Formal Concept Analysis for the Identification of Molecular Fragment Combinations Specific for Active and Highly Potent Compounds

Eugen Lounkine, Jens Auer, and Jürgen Bajorath*

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dahlmannstrasse 2, D-53113 Bonn, Germany

Received May 5, 2008

We introduce fragment formal concept analysis (FragFCA) to study complex relationships between fragments in active compounds taking potency information into account. Fragment combinations that are unique to active or highly potent compounds or that are shared by molecules having different or overlapping activity profiles are systematically identified using chemically intuitive queries of varying complexity. The methodology is applied to analyze fragment distributions in antagonists of seven G protein coupled receptor targets and identify signature fragments. Pairs or triplets of molecular fragments are found to be most specific for different activity profiles and compound potency levels. In addition, we demonstrate the ability of FragFCA to identify selective hits in high-throughput screening data sets.

Introduction

Molecular fragments are an important source of activity information and often used in compound and library design.[1] For example, in fragment based drug design, predefined molecular fragments are combined in order to generate inhibitors for selected targets.[1] For a number of target classes, attempts have been made to identify so-called privileged substructures that are associated with class-directed compound activity.[2,3] For G protein coupled receptor (GPCR[a]) ligands, it has been shown that privileged substructures can often become target-specific through chemical diversification with functional groups.[2] For kinases, analysis of the frequency of fragment occurrence in different inhibitors has also identified sets of privileged substructures.[3]

Fragment frequency analysis typically compares compound sets in a pairwise manner or characterizes fragment distributions in large data sets. For example, frequency analysis has been used to identify combinations of molecular fragments that are highly recurrent in synthetic compounds.[4] Another study by Sutherland et al. has identified individual fragments that occur in ligands of related as well as unrelated targets on the basis of frequency analysis.[5]

Recently, another approach has been introduced to prioritize HTS hits based on the affinity of fragments they contain, as determined by screening of fragment libraries.[6] Similarly, a method has been developed to identify fragments that distinguish highly potent from moderately active compounds.[7] Furthermore, combinations of fragments that are characteristic of different compound activity classes have also been isolated from randomly generated molecular fragment populations.[8]

In addition to analyzing fragment distributions in active or database compounds, molecular fragments have also become focal points in the study of polypharmacology, in particular, the generation of ligands that are active against multiple targets.[9] Fragments that are recurrent in compounds with different activities are relevant as building blocks for polypharmacological ligand design and might also be used to better control compound side effects.[9] In this context, Sheridan has suggested an analysis scheme based on common subgraph mining to extract fragments that are associated with multiple ligand activities.[10]

We have carried out a fragment-based analysis of complex structure−activity relationships between compounds with overlapping biological activities. To these ends, we have adopted formal concept analysis (FCA), a machine learning and data analysis technique originally introduced in information science,[11] in order to identify fragment combinations that specifically occur in compounds having different activity profiles or that are unique to highly potent molecules. Fragment FCA (FragFCA) is an interactive method that makes it possible to identify sets of signature fragments using chemically intuitive queries of varying complexity. Herein, we report FragFCA and its application to a set of biogenic amine GPCR antagonists with overlapping, yet distinct activity profiles and cathepsin L and S HTS data sets.

Methods

Formal Concept Analysis. Formal concept analysis derives ontologies from sets of *objects* and their properties (*attributes*).[11] From these ontologies, subsets of objects with desired properties are extracted. A *formal context* consists of all objects and the associated attributes. In FragFCA, a formal context consists of fragments and properties that indicate in which compounds single fragments and combinations occur and how they are related to activity (active vs inactive) or differences in potency. The context can be represented as simple "occurs in" and "does not occur in" relationships, as illustrated in Figure 1A. In this example, the benzene fragment "occurs in" A, "occurs in" B and "does not occur" in C, whereas the combination of the benzene and the piperazine fragment "occurs in" A, "does not occur in" B, and "does not occur in" C. FragFCA can be used to extract increasingly complex relationships and all fragment combinations they involve.

A *concept* is defined as a set of objects sharing a set of attributes and vice versa. In FragFCA, objects are (combinations of) molecular fragments and attributes include, for example, compound activity or potency information. Concepts are represented in *concept lattices.* A concept lattice reports activity

* To whom correspondence should be addressed. Phone: +49-228-2699-306. Fax: +49-228-2699-341. E-mail: bajorath@bit.uni-bonn.de.

[a] Abbreviations: FCA, formal concept analysis; FragFCA, fragment formal concept analysis; GPCR, G protein coupled receptor; HTS, high-throughput screening.

10.1021/jm800515r CCC: $40.75
Published on Web 08/13/2008

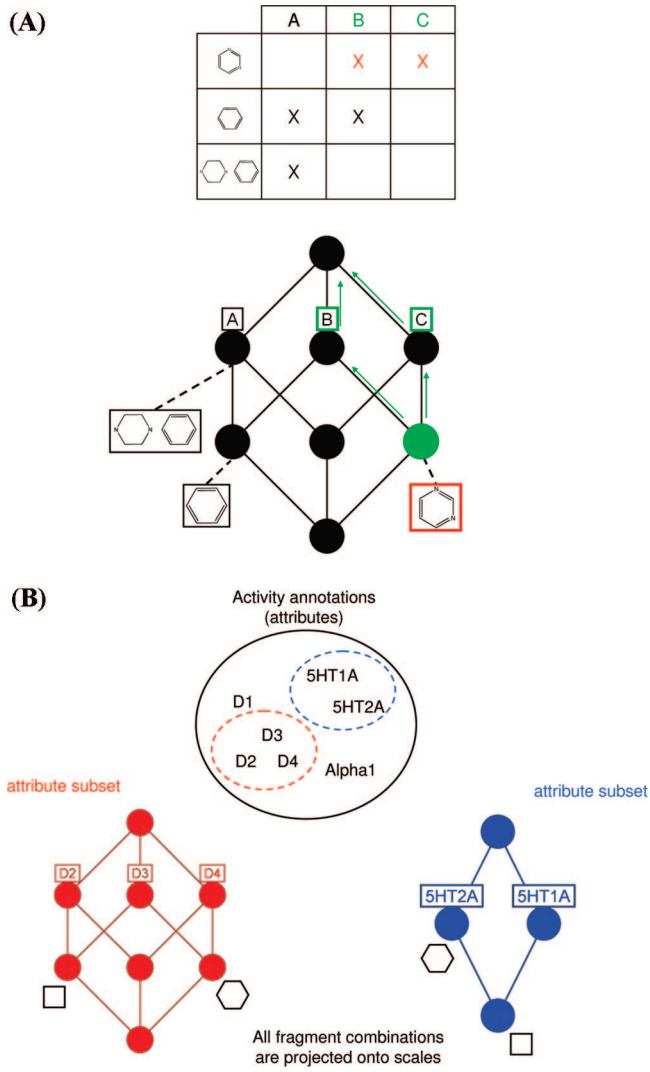

Figure 1. Scales and concept lattices. (A) The context (top) establishes relationships between fragment combinations (objects) and activity annotations (attributes). X indicates that a fragment has a specific activity, e.g., the pyrimidine fragment is found in classes B and C but not in A. The corresponding concept lattice is shown at the bottom. The node representing the concept "B and C but not A" is colored green. Following the green arrows to the top of the lattice yields the attributes "B and C". The pyrimidine fragment associated with this concept is directly shown below the node in a red box. (B) Scales focus on a subset of attributes. At the top, seven attributes corresponding to seven compound activity classes are shown. At the bottom, concept lattices of two scales are shown that capture two and three attributes, respectively. Each scale provides information about all objects (i.e., fragment combinations, indicated by small black icons).

annotations, fragments, and the relationships between them. The lattice is formed by interconnected nodes, each representing a particular concept, i.e., the relationship between selected properties and the fragments that share them (Figure 1A).

How is a concept lattice utilized? In order to extract fragment combinations with desired properties, one first needs to identify the corresponding node. The properties are written above the nodes. If we follow the edges from each node toward the top of the lattice, an associated set of properties is obtained, as illustrated in Figure 1A. Then we select the node that is associated with the desired properties, for example, potency ranges of interest. From the selected node, the corresponding fragment set is extracted (reported below the node). Additional fragment combinations that share the desired but also other properties can be identified by following the edges toward the bottom of the lattice.

Table 1. GPCR Antagonist Set[a]

activity ($\leq$10 μM)	activity annotations
dopamine D1 receptor antagonist	84
dopamine D2 receptor antagonist	216
dopamine D3 receptor antagonist	75
dopamine D4 receptor antagonist	93
serotonin 5HT1A receptor antagonist	95
serotonin 5HT2A receptor antagonist	32
adrenergic α1 receptor antagonist	92

[a] The 267 compounds in this set have multiple activities and represent a total of 687 activity annotations.

Scales are defined that preselect subsets of properties. Focusing on subsets of properties (e.g., activity annotations) is often crucial because concept lattices with too many properties become difficult to navigate. For any given query, an appropriate scale can be selected, for example, a potency scale. Queries of increasing complexity can be assembled by combining different scales. The subset of fragments that is selected on a particular scale is forwarded to the next one, as illustrated in Figure 1B. Thereby, fragment sets with multiple layers of user-defined properties can be identified.

Our concepts and scales were implemented using the publicly available ToscanaJ software.[12] ToscanaJ makes it possible to combine various scales and explore specific subsets of fragments by interactive successive filtering through multiple scales. Concept lattices were drawn using ToscanaJ export functions.

GPCR Antagonist Set. We have analyzed a previously reported[13] and publicly available (www.lifescienceinformatics.uni-bonn.de) set of 267 biogenic amine GPCR antagonists, as summarized in Table 1. Compounds in this set are active against multiple receptors at different potency levels. A compound was assigned to a class if it was active against the target receptor with an IC_{50} value of 10 μM or lower. On the basis of this threshold value, the 267 antagonists received a total of 687 activity assignments, as reported in Table 1.

Fragment Generation. We have applied a hierarchical fragmentation scheme that divides compounds into rings, linkers, and substituents (side chains).[16] As a refinement of conventional hierarchical fragmentation,[16] we not only sample condensated rings as fragments but also further divide them into nonfused individual ring components. This fragmentation scheme was implemented in the Molecular Operating Environment[17] and is illustrated in Figure 2. Fragments were generated from all GPCR antagonists and combined. From the initial set of 701 unique fragments, small fragments with fewer than 4 atoms and large fragments with more than 20 atoms were removed, resulting in a final set of 427 fragments.

Enumeration of Fragment Combinations. From these 427 GPCR fragments, a structural key-type fingerprint was generated and calculated for each of the 267 antagonists. For each compound, all individual fragments, pairs, triplets, and quadruplets were extracted from its fingerprint representation. A total of 231 464 different combinations consisting of one to four fragments were obtained. The enumeration of fragment combinations and FragFCA can be applied to any fragmentation scheme and structural key-type fingerprint representation.

Activity Annotation of Fragment Combinations. Fragment combinations were annotated with qualitative and quantitative compound activity information. An antagonist was considered active against a GPCR target if its IC_{50} was equal to or below 10 μM and inactive if it was above this value. Furthermore, we distinguished between five different potency ranges for active

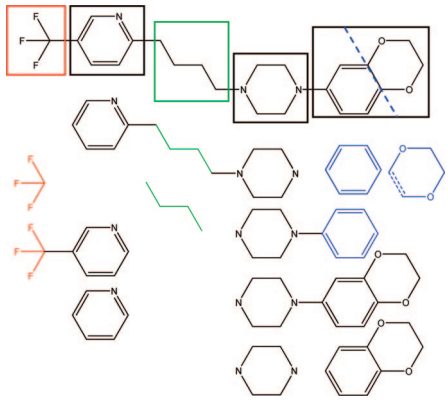

Figure 2. Fragmentation scheme. A model compound and the resulting fragments are shown. The compound is subdivided into rings (black), linkers (green), and side chains (red). Condensated rings are separated into individual ring components that retain shared atoms and bonds. Dashed lines in aliphatic rings (blue) indicate that these rings were fused with an aromatic ring.

compounds: ≤1 nM, 1−10 nM, 10−100 nM, 100 nM to 1 μM, and 1−10 μM. If a fragment combination was found in several active compounds with different potency, it was annotated with multiple potency ranges.

Definition of Scales. Global scales were used to qualitatively compare multiple compound activity classes at different levels of detail. In addition, for each class, three specific scales were defined, a frequency, activity, and potency scale. The frequency scale determines the number of active compounds that contain a particular fragment combination. The activity scale distinguishes between fragment combinations that occur only in active, active and inactive, or only inactive compounds. Potency scales differentiate active compounds according to potency ranges.

Nonredundant and Minimal Fragment Sets. From each fragment set retrieved by a query, redundant fragment information was omitted by removal of fragment combinations that contained selected singletons, duplets, or triplets as subsets. Thereby, nonredundant sets of fragment combinations were obtained that covered the maximal number of compounds with desired attributes. Furthermore, fingerprint overlap calculations revealed minimal fragment sets covering selected molecules.

Cathepsin Data Sets. In addition, we have applied FragFCA to a previously reported[14] and publicly available (www.lifescienceinformatics.uni-bonn.de) collection of inhibitors of cathepsins, a thiol protease family, in order to identify selectivity markers. The fragmentation scheme as described for GPCR antagonists was applied. The extracted fragment combinations were then used to identify selective cathepsin L inhibitors by analyzing two HTS bioassay sets publicly available in Pubchem[15] (cathepsin L, AID 460; cathepsin S, AID 501).

Results

Defining Scales. For large formal contexts such as the fragment combinations and attributes used here, visualizing all relationships in a single concept lattice is not feasible because there are too many possible concepts. Therefore, scales are defined that focus on subsets of attributes. Figure 3A illustrates the hierarchical organization of global scales that are used to distinguish between multiple compound classes, and Figure 3B shows the more specific frequency, activity, and potency scales. Combinations of scales define queries that are capable of revealing different types of fragment-based relationships, as demonstrated in the following. Owing to the presence of overlapping activities and differences in potency, the GPCR antagonists analyzed in this study present complicated structure−activity relationships. For our analysis of these compound sets, four global GPCR scales were used and, in addition, five specific scales for each of the seven GPCR targets, resulting in a total number of 39 scales. In the following, we discuss the results of six exemplary FragFCA queries having different levels of complexity.

Fragments Characteristic of Dopamine and α1 Receptor Antagonists. We first compared fragment distributions in different activity classes. As an example, we determined fragment combinations that were characteristic of dopamine antagonists. The D1 activity scale was used to extract 41 049 fragment combinations that occurred in compounds active against D1 (but not in inactive compounds). As reported in the global GPCR scale in Figure 4A, 90% (37 098) of these combinations only occurred in dopamine receptor antagonists, 3367 fragment combinations were shared with serotonin receptor antagonists (but did not occur in α1 ligands), and 584 combinations were shared by 5HT, D, and α1. As shown in Figure 4B, the 3367 and 584 shared fragments were unevenly distributed in serotonin receptor antagonists; they mostly occurred in 5HT1A, rather than 5HT2A antagonists. In addition, none of the fragment combinations found in α1 ligands also occurred in 5HT2A antagonists. We next analyzed the three fragment subsets using the global D2 scale. Most fragment combinations were found to be D2 specific. None of the fragments shared with serotonin receptor ligands were specific for either D3 or D4, and the 584 fragment combinations shared among all classes on the global GPCR scale only occurred in D2 but not D3 or D4 antagonists. By use of these scales, signature fragments and combinations can be easily identified. An example is shown in Figure 4C. The benzimidazol-2-one fragment in the center is found in both D2 and serotonin receptor antagonists. However, in combination with each of the surrounding fragments, it is only present in D2 antagonists.

We next analyzed the distribution of the seven fragment pairs in designated dopamine, serotonin, and α1 antagonists available in the Molecular Drug Data Report (MDDR).[18] Four of seven individual combinations were only found in selective D1 antagonists. The three remaining pairs also occurred in other antagonists. However, together they only occurred in two serotonin receptor antagonists but all selective D1 antagonists.

The same type of FragFCA analysis was carried out for α1 antagonists in our GPCR ligand set. In contrast to dopamine receptor antagonists, no α1-specific fragment combinations were found. However, fragments with dual receptor specificity existed. For example, a subset of 27 777 fragment combinations was identified that only occurred in α1 and serotonin receptor antagonists and that could be further reduced to 246 nonredundant combinations (see Methods). The composition of this nonredundant set is reported in Figure 5. As can be seen, the set is dominated by fragment pairs (63%).

Fragments Distinguishing α1 and D2 from 5HT Receptor Antagonists. Next we determined more complex fragment relationships. Fragment combinations were extracted that were shared by α1 and D2 antagonists but did not occur in serotonin receptor antagonists. Therefore, the α1 and D2 activity scales and the global GPCR scale were applied in a sequential manner. A total of 42 265 fragment combinations were extracted that were shared by α1 and D2 antagonists (and were not present in compounds inactive against these two receptors); 9265 of these combinations did not occur in serotonin receptor ligands. These fragments were reduced to a nonredundant set consisting of

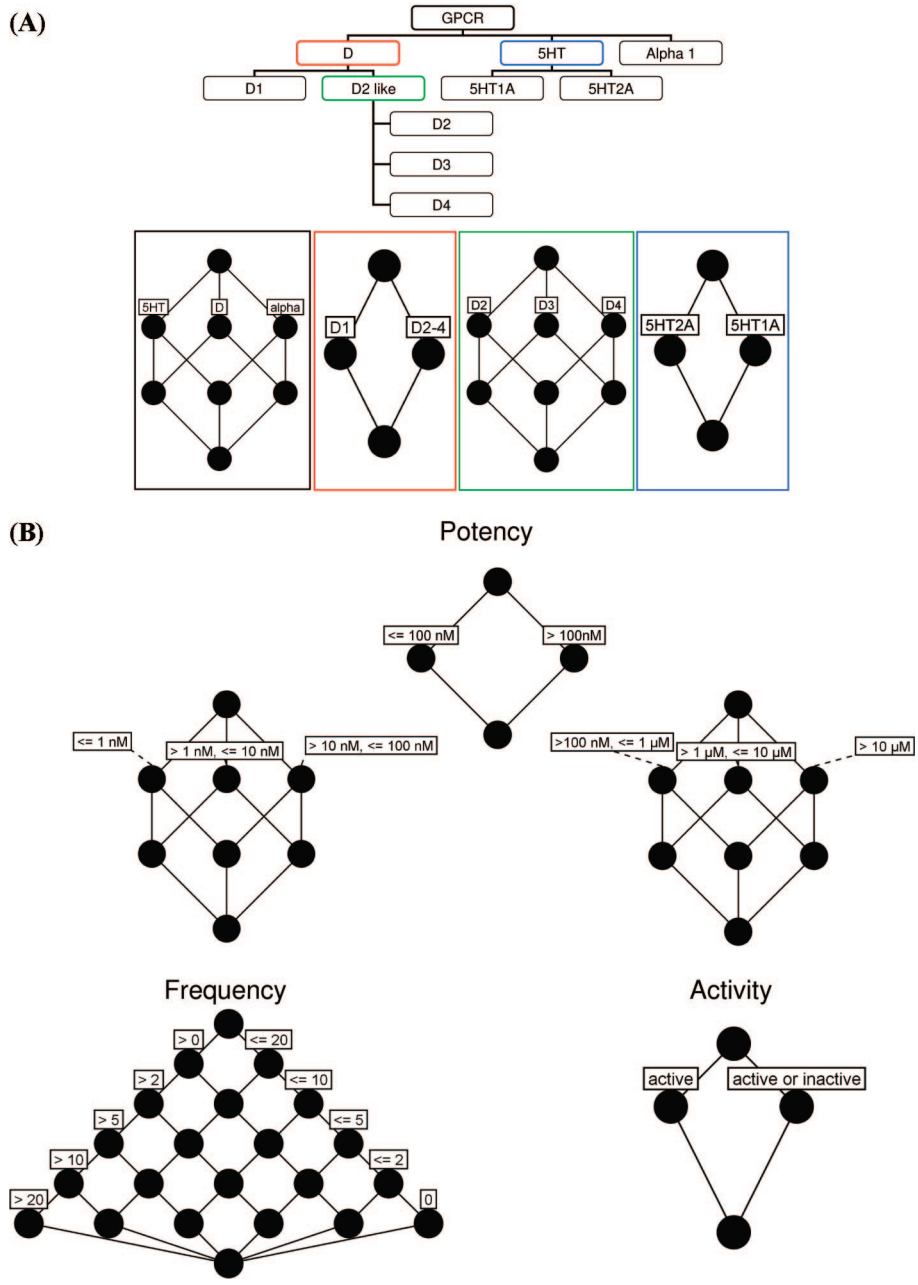

Figure 3. Organization of scales. (A) The tree structure (top) reflects the hierarchical organization of global scales. Each node and its children (bold) correspond to a particular scale (with attributes provided by the children). At the bottom, the different scales are shown. Parents in the tree and corresponding scales are color-coded. (B) Three types of specific scales are defined for each target. The potency scales account for different potency levels. The frequency scales report the number of molecules a fragment combination occurs in. The activity scales distinguish between fragment combinations that are only present in active compounds and fragment combinations that also occur in inactive molecules.

98 fragment combinations. In a database search, these fragment combinations correctly retrieved all 18 compounds having the corresponding activity profile (i.e., active against D2 and α1 but not 5HT). Moreover, the minimal signature set describing all 18 compounds consisted of only two fragment pairs covering 12 and 14 compounds, respectively, as shown in Figure 6.

Both fragment combinations depicted in Figure 6 represent overlapping fragments (due to a shared phenyl moiety) and thus appear to contain redundant information. However, fragment combinations extracted using FragFCA identify compounds containing overlapping or nonoverlapping fragments, as also shown in Figure 6. This versatility permits the identification of structurally diverse selective compounds.

Fragments Specific for 5HT1A Receptor Antagonists. We also attempted to find a minimal set of fragment combinations that distinguish 5HT1A from dopamine and α1 antagonists. First, the activity scale specific for 5HT1A was used and fragment combinations were selected that did not occur in 5HT1A inactive compounds. This query was then further refined using the GPCR global scale. Only 3311 combinations were shared between serotonin and dopamine antagonists but not present in α1 ligands. Most fragment combinations were shared among all three classes or were specific for serotonin receptor antagonists (35% each). A total of 37 641 fragment combinations specific for 5HT1A were selected and reduced to a nonredundant set of 199 unique fragment combinations. Figure 7A reports the composition of this nonredundant set. Again, fragment pairs and triplets constituted the major part (80%) of all specific combinations.

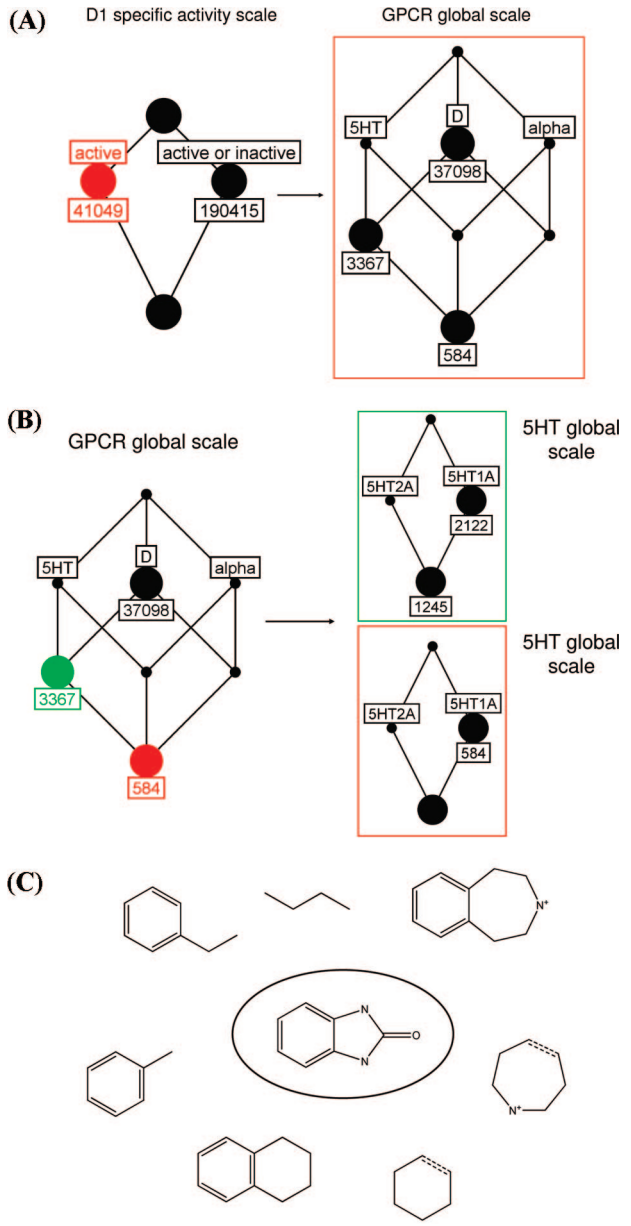

Figure 4. Fragment combinations specific for D1 antagonists. (A) Distribution of D1 specific fragments selected from the D1 activity scale (left) and further analyzed using the global GPCR scale (right). Small nodes indicate that no fragment combinations are directly associated with the corresponding concept. Nodes at the bottom of the concept lattice represent concepts that combine several properties. (B) Distribution of the green and red subsets (left) among serotonin receptor antagonists (right). (C) The encircled fragment in the center is found as a singleton in both D2 and serotonin antagonists. By contrast, the combination with each of the surrounding fragments (yielding pairs) only occurs in D2 antagonists.

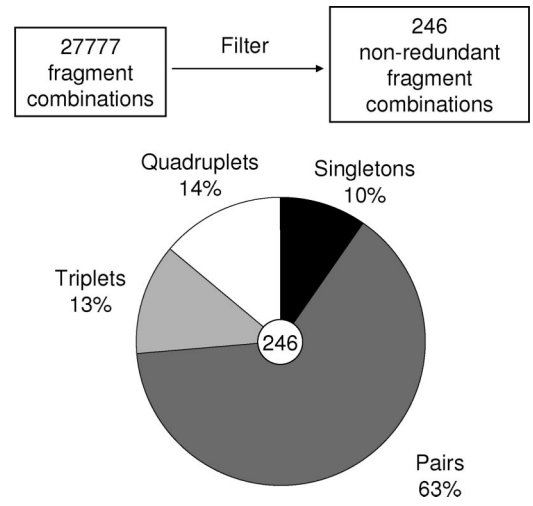

Figure 5. Fragment combinations in α1 and serotonin antagonists. The chart reports the distribution of 246 nonredundant combinations. 63% of all combinations are pairs.

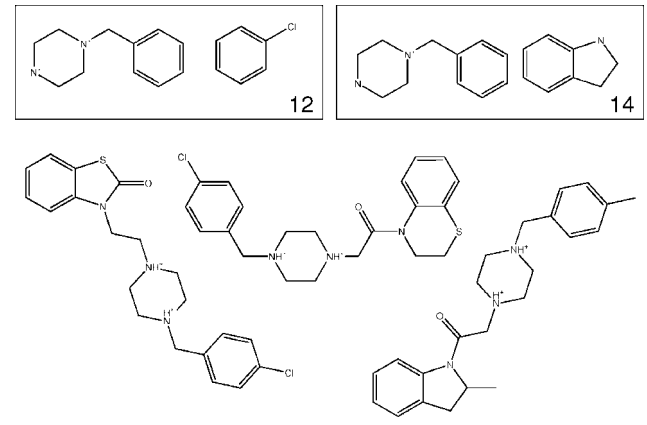

Figure 6. Fragment combinations specific for α1 and D2 against 5HT antagonists. The minimal set of fragment combinations describing all desired compounds is shown at the top. For each fragment pair, the number of molecules it identifies is reported. At the bottom, three representative antagonists are shown that are detected by these fragments.

The minimal set of fragment combinations determined from fingerprint overlap in selected compounds consisted of two fragment pairs and one fragment triplet, shown in Figure 7B. This minimal set identified all nine compounds in the database having the corresponding activity profile (i.e., active against 5HT1A but not dopamine or adrenergic receptors). Representative antagonists are also shown in Figure 7B. The first fragment pair (butaldehyde and cationic phenylpiperazine) described seven of the nine compounds. We further analyzed this pair with respect to the selectivity of its individual fragments. As reported in Figure 7C, the individual fragments were not specific for 5HT1A antagonists. By contrast, the combination of these two fragments was specific.

Fragments in Potent 5HT1A and D4 Receptor Antagonists. The 5HT1A query discussed above was further refined using the 5HT1A potency scale to initially extract 17 225 fragment combinations occurring in compounds with ≤100 nM potency. The GPCR antagonist set contained six specific 5HT1A antagonists at this potency level. We then used the 5HT1A frequency scale to select 3542 fragment combinations that occurred in all of these compounds. The nonredundant set contained 32 fragment combinations, 27 of which were pairs or triplets. Thus, adding two scales to the predefined 5HT1A query made it possible to identify fragment combinations that were specific for a subset of potent 5HT1A antagonists.

In order to search for signature fragments of highly potent D4 antagonists, we directly applied the D4 potency scale to extract 35 593 fragment combinations occurring only in D4 antagonists with ≤100 nM potency. We further refined this query using a D4 high potency scale that selected 588 combinations specific for the highest potency range (≤1 nM). The reduced set consisted of nine fragment combinations depicted in Figure 8. These fragments identified four of six D4

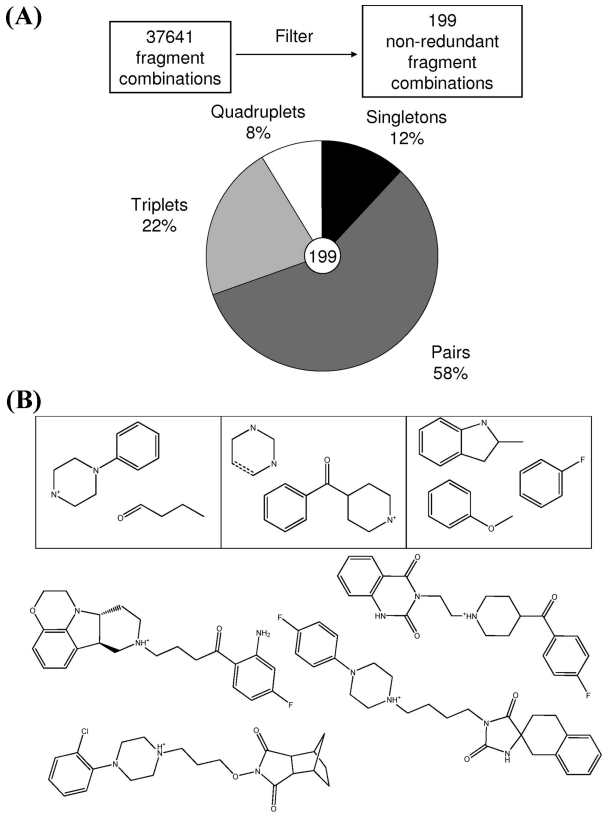

Figure 7. Fragment combinations specific for 5HT1A antagonists. (A) Fragment distribution in the nonredundant subset. (B) The minimal set of fragment combinations identifying all 5HT1A antagonists is shown at the top. At the bottom, four representative 5HT1A antagonists are shown. (C) Individual fragments of a pair. The number of 5HT1A antagonists containing each individual fragment and their combination is reported together with the number of antagonists with different activities. Percentages in parentheses report the fraction of 5HT1A antagonists in the database. The right column reports the fraction of desired compounds in the set identified by the individual fragments and their combination.

antagonists with ≤1 nM potency present in the database and detected no other compounds. The results for all GPCR selectivity queries are summarized in Table 2.

Prediction of Selective HTS Hits Using FragFCA. We also utilized FragFCA for mining of HTS data. Therefore, we extracted 533 fragment combinations from a literature set of cathepsin inhibitors that indicated selectivity of active compounds for cathepsin L compared to cathepsin S. These fragments were then used to analyze a publicly available high-throughput screen for inhibitors of cathepsin L. This set

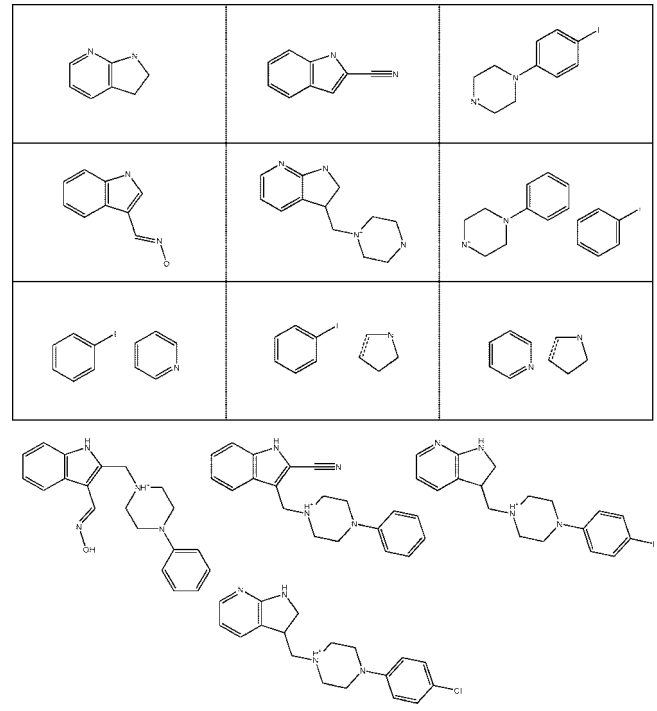

Figure 8. Fragments specific for highly potent D4 antagonists. The minimal set of nine fragment combinations is shown together with four highly potent D4 antagonists detected by this set.

Table 2. Summary of GPCR Antagonist Queries[a]

query	reduced key set size	total number of compounds	number of recovered compounds
α1 and D2 vs 5HT	98	18	18
5HT1A vs D and α1	199	9	9
potent selective 5HT1A	32	6	6
potent selective D4	9	6	4

[a] For each selectivity query, reported are the number of reduced fragment combinations, the total number of available selective compounds, and the number of correctly identified compounds applying the query.

contained a total of 41 active compounds, none of which was also part of the literature set. Because the same compound data set was also screened against cathepsin S, 36 of the 41 active compounds were confirmed to be selective for cathepsin L. Thirteen of these 36 selective compounds were identified using our fragment combinations. By contrast, none of the five inhibitors active against both cathepsin S and L matched any of these fragment combinations. As illustrated in Figure 9, the 13 selective inhibitors identified by FragFCA were structurally diverse.

Discussion

Formal concept analysis is capable of identifying specific patterns within a given context of objects and associated attributes. We introduce the FragFCA approach that can be applied to any fragment scheme and structural key-type fingerprint representations. FragFCA makes it possible to compare multiple compound activity classes and to study complex fragment-based relationships between active compounds. We have developed two types of scales, global and class-specific scales, which represent very versatile and intuitive tools to build highly specific and increasingly complex fragment queries. Global scales enable the comparison of biological activity profiles at different levels of detail, while specific scales provide information, for example, about the potency characteristics or frequency of occurrence of individual fragment

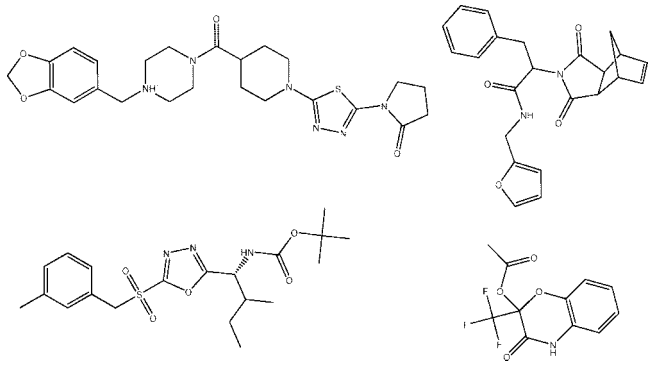

Figure 9. Selective cathepsin L inhibitors. Shown are examples of inhibitors that are selective for cathepsin L over cathepsin S. These structurally diverse inhibitors were identified in a HTS data set using FragFCA.

combinations. The utility of FragFCA goes well beyond fragment frequency analysis that has been applied in a number of previous studies.[2,4,5]

FragFCA is in principle applicable to fragments derived by any fragmentation scheme. Here, we have applied hierarchical fragmentation, which is most commonly used. For the generation of global scales, biological activity has been defined in an "all or nothing" manner, i.e., through application of a 10 μM threshold level. Furthermore, fragment combinations associated with active compounds were not permitted to occur in any inactive molecules. However, for FragFCA, other criteria can be readily applied. For example, by combination of potency and frequency scales, queries with varying threshold levels for activity and/or fragment occurrence can be designed. Here, FragFCA has been applied to sets of approximately 300 active compounds but FragFCA is in general not limited by database size.

A major goal of FragFCA is the identification of molecular fragments and fragment combinations that are specific for compound activity classes or subsets of active compounds at different potency levels. We have shown that specific fragment combinations can be reduced to nonredundant and minimal sets that are highly descriptive. Such fragment combinations can be applied to differentiate compound classes from each other or to search for active and also highly potent compounds via simple substructure queries. It is conceivable that specific fragment combinations identified by FragFCA might also aid in fragment-based ligand or focused library design.

The GPCR antagonist set analyzed herein represents complicated structure−activity relationships because test compounds have overlapping activity profiles. On the basis of a conventional hierarchical fragmentation scheme, we have been able to identify signature fragments for a variety of GPCR antagonist subsets with different characteristics. A major finding has been that most GPCR antagonist-specific information is contained in fragment pairs and triplets rather than individual fragments. These findings indicate that small fragment combinations are most relevant for distinguishing between different activity profiles or compound potency levels. We also applied the FragFCA approach to another target family, cathepsins, and identified fragment combinations capable of extracting selective cathepsin L inhibitors from HTS data.

Conclusions

Fragment formal concept analysis is based on principles from information science and designed to systematically identify signature fragments that are specific for different compound classes or potent molecules. FragFCA permits fragment searching in a flexible and interactive manner using chemically intuitive scales. Combination of scales results in queries of varying complexity. FragFCA analysis has successfully identified fragment combinations specific for subsets of GPCR antagonists with different activity profiles and potency and for cathepsin L inhibitors in screening data sets. Signature fragment combinations have identified structurally diverse molecules having similar activity. In general, fragment pairs and triplets were found to capture most activity-specific information.

References

(1) Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* **2004**, *47*, 3463–3482.
(2) Schnur, D. M.; Hermsmeier, M. A.; Tebben, A. J. Are Target-Family-Privileged Substructures Truly Privileged? *J. Med. Chem.* **2006**, *49*, 2000–2009.
(3) Aronov, A. M.; McClain, B.; Moody, C. S.; Murcko, M. A. Kinase-Likeness and Kinase-Privileged Fragments: Toward Virtual Polypharmacology. *J. Med. Chem.* **2008**, *51*, 1214–1222.
(4) Lameijer, E.; Kok, J. N.; Bäck, T.; Ijzerman, A. P. Mining a Chemical Database for Fragment Co-Occurrence: Discovery of "Chemical Clichés". *J. Chem. Inf. Model.* **2007**, *46*, 553–562.
(5) Sutherland, J. J.; Higgs, R. E.; Watson, I.; Vieth, M. Chemical Fragments as Foundations for Understanding Target Space and Activity Prediction. *J. Med. Chem.* **2008**, *51*, 2689–2700.
(6) Crisman, T. J.; Bender, A.; Milik, M.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C. K.; Fejzo, J.; Hommel, U.; Davies, J. W.; Glick, M. "Virtual Fragment Linking": An Approach To Identify Potent Binders from Low Affinity Fragment Hits. *J. Med. Chem.* **2008**, *51*, 2481–2491.
(7) Wolohan, P. R. N.; Akella, L. B.; Dorfman, R. J.; Nell, P. G.; Mundt, S. M.; Clark, R. D. Structural Unit Analysis Identifies Lead Series and Facilitates Scaffold Hopping in Combinatorial Chemistry. *J. Chem. Inf. Model.* **2006**, *46*, 1188–1193.
(8) Batista, J.; Bajorath, J. Mining of Randomly Generated Molecular Fragment Populations Uncovers Activity-Specific Fragment Hierarchies. *J. Chem. Inf. Model.* **2007**, *47*, 1405–1413.
(9) Morphy, R.; Rankovic, Z. Designed Multiple Ligands. An Emerging Drug Discovery Paradigm. *J. Med. Chem.* **2005**, *48*, 6523–6543.
(10) Sheridan, R. P. Finding Multiactivity Substructures by Mining Databases of Drug-like Compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1037–1050.
(11) Priss, U. Formal Concept Analysis in Information Science. *Annu. Rev. Inf. Sci. Technol.* **2006**, *40*, 521–543.
(12) Becker, P.; Hereth, J.; Stumme, G. ToscanaJ. An Open Source Tool for Qualitative Data Analysis. In *Advances in Formal Concept Analysis for Knowledge Discovery in Databases*, Proceedings Workshop FCAKDD of the 15th European Conference on Artificial Intelligence (ECAI), Lyon, France, 2002; Duquenne, V., Ganter, B., Liquiere, M., Nguifo, E. M., Stumme, G., Eds.; IOS Press: Amsterdam, 2002.
(13) Vogt, I.; Ahmed, H.; Auer, J.; Bajorath, J. Exploring Structure−Selectivity Relationships of Biogenic Amine GPCR Antagonists Using Similarity Searching and Dynamic Compound Mapping. *Mol. Diversity*, in press; available online in the EarlyView section of the journal.
(14) Stumpfe, D.; Ahmed, H. E.; Vogt, I.; Bajorath, J. Methods for computer-aided chemical biology. Part 1: Design of a benchmark system for the evaluation of compound selectivity. *Chem. Biol. Drug Des.* **2007**, *70*, 182–194.
(15) Pubchem. http://pubchem.ncbi.nlm.nih.gov.
(16) Bemis, G.; Murcko, M. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.
(17) *Molecular Operating Environment (MOE)*, version 2007.09; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2007; http://www.chemcomp.com (accessed Nov 1, 2007).
(18) *Molecular Drug Data Report (MDDR)*, version 2005.2; Symyx Software: San Ramon, USA.

JM800515R