

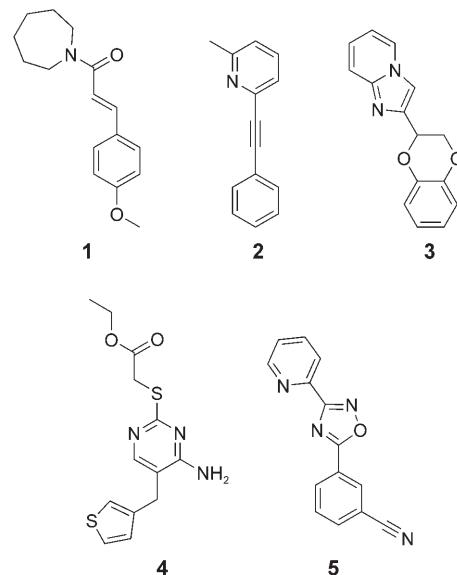
Suche nach Wirkstoff-Grundgerüsten mit 3D-Pharmakophor-hypothesen und Ensembles neuronaler Netze**

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

Lena Maria Renner gewidmet

Die Suche nach neuen Grundgerüsten („Scaffold-Hopping“) stellt eine große Herausforderung in der Wirkstoffentwicklung dar.^[1] Ausgehend von einer biologisch aktiven Substanz besteht das Ziel darin, Moleküle anderer Struktur mit gleicher Funktion zu finden, die ihre Wirkung über denselben Rezeptor ausüben. Angesichts der Größe heutiger Substanzbibliotheken kann ein Vorauswahl mithilfe von Computerverfahren die Effizienz und Erfolgsquote in biologischen und biochemischen Tests wesentlich erhöhen.^[2]

Maschinelle Lernverfahren wie künstliche neuronale Netze (KNN) sind effektive Methoden zur Priorisierung des chemischen Raums.^[3] Hier stellen wir eine Kombination von dreidimensionalen (3D) Pharmakophor-Deskriptoren und KNN zur Vorhersage von neuen negativen allosterischen Modulatoren (NAM) wie **1–5** des metabotropen Glutamatrezeptors 5 (mGluR5) vor. Therapeutische Anwendungen solcher Verbindungen liegen in der Behandlung von Schmerz, Angstzuständen oder der Parkinson-Erkrankung.^[4] Metabotrope Glutamatrezeptoren gehören zur Klasse III der G-Protein-gekoppelten Rezeptoren (GPCR).^[5] Die Bindetasche der mGluR5-NAMs liegt in der Transmembranregion des Rezeptors, für die noch keine hochauflöste Struktur bekannt ist.^[6] Ligandenbasierte virtuelle Screening-Verfahren sind daher für diese Rezeptorklasse besonders vielverspre-



chend. Das virtuelle Screening auf Pharmakophorähnlichkeit führte bereits zur Identifizierung neuer Grundgerüste von mGluR5-NAMs,^[7] bei dem sich **1** als der beste Inhibitor erwies ($K_i = 12 \mu\text{M}$). Hier demonstrieren wir den Einfluss zweier Typen neuronaler Netze auf das virtuelle Screening und die Suche nach Grundgerüsten.

Die Repräsentation und Kodierung eines Moleküls hat einen entscheidenden Einfluss auf das Ergebnis eines virtuellen Screenings.^[8,9] Autokorrelations-Deskriptoren wurden bereits erfolgreich in der Wirkstoffentwicklung eingesetzt.^[10] Hier wurde mit CATS3D^[7,8b] ebenfalls ein solcher Deskriptor verwendet. CATS3D beschreibt die Häufigkeitsverteilung von Paaren potenzieller Pharmakophorpunkte eines Moleküls in einem dreidimensionalen kartesischen Koordinatensystem (für Details siehe die Hintergrundinformationen). Anstelle der Repräsentation des vollständigen Konformationsraums eines Moleküls wurde nur eine einzige heuristische Konformation für jedes Molekül mit CORINA^[11a] erzeugt. Ein solcher Ansatz hat bereits zu guten Ergebnissen im virtuellen Screening geführt.^[11b]

Zwei Klassen neuronaler Netze wurden eingesetzt: überwachte mehrschichtige Netze und selbstorganisierende Karten (self-organizing maps, SOMs). Mit überwachten Netzen wurde eine fokussierte Bibliothek von Molekülen mit vorhergesagter mGluR5-Aktivität erstellt. Eine repräsentative und diverse Auswahl hieraus wurde über SOMs getroffen. Überwachte Netze lieferten zwei Vorhersagen: 1) die

[*] A. Böcker, M. Schmuker, Prof. Dr. G. Schneider
Institut für Organische Chemie und Chemische Biologie
Johann-Wolfgang-Goethe-Universität
Siesmayerstraße 70, 60323 Frankfurt am Main (Deutschland)
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 (Deutschland)
Fax: (+49) 69-1503-188
E-Mail: tanja.weil@merz.de

[†] Beide Autoren trugen zu gleichen Teilen zu dieser Arbeit bei.

[**] Die Autoren danken Tanja Bauer, Andrea Baude und Sylvia Willi für technische Unterstützung bei den mGluR1/5-Messungen sowie Dr. Karl-Heinz Eisele für Hilfe bei den Dynamic-Light-Scattering-Experimenten. Für finanzielle Unterstützung danken wir dem Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt am Main.

Hintergrundinformationen zu diesem Beitrag sind im WWW unter <http://www.angewandte.de> zu finden oder können beim Autor angefordert werden.

Trennung von mGluR5-NAMs von Molekülen mit Aktivität gegen andere Zielstrukturen („mGluR5-Artigkeit“) und 2) die Trennung von mGluR5-NAMs von mGluR1-NAMs („mGluR5/mGluR1-Selektivität“). Innerhalb der mGlu-Rezeptoren ist mGluR1 dem mGluR5 am ähnlichsten. Zur Erhöhung der Vorhersagegenauigkeit wurden statt einzelner neuronaler Netze Ensembles von neuronalen Netzen eingesetzt (gemittelte „Jury“-Vorhersage).^[12]

Als Datensatz für das Netztraining wurden 68 allosterische mGluR5-Antagonisten mit funktionalem $IC_{50} < 1 \mu\text{M}$ aus der Literatur und Merz-internen Daten ausgewählt (z. B. 2–5). Fünf unterschiedliche diverse Datensätze von 100 inaktiven Molekülen (mit Aktivität gegen andere Rezeptoren) wurden aus der COBRA-Datenbank gewählt.^[13] Der Trainingsdatensatz für die Selektivitätsklassifizierung bestand aus den 68 mGluR5-aktiven und 158 allosterischen Antagonisten von mGluR1 mit $IC_{50} < 1 \mu\text{M}$. Insgesamt wurden zehn Netze auf „mGluR5-Artigkeit“ trainiert (fünf durch Kombination der Datensätze und fünf weitere durch Berechnung der Hauptkomponenten der CATS3D-Deskriptoren (Eigenwerte > 1) der fünf Datensätze). Zwei Netze wurden auf Selektivität trainiert.

SOMs wurden zur Darstellung der Verteilung der Moleküle im hochdimensionalen CATS3D-Raum eingesetzt. Eine SOM projiziert diese Verteilung auf eine zweidimensionale Karte. Auf der Grundlage dieser Darstellung wurde eine repräsentative Auswahl an Molekülen für Messungen der biologischen Aktivität getroffen (siehe die Hintergrundinformationen).

Zum virtuellen Screening wurde die Enamine-Datenbank^[19] eingesetzt (1022483 Moleküle). Die Anwendung des „mGluR5-Artigkeit“-Filters resultierte in 41633 Molekülen (Jury-Netz-Score > 0.885). Der Selektivitätsfilter reduzierte diese Zahl weiter auf eine fokussierte Bibliothek mit 8403 Molekülen (Jury-Netz-Score > 0.99).

Bei dieser Untersuchung wurden drei Definitionen eines „Grundgerüsts“ herangezogen: 1) das „Murcko-Grundgerüst“,^[14] definiert durch Beibehaltung aller Informationen über Atom- und Bindungsarten, 2) ein „Kohlenstoff-Grundgerüst“,^[1c,d] definiert durch Vernachlässigung dieser Information, und 3) ein „reduziertes Grundgerüst“,^[1c,d] bei dem zusätzlich die Größe von Ringen variabel ist.

Die Moleküle des mGluR5-Trainingsdatensatzes und die vorhergesagte fokussierte Bibliothek überlappten nur geringfügig hinsichtlich ihrer Murcko-Grundgerüste (Trainingsdatensatz: 34 Gerüste, fokussierte Bibliothek: 1586 Gerüste, gemeinsam: 1). Partielle (22, 499, 11) und fast vollständige Überlappung (14, 75, 12) wurde für Kohlenstoff-Grundgerüste bzw. reduzierte Grundgerüste gefunden. Nach allen drei Definitionen konnte jeweils eine signifikante Zahl neuer Grundgerüste gefunden werden.

Zur Auswahl eines diversen Satzes von Molekülen für biologische Tests wurden mit den Molekülen der fokussierten Bibliothek und den Molekülen des mGluR5-Trainingsdatensatzes zwei SOMs mit jeweils 5×5 -Neuronen trainiert, eine mit CATS3D-Deskriptoren und eine mit MACCS-Substruktur-Fingerprints (166 Bits gemäß der Implementierung in MOE^[20]). Die Verteilung der Moleküle auf der CATS3D-SOM zeigt zwei stark besiedelte (E4, E5) und viele dünn

besiedelte Cluster (Abbildung 1 a). Dieselben Moleküle liegen in der MACCS-SOM deutlich gleichmäßiger verteilt vor. Dies deutet auf eine höhere chemische Variabilität der Moleküle im Vergleich zu ihrem Pharmakophormuster hin,

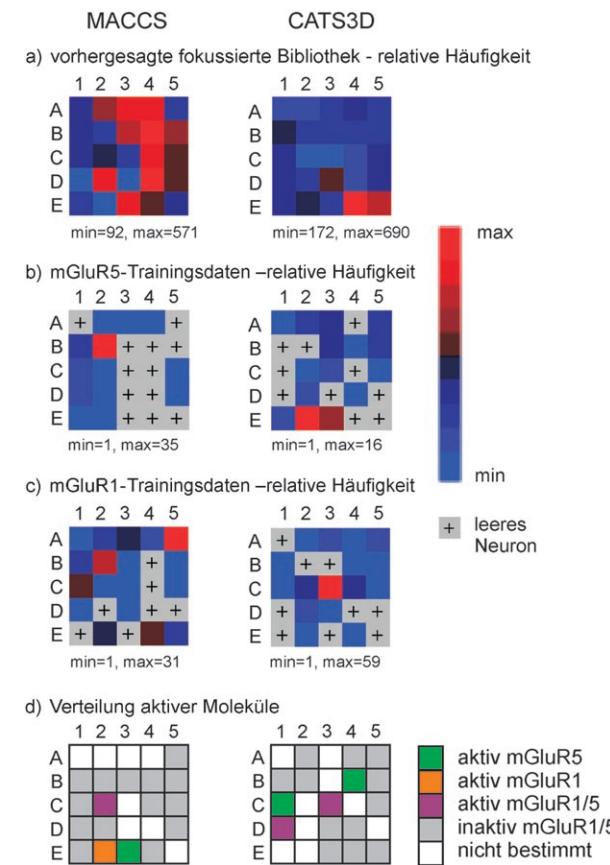


Abbildung 1. SOM-Projektionen der als aktiv vorhergesagten Moleküle. Jedes Quadrat repräsentiert einen Datencluster („Neuron“). Linke Spalte: MACCS-Deskriptor. Rechte Spalte: CATS3D-Deskriptor. a) Verteilung der als aktiv vorhergesagten Moleküle (fokussierte Bibliothek). b) Verteilung der mGluR5-Trainingsdaten. c) Verteilung der mGluR1-Trainingsdaten. d) Verteilung der gemessenen Aktivität der getesteten Moleküle.

was genau der Erwartung an eine fokussierte Substanzbibliothek mit vielen verschiedenen Grundgerüsten entspricht.

Die Projektionen der mGluR5- und mGluR1-Liganden (Abbildung 1 b bzw. 1 c) aus dem Trainingsdatensatz auf die SOMs veranschaulichen zwei charakteristische Eigenschaften der erstellten Bibliothek.

Die erste Eigenschaft betrifft die bereits früher beobachtete Schwierigkeit bei der Vorhersage einer mGluR5- oder mGluR1-Selektivität mit CATS3D.^[7] Diese wurde mit der unscharfen Beschreibung der Moleküle sowie der Ähnlichkeit der Bindetaschen der beiden Rezeptoren erklärt. Viele Neuronen in den SOMs sind von mGluR1- und mGluR5-Liganden besetzt, unabhängig von deren Beschreibung mit CATS3D oder MACCS. Es ist nicht überraschend, dass mit einer vereinfachten Pharmakophorbeschreibung wie CATS3D auch Moleküle für verwandte Rezeptoren gefunden werden.^[15] Dieser Effekt wurde bereits in einer Liganden-

design-Strategie genutzt: Die Variation von Liganden eines Rezeptors führte zu potenteren Liganden verwandter Rezeptoren.^[16]

Die zweite Eigenschaft betrifft die Erkundung neuen chemischen Raumes für mGluR5-Inhibitoren: Einige Cluster (MACCS: 12, CATS3D: 10) enthalten kein Referenzmolekül aus dem mGluR5-Trainingsdatensatz, und wenige Cluster (MACCS und CATS3D jeweils 3) enthalten weder mGluR5- noch mGluR1-Referenzmoleküle. Moleküle aus solchen Clustern sind besonders interessante Kandidaten für das Screening. Basierend auf den Vorhersagen der neuronalen Netze fallen diese Moleküle nicht mehr in den bekannten chemischen Raum von mGluR1- oder mGluR5-Liganden.

Aus unseren Leitstrukturoptimierungsprogrammen war uns bekannt, dass wenig aktive mGluR5-Antagonisten ($IC_{50} > 10 \mu\text{M}$) einen hoch aktiven Radioliganden generell nur schwach aus der Bindetasche verdrängen (neutrale Liganden für allosterische Bindungsstellen, z.B. 5-Methyl-6-(phenylethynyl)pyridin^[17]). Auch funktionell hochaktive Antagonisten können nur schwach verdrängend wirken, wenn sich die Bindetaschen nur partiell überlappen. Um die Genauigkeit unserer Vorhersagen zu beurteilen, wurde daher die funktionelle Aktivität einer diversen Auswahl aus der vorhergesagten Bibliothek gemessen. Da für die Mehrheit der gewählten Substanzen aufgrund der schlechten Löslichkeit oberhalb von $10 \mu\text{M}$ keine Dosis-Wirkungs-Kurve erstellt werden konnte, wurde die biologische Aktivität einiger Substanzen aus Einpunktbestimmungen abgeleitet (siehe die Hintergrundinformationen).

Aus jedem SOM-Cluster wurde ein repräsentatives Molekül der Enamine-Bibliothek gekauft. 33 der 50 Moleküle wurden geliefert und getestet (16 aus der CATS3D-Karte und 17 aus der MACCS-Karte). Sieben Moleküle zeigten einen geschätzten IC_{50} -Wert unter $50 \mu\text{M}$ in mGluR1- oder mGluR5-FLIPR-Assays (siehe die Hintergrundinformationen). Die potentesten mGluR5-Inhibitoren **6**, **7** und **8** hatten funktionelle IC_{50} -Werte zwischen 9.6 und $20.7 \mu\text{M}$ (Abbildung 2). Vier Moleküle zeigten IC_{50} -Werte $< 50 \mu\text{M}$ in mGluR1-FLIPR-Assays, **6** und **8** schnitten mit 2.5 bzw. $4.6 \mu\text{M}$ am besten ab (Abbildung 3). In dynamischen Lichtstreuungsexperimenten wies keine der aktiven Verbindungen unspezifische Aggregation und daraus resultierende unspezifische Aktivität auf.^[18]

Bemerkenswerterweise konnten zwei aktive Moleküle in Clustern ohne bekannte Referenzstruktur gefunden werden (MACCS: E3; CATS3D: D1). Keiner der Treffer hatte ein Murcko- oder Kohlenstoff-Grundgerüst eines bekannten mGluR5- oder mGluR1-Referenzmoleküls oder wurde bereits als aktiv gegen eine der anderen mGluR-Unterklassen (mGluR2–4, mGluR6–8) beschrieben.

Hier wurde ein Ansatz für das virtuelle Screening mit überwachten und nichtüberwachten neuronalen Netzen in Kombination mit einem Alignment-unabhängigen 3D-Pharmakophor-Deskriptor vorgestellt und evaluiert. Ein zweistufiger Filter wurde eingesetzt: zunächst durch überwachte neuronale Netze, dann mit einer SOM-basierten Diversitätsanalyse. Es konnten andere Grundgerüste als bei anderen Studien (Ähnlichkeitssuche, SOM)^[7,16b] identifiziert werden. Unterschiedliche Methoden (auch unterschiedliche KNN-

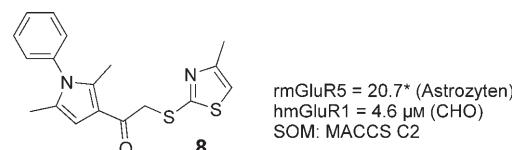
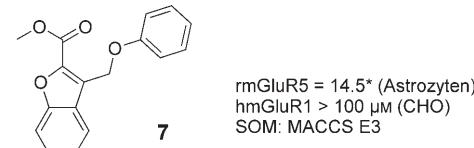
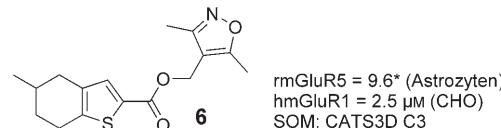


Abbildung 2. Aktive Moleküle aus dem virtuellen Screening mit Aktivitätsdaten für mGluR5 und mGluR1 sowie SOM-Cluster-Nummer (siehe Abbildung 1). Sterne (*) bezeichnen auf der Grundlage von Einpunktbestimmungen geschätzte IC_{50} -Werte.

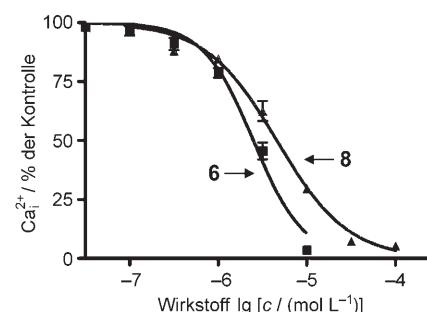


Abbildung 3. Funktionelle Antwort des hmGluR1-Rezeptors für die Moleküle **6** und **8**. Mit einem FLIPR-Assay wird der intrazelluläre Ca^{2+} -Spiegel ermittelt.

Klassen) führen somit zu abweichenden Ergebnissen. Die vorgestellte Strategie ist geeignet für die frühe Phase der Wirkstoff- und Leitstruktursuche, in der Kandidaten aus einer großen Menge an Substanzen herausgefiltert werden müssen. Eine weitere vielversprechende Anwendung könnte im systematischen Auffüllen von Lücken in Molekülbibliotheken für das biologische Screening liegen.

Eingegangen am 7. Oktober 2006,
veränderte Fassung am 15. Februar 2007

Stichwörter: Funktionsassay · G-gekoppelte Rezeptoren · Selektivität · Virtuelles Screening · Wirkstoff-Design

- [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, *Chem.-MedChem* **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, Deutschland, **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. Tammang, 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-Datenbank: Enamine Ltd., Kiew, Ukraine.
- [20] MOE v2004.03, Chemical Computing Group Inc., Montreal, Kanada, **2004**.