View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Springer - Publisher Connector

Mol Divers (2010) 14:401–408
DOI 10.1007/s11030-009-9187-z

FULL-LENGTH PAPER

Design of chemical libraries with potentially bioactive molecules applying a maximum common substructure concept

Michael Lisurek · Bernd Rupp · Jörg Wichard · Martin Neuenschwander · Jens Peter von Kries · Ronald Frank · Jörg Rademann · Ronald Kühne

Received: 22 April 2009 / Accepted: 26 July 2009 / Published online: 15 August 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Success in small molecule screening relies heavily on the preselection of compounds. Here, we present a strategy for the enrichment of chemical libraries with potentially bioactive compounds integrating the collected knowledge of medicinal chemistry. Employing a genetic algorithm, substructures typically occurring in bioactive compounds were identified using the World Drug Index. Availability of compounds containing the selected substructures was analysed in vendor libraries, and the substructure-specific sublibraries were assembled. Compounds containing reactive, undesired functional groups were omitted. Using a diversity filter for both physico-chemical properties and the substructure composition, the compounds of all the sublibraries were ranked. Accordingly, a screening collection of 16,671 compounds was selected. Diversity and chemical space coverage of the collection indicate that it is highly diverse and well-placed in the chemical space spanned by bioactive compounds. Furthermore, secondary assay-validated hits presented in this study show the practical relevance of our library design strategy.

Keywords Bio informatics · Drug design · High throughput screening · Library design · Molecular diversity

Electronic supplementary material The online version of this article (doi:10.1007/s11030-009-9187-z) contains supplementary material, which is available to authorized users.

M. Lisurek · B. Rupp · J. Wichard · M. Neuenschwander · J. P. von Kries · J. Rademann (✉) · R. Kühne (✉)
FMP Leibniz Institut für Molekulare Pharmakologie, Robert-Roessle Straße 10, 13125 Berlin, Germany
e-mail: rademann@fmp-berlin.de

R. Kühne
e-mail: kuehne@fmp-berlin.de

R. Frank
Department of Chemical Biology, HZI Helmholz Centre for Infection Research, Inhoffenstraße 7, 38124 Braunschweig, Germany

J. Rademann
Institut für Chemie und Biochemie, Freie Universität Berlin, Takusstraße 3, 14195 Berlin, Germany

Introduction

Over the recent years, screening of chemical libraries has developed into a broadly used methodology to identify and optimise small molecules as research tools in chemical biology. Efficient screening requires the use of compound libraries reflecting the biologically relevant chemical space. Since the number of chemically accessible low molecular weight compounds is vast [1], every screening collection can represent only a selection of the enormous number of stable, potentially bioactive compounds. Therefore, the rational preselection of compounds to be screened is of utmost importance, especially if a 'general purpose' library for a broad range of targets is to be designed and the number of compounds in the final screening collection has to be limited for reasons of costs.

In recent years, protein-family-focused libraries have been increasingly described as libraries of choice to produce better hit rates in high throughput screening [2]. However, in our case, the task was to compose a library not restricted to certain target families, but rather a 'general purpose' library serving the academic community through the screening centres of the ChemBioNet (ChemBioNet, http://www.chembionet.de, is an interdisciplinary consortium of chemists and biologists who exploit small molecules to study biological systems). Such a library should fulfill the following basic requirements: (a) it should be enriched with puta-

tive bioactive compounds, (b) it should exhibit a high degree of chemical diversity, (c) it should permit the extraction of structurally related hit clusters, (d) it must be free of artefact-causing reactive or unstable compounds [3], and (e) it must be physically available. Traditionally, screening collections are based on analysis of a descriptor space spanned by sets of physicochemical and topological descriptors to evaluate the diversity and similarity within the compound collection [4,5]. The main disadvantage of these descriptor-based strategies is their abstract formulations of molecular structures and properties. Thus, applications of such strategies are often disappointing because a single class of compounds defined by their similarity within a descriptor space can contain different substructures and does not integrate the collected experience of medicinal chemists. The occurrence of defined substructure elements, however, is crucial not only for the biological activities of drugs but even more for their pharmacokinetic or absorption, distribution, metabolism and excretion properties. Recently, fragment-based screening approaches have been applied successfully [6]. However, the binding of small fragments is weak and needs highly sensitive methods such as X-ray crystallography [7], NMR spectroscopy [8] or the amplification of binding, e.g. by dynamic ligation screening [9].

Thus, identification of biologically relevant substructures is an important prerequisite of all the screening libraries. Different methods to search and identify substructures in compound databases have been developed previously [10–16]. Analysis of substructure contents of ligands for a target family led to the formulation of the privileged substructure concept by Evans et al. and Patchett et al. [17,18]. These substructures are often used to build up target family-selective screening libraries. Schnur et al. performed a substructure analysis of ligand sets from five different target families and examined the occurrence of potential privileged substructures [19]. They found a remarkable promiscuity of the substructures and concluded that most of these substructures might be better described as drug-like and not as target family specific. In other words, the use of substructures is suitable for both the design of general purpose screening libraries and focused libraries.

Here, we present a concept for library design that combines substructures derived from bioactive compounds with a diversity-driven compound selection. This concept consists of the following steps: (a) identification of the bioactive substructures from the Derwent World Drug Index (WDI) [20], (b) assembly of sublibraries for every commercially available substructure, and (c) sublibrary-specific diversity calculations considering physicochemical properties and combinations with other identified substructures. The main advantage of the use of substructures derived from a bioactive database compared to the use of molecular descriptors is that it simulates an experienced chemist's reasoning when classifying and considering the molecules. As a result, more structurally homogeneous classes are obtained which are preselected by a track record of reported bioactivities.

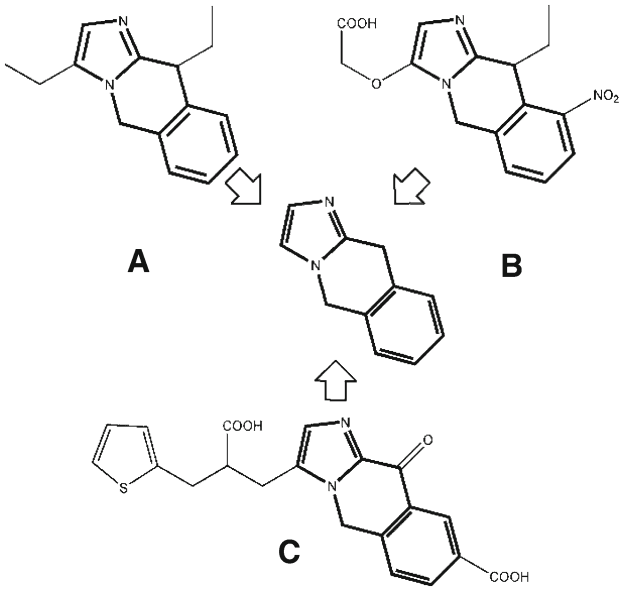

Fig. 1 Graphical representation of the MCS concept using the software ClassPharmer. Given, for example, the three molecules A, B and C, the genetic algorithm of this software identifies the maximum common substructure of these three molecules shown in the middle on the basis of their two-dimensional structure

Results and discussion

Identification of bioactive substructures

The basis for the composition of the screening collection was the identification of substructures frequently occurring in bioactive, non-peptidic organic molecules extracted form the WDI [20]. In order to identify substructures, the maximum common substructure (MCS) algorithm of ClassPharmer version 3.2 was applied (ClassPharmer, Simulations Plus Inc., Santa Fe, New Mexico, USA). The MCS algorithm is based on a genetic approach published by Gasteiger [16]. It performs a graph-based analysis of compounds from a database to classify them topologically and to place them into classes based on large, significant substructures learned on the fly. The principle of the MCS approach is illustrated in Fig. 1. Using the parameters described in computational methods, 1305 classes of substructures were identified using the 35,000 small molecules of the WDI as input. Since the most promising substructures were desired, only those that occurred in at least five different compounds within the WDI were considered. This restriction was introduced to reduce the redundancy in the substructures and to avoid singletons. As a result,

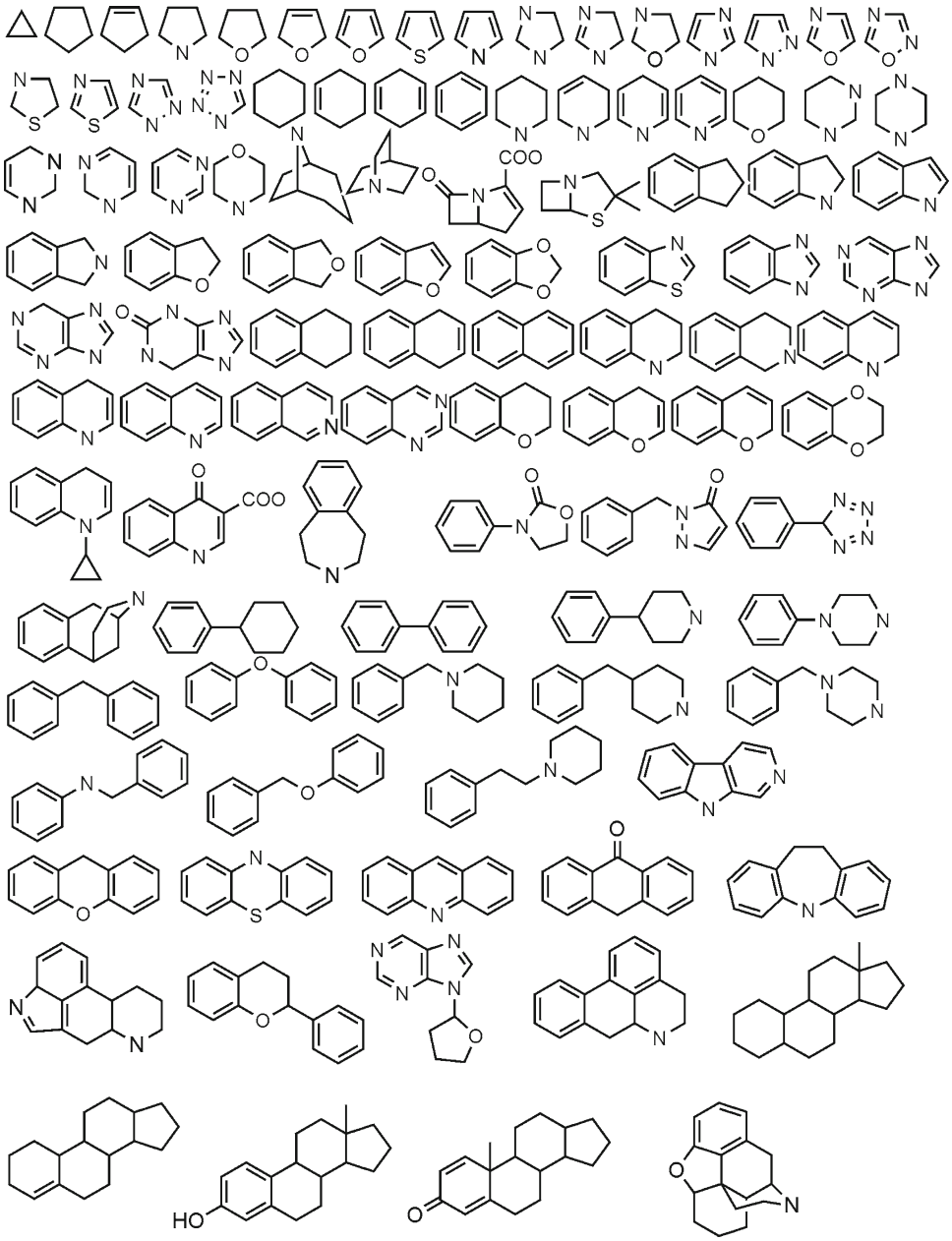

Fig. 2 Overview of the 100 substructure classes with the highest number of compounds. The classes were derived by ClassPharmer using the WDI and are ordered by increasing complexity. A complete list of the 561 identified biologically active substructure classes is contained in the supplemental material

a list of 561 substructures was established. We found that at least one of the 561 selected substructures is present in almost 97% of all the small molecules in the WDI. An overview of the 100 most frequently occurring MCSs derived from the WDI is presented in Fig. 2. A complete list of all the 561 substructures is provided in the supplemental material.

Relevance of the maximum common substructure concept

In order to demonstrate the relevance of our substructure-based concept, we analysed the substructure composition of some available target-focused libraries [21]. These libraries are focused on GPCR targets (10,605 compounds), protein kinases (7,129 compounds), ion channels (5,026 compounds), antibacterial (3,703 compounds) and antiviral activities (1,735 compounds). We found that only one out of these 28,198 compounds did not contain at least one of our 561 WDI-derived substructures. The average number of substructures per compound in the focused libraries is 3.8.

Analysis of commercially available compound libraries

More than 2.1 million compounds from approximately 20 vendors were gathered and analysed. First, in order to exclude covalent binders from the screening collection, a self-written filter program was used, which identifies and removes

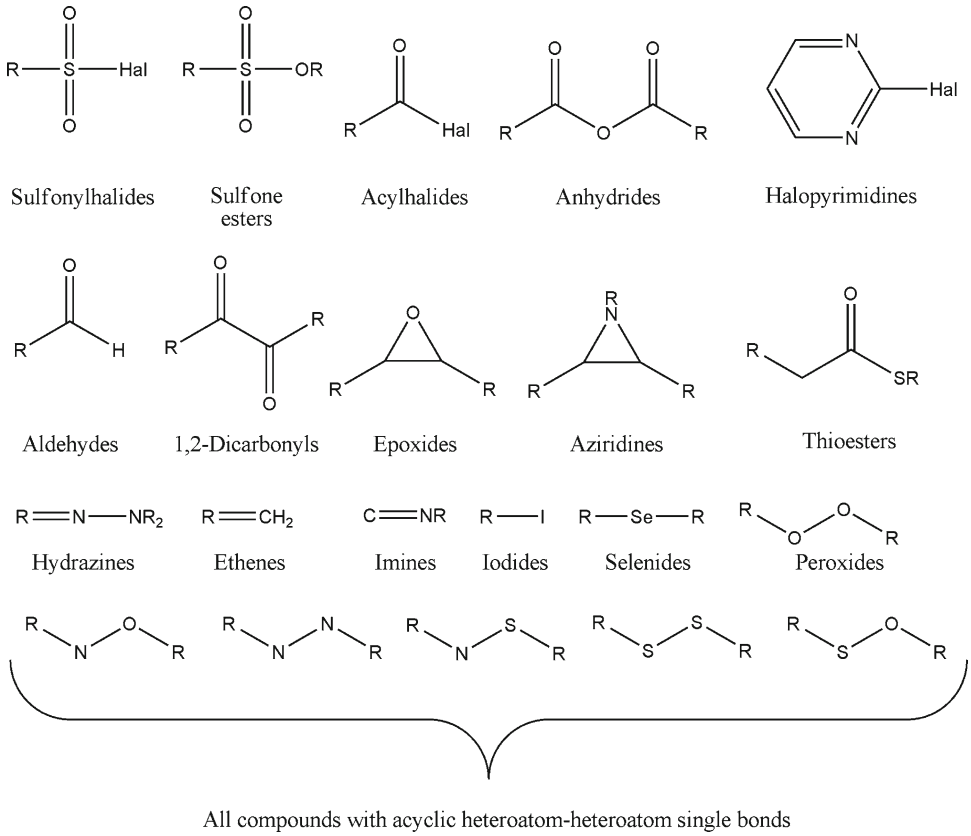

Fig. 3 Graphical representation of reactive and unstable functional groups that should not be contained in a screening collection. Compounds exhibiting these functional groups were excluded from the final library

compounds containing reactive or hydrolytically unstable functional groups from the vendor libraries (Sybyl 7.0, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO, 63144, USA). The whole list of functional groups that were eliminated during this step is shown in Fig. 3.

Afterwards, the vendor libraries were inspected for their content of the WDI-derived 561 substructures. As a maximum, 366 out of the 561 requested substructures were identified in a vendor library of approximately 550,000 compounds [21], corresponding to a substructure coverage of 66% of the WDI-derived bioactive substructures.

Substructure-based molecular diversity

Next, if it was possible, a set of the 100 most diverse compounds that belong to each of the 366 commercially available substructure classes were calculated and selected. The quantity of 100 compounds per series was chosen to obtain a final library size of approximately 20,000 compounds. This step was performed using the MOE subroutine 'Diverse Subset' and the descriptors described in computational methods [22]. One important parameter used in the calculation of the diversity set is the number of the WDI-derived substructures found in the compound. This parameter, used as a numerical descriptor, ensures that the compounds within the sublibraries selected for each of the available substructure classes contain not only the substructure by itself but also the substructure in combination with an increasing number of other identified substructures. Thus, each sublibrary is composed out of compounds containing a certain substructure, and they are, therefore, structurally related, but only the compounds with the most possible diverse decoration within the sublibraries are chosen for the final library. We think that this approach is optimal for extracting hit clusters after a successful screening.

During the selection, the Lipinski's 'Rule-of-five' [23] was considered. In the context of the ChemBioNet library, this rule was used as a simple tool to select molecules which are capable of passing through biological membranes and thus are suitable for cellular assays. Finally, all the compounds from each diverse sublibrary of all the substructures were combined, checked for double occurrences and all the duplicates were deleted. As a result, we ended up with a screening collection for the ChemBioNet consisting of 16,671 compounds.

Properties of the ChemBioNet screening collection

The final screening collection was examined according to the distribution of physicochemical properties. The average

molecular weight of a compound in the screening collection is 388 g/mol, and 80% of all the compounds have a molecular weight between 300 and 500 g/mol. There is no compound with more than five hydrogen bond donors, and none of the compounds contains more than ten hydrogen bond acceptor. The average logP of the compounds is 3.2, and only a very small fraction of compounds with a logP greater than five is present in the newly assorted screening collection. In addition, 90% of the compounds contain ten or fewer rotatable bonds. Thus, the screening library fulfils all the criteria that are usually applied to roughly estimate the bioavailability potential of small molecules.

In order to represent the chemical space of the screening collection, we calculated 32 P_VSA descriptors [24] for the compounds of the ChemBioNet library, the ChemDiv stock collection, the WDI and the ChemACX. These descriptors capture hydrophobic and hydrophilic effects, polarizability and electrostatic interactions, and therefore represent a meaningful chemical space and pay attention on as many properties of the molecules as possible. Moreover, these descriptors are independent of our substructure approach. In order to reduce the dimensionality and to avoid collinearity, a principal component analysis was performed on the basis of these descriptors. Figure 4a represents the chemical space spanned by the ChemBioNet collection in comparison to the vendor stock collection. It can be observed that the 16,671 compounds occupy almost the same chemical space as the 550,000 compounds of the ChemDiv stock collection, although they constitute only 3% of the compounds. Therefore, our strategy for compiling a screening collection did not lead to a major loss of chemical diversity compared to the vendor stock collection. Furthermore, the chemical space of our screening collection was compared to the chemical space of the annotated drug molecules from the WDI [20] and an available chemical database of 237,046 compounds (ChemACX) [25]. The ChemACX library was chosen as an example of a typical database with commercially available chemical compounds. Although, of course, it also contains bioactive molecules, this is not the emphasis of this database. The result can be seen in Fig. 4b. Comparison of the ChemBioNet compounds with those of the WDI shows that both libraries occupy a similar region of the chemical space represented by the first ten principal components. Furthermore, it can be seen that this region of the chemical space is only a small portion of the chemical space spanned by the ChemACX compounds. Though normalized, the chemical space of the 237,046 compounds of the ChemACX spans a much larger space than that of the ChemDiv stock compounds, despite the fact that these are only roughly half the number of compounds contained in the ChemDiv stock collection. Assuming that the selected drug molecules from the WDI represent the biologically relevant chemical space, the comparison of the ChemBioNet chemical space with that of the WDI shows that the ChemBioNet library is well situated in the bioactive chemical space.

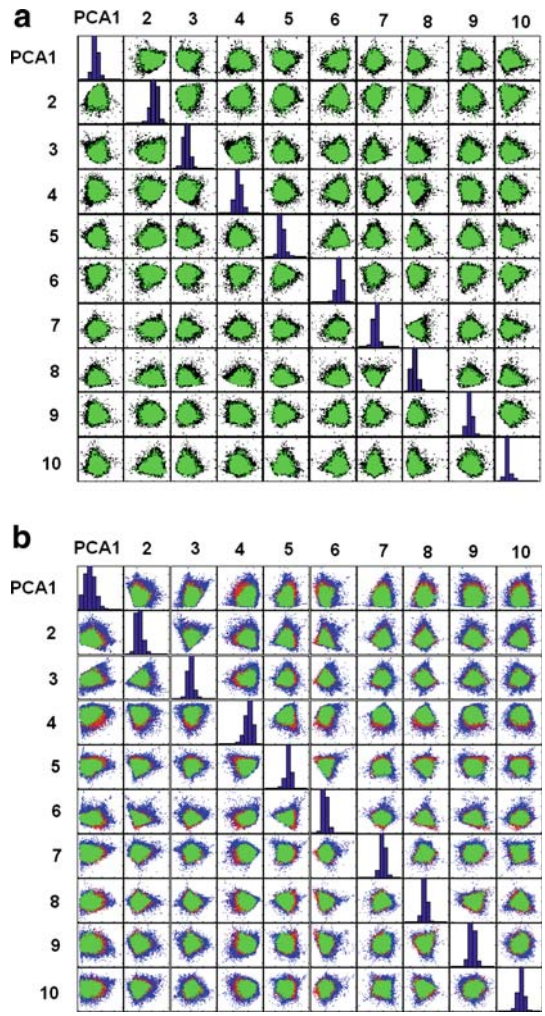

Fig. 4 Representation of the chemical space of our screening collection in comparison to the ChemDiv stock collection, the WDI and the ChemACX 32 P_VSA descriptors were calculated, and then principal component analysis was performed [24]. The figure shows the scatter plot of the first ten principal components, which are plotted against each other. The principal components are scaled from −1 to 1. The diagonal shows the Gaussian shape histogram of the principal components. **a** Principal component analysis of the ChemDiv library (531,770 compounds, *black*) and the ChemBioNet collection (16,671 compounds, *green*). The first ten principal components explain 75% of the variance. The chemical space of the ChemBioNet collection is almost the same as that of the stock collection, although it contains only 3% of the compounds of the stock collection. **b** Principal component analysis of the ChemBioNet screening collection (16,671 compounds, *green*), the bioactive small molecules from the WDI (35,345 compounds, *red*) and the compounds from an available chemical database ChemACX (CambridgeSoft) (214,232 compounds, *blue*). The first ten principal components explain 80% of the variance. It can be seen clearly, that the chemical space of the ChemBioNet collection is roughly the same as that of the WDI, whereas the chemical space of the ChemACX compounds containing many non-bioactive molecules is much larger

Conclusion

The novelty of our strategy consists in a combination of the identification of frequently occurring putative bioactive substructures in annotated bioactive compounds with a chemical diversity approach of substructure-specific sublibraries using both physicochemical descriptors and an MCS based parameter.

The calculation of MCSs was done using the compounds contained in the WDI [20]. Since this database contains compounds with known biological activity which are either available as drugs or have been submitted for clinical development, it enabled the identification of frequently occurring substructures that are characteristic for bioactive molecules. This approach resembles a medicinal chemist's reasoning and should increase the probability of selecting bioactive compounds from vendor libraries. A subset of 561 frequently occurring substructures was derived using the MCS algorithm.

As can be seen in Fig. 2, the smaller substructures represent the classical carbocyclic and heterocyclic ring chemistry. A complete list of the 561 identified biologically active substructure classes is contained in the supplemental material. It is remarkable that when the substructures grow larger, they are usually a combination of the smaller substructures. For example, adamantane, one of our WDI-derived substructures, may be interpreted as a combination of several cyclohexane rings or a combination of the decaline ring with another cyclohexane ring. Of course, one might argue that there is a certain risk in only identifying substructures from the WDI which contains no new chemical entities and therefore no novel ones can be found using our strategy. It should be noted, however, that novel drugs are usually combinations of known bioactive substructures. Moreover, the mentioned risk was further minimized by using the content of substructures as a numerical descriptor for the diversity set of 100 compounds per substructure. As a result, it is possible that combinations of substructures are provided in the final library, which have not been found before in bioactive databases. Using this approach, we selected the compounds with our WDI-derived substructure alone (though with differing side chains) and also in combination with other substructures, where both the number and the combinations with the other substructures varied heavily. For example, two or more small substructures may be condensed, connected linearly by various linkers or linked with each other via another substructure.

Our identified substructures may also be valuable for the design of novel compounds and compound libraries in the future. Since it can be expected that only few new small building blocks will be synthesized, it will primarily be the varying combinations of known substructures which will lead to new chemical entities with new biological activities. The predominant combinations of two or more substructures are: (a) condensed, (b) linked by a single bond, (c) linked by a methylene or ethylene group, (d) linked via an amine group, an ether group or a peptide-, ester- or sulfonamide bond, (e) linked via other substructures, such as 5-, 6-, or 7-membered rings, and finally, (f) two substructures may also be linked via a spiro bond.

The 35,000 small molecules of the WDI consist of up to 28 of our 561 substructures per molecule. Only 3.3% of the compounds in the WDI subset are not composed of any of them. In other words, although we have used only substructures that are present in at least five WDI compounds, the resulting substructures represent almost 97% of the WDI subset. The average number of substructures of a typical compound in the WDI is 4.5. By far the most prominent substructure class derived from the WDI is the benzene ring (9,052 compounds out of 35,345), followed by the piperidine ring (1,008 compounds within this class) and the pyridine ring (924 compounds). The 100 most abundant substructures constitute 25,762 out of the 35,345 small molecules of the WDI (73%). The compounds of the ChemBioNet library contain up to 21 substructures per molecule. The average number of substructures is 4.7 per compound. It is also interesting to compare our WDI-derived substructures with substructures derived from the CRC Dictionary of Natural Products [26]. Although a completely different approach was used [16], their substructures closely resemble the ones derived by our strategy applying the MCS algorithm to the WDI [20].

While we were analysing the compound libraries of various vendors, it turned out that at most 66% of these substructures were available in one vendor stock library. Presumably the reason why not all of the WDI-derived substructures can be obtained commercially might be that some of the substructures are protected by patents or are too costly to synthesize. In other words, with the use of our MCS-based strategy, it is easy to identify and to fill these gaps within the screening collection by adding compounds from different sources. Another advantage of the concept of building diverse subsets around substructures is that it generates a high probability of finding a congeneric series around a hit compound, thus making the subsequent hit-to-lead process more straightforward.

In summary, in this contribution, we have described and evaluated an approach for the composition of a screening collection based on a MCS concept. The approach has been validated by comparing the chemical space of this library with the chemical space of an available chemical database (ChemACX) [25] and the WDI [20] (Fig. 4). It could clearly be shown that the chemical space of the ChemBioNet collection resembles the chemical space of the biologically active small molecules from the WDI.

The successful experimental verification of the designed library showed that it may contain potent and selective ligands for a range of targets. In addition, due to the design concept of this library, usually small congeneric series around

Table 1 Results of various exemplarily in vitro high throughput screens using the ChemBioNet library

Target	Hit type	Assay	Validation screen
TPH1	Activator	Fluorescence-based kinetics	18
	Inhibitor		7
TPH2	Activator	Fluorescence-based kinetics	14
	Inhibitor		21
STAT5	Inhibitor	Fluorescence polarization	2
Shank PDZ	Inhibitor	Fluorescence polarization	2
HGFSF	Inhibitor	Light absorption	48
B-catenin	Inhibitor	Light absorption	45
SARS CoV Mpro	Inhibitor	Mobility shift (Labchip)	57

TPH1/2: tryptophan hydroxylase isoform 1 and 2; STAT5: signal transducer and activator of transcription 5; the Shank PDZ belongs to the proline-rich synapse-associated protein family of multidomain proteins known to play an important role in the organization of synaptic multiprotein complexes; HGFSF is a hepatocyte growth factor/scatter factor (a pleiotropic effector of cells expressing the Met tyrosine kinase receptor); B-catenin is a subunit of the cadherin protein complex; SARS CoV Mpro is the severe acute respiratory syndrome coronavirus main protease

each hit are obtained that help to decide what to synthesize next in an early step of drug development. Since the ChemBioNet is an open resource network supporting academic chemical and biological research, the mission is to provide a real link between biologists, chemists and other scientific disciplines required for high throughput screening, and data documentation and analysis in Europe to explore biological functions.

The ChemBioNet library was purchased and is formatted in 384-well microtiter plates. This collection is open for screening by the academic public in several screening centres in Germany and also in other European countries.

So far, numerous screens of the ChemBioNet collection for a broad range of targets have been conducted. They usually resulted in the identification of chemically diverse hits and, even more promising, whole congeneric hit series. Numerous exemplary screening results of the ChemBioNet library are given in Table 1. The validated hit rates in these examples were 0.01 up to 0.34%.

Experimental section

Cheminformatics

As the source database for bioactive compounds, the WDI was obtained as part of the Catalyst Software package (Catalyst 10, Accelrys Inc., San Diego, CA, 92121, USA) [20]. The MCSs of the compounds within the WDI were calculated using the genetic algorithm contained in the Software ClassPharmer version 3.2 from Simulations Plus Inc [22]. The parameters were set to minimal homogeneity, high redundancy and only exact atom and ring matches. For our list of bioactive substructures, only those substructures were considered, which occurred at least five times within the WDI. The SD-files of approximately 20 vendor compound collections were downloaded or received from the suppliers and stored in a MOE database or Sybyl [22]. In total, more than 2.1 million compounds were surveyed. The substructures and the occurrences of undesired functional groups contained in the vendor libraries were identified using self-written filter programs running in Sybyl.

The diverse subsets were calculated with MOE using the molecular weight, the number of hydrogen bond donors and acceptors, the number of N- and O-heteroatoms, the number of rotatable bonds and the physicochemical parameters logP, logS and logD as descriptors. In addition, the content of substructures contained in each compound (a single number, calculated by the self-written Sybyl filter) was used as a descriptor. Other descriptors were provided by the supplier or calculated using MOE [22].

The 16,671 compounds were purchased from ChemDiv [21] and solubilized in 10-mM stock concentrations in dimethyl sulfoxide.

The set of 32 P_VSA descriptors was calculated for all the compounds [24]. Then, a principal component analysis was performed on this data set in order to compare the chemical space of the selected screening collection, the ChemACX and the bioactive molecules of the WDI [20,25]. The calculation was performed using MOE [22]. The chemical space is represented by the first ten principal components projected against each other in two-dimensional space.

The analysis of the screening results has been done using the SciTegic Pipeline Pilot package (Pipeline Pilot V7.0.1.0., SciTegic Inc., San Diego, CA, 92121, USA).

Acknowledgements We thank Hans-Dieter Höltje and Victoria Higman for critical reading of the manuscript. The screening data were provided by Jörn Saupe, Samuel Beligny, Svantje Behnken and Susann Matthes. Three institutes, namely the Helmholtz Centre for Infec-tion Research (HZI), the Max Delbrück Centrum (MDC), and the Leibniz Institut für Molekulare Pharmakologie (FMP) co-financed the described screening library, which is now ready to use for supporting screening projects. Extensions to this library are currently made by the MPI in Dortmund, the University of Oslo and the University of Konstanz.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Villar HO, Koehler RT (2000) Comments on the design of chemical libraries for screening. Mol Divers 5: 13–24. doi:10.1023/A:1011326914800
2. Miller JL (2006) Recent developments in focused library design: targeting gene-families. Curr Top Med Chem 6:19–29
3. Irwin JJ (2006) How good is your screening library. Curr Opin Chem Biol 10: 352–356. doi:10.1016/j.cbpa.2006.06.003
4. Xue L, Bajorath J (2000) Molecular descriptors for effective classification of biologically active compounds based on principal component analysis identified by a genetic algorithm. J Chem Inf Comput Sci 40: 801–809. doi:10.1021/ci000322m
5. Brenk R, Schipani A, James D, Krasowski A, Gilbert IH, Frearson J, Wyatt PG (2008) Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. ChemMedChem 3: 435–444. doi:10.1002/cmdc.200700139
6. Zartler ER, Shapiro MJ (2005) Fragonomics: fragment-based drug discovery. Curr Opin Chem Biol 9: 366–370. doi:10.1016/j.cbpa.2005.05.002
7. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005) Fragment-based lead discovery using X-ray crystallography. J Med Chem 48: 403–413. doi:10.1021/jm0495778
8. Jacoby E, Davies J, Blommers MJ (2003) Design of small molecule libraries for NMR screening and other applications in drug discovery. Curr Top Med Chem 3: 11–23. doi:10.2174/1568026033392606
9. Schmidt MF, Isidro-Llobet A, Lisurek M, El-Dahshan A, Tan J, Hilgenfeld R, Rademann J (2008) Sensitized detection of inhibitory fragments and iterative development of non-peptidic protease inhibitors by dynamic ligation screening. Angew Chem Int Ed Engl 47: 3275–3278. doi:10.1002/anie.200704594
10. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39: 2887–2893. doi:10.1021/jm9602928
11. Xu YJ, Johnson M (2002) Using molecular equivalence numbers to visually explore structural features that distinguish chemical libraries. J Chem Inf Comput Sci 42: 912–926. doi:10.1021/ci0255351
12. Martin YC (1990) Computer design of potentially bioactive molecules by geometric searching with ALADDIN. Tetrahedron Comput Methodol 3: 15–25. doi:10.1016/0898-5529(90)90117-Q
13. Martin YC (1992) 3D database searching in drug design. J Med Chem 35: 2145–2154. doi:10.1021/jm00090a001
14. Abel U, Koch C, Speitling M, Hansske FG (2002) Modern methods to produce natural-product libraries. Curr Opin Chem Biol 6: 453–458. doi:10.1016/S1367-5931(02)00338-1
15. Koch MA, Schuffenhauer A, Scheck M, Wetzel S, Casaulta M, Odermatt A, Ertl P, Waldmann H (2005) Charting biologically relevant chemical space: a structural classification of natural products (SCONP). Proc Natl Acad Sci USA 102: 17272–17277. doi:10.1073/pnas.0503647102
16. Wagener M, Gasteiger J (1994) The determination of maximum common substructures by a genetic algorithm: application in synthesis design and for the structural analysis of biological activity. Angew Chem Int Ed Engl 33: 1189–1192. doi:10.1002/anie.199411891
17. Evans BE, Rittle KE, Bock MG, DiPardo RM, Freidinger RM, Whiter WL, Lundell GF, Veber DF, Anderson PS, Chang RSL, Lotti VJ, Cerino DJ, Chen TB, Kling PJ, Kunkel KA, Springer JP, Hirshfield J (1988) Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. J Med Chem 31: 2235–2246. doi:10.1021/jm00120a002
18. Patchett AA, Nargund RP (2000) Privileged structures—an update. Annu Rep Med Chem 35: 289–298. doi:10.1016/S0065-7743(00)35027-8
19. Schnur DM, Hermsmeier MA, Tebben AJ (2006) Are target-family-privileged substructures truly privileged. J Med Chem 49: 2000–2009. doi:10.1021/jm0502900
20. WDI, Derwent World Drug Index, Release 2005, Derwent Information Ltd., London
21. ChemDiv Inc., ChemDiv Chemical Database, http://www.chemdiv.com, 6605 Nancy Ridge Drive, San Diego, CA, 92121, USA
22. MOE Molecular Operating Environment, version 2005.06, Chemical Computing Group Inc., Montreal, Quebec, Canada
23. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23: 3–25. doi:10.1016/S0169-409X(96)00423-1
24. Labute P (2000) A widely applicable set of descriptors. J Mol Graph Model 18: 464–477. doi:10.1016/S1093-3263(00)00068-1
25. ChemACX, CambridgeSoft, http://www.chemacx.com, 100 CambridgePark Drive, Cambridge, MA, 02140, USA
26. Dictionary of Natural Products, version 14.1 (2005). Chapman & Hall/CRC Informa, London