doi:10.2533/chimia.2012.120 Chimia 66 (2012) 120–124

From Theory to Bench Experiment by Computer-assisted Drug Design

Gisbert Schneider*

Abstract: Tight integration of computer-assisted molecular design with practical realization by medicinal chemistry will be essential for finding next-generation drugs that are optimized for multiple pharmaceutically relevant properties. ETH Zürich has established an interdisciplinary research group devoted to exploring the potential of this scientific approach by combining expertise from pharmaceutical chemistry and computer sciences. In this article, some of the group's activities and projects are presented. A current focus is on machine-learning applications aiming at hit and lead structure identification by virtual screening and *de novo* design. The central concept of 'adaptive fitness landscapes' is highlighted along with practical examples from drug discovery projects.

Keywords: Drug discovery · Machine-learning · Medicinal chemistry · Protein structure · Virtual screening

Gisbert Schneider studied biochemistry and computer science at the Free University of Berlin, Germany, where he received his doctoral degree in 1994. After several post-doctoral research activities he joined F. Hoffmann-La Roche Pharmaceuticals in Basel, Switzerland, where he headed the cheminformatics group. During this time in industry he received his habilitation and *venia legendi* in biochemistry and bioinformatics from the University of Freiburg, Germany. In 2002 he became a full professor of chem- and bioinformatics (Beilstein Endowed Chair) at Goethe-University Frankfurt, Germany, where he now is a distinguished adjunct professor. In 2010 he joined ETH Zürich as a full professor of computer-assisted drug design.

Online information about the Schneider group at ETH can be found at URL: *http://www.modlab.ethz.ch*

*Correspondence: Prof. Dr. G. Schneider
Eidgenössische Technische Hochschule
Department of Chemistry and Applied Biosciences
Institute of Pharmaceutical Sciences
CH-8093 Zürich
Tel.: +41 44 633 74 38
E-mail: gisbert.schneider@pharma.ethz.ch

A Role for Computational Medicinal Chemistry

Pharmaceutical drug discovery has been fuelled to a large extent by high-throughput screening (HTS),[1] fragment-based approaches,[2] and serendipitous findings.[3] The field prospers with each technological breakthrough in synthetic organic chemistry, *e.g.* combinatorial 'click-type' chemistry and ring-closing metathesis reactions, miniaturized flow systems,[4,5] as well as biochemical and biophysical activity determination,[6] *e.g.* by innovative whole-cell assays[7] and advanced spectroscopic methods,[8] to just name some prominent examples. Despite such outstanding technological advances the productivity of pharmaceutical industry is currently being perceived as stalled, with only small numbers of new drugs being approved by the authorities.[9] It has become apparent that the traditional model of drug discovery and development, in which primary screening of ever-increasing numbers of compounds generates many failing candidate drugs, urgently needs revision.[10] Progress may be possible by adequately considering the multi-dimensional nature of drug discovery,[11] including compound toxicity and aqueous solubility,[12,13] and extending the chemical diversity that is currently addressed and exploited by HTS. Here, consequent exploration of computer-assisted drug design can play a formative role (Fig. 1).[14,15]

Already now computational approaches in medicinal chemistry, particularly in molecular design, serve as an additional entry point to achieving sustained success in lead discovery.[16] Modern HTS is complemented by structure- (receptor-) and ligand-based virtual screening, *i.e.* computationally sieving through large virtual libraries of druglike small molecules and predicting bioactivity profiles for prioritized sets of screening compounds.[17] Focused compound libraries with signifi-

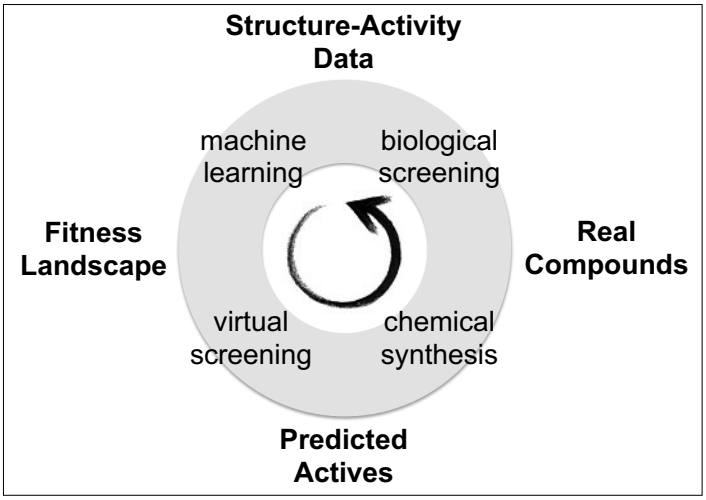

Fig. 1. A molecular design cycle. Adaptive molecular design includes computer-assisted model building, virtual screening and *de novo* design, chemical synthesis and testing of candidate compounds. It may be entered at any stage depending on available project knowledge.

cant hit rates in biochemical assays may be obtained through *in silico* property profiling and chemotype selection.[18] Similarly, *de novo* design attempts to generate novel ligands by virtually assembling molecular building-blocks, guided by multi-objective 'fitness functions' that help navigating in chemical space towards regions of desired bioactivity.[19] In fact, the number of organic molecules that could theoretically be synthesized is estimated to exceed 10^{60} – a mind-bogglingly large number.[20] Compared to the typical number of compounds tested in an HTS campaign – approximately 1–2 million – the sheer size of chemical space clearly prohibits any *global*, exhaustive exploration. What seems plausible is to pursue *local* optimization tactics by computer-assisted compound profiling. A working approach to address this challenge is to perform extensive *negative design*, that is, prior to selecting desirable candidates (*positive design*) one eliminates molecules that have predicted adverse properties and features. This idea is extended by limiting the search space to synthetically feasible compounds that could be assembled from few building blocks with established chemical reactions. Of note, small to medium-sized fragment collections have been proven to yield numerous valid starting points for hit-to-lead optimization in various drug discovery projects.[21] Recently, software tools have been developed for the purpose of fully automated fragment growing and linking using virtual reaction schemes, and several pioneering practical applications have been published.[22]

Computer-assisted Drug Design at ETH Zürich

These developments corroborate *trans*-disciplinary research at the interface between theory and laboratory experiment as appropriate and essential for finding inventive solutions to pressing issues in medicinal chemistry. Consequently, research activities of the Computer-Assisted Drug Design group in the Department of Chemistry and Applied Biosciences at ETH Zürich concentrate on the development and implementation of innovative concepts, algorithms and software for rapid identification of bioactive tool compounds and pharmaceutical lead structures. At the heart of these studies lies the machine-driven *de novo* design and virtual screening of both individual candidate molecules and small focused compound libraries that exhibit a desired pharmacological activity profile.[23] Our research includes drug re-purposing, *in silico* polypharmacology and chemogenomics projects, analysis of protein structure and modulation of protein–protein interaction, as well as the de-orphanization of drugs and their macromolecular receptors. The group runs own synthesis and testing facilities and a service point for virtual screening (SerViS). A current focus is on the design of innovative immunomodulatory agents and anti-infective lead structure candidates including host-defense peptides and natural-product mimicking compounds.[24] In tight cooperation with leading groups from academia and pharmaceutical industry, molecular design concepts are evaluated and applied to drug discovery projects. As the molecular design cycle involves multiple scientific disciplines and requires rigorous inter-disciplinary thinking, the ETH team consists of students and researchers with different scientific skills and background. First-rate equipment is available to support computer scientists, bio/cheminformaticians, pharmaceutical chemists, biochemists, and engineers alike. We strive for a stimulating and idea-provoking research environment to enable and facilitate computational medicinal chemistry and break down potential barriers between the scientific disciplines involved.

Model-building by Adaptive Learning

Iterative synthesize-and-test cycles are key to optimization of compound properties.[25] We recently demonstrated that there are optimal combinations of the size of a screening library and the number of iterative screening rounds with the aim to keep experimental efforts minimal.[26] Accordingly, machine-learning methods may guide an evolutionary design process that constantly adapts to a dynamic structure-activity relationship model (active learning concept). New compounds are first generated virtually from available molecular building blocks, but then synthesized in analytical or semi-preparative amounts, and finally tested for target binding *in vitro*.

The central idea of our drug design concept is an adaptive fitness landscape as a mathematical model of the underlying structure–activity relationship (SAR) for a given drug target or design objective (Fig. 2). Such a model structures parts of chemical space (that is, all compounds that can be synthesized with a given set of chemical reactions and molecular building blocks) into regions of high ('activity islands') and low ('tabu zones') predicted bioactivity. At the beginning of a drug discovery project, in absence of many known active and inactive compounds, the model landscape is largely unbiased (Fig. 2a). With increasing numbers of active and inactive compounds being discovered, machine-learning algorithms incorporate this gained knowledge in the adapting landscape (Fig. 2b, 2c), and thereby guide the next round of compound synthesis and testing. Visualization of fitness landscapes can help in compound prioritization and optimization.[28] Several such methods have been developed and implemented by our group. The latest software tool is LiSARD (Ligand Structure-Activity Relationship Display), which was specifically designed for project applications having the medicinal chemist in mind.[27] Selected examples of computationally *de novo* designed compounds that were iteratively optimized are presented in Fig. 3.

From Models to Molecules by Virtual Screening

Computer-based ligand identification by machine-learning shall be presented

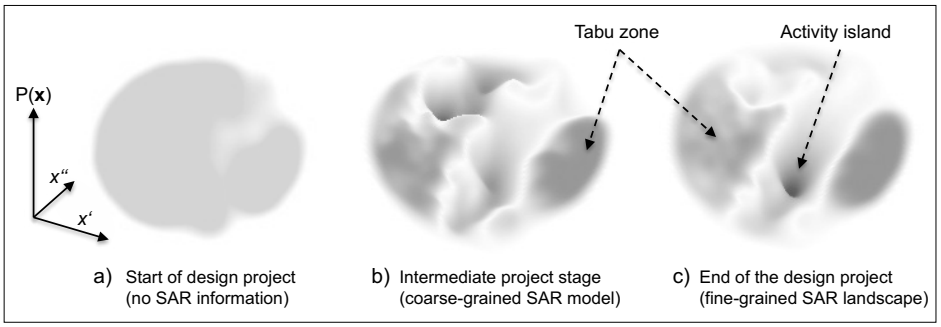

Fig. 2. Evolving fitness landscape. The figure presents the model of an adaptive structure–activity relationship (SAR) landscape at various stages during a molecular design project (a–c). With increasing numbers of tested compounds, the landscape becomes more detailed, thereby allowing for increasingly fine-grained molecular optimization. Both active and inactive compounds contribute to the model. In this way, the automated synthesis and testing of candidate compounds can be controlled by autonomous software. P(x) is a computed *pseudo*-probability function that structures the search space (part of chemical space) in regions with likely success or failure. Areas of chemical space that are associated with low predictive confidence appear transparent in this visualization. Several such SAR landscapes can be combined for multi-dimensional optimization of pharmaceutical lead compounds. Landscapes were generated using the LiSARD software tool.[27]

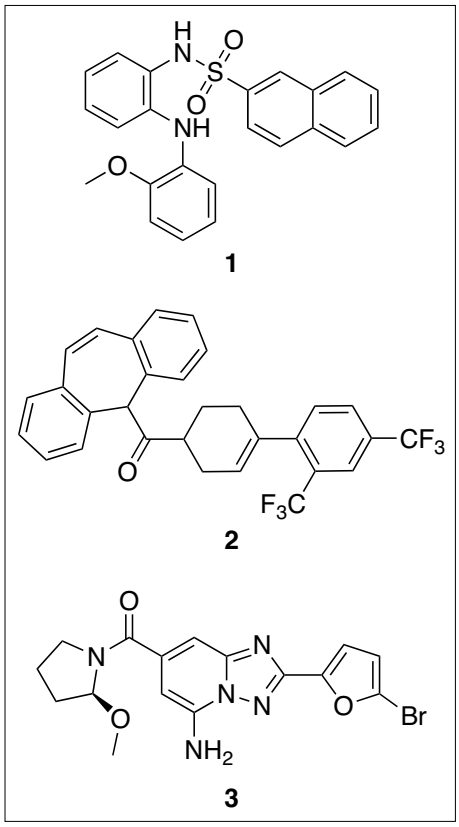

Fig. 3. Selected lead compounds that were computationally designed and adaptively optimized: Kv1.5 potassium channel blocker (**1**, IC_{50} = 0.47 μM),[22a] cannabinoid **1** receptor inverse agonist (**2**, K_i = 0.3 μM),[29] A_{2A} purinergic receptor antagonist (**3**, K_i = 2.4 nM).[30]

in some more detail for two selected examples: Gaussian process (GP) modeling for identification of i) activators of peroxisome proliferator-activated receptor gamma (PPAR-γ), a transcription factor,[31] and ii) prediction of protein-protein interface hot-spots for identification of a small interferon-α antagonist.

Identification of PPAR-γ Agonists by Machine Learning

To obtain a useful quantitative (Q) SAR function $f(x)$ that predicts a molecular property of interest, *e.g.* pK_d, from a molecular representation x (numerical molecular descriptors) contradictory aspects have to be considered: On the one hand, the QSAR function f must be sufficiently complex to accurately express the typically nonlinear relationship between molecular descriptor vectors $x_1, x_2, \ldots, x_n$ and the corresponding experimental measurements $y_1, y_2, \ldots, y_n$. On the other hand, f should not be too complex so that 'over-fitting' is avoided (*i.e.* memorizing the training data at the cost of poor predictive accuracy on new compounds). This tradeoff can be captured mathematically by minimizing the *regularized empirical loss function*[32]

$$\min R_{emp}^{reg}(f) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(f(x_i)-y_i\right)^2}_{\text{quality of fit}} + \underbrace{\lambda \cdot r(f)}_{\text{regularizer}},$$

where, $r(f)$ is a regularization function for f that penalizes the complexity of function f to reduce the risk of over-fitting. Parameter λ adjusts the influence of the regularization function r on QSAR model learning. Following this general concept, we trained a so-called *Gaussian Process* model to predict compound potency using a set of known PPAR ligands.[33] We represented all compounds by several molecular descriptor types – ranging from generic properties (clog*P*, overall charge) to topological and three-dimensional structural descriptors. The idea was to provide sufficiently diverse molecular representations so that the machine-learning model could extract functionally relevant features. The resulting QSAR function had a prediction error in the order of the error of measurement. Then, a large collection of screening compounds was analyzed by this function, and compounds were selected for biological screening that were predicted to strongly activate PPAR-γ binding (Fig. 4). Among several hits, compound **4** – a racemic natural product derivative – turned out to activate PPAR-γ (EC_{50} = 10±0.2 μM) without modulating PPAR-α activity. This PPAR subtype-selective chemotype is ready for hit-to-lead development with the primary aim to improve potency. Notably, for predicting PPAR activation, linear models turned out to be insufficient, but nonlinear SAR modeling resulted in several new bioactive compounds.

Identification of an Interferon-α Antagonist

Predicting protein–protein complex formation and targeting protein–protein interfaces by druglike compounds is an area of intensive research.[34] We have developed an alignment-free computational method predicting interface residues,[35] based on a so-called *knowledge-based scoring function*.[36] Prediction robustness was assessed on more than 1500 diverse proteins forming homo- and hetero-dimer complexes. Functional 'hot-spot' residues are frequent among the predicted interface residues, and, as a unique feature, the technique does not rely on sequence conservation. Blocking the protein–protein interaction between interferon-α and its receptor is directly linked to immune suppression. Therefore, modulating this interaction by small druglike molecules might represent a future therapeutic strategy for diseases associated with excessive production of interferon-α, *e.g.* lupus erythematosus and insulin-dependent diabetes mellitus.[37] As a step in this direction, we applied our interface prediction tool in combination with structure-based virtual screening to finding compounds that efficiently block the interferon–receptor interaction. Starting from an X-ray structure model of interferon-α, we used our software PocketPicker[38] to identify potential ligand-binding pockets on the protein surface. Then, all pockets were computationally investigated for potential hot-spot residues. The most promising candidate pocket served as a template for pharmacophore-based virtual screening of a large collection of commercially available compounds. Among the compounds tested, we found a potent, fragment-like inhibitor (**5**) of cellular interferon production (Fig. 5).[40] This first-in-class hit can now be used as a starting point for lead structure generation by medicinal chemistry, or undergo computational optimization by fragment-based *de novo* design.

Designing Bioactive Compounds *de novo*

Automated computer-based design of bioactive compounds has been an intensively researched area since more than three decades.[21] In 1994, we introduced adaptive machine-learning algorithms to computer-assisted compound generation

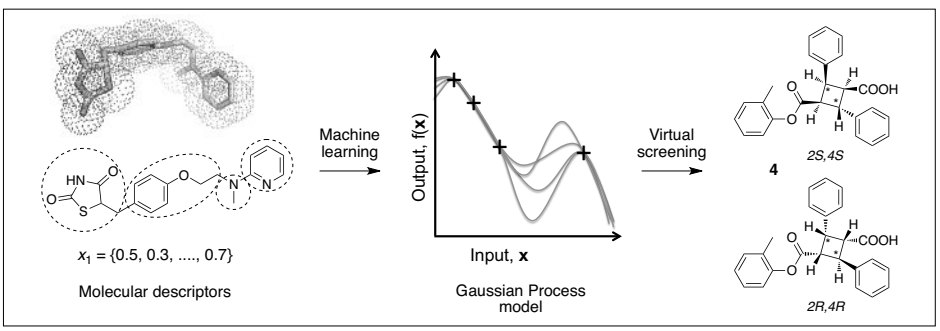

Fig. 4. Machine-learning models for virtual screening. New selective activators of transcription factor PPAR-γ (*e.g.* truxillic acid derivative **4**) were identified by virtual screening using a machine-learning SAR model that was trained on known PPAR-γ agonists represented by several types of molecular descriptors.

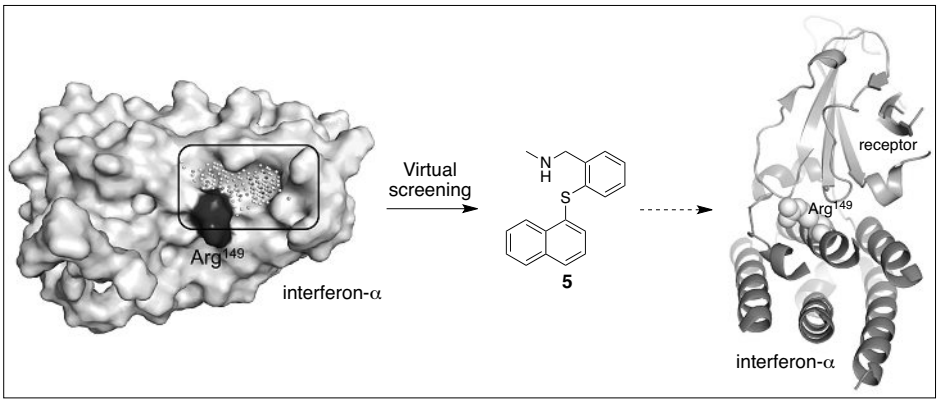

Fig. 5. Virtual screening for protein–protein interaction inhibitors. Computational scanning of the surface of interferon-α and prediction of 'hot-spot' residues involved in protein–protein interaction (here: Arg[149]) led to the identification of a preferred ligand-binding site. An interferon-derived 'fuzzy' pharmacophore model was used for compound screening, which resulted in the retrieval of compound **5**, an inhibitor of the interferon-α–receptor interaction. An X-ray structure of the interferon-α–receptor complex (PDB-ID: 3se3)[39] was published after completion of the virtual screening study.

and have improved these methods ever since.[41] The ligand-based *de novo* design software TOPAS (TOPology Assigning System) was the basis for the first fully automated evolutionary molecular design tool considering *pseudo*-reactions for virtual building block assembly.[42] Its youngest descendant, the software DOGS (Design Of Genuine Structures), employs validated organic reactions for this purpose, mimicking a medicinal chemistry laboratory.[43] With each designed compound the software suggests a straightforward synthesis route and readily available educts. The search algorithm has access to $7.8 \cdot 10^7$ virtual one-step products, approximately $2.6 \cdot 10^{15}$ two-step products, and $7.6 \cdot 10^{30}$ three-step products. In an exploratory study, we used a combination of complementary virtual screening tools for the analysis of compounds that were designed by DOGS with the aim to inhibit polo-like kinase 1 (Plk1), a target for the development of cancer therapeutics.[44] Emphasis was put on the generation of type II inhibitors arresting the inactive enzymatic state.[45] A comparative structural model of the inactive state of Plk1 was constructed, and the nucleotide binding pocket conformations in the DFG (Asp-Phe-Gly)-in (active) and DFG-out (inactive) state were compared using a computational approach to pocket similarity assessment. *De novo* designed compounds were analyzed using pharmacophore matching, fitness landscape analysis, and automated ligand docking. We then synthesized compound (**6**) following the synthetic route suggested by the software (Fig. 6). It turned out that compound **6** indeed arrests inactive Plk1 *in vitro*, and does not exhibit significant inhibition of activated Plk1 and a large panel of other kinases tested. This proof-of-principle study demonstrates that by smart coupling of virtual screening, fragment-based synthesis and activity testing, new bioactive agents with a desired activity profile may be obtained.

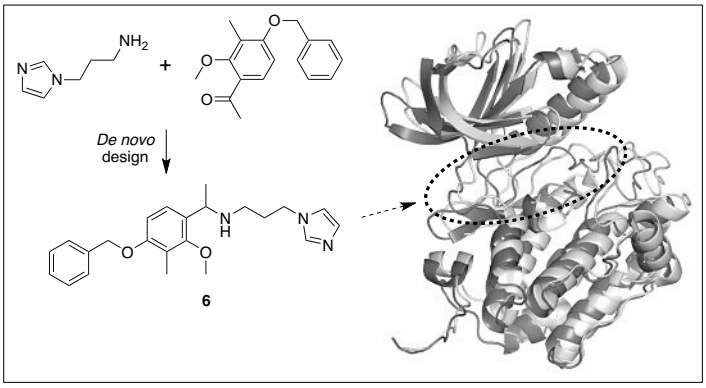

Fig. 6. *De novo* ligand design. Compound **6** potently and selectively inhibits human Polo-like kinase 1 (Plk1) in its inactive state (DFG-out activation-loop conformation). The inhibitor was designed *de novo* using the software DOGS which was developed at ETH Zürich. Compound **6** was synthesized as suggested by DOGS following a one-step protocol (reductive amination). The model on the right presents a superimposition of active Plk1 (PDB ID: 2ou7[46], light gray) with a comparative ('homology') model of inactive Plk1 (dark gray). The presumable ligand-binding site is highlighted. Cartoon structures were generated with MacPyMol [*www.pymol.org*].

Outlook

Similar to bioinformatics databases containing genomic sequence information, large searchable data banks of small molecules and their properties have become freely available – *e.g.* ChEMBL,[47] PubChem,[48] ChemBank,[49] ChEBI,[50] ChemDB[51] – and the known bioactive chemical space is continuously extended and refined. This huge body of chemical structures and literature data will unquestionably help navigating druglike chemical space.[52] These data also create a wealth of intriguing machine-learning challenges as well as opportunities to efficiently and accurately predict properties of small molecules and reactions for drug discovery.[53] Advanced drug design methods must be able to cope with unstructured information on druglike compounds and the vast combinatorial nature of chemical space. In hit-to-lead optimization lies a true challenge for computer-assisted drug design for the years to come.[14] Multiple target functions have to be considered in parallel, with only limited and often noisy reference data for model development available, and context-dependent suitability of molecular representations. In addition, challenging next-generation targets have started to come into focus for computer-based drug design, for example, protein–protein interaction interfaces,[54] RNA,[55] as well as transient and allosteric ligand-binding pockets.[56] Problem solving in such a setting might indeed be a domain of machine-learning, as several innovative algorithmic solutions indicate.[57] Despite all enthusiasm, one has to keep in mind that computational tools do not provide a cure-all recipe to problem solving in medicinal chemistry. Technological advances in many disciplines, with biology, chemistry and computer sciences playing leading roles, and their conceptual amalgamation are necessary for desired progress in drug discovery.[58]

Acknowledgements

The author is grateful to his coworkers and collaborating partners for realizing these studies. The work presented was financially supported by the ETH Zürich, Swiss National Science Foundation (SNF, grant 205321-134783), Deutsche Forschungsgemeinschaft (DFG, grant FOR1406_TP4), and OPO-Foundation Zürich.

Received: January 20, 2012

[1] a) L. M. Mayr, D. Bojanic, *Curr. Opin. Pharmacol.* **2009**, *9*, 580; b) R. Macarrón, R. P. Hertzberg, *Mol. Biotechnol.* **2011**, *47*, 270.

[2] a) C. W. Murray, D. C. Rees, *Nat. Chem.* **2009**, *1*, 187; b) E. Feyfant, J. B. Cross, K. Paris, D. H. Tsao, *Methods Mol. Biol.* **2011**, *685*, 241.

[3] T. A. Ban, *Dialogues Clin. Neurosci.* **2006**, *8*, 335.

[4] M. Colombo, I. Peretto, *Drug Discov. Today* **2008**, *13*, 677.

[5] a) S. Y. Wong-Hawkes, J. C. Matteo, B. H. Warrington, J. D. White, *Ernst Schering Found. Symp. Proc.* **2006**, *3*, 39; b) A. Chighine, G. Sechi, M. Bradley, *Drug Discov. Today* **2007**, *12*, 459.

[6] J. P. Renaud, M. A. Delsuc, *Curr. Opin. Pharmacol.* **2009**, *9*, 622.

[7] a) G. S. Elliott, *Comb. Chem. High Throughput Screen.* **2009**, *12*, 849; b) A. Tárnok, A. Pierzchalski, G. Valet, *Curr. Med. Chem.* **2010**, *17*, 1719.

[8] a) B. Liu, S. Li, J. Hu, *Am. J. Pharmacogenomics* **2004**, *4*, 263; b) M. Kobayashi, K. Retra, F. Figaroa, J. G. Hollander, E. Ab, R. J. Heetebrij, H. Irth, G. Siegal, *J. Biomol. Screen.* **2010**, *15*, 978.

[9] a) Thomson Reuters: International Year of Chemistry (IYC 2011) Report, pp. 1–18, **2011** (*http://ip-science.thomsonreuters.com/info/changing-role-of-chemistry/*); b) D. Weiss, P. Naik, R. Weiss, *Nat. Rev. Drug Discov.* **2009**, *8*, 533.

[10] J. R. Empfield, P. D. Leeson, *IDrugs* **2010**, *13*, 869.

[11] a) K. H. Bleicher, H.-J. Böhm, K. Müller, A. I. Alanine, *Nat. Rev. Drug Discov.* **2003**, *2*, 369; b) J. B. Campbell, *IDrugs* **2010**, *13*, 874; c) G. M. Maggiora, *J. Comput. Aided Mol. Des.* **2011**, *25*, 699.

[12] a) R. Kavlock, D. Dix, *J. Toxicol. Environ. Health B Crit. Rev.* **2010**, *13*, 197; b) A. M. Richard, C. Yang, R. S. Judson, *Toxicol. Mech. Methods* **2008**, *18*, 103.

[13] J. Wang, T. Hou, *Comb. Chem. High Throughput Screen.* **2011**, *14*, 328.

[14] G. Schneider, *Nat. Rev. Drug Discov.* **2010**, *9*, 273.

[15] a) T. I. Oprea, H. Matter, *Curr. Opin. Chem. Biol.* **2004**, *8*, 349; b) O. Roche, W. Guba, *Mini Rev. Med. Chem.* **2005**, *5*, 677.

[16] a) M. Glick, E. Jacoby, *Curr. Opin. Chem. Biol.* **2011**, *15*, 540; b) G. Schneider (Ed.) Special focus issue, parts I-III: 'Computational Medicinal Chemistry', *Future Med. Chem.* **2011**, *3*, issues 4, 6, 7.

[17] a) 'Virtual Screening for Bioactive Molecules', Eds. H.-J. Böhm, G. Schneider, Wiley-VCH, Weinheim, **2000**; b) 'Virtual Screening: Principles, Challenges, and Practical Guidelines', Ed. C. Sotriffer, Wiley-VCH, Weinheim, **2011**.

[18] a) R. E. Dolle, *Methods Mol. Biol.* **2011**, *685*, 3; b) J. Z. Zhou, *Methods Mol. Biol.* **2011**, *685*, 27.

[19] a) V. J. Gillet, *Methods Mol. Biol.* **2004**, *275*, 335; b) J. R. Fischer, U. Lessel, M. Rarey, *J. Chem. Inf. Model.* **2010**, *50*, 1; c) C. A. Nicolaou, C. C. Kannas, *Methods Mol. Biol.* **2011**, *685*, 53.

[20] a) R. S. Bohacek, C. McMartin, W. C. Guida, *Med. Res. Rev.* **1996**, *16*, 3; b) P. Kirkpatrick, C. Ellis (Eds.), *Nature* **2004**, *432*, 823.

[21] a) T. Hesterkamp, M. Whittaker, *Curr. Opin. Chem. Biol.* **2008**, *12*, 260; b) C. W. Murray, T. L. Blundell, *Curr. Opin. Struct. Biol.* **2010**, *20*, 497; c) O. Ichihara, J. Barker, R. J. Law, M. Whittaker, *Mol. Inf.* **2011**, *30*, 298.

[22] a) G. Schneider, O. Clément-Chomienne, L. Hilfiger, P. Schneider, S. Kirsch, H.-J. Böhm, W. Neidhart, *Angew. Chem. Int. Ed.* **2000**, *39*, 4130; b) H. M. Vinkers, M. R. de Jonge, F. F. Daeyaert, J. Heeres, L. M. Koymans, J. H. van Lenthe, P. J. Lewi, H. Timmerman, K. Van Aken, P. A. Janssen, *J. Med. Chem.* **2003**, *46*, 2765; c) P. S. Kutchukian, D. Lou, E. I. Shakhnovich, *J. Chem. Inf. Model.* **2009**, *49*, 1630; d) Q. Hu, Z. Peng, J. Kostrowicki, A. Kuki, *Methods Mol. Biol.* **2011**, *685*, 253; e) M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.-H. Altmann, G. Schneider, E. Jacoby, S. Renner, *J. Chem. Inf. Model.* **2011**, *51*, 3093.

[23] a) G. Schneider, *J. Comput. Aided Mol. Des.* **2012**, *26*, 115; b) G. Schneider, U. Fechner, *Nat. Rev. Drug Discov.* **2005**, *4*, 649.

[24] C. D. Fjell, J. A. Hiss, R. E. Hancock, G. Schneider, *Nat. Rev. Drug Discov.* **2012**, *11*, 37.

[25] a) G. Schneider, S.-S. So, 'Adaptive Systems in Drug Design', Landes Bioscience, Georgetown, **2001**; b) I. Muegge, *Mini Rev. Med. Chem.* **2008**, *8*, 927.

[26] G. Schneider, A. Schüller, *Methods Mol. Biol.* **2009**, *572*, 135.

[27] M. Reutlinger, W. Guba, R. E. Martin, A. I. Alanine, T. Hoffmann, A. Klenner, J. A. Hiss, P. Schneider, G. Schneider, *Angew. Chem. Int. Ed.* **2011**, *50*, 11633.

[28] a) M. Reutlinger, G. Schneider, *J. Mol. Graphics Modell.* **2012**, *34*, 108; b) J. Bajorath, L. Peltason, M. Wawer, R. Guha, M. S. Lajiness, J. H. Van Drie, *Drug Discov. Today* **2009**, *14*, 698; c) T. I. Oprea, J. Gottfries, *J. Comb. Chem.* **2001**, *3*, 157.

[29] a) M. Rogers-Evans, A. I. Alanine, K. H. Bleicher, D. Kube, G. Schneider, *QSAR Comb. Sci.* **2004**, *26*, 426; b) L. Alig, J. Alsenz, M. Andjelkovic, S. Bendels, A. Bénardeau, K. H. Bleicher, A. Bourson, P. David-Pierson, W. Guba, S. Hildbrand, D. Kube, T. Lübbers, A. V. Mayweg, R. Narquizian, W. Neidhart, M. Nettekoven, J. M. Plancher, C. Rocha, M. Rogers-Evans, S. Röver, G. Schneider, S. Taylor, P. Waldmeier, *J. Med. Chem.* **2008**, *51*, 2115.

[30] G. Schneider, M. Nettekoven, *J. Comb. Chem.* **2003**, *5*, 233.

[31] a) J. Berger, D. E. Moller, *Annu. Rev. Med.* **2002**, *53*, 409; b) S. Crunkhorn, *Nat. Rev. Drug Discov.* **2011**, *10*, 814.

[32] A. J. Smola, B. Schölkopf, *Statistics and Computing* **2004**, *14*, 199.

[33] M. Rupp, T. Schroeter, R. Steri, H. Zettl, E. Proschak, K. Hansen, O. Rau, O. Schwarz, L. Müller-Kuhrt, M. Schubert-Zsilavecz, K. R. Müller, G. Schneider, *ChemMedChem* **2010**, *5*, 191.

[34] a) S. H. Khan, F. Ahmad, N. Ahmad, D. C. Flynn, R. Kumar, *J. Biomol. Struct. Dyn.* **2011**, *28*, 929; b) M. N. Wass, A. David, M. J. E. Sternberg, *Curr. Opin. Struct. Biol.* **2011**, *21*, 382; c) N. Tuncbag, A. Gursoy, O. Keskin, *Phys. Biol.* **2011**, *8*, 035006.

[35] T. Geppert, B. Hoy, S. Wessler, G. Schneider, *Chem. Biol.* **2011**, *18*, 344.

[36] H. Gohlke, G. Klebe, *Curr. Opin. Struct. Biol.* **2001**, *11*, 231.

[37] A. N. Theofilopoulos, R. Baccala, B. Beutler, D. H. Kono, *Annu. Rev. Immunol.* **2005**, *23*, 307.

[38] M. Weisel, E. Proschak, G. Schneider, *Chem. Cent. J.* **2007**, *1*, 7.

[39] C. Thomas, I. Moraga, D. Levin, P. O. Krutzik, Y. Podoplelova, A. Trejo, C. Lee, G. Yarden, S. E. Vleck, J. S. Glenn, G. P. Nolan, J. Piehler, G. Schreiber, K. C. Garcia, *Cell* **2011**, *146*, 621.

[40] T. Geppert, S. Bauer, J. A. Hiss, E. Conrad, M. Reutlinger, P. Schneider, M. Weisel, B. Pfeiffer, K.-H. Altmann, Z. Waibler, G. Schneider, *Angew. Chem. Int. Ed.* **2012**, *51*, 258.

[41] a) G. Schneider, J. Schuchhardt, P. Wrede, *Comput. Appl. Biosci.* **1994**, *10*, 635; b) G. Schneider, J. Schuchhardt, P. Wrede, *Biol. Cybern.* **1995**, *73*, 245; c) G. Schneider, W. Schrödl, G. Wallukat, J. Müller, E. Nissen, W. Rönspeck, P. Wrede, R. Kunze, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12179.

[42] G. Schneider, M. L. Lee, M. Stahl, P. Schneider, *J. Comput. Aided Mol. Des.* **2000**, *14*, 487.

[43] M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, S. Weggen, H. Stark, G. Schneider, *PLoS Comput. Biol.* **2012**, e1002380.

[44] G. Schneider, T. Geppert, M. Hartenfeller, F. Reisen, A. Klenner, M. Reutlinger, V. Hähnke, J. A. Hiss, H. Zettl, S. Keppner, B. Spänkuch, P. Schneider, *Future Med. Chem.* **2011**, *3*, 415.

[45] Y. Liu, N. S. Gray, *Nat. Chem. Biol.* **2006**, *2*, 358.

[46] a) M. Kothe, D. Kohls, S. Low, R. Coli, A. C. Cheng, S. L. Jacques, T. L. Johnson, C. Lewis, C. Loh, J. Nonomiya, A. L. Sheils, K. A. Verdries, T. A. Wynn, C. Kuhn, Y. H. Ding, *Biochemistry* **2007**, *46*, 5960; b) H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235.

[47] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res.* **2012**, *40*, D1100.

[48] a) Q. Li, T. Cheng, Y. Wang, S. H. Bryant, *Drug Discov. Today* **2010**, *15*, 1052; b) E. E. Bolton, J. Chen, S. Kim, L. Han, S. He, W. Shi, V. Simonyan, Y. Sun, P. A. Thiessen, J. Wang, B. Yu, J. Zhang, S. H. Bryant, *J. Cheminform.* **2011**, *3*, 32.

[49] K. P. Seiler, G. A. George, M. P. Happ, N. E. Bodycombe, H. A. Carrinski, S. Norton, S. Brudz, J. P. Sullivan, J. Muhlich, M. Serrano, P. Ferraiolo, N. J. Tolliday, S. L. Schreiber, P. A. Clemons, *Nucleic Acids Res.* **2008**, *36*, D351.

[50] P. de Matos, R. Alcántara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner, C. Steinbeck, *Nucleic Acids Res.* **2010**, *38*, D249.

[51] J. Chen, S. J. Swamidass, Y. Dou, J. Bruand, P. Baldi, *Bioinformatics* **2005**, *21*, 4133.

[52] a) R. Gozalbes, A. Pineda-Lucena, *Comb. Chem. High Throughput Screen.* **2011**, *14*, 548; b) L. J. Bellis, R. Akhtar, B. Al-Lazikani, F. Atkinson, A. P. Bento, J. Chambers, M. Davies, A. Gaulton, A. Hersey, K. Ikeda, F. A. Krüger, Y. Light, S. McGlinchey, R. Santos, B. Stauch, J. P. Overington, *Biochem. Soc. Trans.* **2011**, *39*, 1365.

[53] a) T. Scior, P. Bernard, J. L. Medina-Franco, G. M. Maggiora, *Mini Rev. Med. Chem.* **2007**, *7*, 851; b) Q. Li, T. Cheng, Y. Wang, S. H. Bryant, *Drug Discov. Today* **2010**, *15*, 1052; c) M. P. Gleeson, A. Hersey, D. Montanari, J. Overington, *Nat. Rev. Drug Discov.* **2011**, *10*, 197.

[54] C. Reynès, H. Host, A. C. Camproux, G. Laconde, F. Leroux, A. Mazars, B. Deprez, R. Fahraeus, B. O. Villoutreix, O. Sperandio, *PLoS Comput. Biol.* **2010**, *6*, e1000695.

[55] a) T. Tuccinardi, *Future Med. Chem.* **2011**, *3*, 723; b) A. Schüller, M. Suhartono, U. Fechner, Y. Tanrikulu, S. Breitung, U. Scheffer, M. W. Göbel, G. Schneider, *J. Comput. Aided Mol. Des.* **2008**, *22*, 59; c) P. Daldrop, F. E. Reyes, D. A. Robinson, C. M. Hammond, D. M. Lilley, R. T. Batey, R. Brenk, *Chem. Biol.* **2011**, *18*, 324; d) A. C. Stelzer, A. T. Frank, J. D. Kratz, M. D. Swanson, M. J. Gonzalez-Hernandez, J. Lee, I. Andricioaei, D. M. Markovitz, H. M. Al-Hashimi, *Nat. Chem. Biol.* **2011**, *7*, 553.

[56] a) Y. Tanrikulu, E. Proschak, T. Werner, T. Geppert, N. Todoroff, A. Klenner, T. Kottke, K. Sander, E. Schneider, R. Seifert, H. Stark, T. Clark, G. Schneider, *ChemMedChem* **2009**, *4*, 820; b) B. Stauch, H. Hofmann, M. Perkovic, W. Weisel, F. Kopietz, K. Cichutek, C. Münk, G. Schneider, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12079; c) S. Eyrisch, V. Helms, *J. Med. Chem.* **2007**, *50*, 3457.

[57] P. Baldi, K. R. Müller, G. Schneider (Eds.), Special issue: 'Charting Chemical Space: Challenges and Opportunities for Artificial Intelligence and Machine Learning', *Mol. Inf.* **2011**, *30*, issue 9.

[58] a) S. Davey, *Nature* **2009**, *458*, 294; b) C. M. Crews, *Chem. Biol.* **2010**, *17*, 551.