# Elucidation of Structure−Activity Relationship Pathways in Biological Screening Data

Mathias Wawer, Lisa Peltason, 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 November 7, 2008*

A computational molecular network analysis of various high-throughput screening (HTS) data sets including inhibition assays and cell-based screens organizes screening hits according to different local structure−activity relationships (SARs). The resulting network representations make it possible to focus on different local SAR environments in screening data. We have designed a simple scoring function accounting for similarity and potency relationships among hits that identifies SAR pathways leading from active compounds in different SAR contexts to key compounds forming activity cliffs. From these pathways, SAR information can be extracted and utilized to select hits for further analysis. In clusters of hits related by different local SARs, alternative pathways can be systematically explored and ranked according to SAR information content, which makes it possible to prioritize hits in a consistent manner.

## Introduction

The analysis of HTS[a] data and characterization of hits typically involves multiple-step analyses[1] that employ statistical approaches[2] and data visualization techniques.[3] Over the past decade, different visualization tools have been developed[4,5] to graphically analyze HTS data and distributions of active compounds in histograms, scatter plots, or heat maps. It is well appreciated that HTS data analysis is a complicated task that still is far from being routine.[1,2] For example, the large numbers of weakly potent primary hits that are obtained in many screens make it often difficult to understand underlying SARs, if they exist, and select suitable candidates for further analysis. To complement currently available HTS data analysis and visualization tools, it should, in principle, be possible to organize HTS hits in graph representations such as molecular networks that reflect their structural and potency relationships, but this possibility has thus far not been explored.

The design and analysis of biological networks has been, and continues to be, a major topic in bioinformatics and systems biology,[6,7] and network representations have recently also been applied to systematically study relationships between drugs and targets or drug classes.[8−10] Furthermore, network-like similarity graphs (NSGs) have been introduced to systematically analyze multiple SAR features contained in compound activity classes.[11] The analysis of NSGs makes it possible to focus on key compounds that determine SAR characteristics of series of active molecules.[11]

Similar to HTS data analysis, systematic SAR analysis is a challenging task due to the often highly complex nature of SARs and their substantial diversity.[12−14] Only recently, computational models have been introduced to explore SAR characteristics in a systematic manner and quantify them.[15,16] For example, the SAR index (SARI),[15] was originally developed to classify SARs on a large scale and has also been adapted to account for SAR contributions of individual compounds.[11]

From a principle point of view, SAR categories can be best rationalized by considering the topology of "activity landscapes" that account for changes in chemical structures of active compounds and corresponding biological effects.[13,14] Activity landscapes can be visualized as topological maps where compound potency is added as a third dimension to a 2D projection of chemical space.[17] If chemically similar compounds have different potency, the resulting topology is rugged and might contain "activity cliffs".[13] By contrast, chemically diverse compounds having similar potency define regions of smooth topology.

SARI analysis of many different compound activity classes has identified three major SAR categories; "continuous", "discontinuous", and "heterogeneous" SARs.[15] Continuous and discontinuous SARs correspond to smooth and rugged activity landscapes, respectively. Heterogeneous SARs, which are most frequently observed,[14,15] combine continuous and discontinuous elements, e.g., smooth regions in activity landscapes leading to activity cliffs. The SARI formalism has also become an integral part of NSG design in order to complement molecular network representations with quantitative information about local and global SAR features.[11]

Here we have explored whether NSG analysis might help to analyze screening data sets. Compared to the analysis of compounds from hit-to-lead or lead optimization projects, screening set analysis presents additional challenges. For example, compound activity classes available in biologically annotated databases mostly consist of chemically optimized compounds taken from the scientific or patent literature, whereas many screening hits are only weakly active and might contain false positives. Furthermore, hits in cell-based screens might act on different targets, which further complicates SAR analysis.

Despite these potential complications, NSG modeling of diverse screening data sets including inhibition assays and cell-based screens shows that screening hits generally form different local SAR environments. Compound pathways can be identified that contain SAR information and lead to the most potent hits. These findings indicate that molecular network analysis is capable of organizing screening hits on the basis of SAR criteria

* 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: HTS, high-throughput screening; NSG, network-like similarity graph; SALI, structure−activity landscape index; SAR, structure−activity relationship; SARI, structure−activity relationship index; Tc, Tanimoto coefficient.

10.1021/jm8014102 CCC: $40.75 
Published on Web 01/13/2009

**Table 1.** Screening Data Sets[a]

| data set | description | assay type | PubChem AID | hits | SARI |
|---|---|---|---|---|---|
| CYP 2C19 | inhibitors of cytochrome P450 2C19 | inhibition | 899 | 1769 | 0.911 |
| TSHR | agonists of the thyroid stimulating hormone receptor (activators of intracellular cAMP concentrations in parental HEK 293 cells) | cell-based | 938 | 1766 | 0.906 |
| HADH2 | inhibitors of hydroxyacyl CoA dehydroxygenase II | inhibition | 886 | 2434 | 0.964 |

[a] "AID" reports the PubChem bioassay identification code, "Hits" the total number of active compounds per data set, and "SARI" the global SARI score calculated for active compounds in the screening set.

and extracting SAR information from screening data that can be utilized to select series of hits for further analysis.

## Materials and Methods

**Global SAR Index.** SARI consists of two components, the continuity and discontinuity score, which quantify the composition of smooth and rugged parts of an activity landscape, respectively.[8] For a set of active compounds, the continuity score is calculated as a potency-weighted mean of pairwise compound dissimilarity and the discontinuity score as the average potency difference between pairs of similar compounds scaled by their pairwise similarity. Details concerning the score derivation and normalization are provided as Supporting Information. The continuity score assigns high weight to structurally diverse compound pairs having similar potency, whereas the discontinuity score emphasizes structurally similar compounds having significant potency differences (i.e., compounds forming activity cliffs). SARI is calculated as the average between the continuity and the complement of the discontinuity score and ranges from 0 to 1:

$$\mathrm{SARI} = \frac{1}{2}(\mathrm{score}_{\mathrm{cont}} + (1 - \mathrm{score}_{\mathrm{disc}}))$$

High SARI scores reflect a continuous, intermediate scores a heterogeneous and low scores a discontinuous global SAR.

**Compound Discontinuity Scores.** The SARI formalism has also been adapted to assign discontinuity scores on a per compound basis. Thus, to identify compounds that introduce activity cliffs in an activity landscape, a compound discontinuity score is calculated as:

$$\mathrm{raw}_{\mathrm{disc}}(i) = \frac{\sum_{\{j|\mathrm{sim}(i,j)>0.75, i\neq j\}} \mathrm{potdiff}(i,j) \times \mathrm{sim}(i,j)}{|\{j|\mathrm{sim}(i,j) > 0.75, i \neq j\}|}$$

Here potdiff $(i,j)$ is the difference in potency values between compounds $i$ and $j$ and sim$(i,j)$ their similarity, calculated as the MACCS[18] Tanimoto coefficient (Tc).[19] To focus on pairs of compounds that are relevant for the formation of activity cliffs, a threshold of sim$(i,j) > 0.75$ is applied. For each active compound, the score accounts for its individual contribution to overall SAR discontinuity. Molecules with high scores are activity cliff markers. These local discontinuity scores are normalized with respect to all compounds (see Supporting Information) and also yield values within the range [0,1]. The larger the score, the more a compound contributes to SAR discontinuity.

**Network-like Similarity Graphs.** Network representations are used to systematically organize and display similarity and potency relationships within sets of active compounds. Network-like similarity graphs (NSGs) combine five different levels of information (calculation details are provided as Supporting Information). (1) *Pairwise similarity relationships*. Compounds are represented as nodes and connected by an edge if their pairwise similarity exceeds a predefined threshold value (MACCS Tc > 0.75). For screening data sets, inactive compounds can be included in the analysis. (2) *Compound clustering.* Clusters produced by Ward's hierarchical clustering[20] are usually shown on a gray background. For the representation of large screening sets, cluster backgrounds are omitted for clarity. Clustering is based on all pairwise similarity relationships and therefore complements the information provided by binary similarity comparisons leading to edges. Accordingly, edges might exist between compounds belonging to different clusters but compounds within the same cluster are not always connected. (3) *Potency distribution.* A color code is applied to represent the spectrum of available compound potency values ranging from green for lowest to red for highest potency. For screening sets, compound potency values are reported as the negative logarithm of the molar concentration of half-maximal efficacy (termed pAC50). (4) *Local SAR discontinuity*. Nodes are scaled in size according to the magnitude of individual discontinuity scores. (5) *Cluster SARs*. For compound clusters, SARI scores are calculated to distinguish between different local SAR components contained in the data set.

The resulting graph representations are generated with the igraph package of R[21,22] by applying the Fruchterman−Reingold layout algorithm[23] that distributes nodes on the basis of connectivity without scaling distances by similarity values. Thus, the length of edges is varied for representation purposes and has no other meaning. In NSGs, singletons (i.e., compounds without threshold similarity to others) are omitted for clarity.

**SAR Pathway Analysis.** To identify pathways in NSGs that connect compounds in clusters or establish links between regions of low and high SAR discontinuity, a scoring function is introduced that assigns a cost to every edge based on the change in potency between neighboring compounds and their pairwise similarity:

$$\mathrm{cost}(\mathrm{edge}_{ij}) = \frac{(1 + \mathrm{potdiff}(i,j))^a}{\mathrm{sim}(i,j)}$$

By changing the factor $a$, the sensitivity toward activity cliffs can be adjusted. Increasing values of $a$ will result in higher costs for large potency differences. Thus, changes in the factor $a$ will affect the shape of the pathways. If $a$ is set to 1, there will be no difference in costs between a pathway that bridges a given potency difference in a single step and a pathway that gradually ascends to higher potency levels if they are otherwise comparable. Here $a$ is set to 2 in order to detect pathways between different SAR regions and to 4 to systematically explore pathways in compound clusters.

The cost of the entire pathway consisting of $n$ compounds and the edges connecting them is obtained as the sum of the costs of its edges.

$$\mathrm{cost}(\mathrm{path}_{1,n}) = \sum_{i=1}^{n-1} \mathrm{cost}(\mathrm{edge}_{i,i+1})$$

Minimizing the pathway cost function identifies the "lowest cost" path between two selected nodes. These pathways are calculated using the Dijkstra algorithm.[24]

**Screening Data Sets.** The screening data sets utilized in our analysis were assembled from the PubChem[25] database and are

reported in Table 1. Only confirmatory screens with available potency information were selected. Compounds with inconclusive screening results were removed prior to our analysis.

## Results and Discussion

**Network-like Similarity Graphs and Screening Data.** Our graphs represent activity landscapes of compound data sets, elucidate similarity and potency relationships, and identify groups of compounds that are related to each other by different local SARs. This is illustrated in Figure 1 that shows the NSG representation of a set of 146 farnesyltransferase inhibitors taken from the MDDR.[26] With global continuity, discontinuity, and resulting SARI scores of 0.58, 0.71, and 0.44, respectively, this compound set is overall characterized by a heterogeneous SAR. Therefore, one should expect the coexistence of local SAR continuity and discontinuity, which is clearly revealed by the NSG, i.e., clusters of compounds related by continuous SARs are well distinguished from regions containing activity cliffs (Figure 1).

Screening data sets principally differ from compound activity classes extracted from biologically annotated databases because they usually contain a large number of weakly active hits. Consequently, one would expect that global SAR features of screening data should be highly continuous in nature. Consistent with this assumption, we find that SARI scores calculated for screening data sets are generally high. For example, for the data sets analyzed in this study, SARI scores of $>0.9$ were obtained (Table 1). Given the globally continuous nature of screening set SARs, we have systematically analyzed underlying activity landscapes and corresponding local SAR features on the basis of NSGs. In addition, compared to compound sets such as the farnesyltransferase inhibitors discussed above, screening data sets contain inactive compounds, which can also be included in NSG representations. This is shown in Figure 2 that compares NSGs of a compound data set resulting from a screen for inhibitors of cytochrome P450 isoform 2C19 taking active and inactive (Figure 2a) or only active (Figure 2b) compounds into account. Here a total of 4988 compounds including 1769 hits are analyzed and the NSGs are more complex than the one shown in Figure 1 due to the inclusion of many more compounds and pairwise similarity relationships (edges). However, the NSGs in Figure 2 also reveal a notable degree of local SAR heterogeneity in screening data. The most potent hits are distributed in clusters representing different local SAR environments and a number of activity cliffs emerge (i.e., connected larger red and green nodes) These observations are further illustrated by NSGs of various other screening sets shown in Supporting Information Figure S1. Given the finding that NSGs make it possible to focus on regions of local SAR heterogeneity in screening data sets, we next investigated how SAR information might be extracted from these representations.

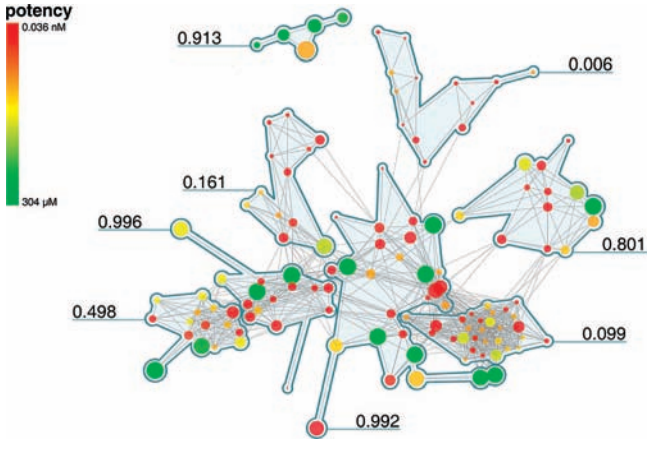


**Figure 1.** Network-like similarity graph of farnesyltransferase inhibitors. Nodes represent compounds and edges similarity relationships. Nodes are color-coded according to their potency using a continuous spectrum from green (lowest potency) to red (highest). Clusters of similar molecules are highlighted and annotated with cluster discontinuity scores. In addition, nodes are scaled in size according to increasing compound discontinuity scores.

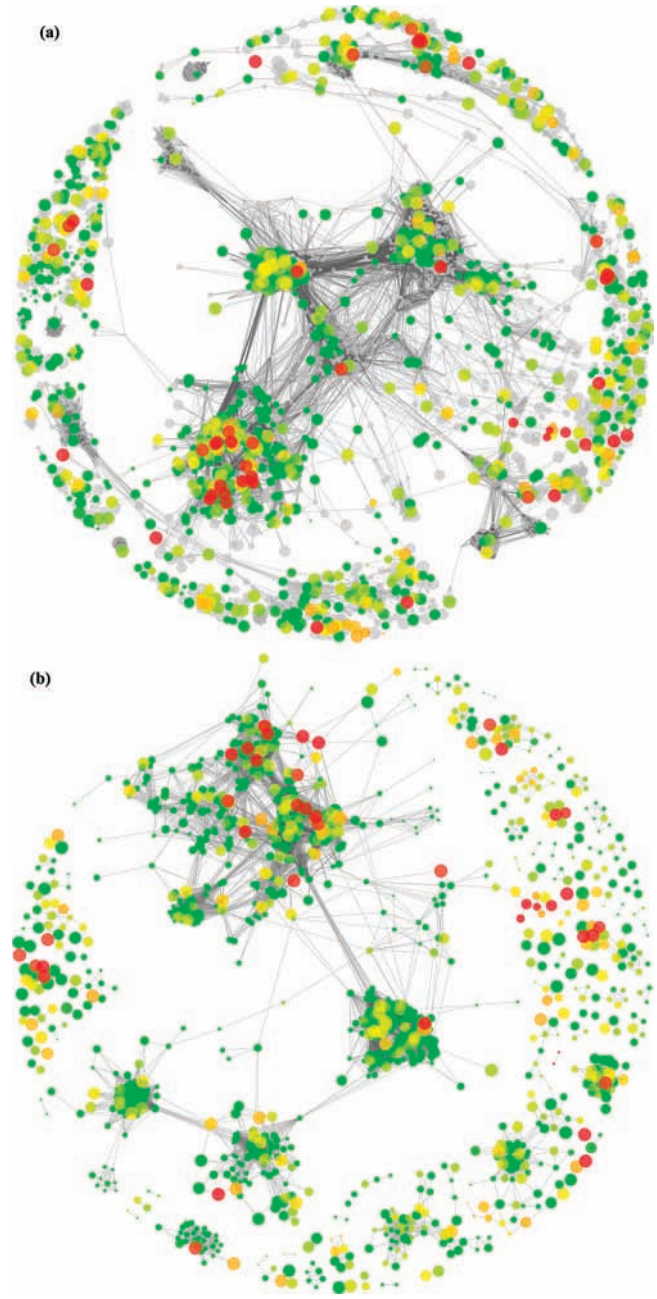


**Figure 2.** NSG representations of a cytochrome P450 screening data set. In (a), the graph for the entire set is shown including active and inactive compounds. Inactive compounds are represented as gray nodes. In (b), only hits are considered, which leads to an NSG with different topology.

**SAR Information and Compound Pathways.** To evaluate local SAR environments revealed by NSGs and relate them to each other, a cost function was designed to screen network components and identify series of compounds with smallest potency differences along the edges. Applying this function

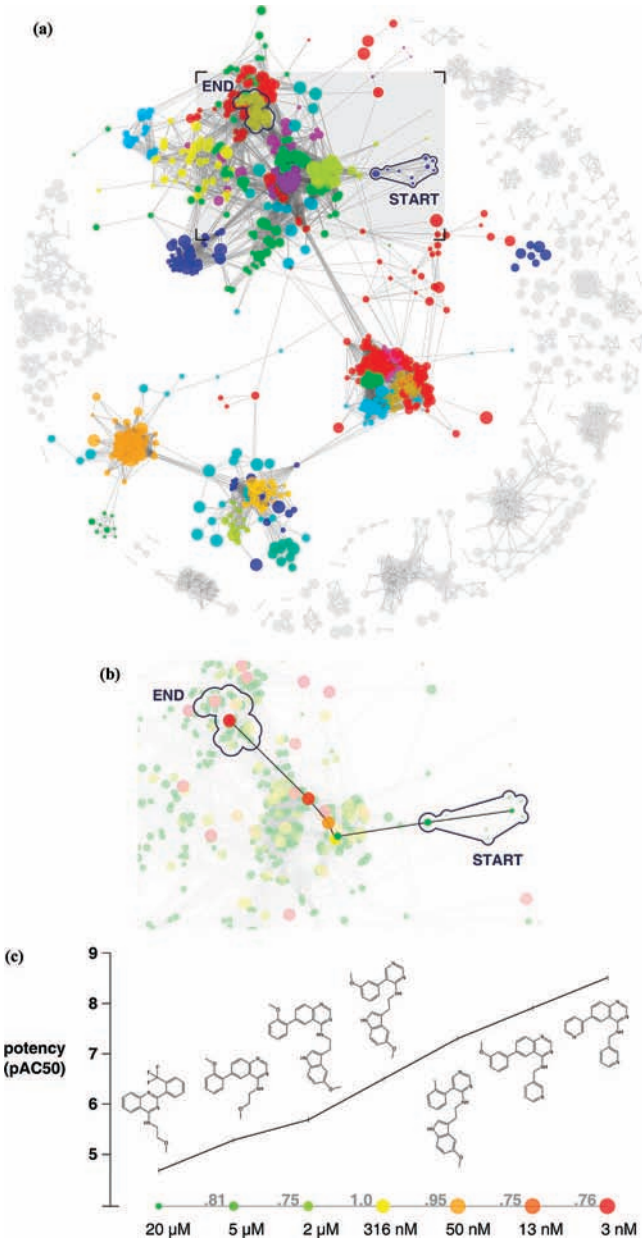


**Figure 3.** Cluster distribution and pathway analysis. In (a), a simplified NSG representation of the P450 hit set (Figure 1b) is shown (without potency information) that highlights the cluster distribution within the largest connected network component. Clusters are displayed in different colors. The gray section is shown in detail in (b). In (b), the lowest cost pathway between a compound in a continuous SAR region and an activity cliff is outlined. The orientation of the graph corresponds to (a). In (c), a two-dimensional representation of this pathway is shown (pathway diagram). The *x*-axis shows scaled nodes of participating compounds with their pairwise similarity values, and the *y*-axis reports their potency ($pAC_{50}$ values). The structures of pathway compounds are displayed in the diagram and their individual potency values are reported.

makes it possible to identify pathways that connect chosen start and end points in the graph, for example, active compounds within a continuous SAR region, i.e., small green or yellow nodes, and other hits forming a local activity cliff, i.e., pairs of large red and green nodes, which are activity cliff markers. Pathways make compound and SAR information immediately available. In Figure 3a, a simplified NSG representation of the P450 data set is shown that highlights the cluster distribution of hits. As can be seen, distinct clusters are present in the data set including compounds that connect different clusters on the

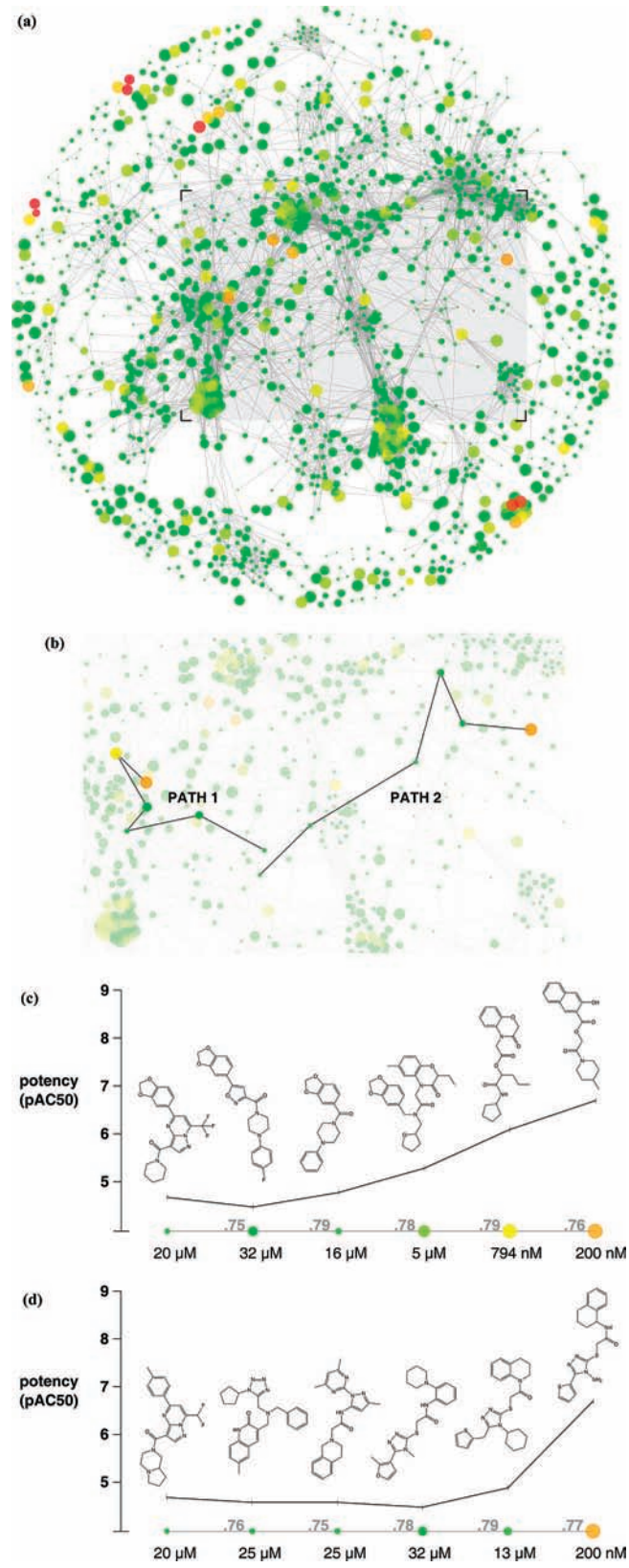


**Figure 4.** Pathway comparison. (a) shows the NSG of active compounds in the thyroid stimulating hormone receptor screening set (the gray section is shown in detail in (b)) and (b) the lowest cost pathways between selected starting compounds and two termini. In (c) and (d), the diagrams of pathways 1 and 2 are shown, respectively.

basis of similarity relationships (thus illustrating the complementarity of global cluster information and binary similarity relationships). As illustrated in Figure 3b, one can define any

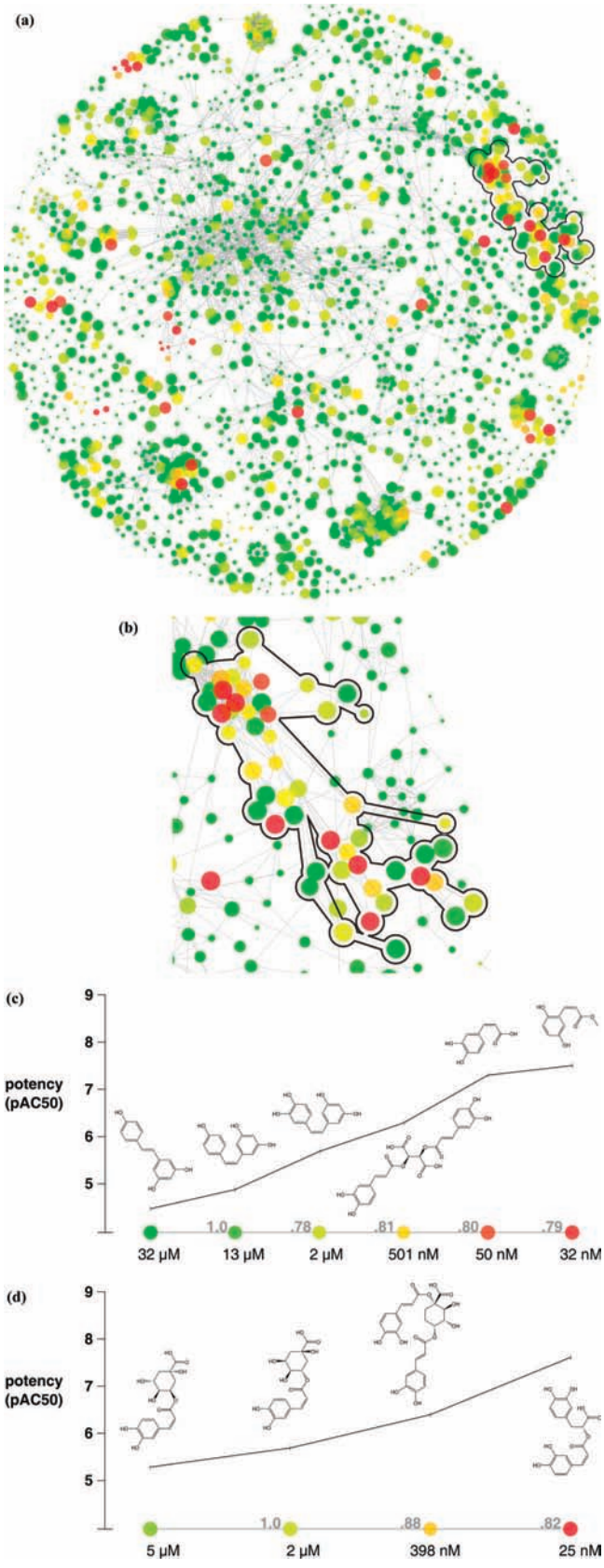


**Figure 5.** Cluster pathways. In (a), the NSG of hits from the hydroxyacyl CoA dehydroxygenase II screening set is shown. A cluster containing activity cliffs is highlighted for which compound pathways were systematically explored. This cluster is displayed in (b) and (c) shows the diagram of a selected cluster pathway with an increase in compound potency spanning 3 orders of magnitude. In (d), the lowest cost pathway leading to alternative activity cliff within the cluster is shown.

compound of interest in the graph as an origin and then calculate the lowest cost pathway to a chosen end point, here an activity cliff. Figure 3c shows the identified pathway and the compounds that form it. In this case, the pathway consists of seven compounds with steadily increasing potency that represent two different chemotypes. Thus, the pathway directly encodes SAR information and alternative chemotypes are available to reach the activity cliff from a region characterized by local SAR continuity.

**Pathways of Different SAR Information Content.** In Figure 4a, the NSG representation of a data set from a cell-based screen for agonists of the thyroid stimulating hormone receptor is shown. The data set is comparable in size to the P450 set discussed above (see Table 1) and spans a similar potency range. However, the NSG illustrates that it is more dominated by weakly active hits (small green nodes) than the P450 set. This cell-based screening data set was also subjected to pathway analysis. In Figure 4b, alternative lowest cost pathways are delineated that capture different SAR information and represent different types of pathways frequently found in screening data. Each pathway is formed by six compounds. Both pathways originate in the most continuous local SAR region but lead to different activity cliffs. Pathway 1 leads to the hit with highest compound discontinuity score in the data set and pathway 2 to the hit with third-highest discontinuity score. Parts c and d of Figure 4 show the pathway diagrams of pathway 1 and 2, respectively. As can be seen, pathway 1 is similar to the P450 inhibitor pathway in Figure 3c; there is an increase in potency along the way, and the pathway compounds display a rather obvious pharmacophore resemblance. By contrast, pathway 2 is essentially flat until it reaches the local activity cliff and its compounds are structurally more diverse. Thus, pathway 1 captures more local SAR information than pathway 2 and the hits forming it would be assigned a higher priority for follow-up studies.

**Systematic Pathway Analysis.** In addition to exploring pathways between chosen start and end points, possible compound pathways in NSG regions can also be systematically explored. For example, one can enumerate all possible pathways within a cluster of hits. Figure 5a shows the NSG representation of a set of 2434 hydroxyacyl CoA dehydroxygenase II inhibitors. A cluster containing apparent activity cliffs is highlighted. This cluster is shown in Figure 5b. It consists of 54 hits for which all possible pathways between compound pairs were calculated. Preselection of pathways containing five or more molecules yielded a total of 867 alternative pathways that were ranked according to potency differences between their start and end points. Seven pathways were found that spanned 3 orders of magnitude in potency. A representative example is shown in Figure 5c. This pathway consists of six compounds representing variations of a dihydroxybenzene core structure and displays a continuous increase in potency. Thus, the identification of such pathways makes it possible to study different series of similar compounds having most significant potency differences.

For practical applications, activity cliffs detected in NSGs present primary targets for hit selection. Systematic calculation of pathways leading to four activity cliffs contained in the targeted compound cluster revealed that in each case multiple pathways were found that originate from different compounds, include different chemotypes, and display steady increases in potency. For example, Figure 5d shows the diagram of the lowest cost pathways leading to one of these activity cliffs. Similar pathways were identified, leading to other cliffs. Thus systematic pathway calculations also sample and organize SAR

information around activity cliffs where compounds are likely selected for further analysis.

## Summary and Conclusions

In this study, we have shown that screening data sets can be organized in network-like similarity graphs that highlight different local SAR environments and reveal the presence of SAR heterogeneity. Compound pathways that connect different local SAR regions or lead to activity cliffs can be identified, on a case-by-case basis or in a systematic manner, and from these pathways, SAR information can be extracted. NSGs calculated for large hit sets with hundreds or thousands of compounds can be readily interpreted. However, when increasingly large numbers of inactive compounds are included in NSG generation, visual interpretation will become increasingly difficult. For practical applications, this might not pose a significant problem because one is mostly interested in exploring SAR relationships among active compounds. However, NSG analysis and the identification of SAR pathways is not dependent on visual interpretation because the required information is available in the form of cluster memberships, similarity relationships, and SARI scores. For example, SARI cluster scores make it possible to organize clusters according to local SAR features and differences in compound discontinuity scores identify activity cliffs. Thus, even highly complex NSGs can be readily mined for SAR information. However, visual interpretation of NSGs provides an intuitive access to the evaluation of hit distributions in screening data and provides a basis for immediately focusing on regions from which compound pathways can be selected as a source of SAR information. The associated cost function can be utilized to systematically scan NSG regions for preferred compound pathways. As discussed, the function can be easily modified in order to balance similarity and potency relationships in different ways and thereby tune pathway analysis for specific applications. Because alternative pathways leading to activity cliffs can be identified in NSGs that contain different compound series and SAR information, systematic graph mining is expected to considerably aid in the selection of hits from screening data sets.

**Supporting Information Available:** Details concerning the SARI methodology and score calculations are provided and NSG representations of various screening data sets are shown. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Gribbon, P.; Lyons, R.; Laflin, P.; Bradley, J.; Chambers, C.; Williams, B. S.; Kighley, W.; Sewing, A. Evaluating Real-Life High-Throughput Screening Data. *J. Biomol. Screen.* **2005**, *10*, 99–107.
(2) Malo, N.; Hanley, J. A.; Cerquozzi, S.; Pelletier, J.; Nadon, R. Statistical Practice in High-Throughput Data Analysis. *Nat. Biotechnol.* **2006**, *24*, 167–175.
(3) Gedeck, P.; Willett, P. Visual and Computational Analysis of Structure−Activity Relationships in High-Throughput Screening Data. *Curr. Opin. Chem. Biol.* **2001**, *5*, 389–395.
(4) Ahlberg, C. Visual Exploration of HTS Databases: Bridging the Gap between Chemistry and Biology. *Drug Discovery Today* **1999**, *4*, 270–485.
(5) Roberts, G.; Myatt, G. J.; Johnson, W. P.; Cross, K. P.; Blower, P. E., Jr. Lead Scope: Software for Exploring Large Sets of Screening Data. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1302–1314.
(6) Alm, E.; Arkin, A. P. Biological Networks. *Curr. Opin. Struct. Biol.* **2003**, *13*, 193–202.
(7) Grigorov, M. G. Global Properties of Biological Networks. *Drug Discovery Today* **2005**, *10*, 365–372.
(8) Paolini, G. V.; Shapland, R. H. B.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. Global Mapping of Pharmacological Space. *Nat. Biotechnol.* **2006**, *24*, 805–815.
(9) Yildirim, M. A.; Goh, K.-I.; Cusick, M. E.; Barabási, A.-L.; Vidal, M. Drug-Target Network. *Nat. Biotechnol.* **2007**, *25*, 1119–1126.
(10) Hert, J.; Keiser, M. J.; Irwin, J. J.; Oprea, T.; Shoichet, B. K. Quantifying the Relationship among Drug Classes. *J. Chem. Inf. Model.* **2008**, *48*, 755–765.
(11) Wawer, M.; Peltason, L.; Weskamp, N.; Teckentrup, A.; Bajorath, J. Structure−Activity Relationship Anatomy by Network-like Similarity Graphs and Local Structure−Activity Relationship Indices. *J. Med. Chem.* **2008**, *51*, 6075–6084.
(12) Kubinyi, H. Similarity and Dissimilarity. A medicinal chemist's view. *Persp. Drug Discovery Des.* **1998**, *9*−*11*, 225–252.
(13) Maggiora, G. M. On Outliers and Activity Cliffs—Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535–1535.
(14) Peltason, L.; Bajorath, J. Molecular Similarity Analysis Uncovers Heterogeneous Structure-Activity Relationships and Variable Activity Landscapes. *Chem. Biol.* **2007**, *14*, 489–497.
(15) Peltason, L.; Bajorath, J. SAR Index: Quantifying the Nature of Structure−Activity Relationships. *J. Med. Chem.* **2007**, *50*, 5571–5578.
(16) Guha, R.; Van Drie, J. H. Structure−Activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model.* **2008**, *48*, 646–658.
(17) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations, and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225–233.
(18) *MACCS Structural Keys*; Symyx Software: San Ramon, CA.
(19) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.
(20) Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.
(21) R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2008.
(22) Csardi, G.; *igraph Library*, version 0.5; Research Institute for Nuclear and Particle Physics: Budapest, 2008; http://cneurocvs.rmki.kfki.hu/igraph(accessed March 7, 2008).
(23) Fruchterman, T. M. J.; Reingold, E. M. Graph Drawing by Force-directed Placement. *Software: Pract. Experience* **1991**, *21*, 1129–1164.
(24) Dijkstra, E. W. A Note on Two Problems in Connection with Graphs. *Numer. Math.* **1959**, *1*, 269–271.
(25) Pubchem; http://pubchem.ncbi.nlm.nih.gov.
(26) *MDL Drug Data Report (MDDR)*, version 2005.2; Symyx Software: San Ramon, CA, 2005.

JM8014102