"Promiscuous" Ligands and Targets Provide Opportunities for Drug Design

Gisbert Schneider[1,*] and Petra Schneider[2]

[1] Johann Wolfgang Goethe-University,
Siesmayerstr. 70, 60323 Frankfurt am Main, Germany,

[2] Schneider Consulting GbR,
George-C.-Marshall Ring 33, 61440 Oberursel, Germany

E-Mail: *gisbert.schneider@modlab.de

Abstract

Flexibility, structural adaptability and non-selective pharmacophoric features of ligands and macromolecular receptors are major challenges in molecular design. These properties can lead to promiscuous binding behavior and often prevent modeling of reasonable structure-activity relationships. We have analyzed neighborhood behavior of bioactive ligands and protein binding pockets in terms of molecular shape and pharmacophoric features. It turned out that there exist certain relationships between the shape and the buriedness of a protein pocket and its ability to accommodate a small molecule ligand. The self-organizing map concept and clustering techniques is presented as a means for predicting potential bioactivities of ligands, and "de-orphanizing" of drugs and receptor proteins. Opportunities for "scaffold-hopping" and "re-purposing" are discussed in the context of systems chemistry.

Introduction

The notion of "frequent hitters" [1, 2], in particular bioactive compounds with a broad ("promiscuous") binding profile to multiple targets [3, 4], has led to systematic investigation of the requirements for ligand binding selectivity. Such approaches often include a distinction between primary drug targets and "anti-targets" [5, 6], preferred and undesirable substructural elements in lead structure optimization [7, 8] and pharmacophoric ligand

features [9, 10]. More recently these methods have been complemented by network analyses of one-to-many and many-to-many relationships between ligands and targets on a systemic level that is accessible by chemical biology and chemogenomics [11 – 14].

Computational methods support lead structure design by predicting target/anti-target binding and pharmacological effects of screening compounds and lead candidates [15]. These methods are either receptor- (relying on a three-dimensional receptor model) or ligand-based (starting from known ligands), and ideally both concepts are applied in parallel [16]. Here, we address the prediction and utilization of ligand "promiscuity" from these perspectives. First, we briefly introduce the "self-organizing map" (SOM) as a computational tool for data visualization and compound encoding [17]. Then, we demonstrate its applicability to "scaffold-hopping" [18] and "re-purposing" [19, 20], as well as the "de-orphanizing" of drugs and target proteins [21, 22]. We conclude that understanding the reasons for binding promiscuity will offer rich opportunities for future drug design.

THE SELF-ORGANIZING MAP

Among the different visualization techniques that are available to the drug designer, the SOM or "Kohonen network" [23, 24] has found particularly successful applications in hit and lead discovery. Clustering and visualization of high-dimensional data distributions are the two main applications of the SOM. This is most relevant for drug design and target prediction, as compounds are usually represented in a high-dimensional pharmacophoric feature space. Such a data projection can be schematically visualized as a two-dimensional map as illustrated in Figure 1. In this representation, high-dimensional compound distributions can be analyzed by visual inspection of the two-dimensional SOM projection [25].
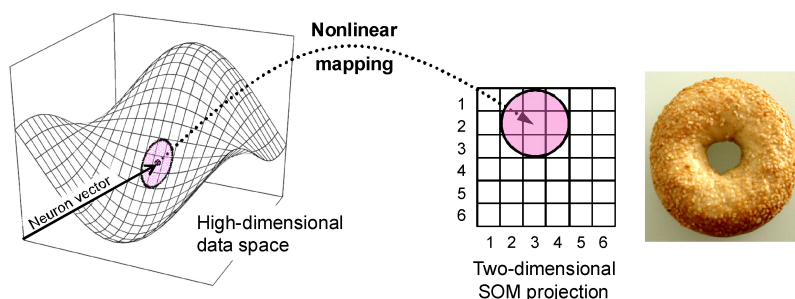

Figure 1. A SOM projection (*right*) can be used to visualize the data distribution in a high-dimensional data space (*left*) (figure adapted from ref. [22]). Here, a SOM containing $6 \times 6 = 36$ clusters (shown as a squared grid) is depicted. Local neighborhoods are conserved, that is, molecules that are close to each other on the SOM are also close in the original high-dimensional space (*e.g.*, a space spanned by many pharmacophoric features). The donut depicts a SOM forming a torus, *i.e.* an "endless two-dimensional space" [25].

The SOM belongs to a class of machine-learning systems which are optimized by "unsupervised learning" techniques [24]. According to this principle, so-called "neuron vectors" (cluster centroids) are positioned in the data space such that their distribution approximates the distribution of the data points. The term "unsupervised" means that no target information or activity values are required for the training process. By help of the SOM algorithm, a projection from a high-dimensional to a low-dimensional space can be obtained, so that clusters of isofunctional molecules can be identified by visual inspection (Figure 2). Briefly, unsupervised learning can be formulated as an iteration of a competitive (Step 2) and a cooperative (Step 3) step:

Step 0:   Define the SOM topology (e. g., rectangular toroidal) and number of clusters to be formed (*i. e.*, centroid vectors = neurons).

Step 1:   Choose a molecule from the high-dimensional descriptor space.

Step 2:   Determine the centroid vector ("winner neuron") that is closest to the molecule in high-dimensional space according to a distance metric.

Step 3:   Move the winner neuron and its neighboring neurons toward the data point.

Step 4:   Go to *Step 1* or terminate.

A software tool demonstrating this algorithmic concept is freely available from our public web server (http://gecco.org.chemie.uni-frankfurt.de/sommer/index.html).

SOM-based Prediction of Ligand Binding Profiles

In the attempt to predict the binding behavior of a small molecule, a first question to answer is how to represent the compound for mathematical analysis. Numerous descriptors have been suggested for this purpose [26]. Here, we used the topological pharmacophore descriptor CATS 2D, which encodes a molecule as a graph-based distribution of potential pharmacophoric points (hydrogen-bond donor or acceptor, lipophilic, positively or negatively charged atoms) and was shown to be suited for virtual compound screening and scaffold-hopping, *i. e.* the identification of isofunctional compounds with a different core structure [27, 28]. A literature-derived compilation of recently published drugs and lead compounds (COBRA version 8.6, 10840 compounds [29]) was encoded by the 150-dimensional CATS 2D descriptor and subjected to SOM clustering and projection (Figure 2). Here, the SOM was arranged as a 15 x 10 rectangular array with toroidal topology. Local clusters of compounds that bind to the same target are referred to as *activity islands* [29, 30]. Figure 2 shows examples of such activity islands. If compounds with a certain desired pharmacological activity, such as enzyme inhibition, or specific target binding, are found to be clustered on the SOM in an activity island, these clusters represent promising target areas for molecular design. Compounds molecules that fall in the desired target area on the SOM are candidates for synthesis and *in vitro* screening.
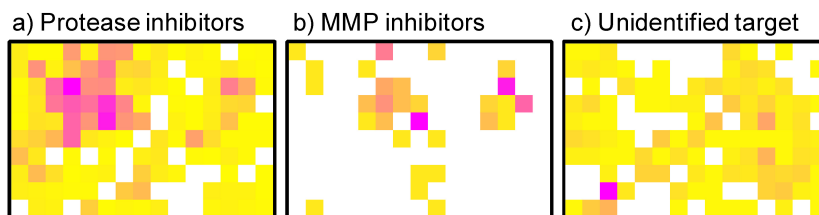
Figure 2. SOM projections of different ligand classes (annotated by target family). Color shading indicates ligand density (*white*: none, *yellow*: few, *magenta*: many). MMP: matrix metalloproteinases. Adapted from ref. [22].

Several hit and lead discovery projects have already successfully employed this method [30 – 35]. Notably, the location of ligands with unknown target protein ("orphan" ligands) on the SOM projection can even be used to identify target candidates by neighborhood analysis [20, 31]: From the known target-binding profiles of co-located drugs one can infer a similar activity for the orphan ligand. Such an example is given by compound **1** (Figure 3), which was predicted to bind to metabotropic glutamate receptor 1 (mGluR1) by selection of candidate compounds that co-located with known mGluR1 ligands on a SOM. This strategy led to the potent and subtype selective coumarine-derived mGluR1 antagonist **1** ($K_i = 24$ nM) [31], which was later developed into a potent lead series. It is important to keep note that SOMs were only used for first-pass compound filtering with the aim to design an activity-enriched focused library, and that subsequent chemical hit-to-lead optimization was still required. Pursuing a similar SOM-based concept, mGluR1 antagonists **2** and **3** were identified by picking structurally diverse compounds from a large screening compound database for bioactivity testing [34]. SOM-guided combinatorial library design yielded compound **4**, a selective antagonist ($K_i = 2.4$ nM) of purinergic receptor A_{2A} [35].
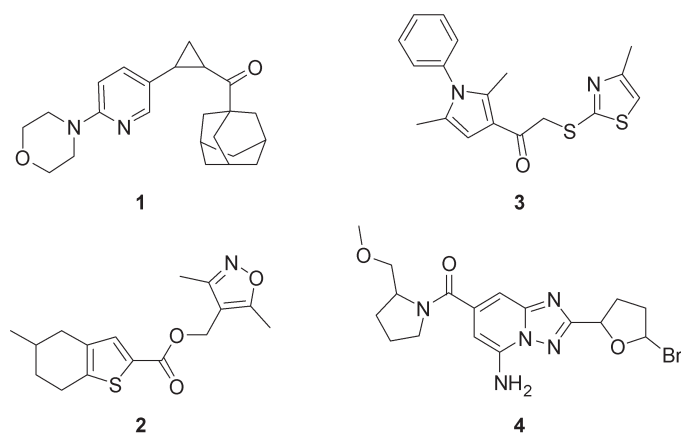

Figure 3. Examples of potent ligands that were identified using the SOM technique for virtual screening and target prediction.

These examples demonstrate the usefulness of ligand clustering by SOMs, with the aim to identify activity islands and exploit promiscuous binding behavior for the design of new bioactive compounds.

Ligand distributions are based on some sort of neighborhood relationship in chemical space, depending on the choice of the molecular descriptor (*e. g.*, pharmacophoric features, physicochemical properties, or substructure composition) [13, 36, 37]. As we have seen, it is possible to infer biological activity (*e. g.* receptor binding) from local ligand clusters. This concept is based on the analysis of a *single* or *few* compounds. An extension is to consider the distribution of all isofunctional ligands (defined, for example, as binding to the same drug target) in some chemical space, and compare *sets* of ligands with each other [38]. This enables the assessment of target similarity as measured by the similarity of the distributions of their respective ligands.

To illustrate the idea, we trained a SOM with the complete COBRA data (10,840 compounds represented by their CATS 2D descriptors). The SOM contained 150 neurons (clusters) arranged as a 15 x 10 rectangular array with toroidal topology. Then, this SOM was used to prepare individual projections for each target-specific ligand set, resulting in a total of 174 projections (*cf.* Figure 2 for example projections). The resulting patterns of the 174 individual ligand sets on the SOM were used as "target fingerprints" for the analysis of drug target similarity. In this study, the fingerprints consisted of $15 \times 10 = 150$ real-valued numbers (ligand density in each of the 150 SOM clusters). Similarity between the target fingerprints was expressed as Pearson correlation coefficient r, and graphically visualized as a target interaction network.

The network resulting from target relations with $r > 0.3$ is presented in Figure 4. We chose this threshold value because it represents a compromise between highly connected networks (obtained for lower values of r) and meagerly connected networks (larger values of r). 165 of the 174 targets analyzed show such a correlation to at least one other target *via* 334 edges. This suggests that most of the known drugs actually have multiple targets, which is in perfect accordance with earlier findings [39, 40]. It is noteworthy that several node clusters emerge, which can be attributed to similar ligand classes (not highlighted in Figure 4 for reasons of clarity). In addition, strongly connected nodes ("hubs") indicate targets by their respective ligands that have promiscuous binding features. Such an analysis complements receptor-based approaches for target comparisons and estimations of target "druggability" [41 – 45].
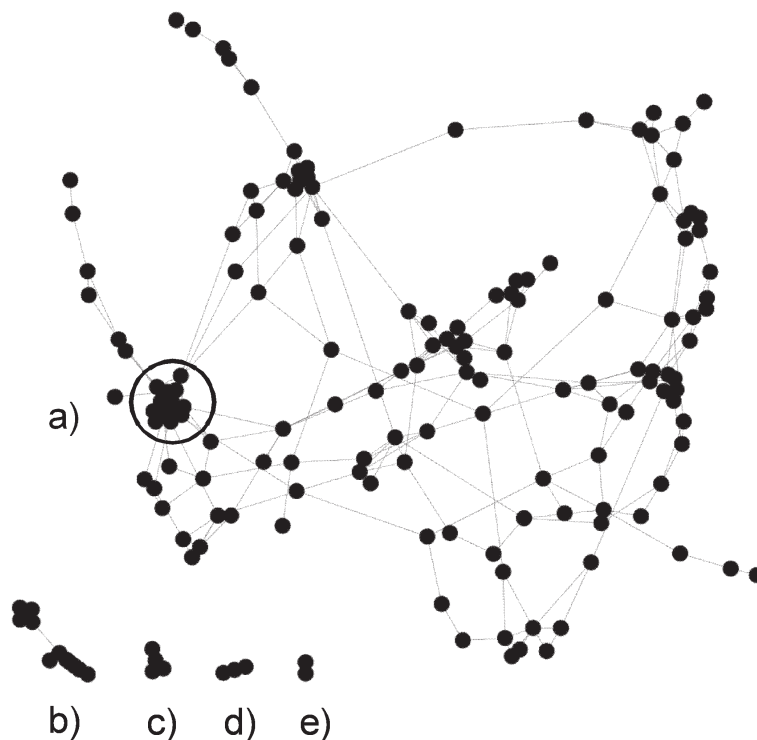
Figure 4. From ligand similarity to target similarity. Network representation of the relationship between ligand sets for 165 drug targets. The network structure results from correlation analysis of ligand properties (SOM-derived pharmacophoric features). Nodes represent drug targets, and the lines are scaled according to their pairwise ligand similarity. Graphical representations of the target correlation network were prepared with the software Cytoscape 2.5.1 (http://www.cytoscape.org/). The network layout was optimized using the spring-embedded layout option provided within Cytoscape. Only targets with a pair-wise ligand correlation of $r > 0.3$ are shown. Selected targets, which potentially share similar ligand features:

a) *within the highlighted circle*: DAT (dopamine transporter DAT), NET (norepinephrine transporter), HTT (serotonine transporter), 5-HT3 (serotonine ion channel), dopamine receptor (GPCR), serotonine receptor (GPCR), histamine receptor (GPCR), potassium ion channel, nicotinic acetylcholine receptor (ion channel), monoamine oxidase (enzyme);

b) dihydrofolate reductase (enzyme), nitric oxide synthase (enzyme), RNA-polymerase (enzyme), P2Y (nuclotide receptor, GPCR), mGluR (metabotropic glutamate receptor, GPCR), AMPA (ionotropic glutamate receptor, ion channel), kainate receptor (ionotropic glutamate receptor, ion channel), P2X (nuclotide receptor, ion channel).

The targets contained in the small clusters **c)**, **d)**, and **e)** have unique ligand features that separate them from the majority of the known drug targets. These might therefore be particularly attractive for the development of novel selective lead compounds.

PREDICTION OF "DRUGGABLE" LIGAND BINDING POCKETS

The assessment of ligand promiscuity immediately poses the question whether there are preferable ligand-binding pockets that are "druggable" with the prospect of allowing for selective ligands to be found [41]. Several studies have been performed addressing this issue, and it turns out that comparably simplistic rules seem to allow for a distinction between druggable and non-druggable surface cavities [44, 45]. We have performed a comprehensive analysis of pocket shapes using known ligand-receptor complexes [45]. Figure 5 presents a SOM projection of liganded and unliganded pockets, which were described by their shape and buriedness [46]. Clearly, there are preferred shape-types that seem to govern the likeliness of ligand binding. This observation suggests that a limited set of preferred pocket shapes have evolved to accommodate substrates and effector molecules. Gaining a deeper understanding of these patterns will certainly help design not only ligands with a high "general likelihood" of binding by incorporation of certain substructure elements, but provide the necessary structural basis for the identification of target-selective ligands.
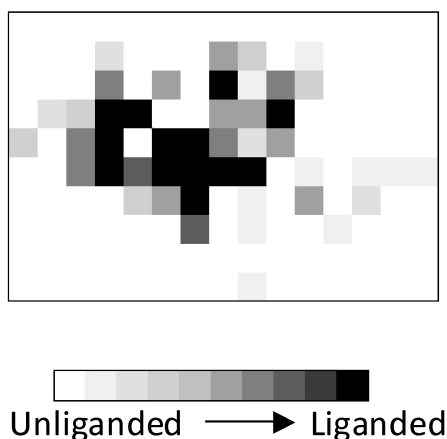

Unliganded ⟶ Liganded

Figure 5. SOM giving the distribution of protein surface cavities that were encoded by a shape descriptor (adapted from ref. [45]). In total, 98 pockets with a ligand bound from structurally diverse proteins (distribution indicated by gray shading) were compared to 2,257 ligand-free cavities. A clear separation between liganded and unliganded pockets is visible.

An example of potential future applications of the pocket shape concept is presented in Figure 6. Angiotensin-converting enzyme (ACE) was analyzed by the software PocketPicker [46], resulting in many surface cavities (Figure 6, left panel). The largest pocket turned out to be the actual active site. Using *de novo* ligand design software, several inhibitor candidates were generated *in machina*, and docked into the active site pocket (Figure 6, right panel). The predicted binding modes are similar to Lisinopril, a known ACE inhibitor [47].
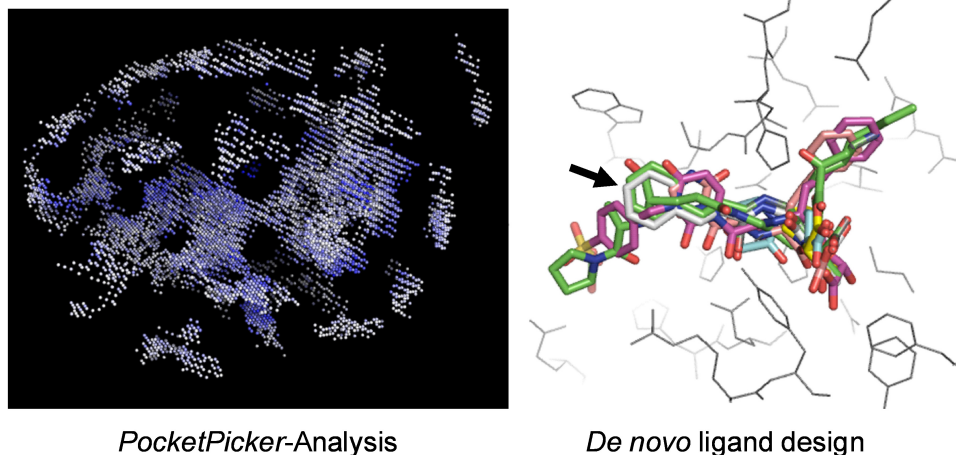
PocketPicker-Analysis   *De novo* ligand design

Figure 6. *Left*: Example of pocket analysis for an X-ray structure of ACE (Protein Data Bank identifier: 1o86 [48]). *Right*: Several *de novo* designed ligand candidates were docked into the active site. They exhibit essentially the same binding mode as the known ACE blocker Lisinopril (*gray-colored* co-crystal structure marked by the arrow).

This preliminary study demonstrates how we envisage systems chemistry might contribute to computer-assisted drug design in the future. Certainly, the SOM and other particular techniques presented here are not the only methods that can be applied. Much more advanced algorithms have been conceived. Computer science and in particular the various fields of technical engineering provide a rich source of pattern recognition and optimization strategies that may – when properly adapted to chemical applications – help understand and explore the potential of ligand "promiscuity" for rational drug design.

REFERENCES

[1] Roche, O., Schneider, P., Zuegge, J., Guba, W., Kansy, M., Alanine, A., Bleicher, K., Danel, F., Gutknecht, E.-M., Rogers-Evans, M., Neidhart, W., Stalder, H., Dillon, M., Sjögren, E., Fotouhi, N., Gillespie, P., Goodnow, R., Harris, W., Jones, P., Taniguchi, M., Tsujii, S., von der Saal, W., Zimmermann, G., Schneider, G. (2002) Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **45**:137 – 142.

[2] Crisman, T.J., Parker, C.N., Jenkins, J.L., Scheiber, J., Thoma, M., Kang, Z.B., Kim, R., Bender, A., Nettles, J.H., Davies, J.W., Glick, M. (2007) Understanding false positives in reporter gene assays: *in silico* chemogenomics approaches to prioritize cell-based HTS data. *J. Chem. Inf. Model* **47**:1319 – 1327.

[3] McGovern, S.L., Caselli, E., Grigorieff, N., Shoichet, B K. (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **45**:1712 – 1722.

[4] Morphy, R., Rankovic, Z. (2007) Fragments, network biology and designing multiple ligands. *Drug Discov. Today* **12**:156 – 160.

[5] Klabunde, T., Evers, A (2005) GPCR antitarget modeling: pharmacophore models for biogenic amine binding GPCRs to avoid GPCR-mediated side effects. *ChemBioChem* **6**:876 – 889.

[6] Raschi, E., Vasina, V., Poluzzi, E., De Ponti, F. (2008) The hERG K+ channel: target and antitarget strategies in drug development. *Pharmacol. Res.* **57**:181 – 195.

[7] Hann, M.M., Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* **8**:255 – 263.

[8] Hann, M.M, Leach, A.R, Harper, G. (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **41**:856 – 864.

[9] Sun, H. (2008) Pharmacophore-based virtual screening. *Curr. Med. Chem.* **15**:1018 – 1024.

[10] Wolber, G., Seidel, T., Bendix, F., Langer, T. (2008) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today* **13**:23 – 29.

[11] Gregori-Puigjané, E., Mestres., J. (2008) A ligand-based approach to mining the chemogenomic space of drugs. *Comb. Chem. High Throughput Screen.* **11**:669 – 676.

[12] Salemme, F.R. (2003) Chemical genomics as an emerging paradigm for postgenomic drug discovery. *Pharmacogenomics* **4**:257 – 267.

[13]   Horvath, D., Jeandenans, C. (2003) Neighborhood behavior of *in silico* structural spaces with respect to in vitro activity spaces – a novel understanding of the molecular similarity principle in the context of multiple receptor binding profiles. *J. Chem. Inf. Comput. Sci.* **43**:680 – 690.

[14]   Mestres, J., Martín-Couce, L., Gregori-Puigjané, E., Cases, M., Boyer, S. (2006) Ligand-based approach to *in silico* pharmacology: nuclear receptor profiling. *J. Chem. Inf. Model* **46**:2725 – 2736.

[15]   Schneider, G., Baringhaus, K.-H. (2008) *Molecular Design – Concepts and Applications*. Wiley-VCH, Weinheim.

[16]   Böhm, H.-J., Schneider, G. (Eds) (2000) *Virtual Screening for Bioactive Molecules*. Wiley-VCH, Weinheim.

[17]   Bauknecht, H., Zell, A., Bayer, H., Levi, P., Wagener, M., Sadowski, J., Gasteiger, J. (1996) Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: dopamine and benzodiazepine agonists. *J. Chem. Inf. Comput. Sci.* **36**:1205 – 1213.

[18]   Schneider, G., Schneider, P., Renner, S. (2006) Scaffold-hopping: how far can you jump? *QSAR Comb. Sci.* **25**: 1162 – 1171.

[19]   Chong, C.R., Sullivan Jr., D.J. (2007) New uses for old drugs. *Nature* **448**:645 – 646.

[20]   Carley, D.W. (2005) Drug repurposing: identify, develop and commercialize new uses for existing or abandoned drugs. Part I. *Drugs* **8**:306 – 309.

[21]   Cavasotto, C.N., Orry, A.J., Abagyan, R.A. (2003) Structure-based identification of binding sites, native ligands and potential inhibitors for G-protein coupled receptors. *Proteins* **51**:423 – 433.

[22]   Schneider, P., Tanrikulu, Y., Schneider, G. (2009) Self-organizing maps in drug discovery: Library design, scaffold-hopping, repurposing. *Curr. Med. Chem.* **16**:258 – 266.

[23]   Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**:59 – 69.

[24]   Kohonen, T (2001) *Self-Organizing Maps*. Springer-Verlag: Berlin.

[25]   Zupan, J., Gasteiger, J. (1999) *Neural Networks in Chemistry and Drug Design*, Wiley-VCH: Weinheim.

[26]   Todeschini, R., Consonni, V. (2000) *Handbook of Molecular Descriptors*, Wiley-VCH: Weinheim.

[27]   Schneider, G., Neidhart, W., Giller, T., Schmid, G. (1999) "Scaffold-Hopping" by topological pharmacophore search: A contribution to virtual screening. *Angew. Chem. Int. Ed.* **38**:2894 – 2896.

[28]   Fechner, U., Schneider, G. (2004) Optimization of a pharmacophore-based correlation vector descriptor. *QSAR Comb. Sci.* **23**:19 – 22.

[29]   Schneider, P., Schneider, G. (2003) Collection of bioactive reference compounds for focused library design. *QSAR Comb. Sci.* **22**:713 – 718.

[30]   Schneider, G., Schneider, P. (2004) Navigation in chemical space: Ligand-based design of focused compound libraries. In: *Chemogenomics in Drug Discovery* (Kubinyi, H., Müller, H., Eds), Wiley-VCH: Weinheim, pp. 341 – 376.

[31]   Noeske, T., Sasse, B.C., Stark, H., Parsons, C.G., Weil, T., Schneider, G. (2006) Predicting compound selectivity by self-organizing maps: cross-activities of metabotropic glutamate receptor antagonists. *ChemMedChem* **1**:1066 – 1068.

[32]   Yan, A. (2007) Application of self-organizing maps in compounds pattern recognition and combinatorial library design. *Comb. Chem. High Throughput. Screen* **9**:473 – 480.

[33]   Hristozov, D., Oprea, T.I., Gasteiger, J. (2007) Ligand-based virtual screening by novelty detection with self-organizing maps. *J. Chem. Inf. Model* **47**:2044 – 2062.

[34]   Renner, S., Hechenberger, M., Noeske, T., Böcker, A., Jatzke, C., Schmuker, M., Parsons, C. G., Weil, T., Schneider, G. (2007) Searching for drug scaffolds with 3D pharmacophores and neural network ensembles. *Angew. Chem. Int. Ed*. **46**:5336 – 5339.

[35]   Schneider, G., Nettekoven, M. (2003) Ligand-based combinatorial design of selective purinergic receptor (A2A) antagonists using self-organizing maps. *J. Comb. Chem.* **5**:233 – 237.

[36]   Bajorath, J. (2008) Computational analysis of ligand relationships within target families. *Curr. Opin. Chem. Biol.* **12**:352 – 358.

[37]   Li, A., Horvath, S. (2007) Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics* **23**:222 – 231.

[38]   Schneider, G., Tanrikulu, Y., Schneider, P. (2009) Self-organizing molecular fingerprints: a ligand-based view on druglike chemical space and off-target prediction. *Future Med. Chem*., in press.

[39]   Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J., Shoichet, B.K. (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **25**:197 – 206.

[40] Park, K., Kim, D. (2008) Binding similarity network of ligand. *Proteins* **71**:960 – 971 (2008).

[41] An, J., Totrov, M., Abagyan, R. (2004) Comprehensive identification of "druggable" protein ligand binding sites. *Genome Inform.* **15**:31 – 41.

[42] Zhang, Z., Grigorov, M.G. (2006) Similarity networks of protein binding sites. *Proteins* **62**:470 – 478 (2006).

[43] Kupas, K., Ultsch, A., Klebe, G. (2008) Large scale analysis of protein-binding cavities using self-organizing maps and wavelet-based surface patches to describe functional properties, selectivity discrimination, and putative cross-reactivity. *Proteins* **71**:1288 – 1306.

[44] Schalon, C., Surgand, J.S., Kellenberger, E., Rognan, D. (2008) A simple and fuzzy method to align and compare druggable ligand-binding sites. *Proteins* **7**:17775 – 1778.

[45] Weisel, M., Proschak, E., Kriegl, J.M., Schneider, G. (2009) Form follows function: Shape analysis of protein cavities for receptor-based drug design. *Proteomics* **9**:451 – 459.

[46] Weisel, M., Proschak, E., Schneider, G. (2007) PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem. Cent. J.* **1**:7.

[47] Krueger, B.A., Dietrich, A. Baringhaus, K.-H., Schneider, G. (2009) Scaffold-hopping potential of fragment-based de novo design: The chances and limits of variation. *QSAR Comb. Sci.*, in press.

[48] Natesh, R., Schwager, S.L., Sturrock, E.D., Acharya, K.R. (2003) Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature* **421**:551 – 554.