Virtual Screening

DOI: 10.1002/anie.200604125

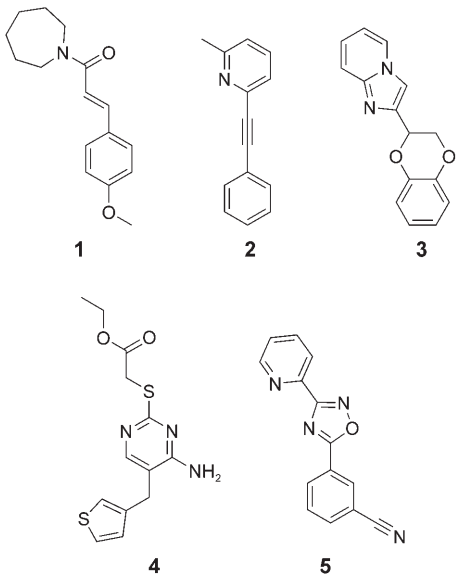
Searching for Drug Scaffolds with 3D Pharmacophores and Neural Network Ensembles**

Steffen Renner, Mirko Hechenberger, Tobias Noeske, Alexander Böcker, Claudia Jatzke, Michael Schmuker, Christopher G. Parsons, Tanja Weil, and Gisbert Schneider**

Dedicated to Lena Maria Renner

The search for new scaffolds, "scaffold hopping", is a challenging goal in drug design.[1] Given at least one active reference molecule, the aim is to identify different chemotypes that effect equal biological responses through the same receptor. Given the size of today's screening libraries, efficient computational evalution of these compounds before biochemical testing is a pivotal step on the way to increase the success rate of screening campaigns.[2]

Machine learning methods like artificial neural networks (ANN) have been shown to be effective for meaningful prioritization of regions in chemical space.[3] Here, we present an approach that combines a three-dimensional (3D) pharmacophore descriptor with ANN for the identification of negative allosteric modulators (NAM), for example **1**–**5**, of metabotropic glutamate receptor 5 (mGluR5). Treatment of pain, anxiety, and Parkinson's disease are therapeutic prospects of such molecules.[4] Metabotropic glutamate receptors are class III G-protein-coupled receptors (GPCR).[5] Currently we lack detailed structural information for the transmembrane region, which contains the binding site for mGluR5 NAMs.[6] As a consequence, ligand-based virtual screening represents an attractive approach for this target class. We have recently reported a ligand-based virtual screening study using pharmacophore similarity searching that already yielded novel active scaffolds for mGluR5 NAMs;[7] the best inhibitor was compound **1** with $IC_{50} = 12\ \mu M$. In the present work, we demonstrate the influence of different ANN models on virtual screening performance and scaffold-hopping.



Molecular representation and encoding is one of the decisive steps that influence the outcome of a virtual screening experiment.[8,9] Autocorrelation descriptors have been successfully used in drug design.[10] We used one such descriptor, CATS3D, for molecule encoding.[7,8b] CATS3D records the distribution of pairs of potential pharmacophore points in a molecule over 3D Cartesian space (for details, see the Supporting Information). Instead of exhaustive conformational sampling, a single heuristic conformer was calculated for each molecule using CORINA.[11a] This approach was demonstrated to be satisfactory for rapid first-pass virtual screening.[11b]

Then, two types of ANN were trained: supervised feed-forward-type networks and self-organizing maps (SOM). The feed-forward networks served as the basis for compilation of an mGluR5-focused library, and the SOM were used for picking structurally diverse compounds for bioactivity testing. Supervised networks were trained on two classification tasks: mGluR5 allosteric antagonists versus molecules acting on other biological targets ("mGluR5-likeness" filter), and mGluR5 versus mGluR1 allosteric antagonists ("mGluR5/mGluR1 selectivity" filter), as mGluR1 is the most similar receptor to mGluR5 among the family of metabotropic glutamate receptors. To enhance prediction accuracy, net-

[*] A. Böcker, M. Schmuker, Prof. Dr. G. Schneider
Institute of Organic Chemistry and Chemical Biology
Johann Wolfgang Goethe University
Siesmayerstrasse 70, 60323 Frankfurt am Main (Germany)
Fax: (+49) 69-798-24880
E-mail: gisbert.schneider@modlab.de
Dr. S. Renner,[+] Dr. M. Hechenberger,[+] T. Noeske, Dr. C. Jatzke, Dr. C. G. Parsons, Dr. T. Weil
Chemical R&D, Medicinal Chemistry/Cheminformatics
Merz Pharmaceuticals GmbH
Eckenheimer Landstrasse 100
60318 Frankfurt am Main (Germany)

[+] S. Renner and M. Hechenberger share first authorship.

[**] We thank Tanja Bauer, Andrea Baude, and Sylvia Willi for technical assistance with mGluR1/5 experiments and Dr. Karl-Heinz Eisele for performing dynamic light-scattering experiments. This research was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt am Main.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

works ensembles instead of single network classifiers were used (averaged "jury" prediction).[12]

For training of the supervised networks, 68 known allosteric mGluR5 antagonists with functional IC_{50} values below 1 μM were used as "active" reference structures; they were compiled from patents, literature, and Merz in-house data (e.g. **2**–**5**). Five different data sets of 100 diverse "inactive" molecules (i.e. molecules that act on different receptors) each were selected from the COBRA collection.[13] The training set for "mGluR5/mGluR1 selectivity" consisted of the 68 mGluR5 antagonists and 158 allosteric mGluR1 antagonists with $IC_{50} < 1$ μM. In total, ten networks were trained on "mGluR5-likeness" prediction: five for the different training set combinations, and five using the principle components (Eigenvalues > 1) of the CATS3D descriptors of the five datasets as input. Two ANNs were trained on selectivity classification.

The SOM were used to visualize the distribution of molecules in high-dimensional CATS3D descriptor space. The result of SOM training is a two-dimensional map indicating the location of the molecules. This map yields molecule clusters based on which we picked a diverse and representative compound subset for testing (see the Supporting Information).

For virtual screening, we used the Enamine compound collection[19] (1 022 483 molecules). Application of the "mGluR5-likeness" filter yielded 41 633 molecules (jury network score > 0.885). This number was further reduced to 8403 molecules by the selectivity filter (jury network score > 0.99), resulting in a focused library of potential mGluR5 antagonists.

We then analyzed scaffold diversity in this library. Three definitions of "scaffold" were employed: 1) the "Murcko scaffold",[14] retaining all information about atom types and bond order, 2) the "carbon scaffold",[1c,d] ignoring information about atom types and bond order, and 3) the "reduced scaffold",[1c,d] ignoring information about ring size as well.

The mGluR5 reference compounds and the predicted focused library only marginally overlapped in their Murcko scaffolds (training set: 34 scaffolds, focused library: 1586 scaffolds, overlap: 1). They partially (no. of scaffolds: 22, 499, 11) and almost entirely (14, 75, 12) overlapped according to the definitions of the carbon scaffold and the reduced scaffold, respectively. Considerable numbers of new chemotypes were identified according to all three scaffold definitions.

For picking a diverse set of screening compounds from the focused library, two SOM with 5 × 5 neurons were trained with the focused library plus the reference ligands of mGluR5: one based on CATS3D descriptors and the other based on MACCS substructure keys (166 bits as implemented in MOE[20]). The compound distribution in the CATS3D SOM displays two highly populated (E4, E5) and many sparsely populated clusters (Figure 1a). The same set of molecules is more evenly distributed over the map resulting from MACCS substructure keys. This hints to a greater variability of chemotypes compared to pharmacophores, which is exactly what one expects from a scaffold-rich, target-focused library.

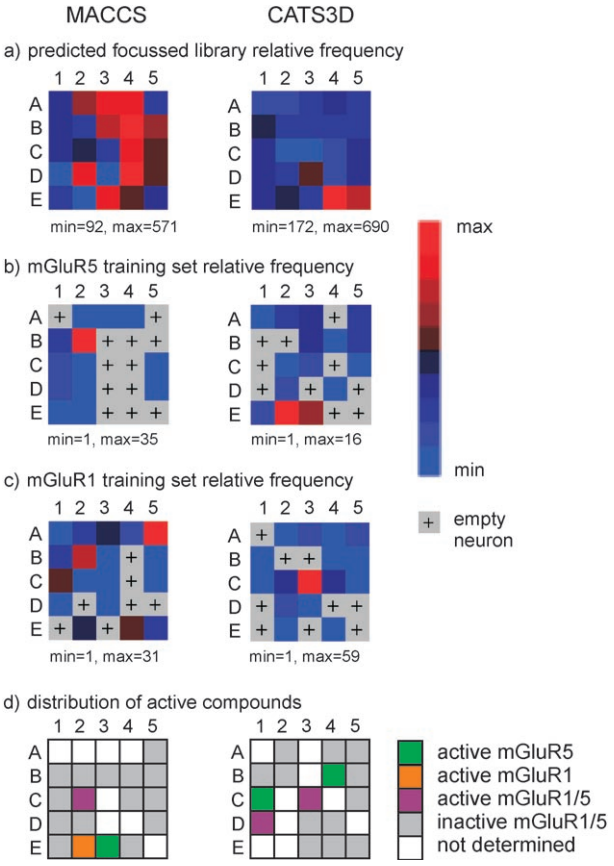


Figure 1. SOM projections of the molecules predicted to be active by the neural network. Each square represents a data cluster ("neuron"). Left column: MACCS key descriptor. Right column: CATS3D descriptor. a) Distribution of predicted active molecules (focused library). b) Distribution of the mGluR5 training set. c) Distribution of the mGluR1 training set. d) Distribution of tested molecules.

Projections of the mGluR5 (Figure 1b) and mGluR1 (Figure 1c) ligands on the maps highlight two characteristics of the predicted active set of molecules: The first refers to the previously observed difficulty of predicting selectivity between mGluR5 and mGluR1 using CATS3D,[7] which can be attributed to the coarse-grained molecular representation and the apparent similarity of the ligand binding sites. Many clusters are populated by both mGluR5 and mGluR1 reference molecules, irrespective of the molecular descriptors used. It is not surprising that using a simplifying pharmacophore descriptor like CATS3D can also yield hits for related receptors.[15] This effect has already been exploited as a ligand design strategy: variations of ligands for one receptor led to potent ligands for related targets.[16]

The second characteristic refers to the exploration of "extended" chemical space: Several clusters (MACCS: 12, CATS3D: 10) do not contain a single known reference molecule for mGluR5, and some (3 for MACCS and CATS3D) do not contain a reference molecule from either mGluR5 or mGluR1. The latter might therefore be particularly "hot" candidates for screening: Based on the neural network predictions, these molecules slightly extend the known ligand space for mGluR1 and mGluR5.

As a general trend, we observed within our lead optimization programs that weakly active mGluR5 antagonists ($IC_{50} > 10$ μM) display only a little or no tendency to displace a highly active radio-ligand from a binding site without showing an effect in functional assays (neutral allosteric site ligands, for example, 5-methyl-6-(phenylethynyl)-pyridine).[17] Other compounds that are functionally highly potent antagonists might have a reduced potency in displacing the radio-ligand, when binding sites are only partially overlapping. For an assessment of the actual value and accuracy of our predictions we therefore relied on functional assays with a diverse subset of the focused library (see the Supporting Information). Nevertheless, such ligands usually represent valuable starting points for an activity optimization program, and potent mGluR5 antagonists always displayed a stronger correlation between functional and binding data (not shown). Since we were not able to measure dose–response curves for the majority of the ligands owing to low solubility above 10 μM, the biological activities of some compounds were estimated from single-point measurements (see the Supporting Information).

From each SOM cluster, a representative molecule was ordered. 33 of 50 molecules were delivered and tested (16 from the CATS3D map, and 17 from the MACCS map). Seven molecules exhibited estimated functional IC_{50} values below 50 μM in mGluR1 or mGluR5 FLIPR assays (see the Supporting Information). The most potent mGluR5 hits **6**, **7**, and **8** yielded functional IC_{50} values between 9.6 and 20.7 μM (Figure 2). Noteworthy, four molecules exhibited $IC_{50} < 50$ μM in our mGluR1 FLIPR assays, with **6** and **8** yielding $IC_{50} =$ 2.5 μM and 4.6 μM, respectively (Figure 3). Dynamic light-scattering experiments for all selected compounds indicated that none of the active molecules seems to be active as a result of unspecific aggregation.[18]

Remarkably, two mGluR5 hits were picked from clusters lacking any of the reference structures (MACCS: E3; CATS3D: D1). None of our hits comprises a Murcko or carbon scaffold of an mGluR5 or mGluR1 reference compound or has been reported to be active on other mGluR subtypes (mGluR2-4, mGluR6-8).

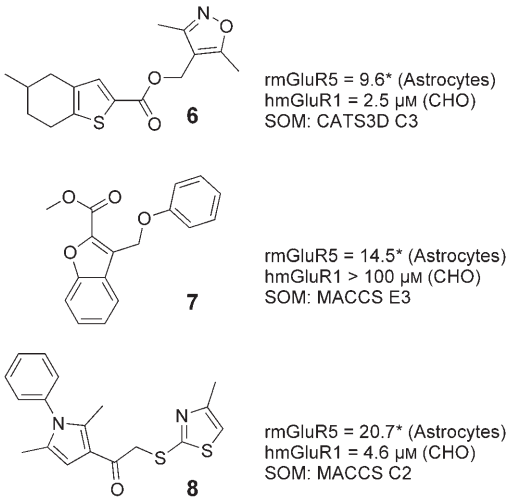


Figure 2. Selected hits that resulted from virtual screening. Activity data for mGluR5 and mGluR1 are given together with the corresponding cluster identifiers of the SOM (cf. Figure 1). The asterisk denotes estimated IC_{50} values based on a single-point determination.

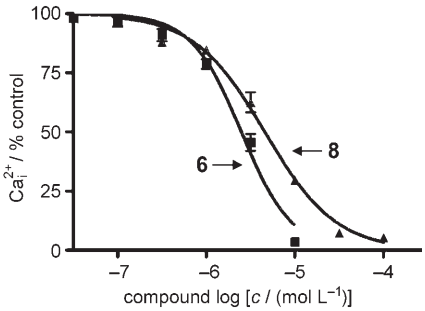


Figure 3. Functional response of the hmGluR1 receptor for compounds **6** and **8** in a FLIPR assay monitoring the intracellular Ca^{2+} level.

Summarizing, we have presented and validated a virtual screening approach for mapping and prioritization of chemical space using supervised and unsupervised neural networks and an alignment-free 3D pharmacophore descriptor. We performed two-step filtering: First by supervised ANN, then by SOM-based diversity analysis and compound selection. Different scaffolds of mGluR ligands were obtained although we followed a similar computational concept (neural networks) as in previous work.[7,16b] It is, therefore, not justified to generally argue that neural network approaches can be expected to lead to similar results. Seemingly minor technical details can have a major impact on the outcome of virtual screening. The presented strategy is suited for early-phase hit and lead discovery since it selects promising screening candidates from a large compound pool prior to testing in experimental assay systems. It might also be useful for systematically filling voids in screening compound collections.

Received: October 7, 2006

Keywords: computer chemistry · drug design · neural networks · selectivity · virtual screening

[1] a) G. Schneider, U. Fechner, *Nat. Rev. Drug Discovery* **2005**, *4*, 649; b) H.-J. Böhm, A. Flohr, M. Stahl, *Drug Discovery Today Technol.* **2004**, *1*, 217; c) J. L. Jenkins, M. Glick, J. W. Davies, *J. Med. Chem.* **2004**, *47*, 6144; d) S. Renner, G. Schneider, *ChemMedChem* **2006**, *1*, 181; e) Q. Zhang, I. Muegge, *J. Med. Chem.* **2006**, *49*, 1536; f) N. Brown, E. Jacoby, *Mini-Rev. Med. Chem.* **2006**, *6*, 1217; g) G. Schneider, P. Schneider, S. Renner, *QSAR Comb. Sci.* **2006**, *25*, 1162.

[2] a) J. Bajorath, *Nat. Rev. Drug Discovery* **2002**, *1*, 882; b) K. H. Bleicher, H.-J. Böhm, K. Müller, A. I. Alanine, *Nat. Rev. Drug Discovery* **2003**, *2*, 369.

[3] a) E. Byvatov, B. C. Sasse, H. Stark, G. Schneider, *ChemBioChem* **2005**, *6*, 997; b) L. Franke, E. Byvatov, O. Werz, D. Steinhilber, P. Schneider, G. Schneider, *J. Med. Chem.* **2005**, *48*, 6997; c) G. V. Paolini, R. H. Shapland, W. P. van Hoorn, J. S. Mason, A. L. Hopkins, *Nat. Biotechnol.* **2006**, *24*, 805; d) Nidhi, M. Glick, J. W. Davies, J. L. Jenkins, *J. Chem. Inf. Model.* **2006**, *46*, 1124; e) P. Selzer, P. Ertl, *J. Chem. Inf. Model.* **2006**, *46*, 2319; f) J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, Weinheim, **1999**.

[4] a) W. P. Spooren, F. Gasparini, T. E. Salt, R. Kuhn, *Trends Pharmacol. Sci.* **2001**, *22*, 331; b) C. J. Swanson, M. Bures, M. P. Johnson, A. M. Linden, J. A. Monn, D. D. Schoepp, *Nat. Rev. Drug Discovery* **2005**, *4*, 131.
[5] T. Noeske, A. Gutcaits, C. G. Parsons, T. Weil, *QSAR Comb. Sci.* **2005**, *24*, 134.
[6] P. Malherbe, N. Kratochwil, M. T. Zenner, J. Piussi, C. Diener, C. Kratzeisen, C. Fischer, R. H. Porter, *Mol. Pharmacol.* **2003**, *64*, 823.
[7] S. Renner, T. Noeske, C. G. Parsons, P. Schneider, T. Weil, G. Schneider, *ChemBioChem* **2005**, *6*, 620.
[8] a) R. D. Brown, Y. Martin, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572; b) U. Fechner, L. Franke, S. Renner, P. Schneider, G. Schneider, *J. Comput.-Aided Mol. Des.* **2003**, *17*, 687; c) J. Hert, P. Willett, D. J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, A. Schuffenhauer, *Org. Biomol. Chem.* **2004**, *2*, 3256.
[9] A. M. Johnson, G. M. Maggiora, *Concepts and Applications of Molecular Similarity*, Wiley, New York, **1990**.
[10] a) M. Wagener, J. Sadowski, J. Gasteiger, *J. Am. Chem. Soc.* **1995**, *117*, 7769; b) R. P. Sheridan, M. D. Miller, D. J. Underwood, S. K. Kearsley, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 128; c) M. Pastor, G. Cruciani, I. McLay, S. Pickett, S. Clementi, *J. Med. Chem.* **2000**, *43*, 3233; d) K. Baumann, *QSAR Comb. Sci.* **2002**, *21*, 505.
[11] a) CORINA, Version 2.64, Molecular Networks GmbH, Erlangen, Germany, **2003**; b) S. Renner, C. H. Schwab, J. Gasteiger, G. Schneider, *J. Chem. Inf. Model.* **2006**, *46*, 2324.
[12] a) S.-S. So, M. Karplus, *J. Med. Chem.* **1996**, *39*, 1521; b) G. W. Kauffman, P. Jurs, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 753; c) A. Givehchi, G. Schneider, *Mol. Diversity* **2005**, *9*, 371.
[13] P. Schneider, G. Schneider, *QSAR Comb. Sci.* **2003**, *22*, 713.
[14] G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887.
[15] A. Schuffenhauer, P. Floersheim, P. Acklin, E. Jacoby, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 391.
[16] a) M. A. Koch, L. O. Wittenberg, S. Basu, D. A. Jeyaraj, E. Gourzoulidou, K. Reinecke, A. Odermatt, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16721; b) T. Noeske, B. C. Sasse, H. Stark, C. G. Parsons, T. Weil, G. Schneider, *ChemMedChem* **2006**, *1*, 1066.
[17] A. L. Rodriguez, Y. Non, N. K. Sekaran, d. Alagille, G. D. Tamagnan, P. J. Conn, *Mol. Pharmacol.* **2005**, *68*, 1793.
[18] S. L. McGovern, E. Caselli, N. Grigorieff, B. K. Shoichet, *J. Med. Chem.* **2002**, *45*, 1712.
[19] Enamine compound collection: Enamine Ltd., Kiev, Ukraine.
[20] MOE v2004.03, Chemical Computing Group Inc., Montréal, Canada, **2004**.

