JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

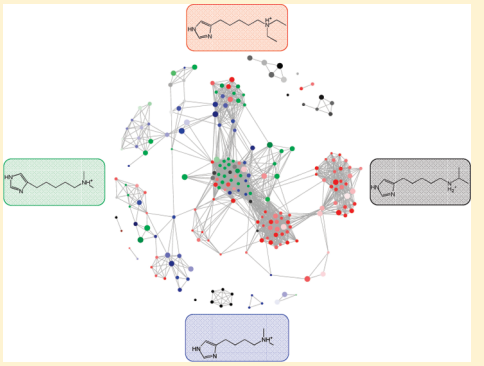
pubs.acs.org/jcim

Molecular Mechanism-Based Network-like Similarity Graphs Reveal Relationships between Different Types of Receptor Ligands and Structural Changes that Determine Agonistic, Inverse-Agonistic, and Antagonistic Effects

Preeti Iyer, Dagmar Stumpfe, 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

ABSTRACT: Receptor ligands might act as agonists, partial agonists, inverse agonists, or antagonists and it is often difficult to understand structural modifications that alter the mechanism of action. In order to compare ligands that are active against a given receptor but have different mechanisms of action, we have designed molecular networks that mirror similarity relationships and incorporate both mechanism of action information and mechanism-specific SAR features. These network representations make it possible to systematically evaluate relationships between different types of receptor ligands and identify communities of structurally very similar ligands with different mechanisms. From a series of such ligands, structural modifications can often be deduced that lead to "mechanism hops".



INTRODUCTION

Pharmacological effects of receptor ligands arise from diverse mechanisms of action.[1,2] For example, ligands might act as agonists, partial agonists, inverse agonists, or antagonists. In general terms, the mechanisms of these types of ligands can be defined as follows: (1) An *agonist* binds to the physiological ligand binding site of a receptor and activates it. (2) An *antagonist* blocks this binding site and thereby prevents receptor activation and signaling. (3) A *partial agonist* also competes with the natural ligand but does not fully activate the receptor. (4) An *inverse agonist* stabilizes an inactive conformation of a receptor and thereby prevents activation and signal transduction.

Mechanistic effects are in general closely linked to different conformational states, or conformational ensembles, of receptors and their ability to interact with effector proteins. G protein-coupled receptors (GPCRs), for which receptor−ligand interactions are just beginning to be understood at the molecular level of detail,[3,4] are prime examples of receptors that engage in highly complex mechanisms controlling functional effects.[3] Differences in the cellular context of GPCR-ligand interactions are also known to alter pharmacological profiles of ligands,[5] giving rise to an intricate network of factors that ultimately determine pharmacological effects. From a medicinal chemistry perspective, it is often difficult to differentiate between different modes of action of GPCR ligands and identify structural determinants of a specific mechanism.[6] Consequently, approaches that help to reveal structural features that influence or determine the mechanism of action of GPCR or other receptor ligands are particularly attractive for medicinal chemistry applications.

In recent years, molecular network representations have been increasingly utilized to systematically account for ligand-target interactions and predict targets of active compounds[7] or analyze structure−activity relationships (SARs).[8] Such molecular networks make it possible to analyze large data sets using a consistent representation frame and often provide graphical access to unexpected ligand-target interactions[7] or complex SAR features.[8] As such, they complement more traditional approaches to analyze ligand-target interactions or SARs.

Given the often complex mechanistic spectrum of receptor ligands, as discussed above, we have been interested in the design of molecular network representations that help to compare ligands with different mechanisms. Therefore, we have generated similarity-based compound networks that incorporate mechanistic and SAR information. These graphical representations make it possible to identify compounds that are related to each other but act by different mechanisms and determine structural features that lead to "mechanism hopping".

MATERIALS AND METHODS

Compound Data Sets. For five different GPCRs, ligand sets were collected from the ChEMBL database.[9] These data sets were assembled to contain ligands having different mechanisms of action including agonists, partial agonists, inverse agonists, and antagonists. The composition of these compound sets is

Received: March 21, 2011

dx.doi.org/10.1021/ci2001378 | *J. Chem. Inf. Model.* XXXX, XXX, 000–000

summarized in Table 1. The sets contained between 148 (AM1) and 307 (AA1) ligands. In two of five cases, S1A and AM1, no inverse agonists were available. All mechanistic annotations for ligands taken from ChEMBL were extracted from original literature sources. Molecules with both agonist and partial agonist annotations were classified as *partial agonists*, and molecules designated as full agonists were classified as *agonists*. In addition, ligands with only inverse agonist or both antagonist and inverse agonist annotations were classified as *inverse agonists*, owing to the observation that compounds that are apparent GPCR antagonists act in many cases by an inverse agonist mechanism.[6] Ligands designated only as antagonists were classified as *antagonists*. As potency annotations, only K_i (or pK_i) were considered. If multiple K_i values were reported for a ligand in original literature sources, the geometric mean was calculated to yield a final potency value.

Network-like Similarity Graphs. The network-like similarity graph (NSG) data structure[10] is a similarity-based molecular network representation that is annotated with additional information layers. Nodes are molecules and edges between them indicate pairwise similarity relationships. Nodes are color-coded using a continuous spectrum to reflect the potency distribution in a compound data set and scaled in size according to the distribution of per-compound discontinuity scores. The per-compound discontinuity score indicates the amount of local SAR discontinuity a compound introduces.[10] Hence, a compound that greatly differs in potency from its immediate structural neighbors makes a large contribution to local SAR discontinuity, and, accordingly, the corresponding node is large. NSGs are generated utilizing the Java implementation in the publicly available SARANEA program[11] and applying a graphical layout algorithm[12] that places densely connected compound subsets in close proximity and separates weakly interconnected regions (compound clusters) from each other. NSGs are usually generated for a set of compounds active (with different potencies) against a given target, i.e. a compound activity class.

Mechanism-Based NSGs. In order to compare receptors ligands with different mechanisms of action using a consistent representation frame, we designed an NSG variant that incorporates mechanism of action information. Therefore, the NSG data structure was modified in different ways, as further detailed in the Results and Discussion section. Pairwise similarity relationships were calculated using the stereochemistry-sensitive Extended Connectivity Fingerprint with bond diameter 4 (ECFP4)[13] as implemented in Pipeline Pilot.[14] An ECFP4 Tanimoto coefficient (Tc) value of 0.4 was applied as the similarity threshold for edges between nodes. This ECFP4 Tc value roughly corresponds to a MACCS structural keys[15] Tc value of 0.8 and indicates the presence of compounds with visible structural similarity. Different from NSGs, nodes were not calculated by potency, but rather by mechanism, using the following color scheme: *agonists*, blue; *partial agonists*, green; *inverse agonists*, gray; *antagonists*, red. For each mechanistic class (compound subset), potency information was conveyed by shading, i.e. for each mechanism color, a continuous shade spectrum from transparent (lowest potency) to opaque (highest potency) was applied. These graphs were implemented in Java, further extending the SARANEA implementation,[11] and termed *Mechanism-based NSGs* (*M-NSGs*). Table 2 summarizes the design elements of M-NSGs.

Table 1. Receptor Ligand Sets[a]

			pK_i	
target receptor	ligand mechanism	no. of compounds	maximum	minimum
adenosine A1	agonist	107	9.8	4.9
receptor (AA1)	partial agonist	54	8.7	5.3
	antagonist	94	9.5	6.0
	inverse agonist	52	9.4	4.2
muscarinic	agonist	26	8.6	3.6
acetylcholine	partial agonist	49	9.5	4.5
receptor M1 (AM1)	antagonist	73	10.0	4.5
dopamine D2	agonist	40	9.0	5.6
receptor (DD2)	partial agonist	44	9.7	6.2
	antagonist	76	9.8	6.6
	inverse agonist	13	11.5	6.1
histamine H3	agonist	44	9.6	4.9
receptor (H3R)	partial agonist	46	9.1	5.0
	antagonist	92	10.0	6.8
	inverse agonist	31	10.1	5.8
serotonin 1a	agonist	46	10.2	5.4
receptor (S1A)	partial agonist	78	10.1	4.9
	antagonist	63	9.4	6.4

[a] Receptor abbreviations are used in the text to designate ligand sets.

RESULTS AND DISCUSSION

Graph Design. M-NSG generation involved different types of calculations, either for an entire ligand set or for each separate mechanism-based subset. First, the graph layout was computed for a complete ligand set after calculating pairwise Tanimoto similarity for all ligands, regardless of their mechanisms of action, thus providing the similarity-based compound network, in analogy to original NSGs. Then, however, similarity- and potency-based discontinuity scores were separately calculated for each ligand subset. On the basis of subset-specific discontinuity scores, the nodes of ligands sharing the same mechanism were scaled in size, hence providing SAR information for each mechanism-based subset. Node scaling is interpreted in the following manner: the larger a node, the higher the degree of SAR discontinuity the compound introduces; combinations of connected medium to large nodes represent discontinuous local SARs and combinations of small nodes continuous local SARs. Finally, mechanism and compound potency information was

Table 2. M-NSG Design Elements

mechanism	color	node size	shading	edge	layout
agonist	blue	mechanism-specific per compound discontinuity score	potency range	Tanimoto similarity >0.4	connectivity-based
partial agonist	green				
antagonist	red				
inverse agonist	gray				

dx.doi.org/10.1021/ci2001378 |J. Chem. Inf. Model. XXXX, XXX, 000–000

incorporated through color and shade coding. Each ligand mechanism was assigned a specific color, and compounds sharing the same mechanism were shaded according to their potency level (within the potency range in the subset). Hence, combinations of the largest transparent and opaque nodes sharing the same color mark the most prominent activity cliffs that occur in each subset. Activity cliffs are structurally similar compounds with very different potency (representing the extreme form of SAR discontinuity).[8] On the basis of these design components, M-NSGs provide similarity and mechanism information across an entire receptor ligand set and, in addition, relative potency and SAR information for each mechanism-specific subset. Thus, M-NSGs also contain all the information provided by NSGs of individual compound sets sharing a specific mechanism of action. Because SAR information can also be obtained from NSGs of individual mechanism-based ligand subsets, we predominantly focus in the following on the exploration of mechanism hopping and underlying structural changes, for which M-NSGs are specifically designed.

Ligand M-NSGs. Figure 1 shows the M-NSG representations for the five receptor ligand sets in Table 1. The M-NSG of the AA1 ligand set in Figure 1a reveals a central graph component (region 1) and several other densely connected compound clusters. In many instances, these clusters mostly, or exclusively, consist of ligands having the same mechanism, e.g. antagonist clusters (red nodes). The central graph component contains partial (green), full agonists (blue), a few inverse agonists (gray), and many differently sized nodes. This indicates the presence of substantial SAR information for agonists in this region. Similarly, the mostly gray-scaled compound community (region 2) at the bottom in Figure 1a contains many inverse agonists with differently sized nodes and a few antagonists having small nodes. Although the AA1 M-NSG displays a notable clustering of compounds by mechanism, there are exceptions, in particular, a densely connected community of ligands with all four mechanisms of action (region 3). Such densely connected mechanistically heterogeneous ligand communities represent prime candidates for further analysis to explore the structural basis of mechanistic changes among similar ligands. In addition, mechanistically more homogeneous regions such as the central graph component in Figure 1a are a source of SAR information for compounds sharing the same or similar mechanisms.

The DD2 M-NSG in Figure 1b represents the smallest of the five data sets and is characterized by the presence of structurally diverse compounds. Structural diversity is mirrored by the low edge density in the graph. The most notable feature of the DD2 M-NSG is its largest ligand community in the center of Figure 1b (region 1) that also contains compounds with all four mechanisms of action.

By contrast, the H3R M-NSG in Figure 1c is characterized by the presence of a densely connected central graph component that essentially consists of four separate ligand communities that are connected via compound bridges. These include two antagonist communities (regions 1 and 2) with different node size and potency distributions and, in addition, two other communities that are characterized by distinct mechanistic heterogeneity (regions 3 and 4). In particular, the community in the center of Figure 1c (region 4) consists of very similar ligands that cover the entire spectrum of mechanisms and thus provides a focal point for further analysis.

Different from the ligand sets discussed so far the remaining AM1 and S1A sets in Figure 1d and 1e, respectively, do not contain inverse agonists. The AM1 M-NSG shows a notable clustering of different series of compounds by mechanism. However, in some cases, individual ligands with different mechanisms occur in an otherwise mechanistically homogeneous cluster including, for example, a weakly potent antagonist found in a partial agonist community (region 1 in Figure 1d), another single antagonist in an agonist/partial agonist community (region 2), and a weakly potent partial agonist within an antagonist community (region 3). Such observations might raise the question as to whether the mechanisms of these individual ligands located in an otherwise mechanistically homogeneous environment have been correctly identified and might thus suggest further experimental evaluation. In addition, another small community (region 4) shows a sequence of agonists with increasing potency where structural neighbors of potent agonists include partial agonists and a weakly potent antagonist, which represents another interesting and perhaps puzzling mechanistic pattern. Furthermore, another ligand community is encircled in Figure 1d (region 5) that contains multiple agonists, partial agonists, and antagonists with different potencies.

The S1A M-NSG in Figure 1e also contains both mechanistically homogeneous and heterogeneous ligand communities. For example, two densely connected ligand communities are observed (regions 1 and 2) that each comprise multiple partial agonists and multiple antagonists. Furthermore, region 3 in the S1A M-NSG contains a pair of compounds representing an agonist/antagonist hop and a small community of structurally related ligands including an antagonist, an agonist, and two partial agonists.

Mechanism Hopping. Selected mechanistically heterogeneous ligand communities were analyzed in detail to explore structural modifications that lead to mechanistic changes. M-NSG regions shown in Figure 2 are labeled with red numbers in Figure 1. In each community, a series of analogs were identified that revealed structural changes altering their mechanisms of action.

The series of AA1 ligands in Figure 2a includes agonists, partial agonists, inverse agonists, and an antagonist. All of these ligands are analogs and only distinguished by different substituents at the same site, i.e. a meta-position of the phenyl ring. An agonist (ligand 1) contains a hydroxyl group at this position. Ligands with a methoxy or methyl group (2 and 3) are partial agonists, and the same mechanism is observed for a fluorine substituent (4). However, changing the fluorine atom to a difluoromethylether group converts an agonist (4) into an antagonist (5). Moreover, changing this group to either a trifluoromethylether or trifluoromethyl substituent generates inverse agonists (6 and 7). Thus, ligands taken from a mechanistically heterogeneous region in the AA1 M-NSG reveal substitutions at a single site that alter the mechanism of action in different ways.

The largest ligand community in the DD2 M-NSG contains three tetraline analogs with different mechanisms that are shown in Figure 2b. These analogs include an agonist (ligand 1), antagonist (2), and inverse agonist (3). Here substitutions at multiple positions in both rings and different stereochemistry at the aliphatic ring distinguish the inverse agonist and antagonist from the agonist.

Five ligands from a mechanistically highly heterogeneous region of the H3R M-NSG shown in Figure 2c include an agonist, a partial agonist, two antagonists, and an inverse agonist. Their structures are distinguished by the length of the aliphatic linker between the imidazole ring and the terminal amine and, in

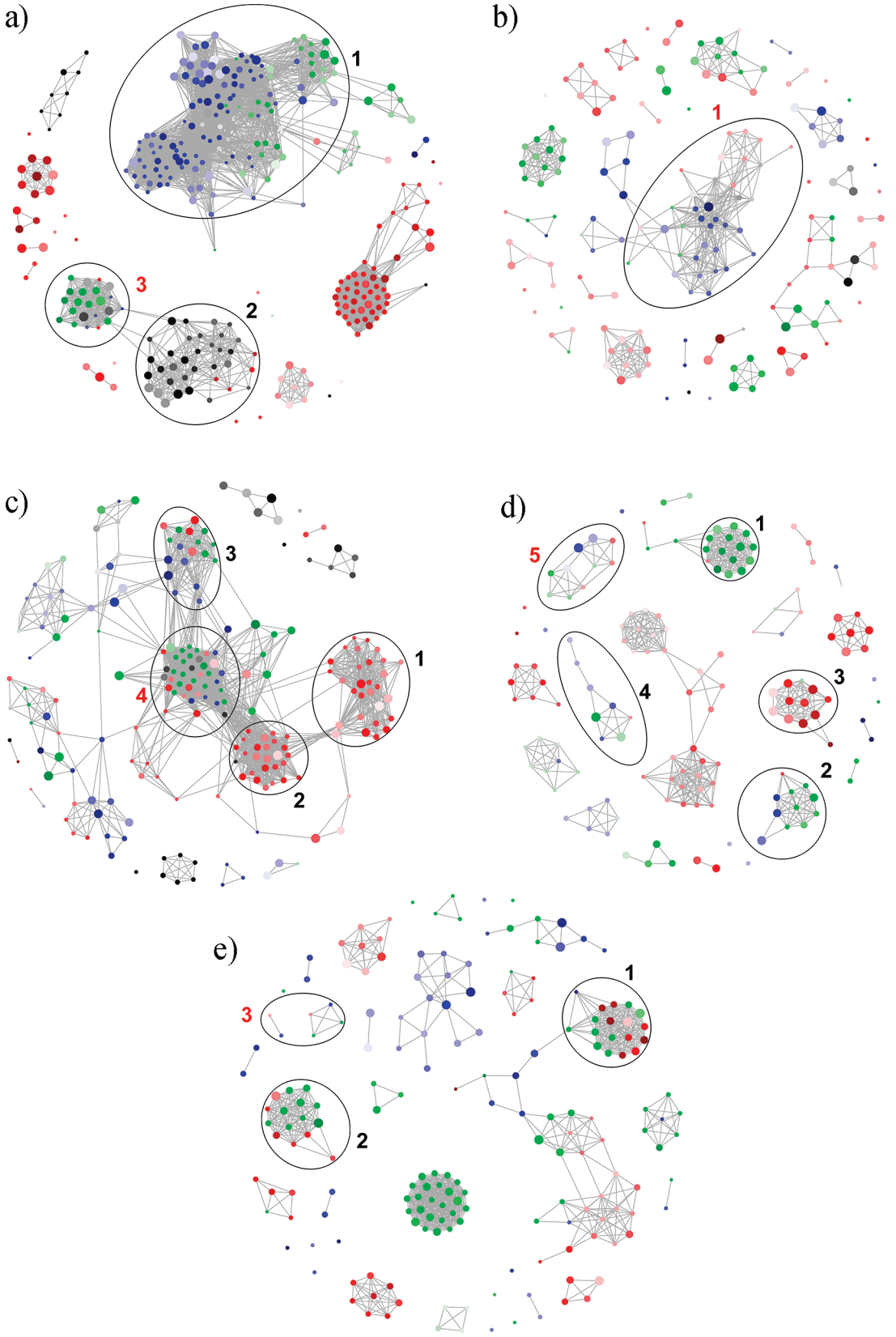


Figure 1. M-NSGs. Graph representations are shown for each complete ligand set according to Table 1 (agonists, blue; partial agonists, green; inverse agonists, gray; antagonists, red). Regions of the graph discussed in the text are encircled and numbered. Regions with red numbers are shown in detail in Figure 2. (a) AA1, (b) DD2, (c) H3R, (d) AM1, (e) S1A. In each M-NSG, selected ligand communities are indicated.

dx.doi.org/10.1021/ci2001378 |*J. Chem. Inf. Model.* XXXX, XXX, 000–000

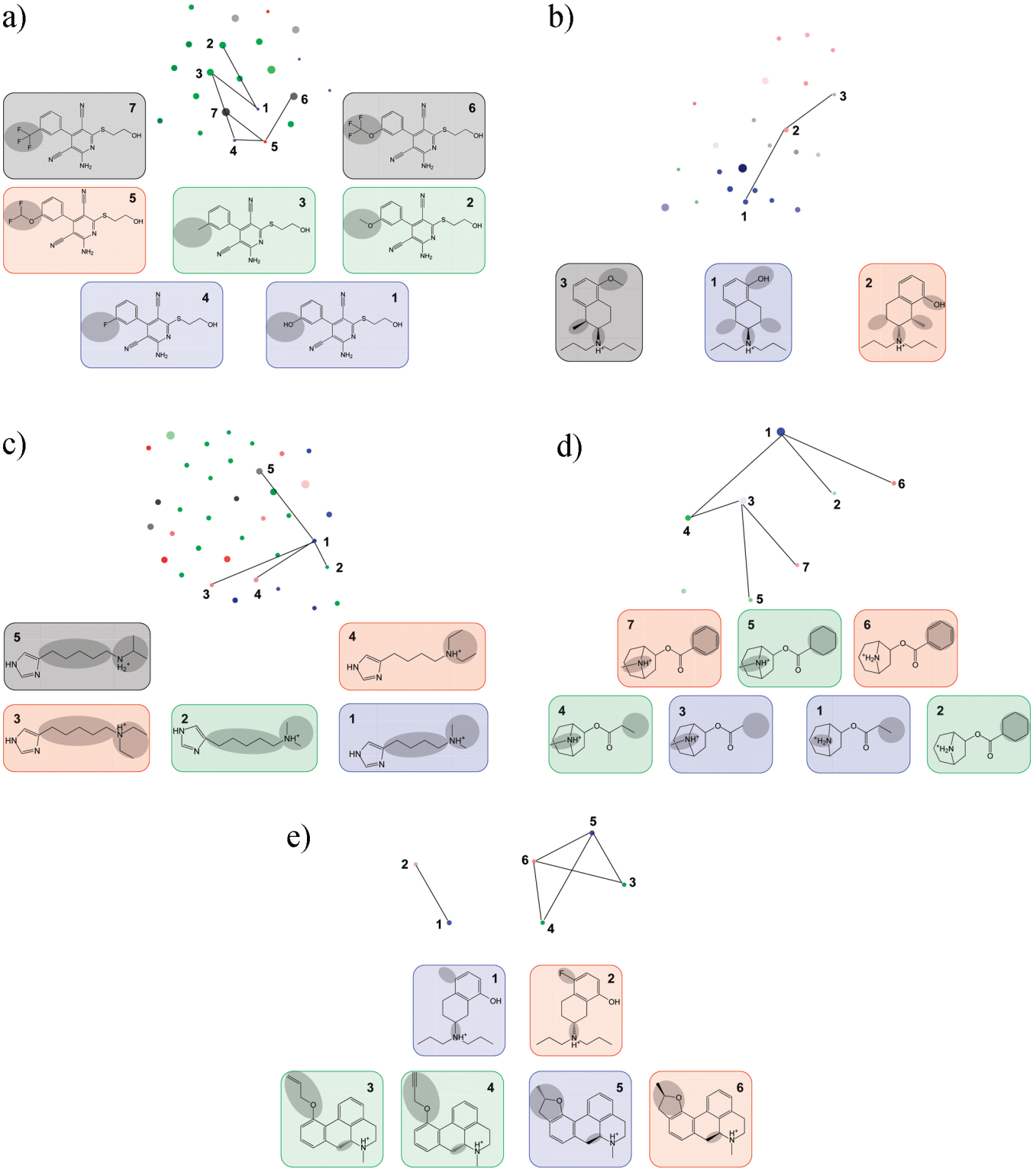


Figure 2. Ligand communities. For each data set, selected ligand communities are displayed (communities with red numbers in Figure 1). For each community, a series of ligands is shown that include mechanism hops. In the community graphs, nodes representing pairs of these ligands that constitute mechanism hops are connected by black edges. Compounds and the corresponding nodes are numbered. Structural modifications in ligands are highlighted, and they are displayed on a background representing their node colors (mechanisms). (a) AA1, (b) DD2, (c) H3R, (d) AM1, (e) S1A.

addition, by substitutions at this amino group. Comparison of ligands 1−4 reveals that the linker length is not responsible for agonistic versus antagonistic effects. Rather, the introduction of an N,N-diethylamino group (in ligands 3 and 4) generates antagonists. If the N,N-diethylamine is changed to an N-tertiary butyl group, an inverse agonist is obtained. Clearly, mechanistic changes can in this case be attributed to the substitutions at the terminal amino group.

dx.doi.org/10.1021/ci2001378 |J. Chem. Inf. Model. XXXX, XXX, 000–000

The series of AM1 ligands in Figure 2d contains five agonists or partial agonists (ligands 1−5) and two antagonists (6 and 7). Both agonists and antagonists contain two forms of a bridged hetereoaliphatic ring, which can thus not be responsible for changes in the mechanism of action. By contrast, mechanism hopping from agonists to antagonists is caused by the introduction of a phenyl substituent at the ester moiety, instead of a small aliphatic group or a cyclohexyl ring. Comparisons of ligands 2 and 6 and of ligands 5 and 7 reveal that the replacement of the cyclohexyl by the benzene ring, i.e. the introduction of an aromatic ring at this position, converts partial agonists into antagonists, another well-defined chemical modification.

In Figure 2e, a pair of S1A ligands is shown (1 and 2) that are also tetraline derivatives, similar to the ligands in Figure 2b. This chemotype is active against both dopamine and serotonin receptors. The two ligands in Figure 2e represent a mechanism hop from an agonist to an antagonist. In both analog series in Figure 2b and 2e, we observe that the stereochemistry at the nitrogen is a differentiating feature between agonists and antagonists. Ligands 1 and 2 in Figure 2e are further distinguished by a fluorine substituent. In addition, Figure 2e also shows another series of S1A ligands (3 to 6) containing a condensated ring system as their core structure. Here different core ring stereochemistry is observed as well as modifications at a phenyl moiety. The two partial agonists (ligands 3 and 4) differ from the full agonist (5) in a ring stereoisomer and a methyl-tetrahydrofuran fused to the phenyl moiety that is only present in the full agonist. However, the agonist and the antagonist (6) in this series are nearly identical; they only differ in the stereoisomer of the methyl substituent at the tetrahydrofuran ring. Thus, in this case, a subtle stereochemical difference involving a single methyl group in chemically complex and rigid receptor ligands triggers a change in the mechanism of action from an agonist to an antagonist.

CONCLUSIONS

Herein we have introduced a graphical analysis tool that incorporates molecular mechanism of action information. The M-NSG data structure makes it possible to graphically analyze sets of receptor ligands with different mechanisms of action and identify mechanistically heterogeneous communities of structurally similar compounds. In the M-NSG implementation, nodes are directly associated with compound structures for interactive display. Hence, compound subsets can be easily selected and further analyzed to explore structural modifications that might lead to mechanistic changes. M-NSG analysis has been carried out for five sets of GPCR ligands acting by three or four different mechanisms. In a number of instances, well-defined structural changes were identified in analog series prioritized on the basis of M-NSG analysis that distinguished between ligands with different mechanisms. For closely related receptors (e.g., isoforms), it might also be possible to pool active compounds by mechanism and study these sets in M-NSGs in order to search for structural changes that might be responsible for similar mechanistic effects across multiple receptors.

AUTHOR INFORMATION

Corresponding Author

*Phone: +49-228-2699-306. Fax: +49-228-2699-341. E-mail: bajorath@bit.uni-bonn.de.

REFERENCES

(1) Kenakin, T. Principles: Receptor Theory in Pharmacology. *Trends Pharmacol. Sci.* **2004**, *25*, 186–192.

(2) Zhu, B. T. Mechanistic Explanation for the Unique Pharmacologic Properties of Receptor Partial Agonists. *Biomed. Pharmacol.* **2005**, *59*, 76–89.

(3) Rosenbaum, D. M.; Rasmussen, S. G.; Kobilka, B. K. The Structure and Function of G Protein-Coupled Receptors. *Nature* **2009**, *459*, 356–363.

(4) Sprang, S. R. Binding the Receptor at Both Ends. *Nature* **2011**, *469*, 172–173.

(5) Nelson, C. P.; Challiss, R. A. 'Phenotypic' Pharmacology: the Influence of Cellular Environment on G Protein-Coupled Receptor Antagonist and Inverse Agonist Pharmacology. *Biochem. Pharmacol.* **2007**, *73*, 737–751.

(6) Greasley, P. J.; Clapham, J. C. Inverse Agonism or Neutral Antagonism at G Protein-Coupled Receptors: a Medicinal Chemistry Challenge worth Pursuing? *Eur. J. Pharmacol.* **2006**, *553*, 1–9.

(7) Keiser, M. J.; Irwin, J. J.; Shoichet, B. K. The Chemical Basis of Pharmacology. *Biochemistry* **2010**, *49*, 10267–10276.

(8) Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity Landscape Representations for Structure-Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209–8223.

(9) *ChEMBL*; European Bioinformatics Institute (EBI): Cambridge, 2010. http://www.ebi.ac.uk/chembl/ (accessed March 2, 2011).

(10) 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.

(11) Lounkine, E.; Wawer, M.; Wassermann, A. M.; Bajorath, J. SARANEA: A Freely Available Program to Mine Structure−Activity and Structure−Selectivity Relationship Information in Compound Data Sets. *J. Chem. Inf. Model* **2010**, *50*, 68–78.

(12) Fruchterman, T. M. J.; Reingold, E. M. Graph Drawing by Force-directed Placement. *Software − Pract. Exper.* **1991**, *21*, 1129–1164.

(13) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.

(14) *Scitegic Pipeline Pilot*, Student ed.; Version 6.1; Accelrys, Inc.: San Diego, CA, 2007.

(15) *MACCS Structural Keys*; Symyx Software: San Ramon, CA, 2005.