBLE APPLICATIONS OF NP-LIKENESS SCORE

An apparent application of NP-likeness score is its use in virtual screening. Pharmaceutical companies are purchasing regularly large number of samples to be screened in their high-throughput screening factories. In addition to standard criteria such as druglike properties, novelty, or no undesirable substructures, the NP-likeness score may be used as a useful prioritization factor to identify samples which should be purchased and screened.

To evaluate the distribution of the NP-likeness score in various compound collections, libraries from 24 commercial compound providers have been downloaded from the ZINC Web site.[36] Additionally we included a set of marketed drugs from the DrugBank.[37] The distribution of NP-likeness for all these collections is shown in Figure 3. While most of the libraries contain typical synthetic molecules, some collections contain also a portion of molecules with high NP-likeness. As expected, the NP-likeness of common drugs from the DrugBank is somewhere in the middle between NPs and

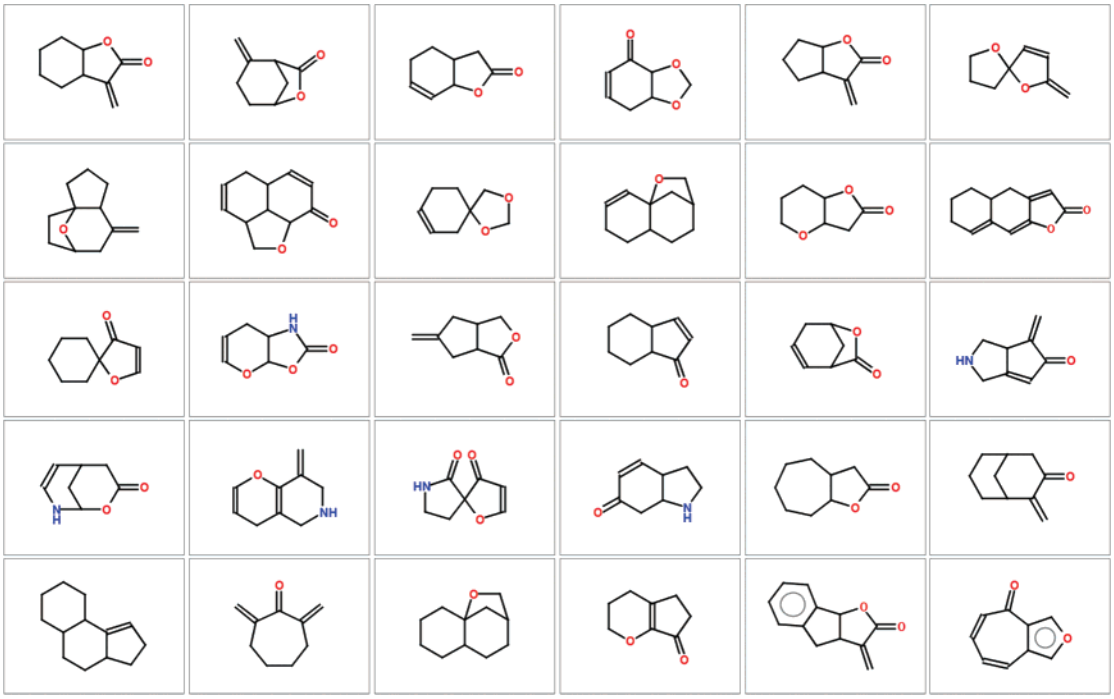

Figure 7. Example of common scaffolds from the PubChem[39] database having a high NP-likeness score.

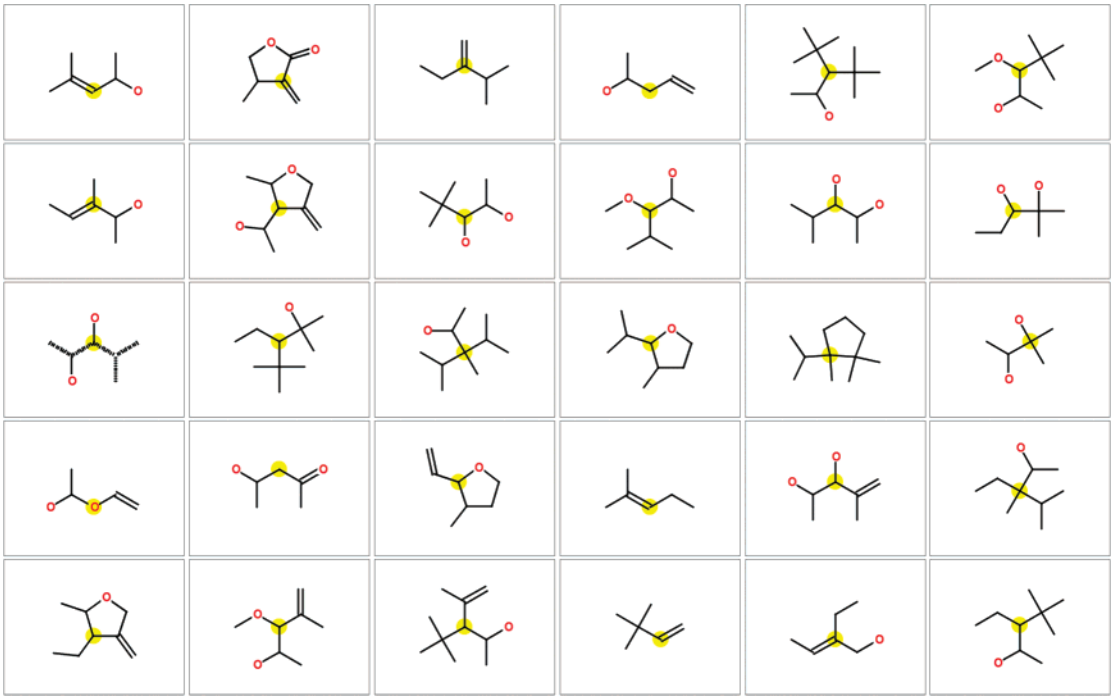

Figure 8. Examples of substructures with high NP-likeness. The yellow disc represents the central atom of the fragment.

SMs. Out of the commercial libraries studied, the MDPI compound collection[38] contained the largest portion of NP-like molecules. MDPI is a very diverse library containing samples collected from different academic sources, including also a number of plant metabolites. The calculated NP-likeness score can efficiently identify this type of molecule. Examples of molecules from the MDPI database with the highest NP-likeness score are shown in Figure 6.

We would like to point out here that the NP-likeness score alone cannot be used as a criterion for the quality of a library, neither it is possible to conclude from it anything about the probability of bioactivity on a specific target of interest. The calculated score is neither a measure of molecular diversity (which can never be the property of an individual molecule, but is always related to an ensemble of molecules). The NP-likeness score is nothing more and nothing less than its names tells us—it is a measure of an overall similarity with the currently known NP structural space.

In the second application example we wanted to demonstrate the applicability of the NP-likeness score for selection of substructures to support combinatorial synthesis. A set of common scaffolds (present in more than 20 molecules) was extracted from the PubChem database.[39] A scaffold is defined here as a single ring or an assembly of fused, bridged,

or spiro rings. For these scaffolds the NP-likeness score was calculated, and some high scoring examples are shown in Figure 7. Of course, in the prospective application not common scaffolds would be scored (which are probably all IP covered), but structures from a proprietary compound database or set of virtual scaffolds generated in silico[40] and then the best-scoring scaffolds would be purchased or synthesized and used as a basis for production of novel, NP-like combinatorial libraries.

Numerous other applications of the NP-likeness score in the drug discovery process are possible. One can think, for example, about a procedure for automatic evolutionary design of molecules optimizing at the same time multiple properties including bioavailability, ease of synthesis, novelty, and, of course, NP-likeness. A list of substructure fragments with the highest NP-likeness score (some examples are shown in Figure 8) may be used by medicinal chemists directly as an "idea generator" helping them to design novel NP-like molecules.

CONCLUSIONS

The NP-likeness score described here is a useful measure which can help to guide the design of new molecules toward interesting regions of chemical space which have been identified as "bioactive regions" by natural evolution. The calculation of the NP-likeness score is simple; once a model is available the calculation consists only of molecule fragmentation, table lookup, and summation of fragment contributions, so millions of molecules may be processed easily. The calculation of NP-likeness is implemented at Novartis as a Web service and is incorporated into several standard processes including virtual screening, selection of compound samples for purchasing, HTS hitlist triaging, and library design.

REFERENCES AND NOTES

(1) Haustedt, L. O.; Mang, C.; Siems, K.; Schiewe, H. Rational approaches to natural-product-based drug design. *Curr. Opin. Drug Discovery. Dev.* **2006**, *9*, 445−462.

(2) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.* **2007**, *70*, 461−477.

(3) Rouhi, A. M. Rediscovering natural products. *Chem. Eng. News* **2003**, *81*, 77−91.

(4) Rouhi, A. M. Betting on natural products for cures. *Chem. Eng. News* **2003**, *81*, 93−103.

(5) Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **2000**, *287*, 1964−1969.

(6) Tan, D. S. Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nat. Chem. Biol.* **2005**, *1*, 74−84.

(7) Firn, R. D.; Jones, C. G. Natural products − a simple model to explain chemical diversity. *Nat. Prod. Rep.* **2003**, *20*, 382−391.

(8) Kingston. D.; Newman, D. Mother nature's combinatorial libraries; their influence on the synthesis of drugs. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 304−316.

(9) Breinbauer, R.; Manger, M.; Scheck, M.; Waldmann, H. Natural Product Guided Compound Library Development. *Curr. Med. Chem.* **2002**, *9*, 2129−2145.

(10) Nören-Müller, A.; Reis-Corrêa, I.; Prinz, H.; Rosenbaum, C.; Saxena, K.; Schwalbe, H. J.; Vestweber, D.; Cagna, G.; Schunk, S.; Schwarz, O.; Schiewe, H.; Waldmann, H. Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10606−10611.

(11) Henkel, T.; Brunne, R. M.; Müller, H.; Reichel, F. Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 643−647.

(12) Stahura, F. L.; Godden, J. W.; Xue, L.; Bajorath, J. Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1245−1252.

(13) Lee, M.-L.; Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* **2001**, *3*, 284−289.

(14) Grabowski, K.; Schneider, G. Properties and Architecture of Drugs and Natural Products Revisited. *Curr. Chem. Biol.* **2007**, *1*, 115−127.

(15) Feher, M.; Schmidt, J. M. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218−227.

(16) Koch, M.; Schuffenhauer, A.; Scheck, M.; Wetzel, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. Charting biologically relevant chemical space: a structural classification of natural products (SCONP). *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17272−17277.

(17) Ertl, P.; Schuffenhauer, A. Cheminformatics Analysis of Natural Products: Lessons from Nature Inspiring the Design of New Drugs. In *Natural Compounds as Drugs Vol 2*; Petersen. F., Amstutz, R., Eds.; Birkhäuser Verlag: Basel, Switzerland, will be published in Spring 2008.

(18) Wetzel, S.; Schuffenhauer, A.; Roggo, S.; Ertl, P.; Waldmann, H. Cheminformatics Analysis of Natural Products and Their Chemical Space. *Chimia* **2007**, *61*, 355−360.

(19) Clark, D. E.; Pickett, S. E. Computational methods for the prediction of 'drug-likeness'. *Drug Discovery Today* **2000**, *5*, 49−58.

(20) Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432*, 855−861.

(21) Gupta, S.; Aires-de-Sousa, J. Comparing the Chemical Spaces of Metabolites and Available Chemicals: Models of Metabolite-likeness. *Mol. Diversity* **2007**, *11*, 23−36.

(22) Eckert, H.; Bajorath, J. Exploring Peptide-likeness of Active Molecules Using 2D Fingerprint Methods. *J. Chem. Inf. Model.* **2007**, *47*, 1366−1378.

(23) *CRC Dictionary of Natural Products, v15.2*; CRC Press: 2006. http://www.crcpress.com/ (accessed May 2007).

(24) Thorson, J. S.; Vogt, T. Glycosylated natural products. In *Carbohydrate-Based Drug Discovery*; Wong, C. H., Ed.; Wiley-VCH Verlag: Weinheim, Germany, 2005; pp 685−711.

(25) *Pipeline Pilot version 6.0*; Scitegic Inc.: San Diego, CA, 2007. http://www.scitegic.com (accessed May 2007).

(26) *Molinspiration Cheminformatics mib package, version 2007.03*; Molinspiration Cheminformatics: Slovensky Grob, Slovak Republic, 2007. http://www.molinspiration.com (accessed May 2007).

(27) Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.

(28) Bremser, W. HOSE - A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355−365.

(29) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 170−178.

(30) Hert, J.; Willett, P.; Wilton D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* **2004**, *2*, 3256−3266.

(31) Japertas, P.; Didziapetris, R.; Petrauskas, A. Fragmental methods in the analysis of biological activities of diverse compound sets. *Mini Rev. Med. Chem.* **2003**, *8*, 797−808.

(32) Rogers, D.; Brown, R. D.; Hahn, M. Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. *J. Biomol. Screen.* **2005**, *10*, 682−686.

(33) Ormerod, A.; Willett, P.; Bawden, D. Comparison of fragment weighting schemes for substructural analysis. *Quant. Struct.-Act. Relat.* **1989**, *8*, 115−129.

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

(35) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: New York, NY, 2001.

(36) Irwin, J. J.; Shoichet, B. K. ZINC-a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 177−182. See also http://blaster.docking.org/zinc/ (accessed May 2007).

(37) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids*

Res. **2006**, *34*, D668-72. See also http://redpoll.pharmacy.ualberta.ca/drugbank/ (accessed May 2007).
(38) *MDPI compound collection v46*. http://www.mdpi.org/molmall/ (accessed May 2007).
(39) *The PubChem Database*. http://pubchem.ncbi.nlm.nih.gov/ (accessed May 2007).
(40) Ertl, P.; Jelfs, S.; Mühlbacher, J.; Schuffenhauer, A.; Selzer, P. Quest for the Rings − In Silico Exploration of Ring Universe to Identify Novel Bioactive Heteroaromatic Scaffolds. *J. Med. Chem.* **2006**, *49*, 4568−4573.

CI700286X