WILEY
Publishers Since 1807

COLOR REPRODUCTION IN YOUR ARTICLE

If color figures were included in your original manuscript, this page applies. Because of the high cost of color printing, we can only print figures in color if authors cover the expense.

Please indicate if you would like your figures to be printed in color or black and white. Color images will be reproduced online in Wiley *Online Library* at no charge, whether or not you opt for color printing.

Failure to return this form will result in the publication of your figures in black and white.

JOURNAL Drug Development Research VOLUME _____ ISSUE _____

TITLE OF MANUSCRIPT _____

MS. NO. _____ NO. OF COLOR PAGES _____ AUTHOR(S) _____

No. Color Pages	Color Charges	No. Color Pages	Color Charges	No. Color Pages	Color Charges
1	**500**	**5**	**2500**	**9**	**4500**
2	**1000**	**6**	**3000**	**10**	**5000**
3	**1500**	**7**	**3500**	**11**	**5500**
4	**2000**	**8**	**4000**	**12**	**6000**

*****Please contact the Production Editor for a quote if you have more than 12 pages of color*****

☐ Please print my figures in black and white

☐ Please print my figures in color $ _____

BILL TO:
Name _____ **Purchase Order No.** _____

Institution _____ Phone _____

Address _____

Fax _____

E-mail _____

Additional Reprint Purchases

Should you wish to purchase additional copies of your article, please click on the link and follow the instructions provided:
https://caesar.sheridan.com/reprints/redir.php?pub=10089&acro=DDR

Corresponding authors are invited to inform their co-authors of the reprint options available.

Please note that regardless of the form in which they are acquired, reprints should not be resold, nor further disseminated in electronic form, nor deployed in part or in whole in any marketing, promotional or educational contexts without authorization from Wiley. Permissions requests should be directed to mail to: permissionsus@wiley.com

For information about 'Pay-Per-View and Article Select' click on the following link: http://www3.interscience.wiley.com/aboutus/ppv-articleselect.html

Please follow these instructions to avoid delay of publication.

☐ **READ PROOFS CAREFULLY**

- This will be your only chance to review these proofs**. Please note that once your corrected article is posted online, it is considered legally published, and cannot be removed from the Web site for further corrections.**
- Please note that the volume and page numbers shown on the proofs are for position only.

☐ **ANSWER ALL QUERIES ON PROOFS** (Queries for you to answer are attached as the last page of your proof.)

- Mark all corrections directly on the proofs. Note that excessive author alterations may ultimately result in delay of publication and extra costs may be charged to you.

☐ **CHECK FIGURES AND TABLES CAREFULLY**

- Check size, numbering, and orientation of figures.
- All images in the PDF are downsampled (reduced to lower resolution and file size) to facilitate Internet delivery. These images will appear at higher resolution and sharpness in the printed article.
- Review figure legends to ensure that they are complete.
- Check all tables. Review layout, title, and footnotes.

RETURN ☐ **PROOFS**
☐ **CTA (If you have not already signed one)**

RETURN ALL RELEVANT ITEMS WITHIN TWO BUSINESS DAYS OF RECEIPT TO:
Jrnlprod.DDR@cenveo.com

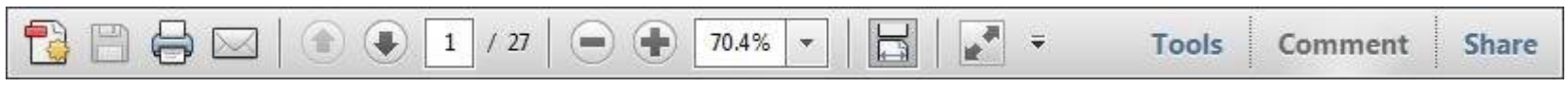
Production Editor, DDR
Journal Production
E-mail: Jrnlprod.DDR@cenveo.com

Refer to journal acronym (DDR) and article production number

USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

WILEY

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:



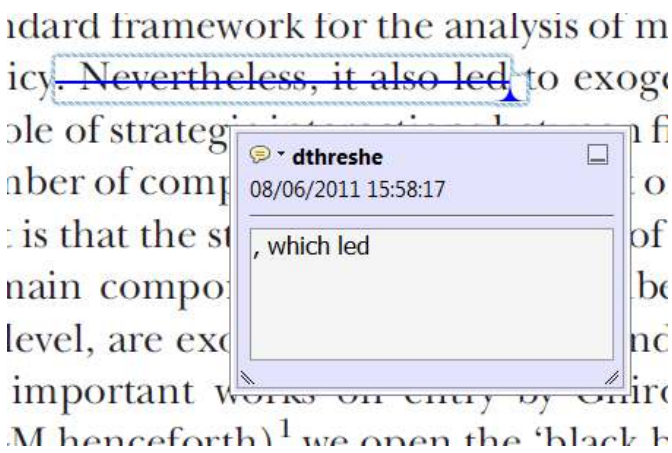
This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:

1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it

- Highlight a word or sentence.
- Click on the Replace (Ins) icon in the Annotations section.
- Type the replacement text into the blue box that appears.

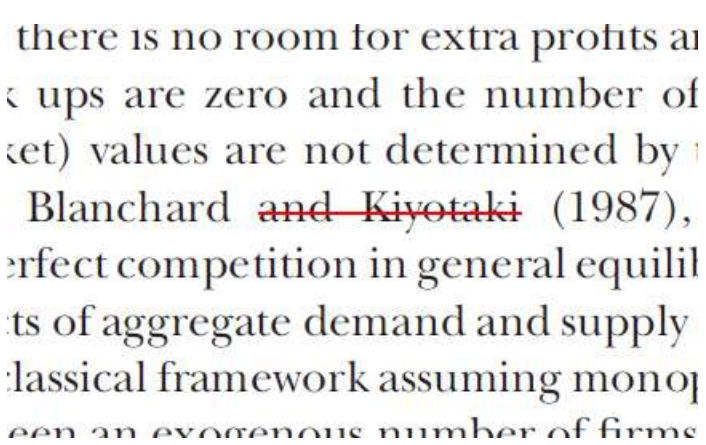


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it

- Highlight a word or sentence.
- Click on the Strikethrough (Del) icon in the Annotations section.

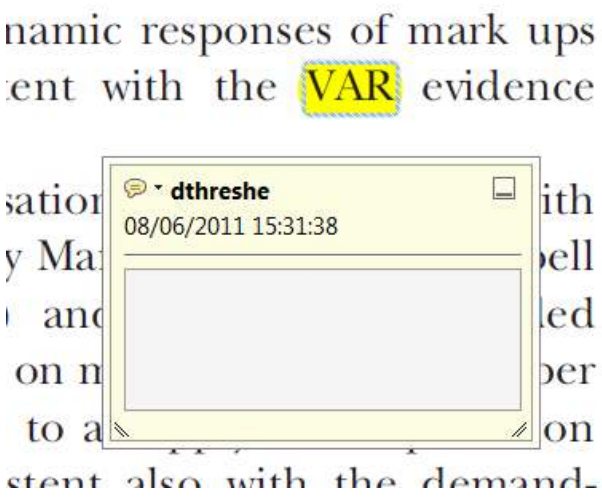


3. Add note to text Tool – for highlighting a section to be changed to bold or italic.

Highlights text in yellow and opens up a text box where comments can be entered.

How to use it

- Highlight the relevant section of text.
- Click on the Add note to text icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.

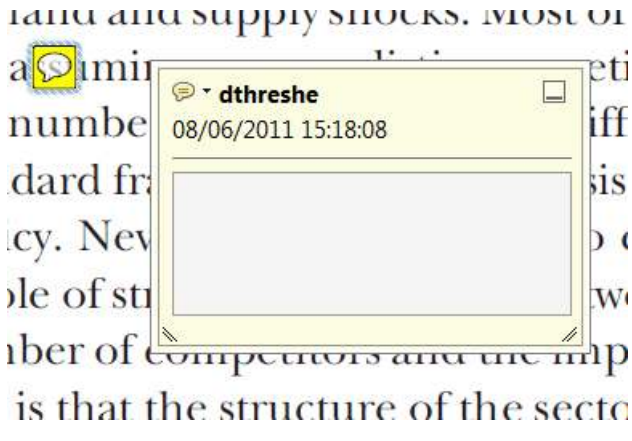


4. Add sticky note Tool – for making notes at specific points in the text.

Marks a point in the proof where a comment needs to be highlighted.

How to use it

- Click on the Add sticky note icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.

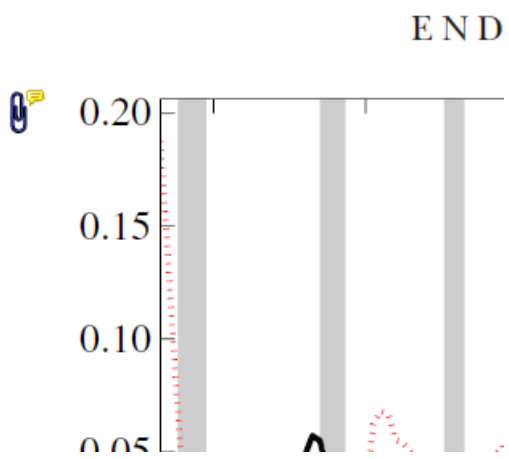


5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate pace in the text.

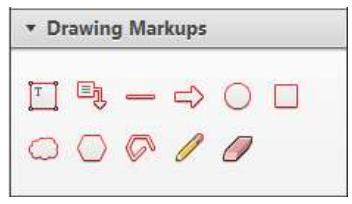
How to use it

- Click on the Attach File icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.



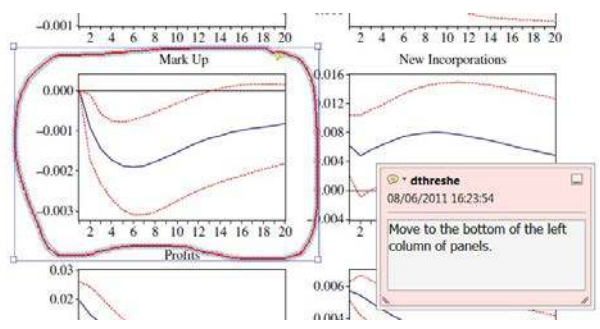
6. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.

Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks.



How to use it

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.



DRUG DEVELOPMENT RESEARCH 75 : ••–•• (2014)

DDR

Research Overview

Network Measures for Chemical Library Design

N. Sukumar,[1,2]* M. P. Krein,[3] G. Prabhu,[1] S. Bhattacharya,[2,4] and S. Sen[1,2]

[1]*Department of Chemistry, Shiv Nadar University, Dadri, Gautam Budh Nagar, U.P. 201314, India*

[2]*Center for Informatics, Shiv Nadar University, Dadri, Gautam Budh Nagar, U.P. 201314, India*

[3]*Rensselaer Exploratory Center for Cheminformatics Research, Department of Chemistry, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

[4]*Department of Mathematics, Shiv Nadar University, Dadri, Gautam Budh Nagar, U.P. 201314, India*

Strategy, Management and Health Policy				
Enabling Technology, Genomics, Proteomics	Preclinical Research	Preclinical Development Toxicology, Formulation Drug Delivery, Pharmacokinetics	Clinical Development Phases I-III Regulatory, Quality, Manufacturing	Postmarketing Phase IV

ABSTRACT In this overview, we examine recent developments in network approaches to drug design. A brief overview of networks is followed by a discussion of how chemical similarity networks and their properties address challenges in drug design. Multiple methods used to assess or enhance chemical diversity for early-stage drug discovery are discussed, as well as methods that can be used for drug repositioning and ligand polypharmacology. Drug Dev Res 75 : ••–••, 2014.

Key words: chemical networks; drug design; diversity; similarity

INTRODUCTION TO MOLECULAR NETWORKS

The last two decades have seen an explosion of interest in the applications of network concepts in the biological and social sciences. While traditional science takes a reductionist, bottom-up approach, seeking explanations of natural phenomena in terms of the interactions of their individual constituents, network science looks for emergent properties of systems, taking the topology of interactions or interconnections as fundamental [Csermely et al., 2013]. A network or graph—in this overview, "graph" and "network" will be used interchangeably—as any real-life system, either physical or postulated, is amenable for study using the mathematics of graph theory. The building blocks of graphs are nodes, usually represented pictorially as circles, and edges, with lines connecting nodes. In its most general sense, a graph is a mathematical object comprising a 3-tuple or triplet, given by $G = (V, E, \psi_G)$, where V is a nonempty set whose elements are termed vertices, E is a set of edges that essentially form links between vertices, and the incidence function, ψ_G, defines the above link by mapping each edge to a pair of vertices. The nodes may all belong to a single class, as in a unipartite graph, or to distinct classes, as in a bipartite graph. The edges may reflect an unsymmetrical relationship between the nodes, leading to a directed graph, or a symmetrical relationship, leading to an undirected graph. Molecular networks may thus be defined as graphs where the nodes represent distinct molecular entities, whether small molecules, macromolecules, or molecular fragments [Csermely et al., 2013; Bolouri, 2014].

Special Issue: Genes, Networks and Bioinformatics in Drug Discovery

*Correspondence to: N. Sukumar, Department of Chemistry, Shiv Nadar University, Dadri, Gautam Budh Nagar U.P. 201314, India.
E-mail: n.sukumar@snu.edu.in

Received ••; Accepted ••

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/ddr.21218

Three distinct classes of molecular networks may be distinguished, based on the specific relationship represented by the edges:

1. Interaction networks
2. Chemical transformation/reaction networks
3. Similarity networks

In *interaction networks*, the edges represent physical contacts between molecules, or intermolecular interactions. Such networks include protein–protein interaction networks (a unipartite graph), drug–target networks [Yildirim et al., 2007], and transcription factor–gene binding site networks [Matys et al., 2003] (bipartite graphs). Metabolic networks belong to the second class, where the nodes represent metabolites and the edges represent enzymatic reactions or chemical transformations connecting them. Both these kinds of networks are widely studied in systems biology. Chemical synthetic schemas, like that in Figure 1, and synthetic accessibility networks [Boda et al., 2007] also belong to this latter class of networks.

The third category includes molecular networks where the edges represent a chemical or mathematical similarity relationship between molecules. Chemical similarity can include obvious similarity in the molecular framework or scaffold, and similarity in molecular properties or biological activities. Often (but not always) the two go together, an expression of the molecular similarity principle [Pearlman and Smith, 1998; Martin et al., 2002]. Molecular similarity is also amenable to quantitative assessment—the subject of molecular similarity analysis. This is accomplished through a set of descriptors or molecular fingerprints (numerical representations of molecular structure), together with a similarity metric (an algorithm that computes a numerical similarity measure from a pair of molecular descriptors or fingerprints). It is this class of network that will be the focus of the present overview. Such networks are said to span a chemical space—an abstract representation of the molecular network. The structure and dynamics of molecular networks have been extensively reviewed by Csermely et al. [2013]. Network information has been employed to predict cross-reactivity assessment [von Eichborn et al., 2011] and for drug repositioning [Das et al., 2010b].

MOLECULAR SIMILARITY ANALYSIS

The goal of molecular similarity analysis is to find molecules that are chemically and structurally similar to known drugs or drug leads. This is because structurally similar molecules are presumed to display similar activities in biological assays (the similarity principle of quantitative structure–activity relationships—QSAR). Similarity can be assessed using any combination of molecular descriptors and similarity metric. There are thousands of descriptors that can be generated with readily available software tools, and many kinds of descriptor representations of molecules, from constitutional and topological descriptors that can be computed simply from the chemical formula or the molecular graph (two-dimensional structure), fragment-based descriptors to surface area descriptors, shape descriptors and electron density-derived descriptors, that are sensitive to the molecular conformation and capture the physics of electrostatic and nonbonded intermolecular interactions. The advantages and shortcomings of different families of molecular descriptors have been the subject of considerable research and debate [Todeschini and Consonni, 2000; Sukumar et al., 2012]. The choice of descriptors sensitively affects the computed similarity—for instance, two molecules might be constructed from the same molecular scaffold and thus be very similar in size and shape, but have very different properties because of the different chemical natures of the functional groups or substituent atoms. Conversely, molecules with very different molecular scaffolds might look similar for binding to a protein, because of a similar shape and electrostatics. The pharmaceutical literature is replete with examples of drugs where one enantiomer is several orders of magnitude more potent than the other, or even where one is an agonist while the other is an antagonist [Sukumar et al., 2012]. Thus, it is not an exaggeration to say that similarity lies in the eye of the

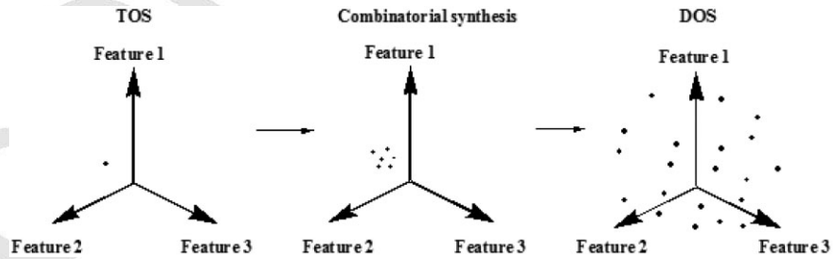


Fig. 1. Distribution of molecules obtained from different synthetic pathways (*left*: target-oriented synthesis TOS, *middle*: combinatorial synthesis, and right: diversity-oriented synthesis DOS) in a three-dimensional chemical space.

beholder. In the final analysis, a choice must be made in dataset representation, in terms of the myriad chemical descriptors and comparison methods available. This decision is usually based on prior knowledge, to strike a balance between interpretation of the significance of changes in that space (low-dimensional structure of the chemical space) and the fidelity of reproduction [Krein et al., 2012].

A good descriptor representation leads to a smooth structure–property or structure–activity relationship, whereas an inappropriate descriptor representation for a particular application leads to a rough structure–activity landscape, or the presence of "activity cliffs" [Maggiora, 2006]. Such activity cliffs represent failures of the similarity principle of QSAR, but they also represent opportunities for medicinal chemists to quantitatively improve the potency of a drug lead through small structural modifications. Much interest has therefore been devoted to the identification and quantification of activity cliffs [Stumpfe and Bajorath, 2012]. One popular measure of the roughness of a structure–activity landscape is the structure–activity landscape index, SALI [Guha and Van Drie, 2008b] (Table 1 and Fig. 2):

$$\mathrm{SALI}_{ij} = \frac{|A_i - A_j|}{\{1 - sim(i,j)\}},$$

where A_i and A_j are the activities of molecules i and j in some bio-assay, and $sim(i,j)$ is a coefficient of similarity between the molecules i and j. The SALI is computed for all pairs of molecules in a dataset, and a cutoff value for SALI is chosen for constructing the SALI graph: any pair of molecules i, j with $\mathrm{SALI}_{i,j}$ greater than the cutoff is connected by a SALI edge. The SALI graph thus focuses attention on the steepest activity cliffs [Guha and Van Drie, 2008a], i.e., pairs of structurally similar molecules displaying large differences in their activities toward some biological target. Plotting the SALI value against the similarity threshold (the SALI curve) gives a measure of the ability of a QSAR model to correctly rank order pairs of molecules by activity. The value of the SALI curve at the highest similarity threshold is a measure of the model's ability to correctly identify the steepest activity cliffs; the value at zero similarity threshold measures the ability of the model to correctly rank order all the molecules.

From the above discussion, it should be clear that activity cliffs also represent boundaries between the domains of applicability of different structure–activity relationships or QSAR models. Any QSAR model is valid for predictions within its domain of applicability, but fails when attempting to make predictions across an activity cliff. A cluster of k molecules connected through a similarity relationship forms an undirected graph that forms the basis of the k-nearest neighbor prediction method. The SALI graph, on the other hand, is a directed network, connecting molecules across activity cliffs, with each edge leading (by convention) from the molecule of lower activity in the pair to that of higher activity (Fig. 2).

Network-like similarity graphs [Wawer et al., 2008] represent another way in which to display structure–activity landscapes and identify discontinuities therein. Here, nodes are color-coded by the potency, and the size of each node is proportional to the local discontinuity in potency. Such network measures are important in the analysis and design of molecular libraries, and for construction of structure–activity relationships.

DESIGNING FOR MOLECULAR DIVERSITY

Strategies for designing molecular screening libraries differ based on the available information and objectives. If the designer already has one or more good lead(s), the objective might be to look at molecules that are chemically similar to known actives, i.e., to look within the same chemical space to exploit the similarity principle—this is known as *focused library design*. If, however, there are no good leads, or if the known drugs exhibit adverse side effects, the designer might be better off casting the net wide in the hopes of finding molecules that exploit a different mode of action. Given the enormously high dimensions of the chemical space spanned by all possible drug-like molecules, identifying drug leads in this manner is an extremely difficult task with no guarantee of success [Bohacek et al., 1996; Triggle, 2009; Virshup et al., 2013]. There has been much debate over the utility of combinatorial synthetic strategies to increase molecular diversity or sample unexplored areas of chemical space [Kodadek, 2011].

To overcome the shortcomings of traditional combinatorial libraries, Schreiber [2000] proposed the concept of diversity-oriented synthesis (DOS), a novel synthetic approach to design and generate molecular libraries possessing structural complexity and diversity. The concept relies on the fact that small molecules interact with biomacromolecules like DNA, RNA, and protein at the surface. Hence, diversity in structural framework will improve the scope of binding of these molecules to various targets (protein–protein interactions, protein–DNA/RNA interactions, and many more), thereby improving the biological activity of the library. Among the several strategies exploited by DOS, one of the most effective and popular is an old idea: to "build" the starting material and "couple" with various

TABLE 1. Similarity Coefficients and Structure–Activity Landscape Index (SALI) Values for the Graph in Figure 2

Tanimoto coefficient matrix generated from Marvin Sketch

Structure	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.00	0.97	0.50	0.58	0.47	0.56	0.56	0.54	0.58	0.54	0.55	0.66	0.45
2	0.97	1.00	0.50	0.56	0.57	0.55	0.55	0.53	0.61	0.52	0.54	0.65	0.44
3	0.50	0.50	1.00	0.56	0.92	0.52	0.53	0.51	0.48	0.48	0.53	0.68	0.88
4	0.58	0.56	0.56	1.00	0.53	0.97	0.97	0.95	0.34	0.39	0.95	0.54	0.52
5	0.47	0.57	0.92	0.53	1.00	0.52	0.51	0.52	0.44	0.50	0.53	0.63	0.89
6	0.56	0.55	0.52	0.97	0.52	1.00	0.95	0.92	0.33	0.38	0.93	0.53	0.51
7	0.56	0.55	0.53	0.97	0.51	0.95	1.00	0.96	0.35	0.40	0.92	0.56	0.50
8	0.54	0.53	0.51	0.95	0.52	0.95	0.96	1.00	0.36	0.36	0.90	0.57	0.49
9	0.58	0.61	0.48	0.34	0.44	0.33	0.35	0.36	1.00	0.55	0.32	0.40	0.43
10	0.54	0.52	0.48	0.39	0.50	0.38	0.40	0.36	0.55	1.00	0.38	0.51	0.45
11	0.55	0.54	0.53	0.95	0.53	0.93	0.92	0.90	0.32	0.38	1.00	0.52	0.55
12	0.66	0.65	0.68	0.54	0.63	0.52	0.56	0.57	0.40	0.51	0.52	1.00	0.59
13	0.45	0.44	0.88	0.52	0.89	0.51	0.50	0.49	0.43	0.45	0.55	0.59	1.00
MCF-7 activity	1	2	3	4	5	6	7	8	9	10	11	12	13
IC50	3.2E-08	6.3E-08	9.0E-09	4.7E-06	4.8E-08	2.4E-08	1.2E-06	2.4E-07	1.7E-05	1.6E-05	1.5E-07	4.8E-04	5.9E-05
PIC50	7.49	7.20	8.05	5.32	7.32	7.62	5.93	6.62	4.77	4.80	6.83	3.32	4.23
SALI between pair of molecules													
		2	3	4	5	6	7	8	9	10	11	12	13
	1	8.91	−1.11	5.11	0.33	−0.28	3.55	1.90	6.46	5.84	1.46	12.20	5.90
	2		−1.68	4.29	−0.27	−0.93	2.81	1.24	6.21	5.02	0.79	10.98	5.34
	3			6.16	8.87	0.88	4.54	2.90	6.27	6.23	2.56	14.76	31.27
	4				−4.23	−85.00	−20.13	−24.43	0.83	0.86	−27.41	4.37	2.28
	5					−0.62	2.82	1.46	4.57	5.04	1.04	10.74	29.13
	6						31.31	13.16	4.24	4.55	11.09	9.22	6.86
	7							−16.07	1.78	1.88	−11.44	5.86	3.36
	8								2.89	2.86	−2.09	7.72	4.69
	9									−0.05	−3.04	2.43	0.95
	10										−3.30	2.98	1.02
	11											7.25	5.79
	12												−2.19

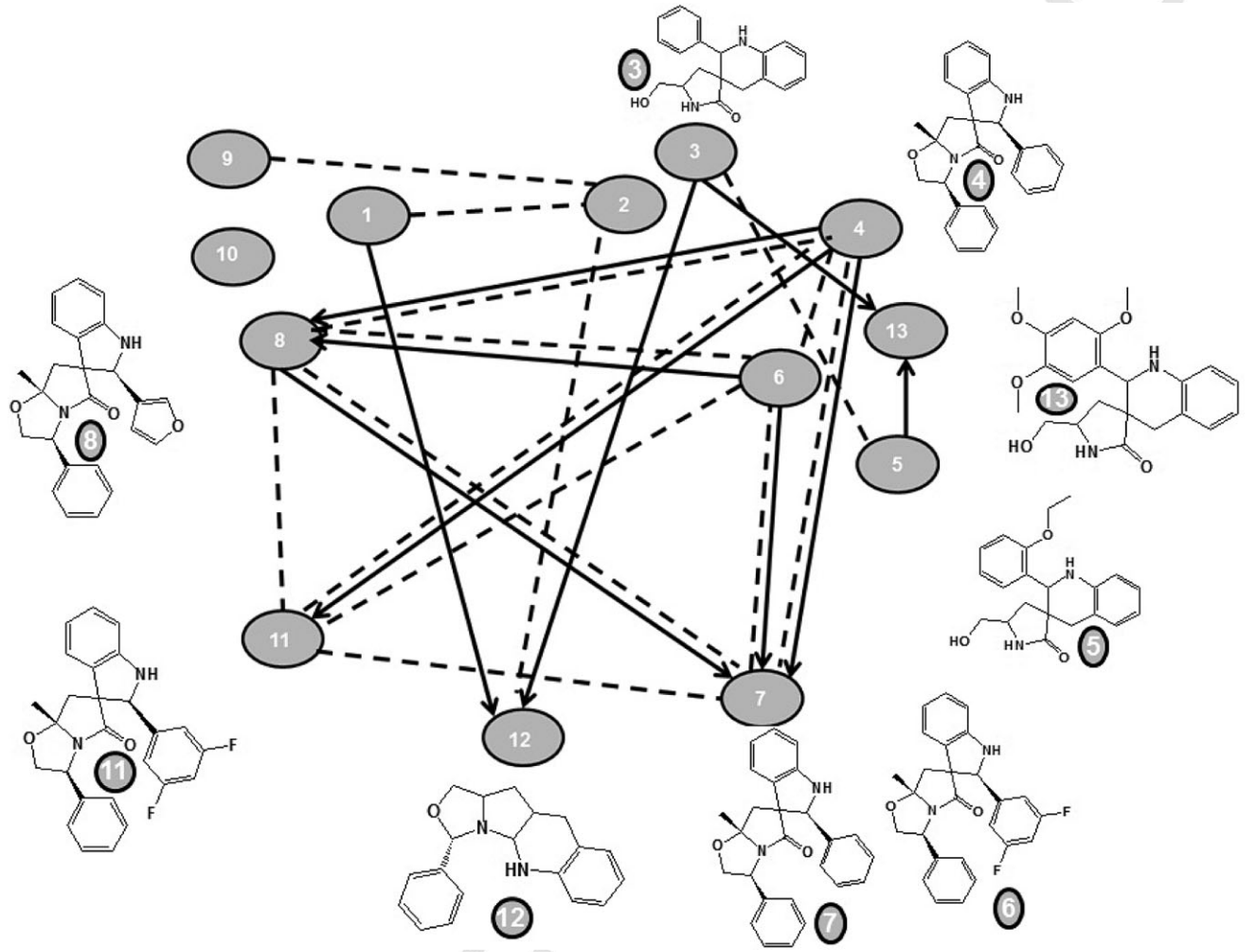


Fig. 2. Similarity network (dotted lines) and structure–activity landscape index (SALI) edges (thick arrows) computed from the data in Table 1. The similarity network was generated using Tanimoto similarity cutoff >0.6 to define an edge between a pair of nodes. The cutoff for the SALI graph was chosen such that all nodes with SALI > 12 are connected with a SALI edge pointing from the molecule of lower activity (IC50) to that of higher activity. This demonstrates the existence of activity cliffs in the network due to variation in the activity between pairs of structurally similar molecules, e.g., molecules 7 and 8. It is also seen that not all structurally similar compounds are connected by SALI edges.

reagents to design a diverse set of structurally unique compounds. The compounds thus obtained from the coupling step are again paired with another set of diverse reagents to end up with skeletal, stereochemical, and biologically diverse compounds. This approach has recently been applied to the synthesis of natural product-inspired compounds [Galloway et al., 2010] and privileged scaffolds [Evans et al., 1988; Welsch et al., 2010; Surakanti et al., 2013].

Computational strategies for the design of diversity libraries have exploited substructural descriptors known as BCUT descriptors [Pearlman and Smith, 1998; Stanton, 1999] in a cell-based representation of chemistry space designed to optimize chemical diversity [Mason and Beno, 2000]. Painter et al. [2011] have used BCUT metrics [Pearlman and Smith, 1998; Stanton, 1999] to analyze the structural diversity of a virtual library, with the objective of designing molecules that explored new regions of chemical space in a DOS strategy. They employed a large virtual library of 11,748 compounds constructed from different scaffolds obtained from a DOS strategy. Analysis of the structural diversity showed that the designed virtual library occupied new chemical space as compared to 32,700 compounds from the Mmolecular Libraries Small Molecule Repository (MLSMR: http://mli.nih.gov/mli/secondary-menu/mlscn/ml-small-molecule-repository/). This allowed the authors to select 53 chemically diverse compounds from the virtual library for synthesis and biological activity testing.

NETWORK MEASURES IN DRUG DESIGN AND DISCOVERY

Networks (including chemical, biological, computer, and social) fall into distinct classes, characterized by well-defined measures that are manifested in differences in the network organization, communication within the network, and evolution of the network. Small-world networks are graphs in which most vertices are not neighbors of one another, but most vertices can be reached from every other vertex in V by a small number of hops. Such classes of networks demonstrate an architecture that falls between that of the classical random graphs and the regular lattices. Random graphs are graphs constructed from randomly connected nodes. The simplest example is the Erdös–Rényi graph [Erdös and Renyi, 1959, 1960], which can be constructed from a complete graph by randomly delinking the nodes [Gilbert, 1959] or by inserting a link between randomly selected pair of nodes [Erdös and Renyi, 1960]. Random graphs characteristically follow Poisson's distribution, with high entropy and low clustering coefficients compared with scale-free networks. The probability that a randomly selected node has degree k is given by $p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$, where $\lambda = \bar{k}$ is the average degree of all nodes in the network.

The structural properties of small-world networks are quantified by prescribing two metrics:

(i) the characteristic path length $L(p)$, which gives the number of edges in the shortest path connecting two vertices from V, averaged over all pair of vertices. $L(p)$ measures the typical separation between two vertices in the network, and thus is a global property of the graph. If n be the number of vertices in the network, then the small-world property is characterized by $L \propto \log n$.

(ii) the clustering coefficient $C(p)$, which measures the cliquishness of a typical vertex neighborhood, thus characterizing a local property. If a vertex in G has k_v neighbors, then at most $k_v(k_v - 1)/2$ edges can exist between them, implying that each pair of neighbors of v is in turn neighbors of each other. If C_v be the fraction of these allowable edges that actually exist, then clustering coefficient C is defined as the average of C_v over all v. Any small-world network is characterized by a $C(p)$ that is much larger than that for a corresponding random network.

The shape of the degree distribution of a small-world network is Poisson like, which is similar to that of a random network. The topology of such a network is relatively homogeneous, with all vertices having nearly the same degree [Watts and Strogatz, 1998].

Scale-free networks are graphs whose vertex degree distribution follows (often asymptotically) a power law. If $P(k)$ be the fraction of vertices in the network that have degree k, then $P(k) \approx k^{-\gamma}$, where γ is the scaling parameter, typically having a value in the range [2,3] [Barabási and Albert, 1999]. The most important characteristic in a scale-free network is the occurrence of a subset of V containing vertices that have degree vertices significantly higher than the average in the graph. The highest degree vertices are the hubs in the network, and often play a central role in determining the properties of the network, and in determining the dynamics and evolution of such networks [Albert and Barabási, 2002; Park and Barabási, 2007]. The presence of hubs in biological and computer networks render such networks resistant to random failure of even a large proportion of nodes or edges [Albert and Barabási, 2002; Park and Barabási, 2007]. The global properties of such chemical space networks has been explored [Benz et al., 2008; Tanaka et al., 2009; Krein and Sukumar, 2011]. It has been demonstrated that large chemical space networks constructed with different similarity metrics show the small-world property, with power law scaling of the degree distribution at high degree [Krein and Sukumar, 2011]. The local properties of chemical space networks have also been the subject of considerable investigation [Guha and Van Drie, 2008b; Bajorath et al., 2009; Stumpfe and Bajorath, 2012] due to their implications for structure–activity relationships and lead optimization, as mentioned above.

Clusters of molecules belonging to different structure–activity relationships are often not isolated into disconnected networks, but connected by "chemical bridges," i.e., molecules that belong to both clusters [Wawer et al., 2008]. These bridging molecules possess high "betweenness centrality" (Fig. 3), essentially describing the frequent occurrence of a vertex on the shortest paths between pair of vertices throughout the network [Freeman, 1977]:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where

$\sigma_{st} = \sigma_{ts}$ denotes the number of shortest paths from $s \in V$ *to* $t \in V$ (set of vertices),

$\sigma_{st}(v)$ = number of shortest paths from s to t that some $v \in V$ lies on.

Raaf and Messabih [2010] have employed betweenness centrality as an algorithm to predict the influential reactions in a metabolic reaction network (described as an "elementary network system"), and

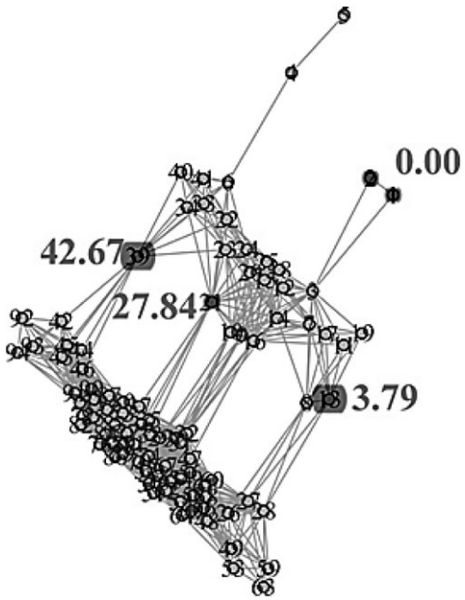


Fig. 3. Similarity network of a of a polymer dataset. A few nodes with high and low betweenness centrality values are identified: these bridge clusters of polymers.

also to identify the decomposition of the metabolic system to subnetwork structure. In the pertinent study, the metabolic reactions were treated as nodes, and the reaction conditions as edges.

Networks wherein clusters combine in a hierarchical manner are characterized by coexistence of modularity, local clustering, and scale-free topology. In hierarchical networks, the clustering coefficient follows $C(k) \sim k\text{-}1$ (a straight line of slope −1 on a log–log plot). Relating local and global network measures thus enables identification of key molecules responsible for "scaffold hopping" [Rush et al., 2005; Barker et al., 2006; Renner and Schneider, 2006; Nettles et al., 2007; Vogt et al., 2010; Sun et al., 2012].

Sheftel et al. [2013] have reported the social network behavior of interconnected aminoacid residues of the β-adrenergic receptor (GPCR family) using "closeness centrality" measured. Closeness centrality describes the "neighborhoodness" [Sabidussi, 1966] of the clustered vertices in the graph:

$$C_c(v) = \frac{1}{\sum_{t \in V} d_G(v,t)},$$

where d_G = distance between the vertices v and t in graph G and v, t ϵ V (set of vertices).

Centrality measures based on eigen-decomposition of the matrix form of networks, often referred to as graph spectral analysis, have been used to identify clusters and subclusters in complex biological networks [Kannan and Vishveshwara, 1999; Vishveshwara et al., 2002]. The eigenvalues represent the flow of information between vertices within clusters or subclusters of the protein domain. Graph spectral analysis [Balasubramanian et al., 2006] is also used to correlate the structure and functions of protein chains in clusters.

Clique-based methods are also used to assess clustering [Barker et al., 2006; Nettles et al., 2007; Milletti and Vulpetti, 2010] where cliques are network subsets where every pair of vertices in the subset is connected by an edge. Thus, those molecules belonging to cliques have a high degree of structural similarity to one another.

Koutsoukas et al. [2014] recently conducted a benchmark study to determine the effectiveness of different descriptor sets in assessing the diversity of a molecular library in bioactivity space. These authors analyzed several popular families of descriptors, including those for pharmacophore-based descriptors, extended connectivity fingerprint descriptors [Rogers and Hahn, 2010] (ECFP4 and FCFP4), MACCS keys [Durant et al., 2002], shape-based descriptors ROCS [Grant et al., 1996; Rush et al., 2005]) and principal moments of inertia (PMI), BCUT descriptors [Pearlman and Smith, 1998; Stanton, 1999], physicochemical property descriptors and Bayes affinity fingerprints [Koutsoukas et al., 2013]. They found that a higher coverage of bioactivity space was achieved using Bayes affinity fingerprints; the latter being descriptors that represent molecules in terms of their in silico predicted bioactivity profiles against a broad spectrum of human protein drug targets. Descriptors such as PMI showed no correlation between diversity in the PMI space and diversity in bioactivity space. Studies such as this underscore the importance of the choice of descriptors in molecular diversity assessment, and that different measures of chemical diversity may not necessarily correlate with diversity in bioactivity space.

Chemogenomic approaches that consider protein target space in addition to the ligand space [Bredel and Jacoby, 2004; Mestres, 2004; Klabunde, 2007; Rognan, 2007; Das et al., 2009, 2010a; Kinningss and Jackson, 2009; Kinnings et al., 2009; Sheridan et al., 2009; Milletti and Vulpetti, 2010; Sukumar and Das, 2011] have been employed to get around the limitation that ligands that are dissimilar on the basis of structural similarity may nevertheless bind to the same protein target. Network pharmacology is thus emerging as a new paradigm in drug discovery [Hopkins, 2008; Zhao and Iyengar, 2012].

EVOLUTION OF CHEMICAL SPACE NETWORKS

One of the most important mechanisms for the dynamic evolution of networks is via preferential attachment—essentially a stochastic process. As the essence of this process, the higher the degree of a vertex in the network, the more likely it is to receive new edges. Thus, the hubs show a significantly stronger ability compared with other vertices to acquire new links added to the network during its dynamic evolution [Albert and Barabási, 2002; Park and Barabási, 2007].

Molecular libraries are not designed by a series of random acts. New molecules are synthesized from existing precursors using well-established synthetic schemes. The new molecules bear a degree of chemical similarity to their precursors, but differ in important ways. Thus, the process of molecular library design is that of evolution of a network, where new nodes are added and connected to existing nodes. Preferential attachment of new nodes to existing nodes of high degree (hubs) leads to the formation of a scale-free network [Albert and Barabási, 2002; Park and Barabási, 2007], as in social networks and computer networks. The power law degree distribution and small-world nature of chemical space networks [Benz et al., 2008; Tanaka et al., 2009; Krein and Sukumar, 2011] lead one to speculate that such networks are likewise formed by preferential reuse of favored precursors or well-studied molecular scaffolds. In practice, awareness of a network's limited diversity can suggest a modification to synthetic strategy that explicitly seeks increasing chemical diversity when there is a failure to discover new leads with novel molecular scaffolds. Analysis of the network properties of chemical libraries and assessment of their molecular diversity thus assume increased significance for the design of new chemical entities.

The temporal characteristics of chemical libraries (and associated network characteristics) are also noteworthy. Public databases such as PubChem [Wang et al., 2010], ZINC [Irwin and Shoichet, 2005], and PDB [Berman et al., 2002] have grown significantly within a very short span of time. Temporal behaviors of complex systems, such as e-mail patterns and earthquakes, have been studied using measures such as memory (a measure of temporal correlations) and burstiness (a measure of intervening time distribution between consecutive events) [Goh and Barabási, 2008]. Such measures may suggest research directions for cheminformatics and provide quantification of historical biases and assay performance.

CONCLUSIONS

We have briefly reviewed the basic characteristics of molecular similarity and dissimilarity networks, focusing on their use in the design of chemical libraries. Large chemical space networks have been shown to possess the small-world property [Krein and Sukumar, 2011] characteristic of many physical, biological, and social networks. Exploring new areas of chemical space in a finite library demands maximization of chemical diversity. Assessment of molecular similarity and chemical diversity are nontrivial problems, due to the large number of possible descriptors, and due to the presence of activity cliffs. SALI networks [Guha and Van Drie, 2008a, 2008b] quantify the roughness of a structure–activity landscape. Similarity and diversity depend sensitively upon the choice of descriptor representation and choice of similarity metrics. Chemical diversity may not translate into diversity in the bioactivity space, due to the presence of activity cliffs [Maggiora, 2006], and the efficacy of a descriptor representation in quantifying diversity in bioactivity space may further depend upon the size and nature of the chemical library [Koutsoukas et al., 2014]. Targeting diversity in bioactivity space is the objective behind DOS.

Chemogenomic and network pharmacology [Klabunde, 2007; Rognan, 2007; Hopkins, 2008; Zhao and Iyengar, 2012] are attempting to overcome the limitations of traditional structure-based and ligand-based drug discovery by utilizing structural information about both the ligands and their target proteins, where available, and exploiting structural and chemical similarities between molecules known to target a protein, as well as structural and sequence similarities between protein binding sites. Such approaches explicitly employ a network formalism.

Local graph invariants are also of interest. Hubs in a similarity network identify clusters that are well defined by a structure–activity relationship. Such molecules are expected to confer robustness to a QSAR model. Molecules that possess high betweenness centrality in a graph act as chemical bridges connecting local clusters belonging to different structure–activity relationships. Inclusion of such molecules in a training set renders a QSAR model more capable of scaffold hopping. The evolution of chemical libraries and their network invariants with time are interesting subjects that have not as yet received significant attention from the cheminformatics community, but they may hold the keys to understanding the social aspects of the discipline at designing new routes to chemical synthesis and drug discovery.

REFERENCES

Albert R, Barabási A-L. 2002. Statistical mechanics of complex networks. Rev Mod Phys 74:47.

Bajorath J, Peltason L, Wawer M, Guha R, Lajiness MS, Van Drie JH. 2009. Navigating structure–activity landscapes. Drug Discov Today 14:698–705.

Balasubramanian K, Khokhani K, Basak SC. 2006. Complex graph matrix representations and characterizations of proteomic maps and chemically induced changes to proteomes. J Proteome Res 5:1133–1142.

Barabási A-L, Albert R. 1999. Emergence of scaling in random networks. Science 286:509–512.

Barker EJ, Buttar D, Cosgrove DA, Gardiner EJ, Kitts P, Willett P, Gillet VJ. 2006. Scaffold hopping using clique detection applied to reduced graphs. J Chem Inf Model 46:503–511.

Benz RW, Swamidass SJ, Baldi P. 2008. Discovery of power-laws in chemical space. J Chem Inf Model 48:1138–1151.

Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S. 2002. The protein data bank. Acta Crystallogr D Biol Crystallogr 58:899–907.

Boda K, Seidel T, Gasteiger J. 2007. Structure and reaction based evaluation of synthetic accessibility. J Comput Aided Mol Des 21:311–325.

Bohacek RS, McMartin C, Guida WC. 1996. The art and practice of structure-based drug design: a molecular modeling perspective. Med Res Rev 16:3–50.

Bolouri H. 2014. Modeling genomic regulatory networks with big data. Trends Genet 30:182–191.

Bredel M, Jacoby E. 2004. Chemogenomics: an emerging strategy for rapid target and drug discovery. Nat Rev Genet 5:262–275.

Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R. 2013. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol Ther 138:333–408.

Das S, Kokardekar A, Breneman CM. 2009. Rapid comparison of protein binding site surfaces with property encoded shape distributions. J Chem Inf Model 49:2863–2872.

Das S, Krein MP, Breneman CM. 2010a. Binding affinity prediction with property-encoded shape distribution signatures. J Chem Inf Model 50:298–308.

Das S, Krein MP, Breneman CM. 2010b. PESDserv: a server for high-throughput comparison of protein binding site surfaces. Bioinformatics 26:1913–1914.

Durant JL, Leland BA, Henry DR, Nourse JG. 2002. Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci 42:1273–1280.

Erdös P, Renyi A. 1959. On random graphs. Publ Math Debrecen 6:290.

Erdös P, Renyi A. 1960. On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5:17.

[15] Evans BE, Rittle KE, Bock MG, DiPardo RM, Freidinger RM, Whitter WL, et al. 1988. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. J Med Chem 31:2235–2246.

[16] Freeman LC. 1977. A set of measures of centrality based on betweenness. Sociometry ••:35–41.

Galloway WRJD, Isidro-Llobet A, Spring DR. 2010. Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. Nat Commun 1:80.

Gilbert EN. 1959. Random graphs. Ann Math Stat 30:1141–1144.

Goh KI, Barabási AL. 2008. Burstiness and memory in complex systems. Europhys Lett 81:48002.

Grant JA, Gallardo MA, Pickup BT. 1996. A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. J Comput Chem 17:1653–1666.

Guha R, Van Drie JH. 2008a. Assessing how well a modeling protocol captures a structure—activity landscape. J Chem Inf Model 48:1716–1728.

Guha R, Van Drie JH. 2008b. Structure-Activity Landscape Index: identifying and quantifying activity cliffs. J Chem Inf Model 48:646–658.

Hopkins AL. 2008. Network pharmacology: the next paradigm in drug discovery. Nature Chem Biol 4:682–690.

Irwin JJ, Shoichet BK. 2005. ZINC-a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182.

Kannan N, Vishveshwara S. 1999. Identification of side-chain clusters in protein structures by a graph spectral method. J Mol Biol 292:441–464.

Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. 2009. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. PLoS Comput Biol 5:e1000423.

Kinningss SL, Jackson RM. 2009. Binding site similarity analysis for the functional classification of the protein kinase family. J Chem Inf Model 49:318–329.

Klabunde T. 2007. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. Br J Pharmacol 152:5–7.

Kodadek T. 2011. The rise, fall and reinvention of combinatorial chemistry. Chem Commun (Camb) 47:9757–9763.

Koutsoukas A, Lowe R, Kalantarmotamedi Y, Mussa HY, Klaffke W, Mitchell JBO, Glen RC, Bender A. 2013. In silico target predictions: defining a benchmarking data set and comparison of performance of the multiclass Naive Bayes and Parzen-Rosenblatt window. J Chem Inf Model 53:1957–1966.

Koutsoukas A, Paricharak S, Galloway WRJD, Spring DR, Ijzerman AP, Robert C, Glen RC, Marcus D, Bender A. 2014. How diverse are diversity assessment methods? A comparative analysis and benchmarking of molecular descriptor space. J Chem Inf Model 54:230–242.

Krein M, Huang TW, Morkowchuk L, Agrafiotis DK, Breneman CM. 2012. Developing best practices for descriptor-based property prediction: appropriate matching of datasets, descriptors, methods, and expectations. Stat Model Mol Descriptors QSAR/QSPR 2:33–64.

Krein MP, Sukumar N. 2011. Exploration of the topology of chemical spaces with network measures. J Phys Chem A 115:12905–12918.

Maggiora GM. 2006. On outliers and activity cliffs—why QSAR often disappoints. J Chem Inf Model 46:1535.

Martin YC, Kofron JL, Traphagen LM. 2002. Do structurally similar molecules have similar biological activity? J Med Chem 45:4350–4358.

Mason JS, Beno BR. 2000. Library design using BCUT chemistry-space descriptors and multiple four-point pharmacophore fingerprints: simultaneous optimization and structure-based diversity. J Mol Graph Model 18:438–451.

Matys V, Fricke E, Geffers R, Gößling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV. 2003.

TRANSFAC®: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31:374–378.

Mestres J. 2004. Computational chemogenomics approaches to systematic knowledge-based drug discovery. Curr Opin Drug Discov Devel 7:304–313.

Milletti F, Vulpetti A. 2010. Predicting polypharmacology by binding site similarity: from kinases to the protein universe. J Chem Inf Model 50:1418–1431.

[17] Nettles JH, Jenkins JL, Williams C, Clark AM, Bender A, Deng Z, Davies JW, Glick MJ. 2007. ••. Mol Graph Model 26:622.

Painter TO, Wang L, Majumder S, Xie X-Q, Brummond KM. 2011. Diverging DOS strategy using an allene-containing tryptophan scaffold and a library design that maximizes biologically relevant chemical space while minimizing the number of compounds. ACS Comb Sci 13:166–174.

Park J, Barabási A-L. 2007. Distribution of node characteristics in complex networks. Proc Natl Acad Sci U S A 104:17916–17920.

[18] Pearlman RS, Smith KM. 1998. Novel software tools for chemical diversity, 3D QSAR in drug design. ••, Springer. p 339–353.

Raaf H, Messabih B. 2010. Betweenness centrality of event graph application to metabolic network modelled by elementary net system. J Appl Sci 10:1610–1615.

Renner S, Schneider G. 2006. Scaffold-hopping potential of ligand-based similarity concepts. ChemMedChem 1:181–185.

Rogers D, Hahn M. 2010. Extended-connectivity fingerprints. J Chem Inf Model 50:742–754.

Rognan D. 2007. Chemogenomic approaches to rational drug design. Br J Pharmacol 152:38–52.

Rush TS, Grant JA, Mosyak L, Nicholls A. 2005. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. J Med Chem 48:1489–1495.

Sabidussi G. 1966. The centrality index of a graph. Psychometrika 31:581–603.

Schreiber SL. 2000. Target-oriented and diversity-oriented organic synthesis in drug discovery. Science 287:1964–1969.

Sheftel S, Muratore K, Black M, Costanzi S. 2013. Graph analysis of beta2 adrenergic receptor structures: a "social network" of GPCR residues. In Silico Pharmacol 1:16.

Sheridan RP, Nam K, Maiorov VN, McMasters DR, Cornell WD. 2009. QSAR models for predicting the similarity in binding profiles for pairs of protein kinases and the variation of models between experimental data sets. J Chem Inf Model 49:1974–1985.

Stanton DT. 1999. Evaluation and use of BCUT descriptors in QSAR and QSPR studies. J Chem Inf Comput Sci 39:11–20.

[19] Stumpfe D, Bajorath •• Jr. 2012. Exploring activity cliffs in medicinal chemistry: miniperspective. J Med Chem 55:2932–2942.

[20] Sukumar N, Das S. 2011. Current trends in virtual high throughput screening using ligand-based and structure-based methods. Comb Chem High Throughput Screen 14:872–888.

Sukumar N, Das S, Krein M, Godawat R, Vitol I, Garde S, Bennett KP, Breneman CM. 2012. Molecular descriptors for biological systems. In: Guha R, Bender A, editors. Computational approaches in cheminformatics and bioinformatics. Hoboken, NJ: Wiley. p 107–143.

Sun H, Tawa G, Wallqvist A. 2012. Classification of scaffold-hopping approaches. Drug Discov Today 17:310–324.

Surakanti R, Sanivarapu S, Thulluri C, Iyer PS, Tangirala RS, Gundla R, Addepally U, Murthy YLN, Velide L, Sen S. 2013. Synthesis of privileged scaffolds by using diversity-oriented synthesis. Chem Asian J 8:1168–1176.

Tanaka N, Ohno K, Niimi T, Moritomo A, Mori K, Orita M. 2009. Small-world phenomena in chemical library networks: application to fragment-based drug discovery. J Chem Inf Model 49:2677–2686.

Todeschini R, Consonni V. 2000. ••. In: Mannhold R, Kubinyi H, [21] Timmerman H, editors. Handbook of molecular descriptors. ••, WILEY—VCH. p ••–••.

Triggle DJ. 2009. The chemist as astronaut: searching for biologically useful space in the chemical universe. Biochem Pharmacol 78:217–223.

von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R. 2011. PROMISCUOUS: a database for network-based drug-repositioning. Nucleic Acids Res 39 (Suppl. 1):D1060–D1066.

Virshup AM, Contreras-García J, Wipf P, Yang W, Beratan DN. 2013. Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. J Am Chem Soc 135:7296–7303.

Vishveshwara S, Brinda KV, Kannan N. 2002. Protein structure: insights from graph theory. J Theor Comput Chem 1:187–211.

Vogt M, Stumpfe D, Geppert H, Bajorath J. 2010. Scaffold hopping using two-dimensional fingerprints: true potential, black magic, or a hopeless endeavor? Guidelines for virtual screening. J Med Chem 53:5707–5715.

Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO, Wang J, Xiao J, Zhang J, Bryant SH. 2010. An overview of the PubChem BioAssay resource. Nucleic Acids Res 38 (Suppl. 1):D255–D266.

Watts DJ, Strogatz SH. 1998. Collective dynamics of "small-world" networks. Nature 393:440–442.

Wawer M, Peltason L, Weskamp N, Teckentrup A, Bajorath •• Jr. 2008. Structure–activity relationship anatomy by network-like similarity graphs and local structure–activity relationship indices. J Med Chem 51:6075–6084.

Welsch ME, Snyder SA, Stockwell BR. 2010. Privileged scaffolds for library design and drug discovery. Curr Opin Chem Biol 14:347–361.

Yildirim MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M. 2007. Drug–target network. Nat Biotechnol 25:1119–1126.

Zhao S, Iyengar R. 2012. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annu Rev Pharmacol Toxicol 52:505–521.

JOBNAME: No Job Name PAGE: 11 SESS: 13 OUTPUT: Fri Jul 25 11:43:46 2014 SUM: 50CDFE41
/Xpp84/wiley_journal_D-E/DDR/ddr_v0_i0/ddr_21218

AUTHOR QUERY FORM

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.

Many thanks for your assistance.

Query References	Query	Remarks
1	AUTHOR: Please supply the full forename of the correspondence.	
2	WILEY BLACKWELL: Please supply the received and accepted dates for this article.	
3	WILEY BLACKWELL: Note that "Special Issue: Genes, Networks and Bioinformatics in Drug Discovery" has been set as title page footnote. Please check and confirm this is correct.	
4	AUTHOR: Please supply the full forename of the authors.	
5	AUTHOR: Das et al. 2010 has been changed to Das et al. 2010b so that this citation matches the Reference List. Please confirm if this is correct.	
6	AUTHOR: Please confirm if Table 1 has been formatted correctly.	
7	AUTHOR: Should BCUT be spelled out? If yes, please supply its full form.	
8	AUTHOR: Please check this website address and confirm that it is correct. (Please note that it is the responsibility of the author(s) to ensure that all URLs given in this article are correct and useable.)	
9	AUTHOR: Erdös 1959, 1960 has been changed to Erdös and Renyi 1959, 1960 so that this citation matches the Reference List. Please confirm if this is correct.	
10	AUTHOR: Erdös 1960 has been changed to Erdös and Renyi 1960 so that this citation matches the Reference List. Please confirm if this is correct.	
11	AUTHOR: The global properties . . . explored. This sentence has been reworded for clarity. Please check and confirm it is correct.	
12	AUTHOR: Please supply the full form for GPCR.	
13	AUTHOR: The evolution of chemical . . . discovery. This sentence has been reworded for clarity. Please check and confirm it is correct.	

Query References	Query	Remarks
14	AUTHOR: For references Albert & Barabási, 2002, Erdös & Renyi, 1959, Erdös & Renyi, 1960, Galloway et al., 2010, Goh & Barabási, 2008, Maggiora, 2006, Nettles et al., 2007 and Sheftel et al., 2013. If these are not one-page articles please supply the first and last pages for these articles.	
15	AUTHOR: In Evans et al., 1988, if there are fewer than 11 authors, please supply all of their names. If there are 11 or more authors, please supply the first 10 authors' names then et al.	
16	AUTHOR: Please supply the volume number for Freeman, 1977.	
17	AUTHOR: Please supply the article title for Nettles et al., 2007.	
18	AUTHOR: Please supply the city location of publisher for Pearlman & Smith, 1998.	
19	AUTHOR: For Stumpfe & Bajorath, 2012 and Wawer et al., 2008, Please supply the forename of author Bajorath Jr.	
20	AUTHOR: Please confirm if the year of publication "2011" for Sukumar & Das is correct.	
21	AUTHOR: Please supply the chapter name, city location of the publisher and page range for Todeschini & Consonni, 2000.	
22	AUTHOR: Figures 1–3 are of poor quality (labels and lines are blurry). Please check required artwork specifications at http://authorservices.wiley.com/bauthor/illustration.asp	