Neural networks and genetic algorithms in drug design

Lothar Terfloth and Johann Gasteiger

Neural networks and genetic algorithms are versatile methods for a variety of tasks in rational drug design, including analysis of structure–activity data, establishment of quantitative structure–activity relationships (QSAR), gene prediction, locating protein-coding regions in DNA sequences, 3D structure alignment, pharmacophore perception, docking of ligands to receptors, automated generation of small organic compounds, and the design of combinatorial libraries. Here, we give a brief overview of these applications of neural networks and genetic algorithms in drug design, and provide an insight into the underlying principles of such methods.

Lothar Terfloth and Johann Gasteiger*
Computer-Chemie-Centrum
und Institut für
Organische Chemie
Universität Erlangen-Nürnberg
Nägelsbachstraße 25
D-91052 Erlangen
Germany
*tel: +49 9131 852 6570
fax: +49 9131 852 6566
e-mail: gasteiger@chemie.uni-erlangen.de

▼ The development of a new drug is still a challenging, time-consuming and cost-intensive process. Computational methods can be used to assist and speed up the drug discovery process. This article focuses on the use of neural networks and genetic algorithms in drug design. In contrast to classical statistical methods such as regression analysis or partial least squares analysis (PLS), artificial neural networks enable the investigation of complex, nonlinear relationships. Neural networks are therefore ideally suited for use in drug design and QSAR.

Following an introduction to neural networks, we discuss several examples of their application. We then go on to give a general outline of genetic algorithms. Finally, we report the applications of genetic algorithms to structure alignment, variable selection in QSAR studies, structure generation, design of combinatorial libraries, and docking of ligands to proteins.

Neural network methods

Mathematical models and algorithms, designed to mimic the information processing and knowledge acquisition of the human brain, are called neural networks. They consist of many basic units, called artificial neurons (or simply neurons), which perform identical tasks. A neuron collects a series of input signals and transforms them into the output signal via a transfer function.

In the course of training, such a network of neurons 'learns' by changing the weights of its neurons. Two different learning methods can be distinguished: supervised and unsupervised learning. When learning is unsupervised, the neural network is provided with the input patterns. After some iterations, it should settle to a stable state. All unsupervised learning systems have a general optimization criterion, such as the minimization of energy or distance, maximization of profit, and so on, which is used for the evaluation of the result at the end of each cycle.

Considering that responses in a given system are known for some objects (data-points), the goal of supervised learning methods is to find a model that correctly associates the inputs (representation of the objects) with the targets (representation of the responses). The targets serve not only as a criterion for how well the system has been trained, but also influence the correction of each weight.

Neural networks can be applied to four basic types of applications:

- association;
- classification (clustering);
- transformation (different representation);
- modeling.

A variety of artificial neural network methods have been developed. Here, we single out some of the more prominent methods. For a list of online resources for neural networks, see Box 1.

Back-propagation networks

The term back-propagation (BPG) network refers to a multi-layer feed-forward (MLFF) network[1,2], trained with back-propagation learning, which is

PII: S1471-1931(01)00173-2

Box 1. Online resources for neural networks and genetic algorithms

Tool	**URL**
Artificial neural networks	
The Battelle Pacific Northwest National Laboratory (Richland, WA, USA) (descriptions of commercial neural network programs and shareware packages)	*http://www.emsl.pnl.gov:2080/proj/neuron/neural/systems*
Websites containing links to neural network resources	*http://www.ccs.neu.edu/groups/honors-program/freshsem/19951996/cloder/ myNNlinks.html* *http://www-sci.sci.kun.nl/cac/www-neural-links.html*
The USENET newsgroup (for discussion of neural networks)	*http://groups.google.com/groups?q=comp.ai.neural-nets*
Stuttgart Neural Network Simulator	*http://www-ra.informatik.uni-tuebingen.de/SNNS*
Online material relating to the book *Neural Networks in Chemistry and Drug Design*[a]	*http://www2.chemie.uni-erlangen.de/publications/ANN-book/index.html*
Genetic algorithms	
The Genetic Algorithms Archive	*http://www.aic.nrl.navy.mil/galist*
The ILLiGAL Home page	*http://gal4.ge.uiuc.edu/index.html*
The PGAPack Parallel Genetic Algorithm Library	*ftp://ftp.mcs.anl.gov/pub/pgapack*
GAlib (a C++ library of genetic algorithm components)	*http://lancet.mit.edu/ga*
Evolving Objects (a C++ library for evolutionary computation)	*http://sourceforge.net/projects/eodev*
Newsgroup discussions related to genetic algorithm research	*http://groups.google.com/groups?q=comp.ai.genetic*

Reference

[a] Zupan, J. and Gasteiger, J. (1999) *Neural Networks in Chemistry and Drug Design* (2nd Edn), Wiley

a supervised learning method (Fig. 1). BPG networks have frequently been applied in QSAR studies because of their modeling power. However, BPG networks are not without their problems, which include the issues of proper network design, overfitting and overtraining.

The architecture or design of the network is given by:

- the number of inputs and outputs;
- the number of layers;
- the number of neurons in each layer;
- the number of weights in each neuron;
- the way in which the weights are linked together, within or between the layer(s);
- which neurons receive the correction signals.

For further information concerning back-propagation networks and a more detailed overview on neural networks and their use in drug design, see Refs 3–5.

Kohonen neural networks

Kohonen introduced an artificial neural network that he called the self-organizing network[6,7]. Learning in a Kohonen network is unsupervised, that is, the property to be investigated is not used in the training process. In essence, a Kohonen network projects points from a multi-dimensional space into a space of lower-dimensionality, usually into a 2D plane (Fig. 2). In this projection, the topology of the high-dimensional space is conserved in the lower-dimensional space (i.e. objects that are close in the high-dimensional space are also close in the lower-dimensional space). Thus, in principle, Kohonen networks can be used for similarity perception or for the clustering of objects.

A package for the self-organizing map algorithm (SOM) is publicly available from the Kohonen group (http://www.cis.hut.fi/research/som_pack/ and http://www.cis.hut.fi/research/som-research/nnrc-programs.shtml). We have implemented the Kohonen algorithm in the KMAP (Kohonen Mapping) program (Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, Germany; http://www2.chemie.uni-erlangen.de/software/kmap/index.html; and Molecular Networks GmbH, Computerchemie, Erlangen, Germany; http://www.mol-net.de/products/kmap/index.html), primarily for chemical applications.

Counter-propagation networks

The Kohonen learning algorithm can also be utilized for supervised learning. For this, objects consist of pairs of information: the input vector for characterizing the objects, and an output vector for representing the property of the object to be studied (Fig. 3). The output property can be a single value or a vector of properties. Such supervised networks using Kohonen's training algorithm are called counterpropagation (CPG) networks[3,8–10]. In the training phase, only the input layer is taken into account for the distance calculation, whereas during the adaption steps, the weights of both the input and the output layer are modified. The trained network can be used to predict unknown property vectors.

Use of neural networks in drug design

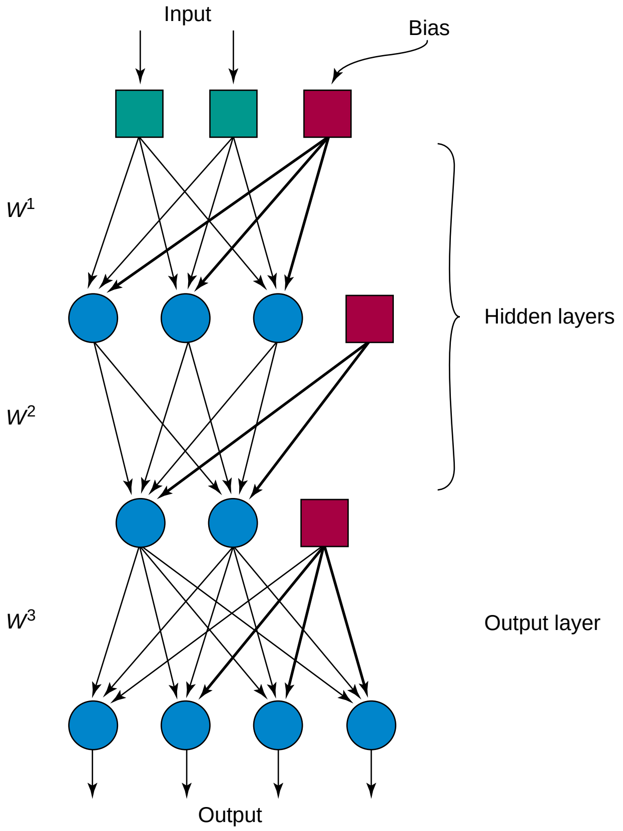
Neural networks can be employed for the following applications in drug design:

- analysis of multi-dimensional data;
- classification and prediction of biological activity and ADME-Tox (absorption, distribution, metabolism, excretion, and toxicity) properties;
- lead discovery;
- comparison of compound libraries;
- analysis of the similarity and diversity of combinatorial libraries;
- analysis of data from HTS.

Anzali *et al.*[11] present a variety of application fields for self-organizing neural networks in drug design, including the mapping of a molecular surface into a 2D plane. The comparison of Kohonen maps for a set of ligands shows common characteristics for ligands binding to the same receptor. Modeling of the affinity of steroids binding to the corticosteroid-binding globulin (CBG) receptor, and the comparison of libraries is also reported.

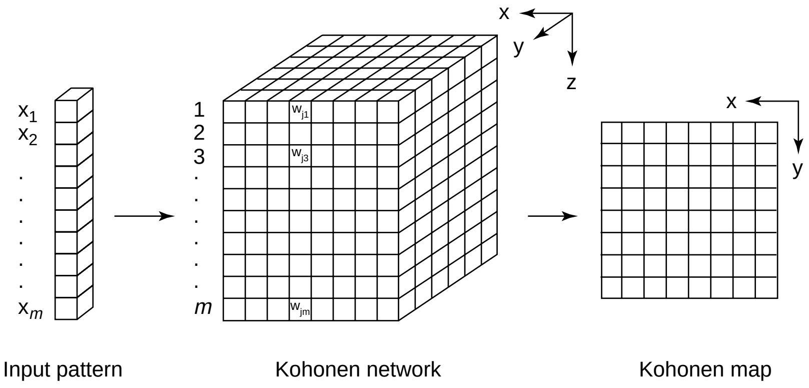
A method called comparative molecular surface analysis (COMSA), developed by Polanski and Gasteiger, and further elaborated by Polanski and Walczak[12], combines the mapping of the mean electrostatic potential on the molecular surface (by a Kohonen self-organizing network) with a PLS analysis to establish a QSAR model.

Typical investigations in drug design that use neural networks include: prediction of the aqueous solubility of drugs[13,14]; prediction of hepatic drug clearance[15]; the QSAR study of HIV-1 reverse transcriptase[16]; the application of Bayesian regularized neural networks to the development of QSAR models[17,18]; and the classification and modeling of chemotherapeutic agents, anti-bacterials, anti-neoplastics and anti-fungals[19].



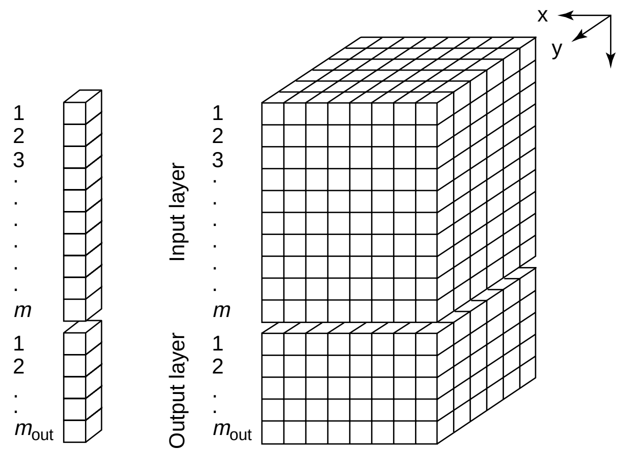
Drug Discovery Today: genomics supplement

Figure 1. Architecture of a back-propagation network with one input layer (green boxes) and three active layers (two hidden layers and the output layer; blue circles). The layers are fully connected, and connections to the biases (red squares) are indicated by heavier lines. The biases influence the adaption of the weights (*W*; represented by the arrows between the layers) during the training phase. Depending on the application for which the neural network is set up, both the number of layers and the number of neurons in each layer has to be determined.



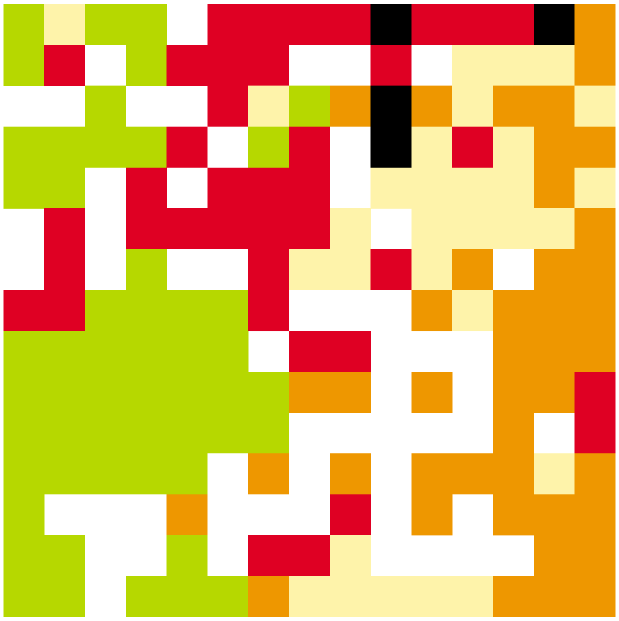
Drug Discovery Today: genomics supplement

Figure 2. Architecture of a Kohonen neural network. Here, a network with 8 × 6 neurons is depicted [the first subscripted index (j) refers to the number of the neuron]. During training, the *m*-dimensional input patterns [$X = (x_1, x_2, x_3, \ldots x_m)$] iteratively enter the network and the responses of all neurons are calculated. The winning neuron (c), which is most similar to the current input pattern, is determined. In order to improve the response to the same input in the next epoch, the weights of neuron c [$W_c = (w_{c1}, w_{c2}, w_{c3}, \ldots w_{cm})$] and of all neurons in the neighborhood of neuron c are corrected. After training, the response of the network is calculated for each input pattern of the dataset. The projection of the dataset into the 2D space is then performed by mapping each input pattern into the coordinates of its winning neuron.



Drug Discovery Today: genomics supplement

Figure 3. Architecture of a counter-propagation network. In addition to a Kohonen neural network, counter-propagation networks also have an output layer. To determine the winning neuron, only the input layer is taken into account. In the adaption step, the weights in the input and ouput layer are corrected. Unknown property vectors can be predicted with the trained network.



Drug Discovery Today: genomics supplement

Figure 4. Classification of a dataset into four different classes with a Kohonen neural network having a rectangular topology. Neurons are colored as follows: red, 5-HT_{1a}-receptor agonists; orange, H_2-receptor antagonists; yellow, MAO_A inhibitors; green, thrombin inhibitors; black, conflict neurons; and white, unoccupied neurons.

Another application of neural networks in the field of genomics is gene prediction, as reported by Cai and Bork[20]. They perform homology searches in protein and expressed sequence tag databases with neural networks in order to determine start codons, stop codons, and splice sites. Uberbacher and Mural describe a method to locate protein-coding regions in DNA sequences by means of a back-propagation network[21]. Neural networks can also be used to predict protein folding[22,23] and to analyse gene expression data[24].

Schneider[25], as well as Manallack and Livingstone[26], have reached the conclusion that neural networks are valuable tools for drug design and have the potential for further development.

In a study of our own, we investigated the classification of a dataset containing 299 compounds, namely 75 5-hydroxytryptamine 5-HT_{1a}-receptor agonists, 75 histamine H_2-receptor antagonists, 74 monoamine oxidase MAO_A inhibitors, and 75 thrombin inhibitors, using the KMAP program (Fig. 4).

Definition of a genetic algorithm

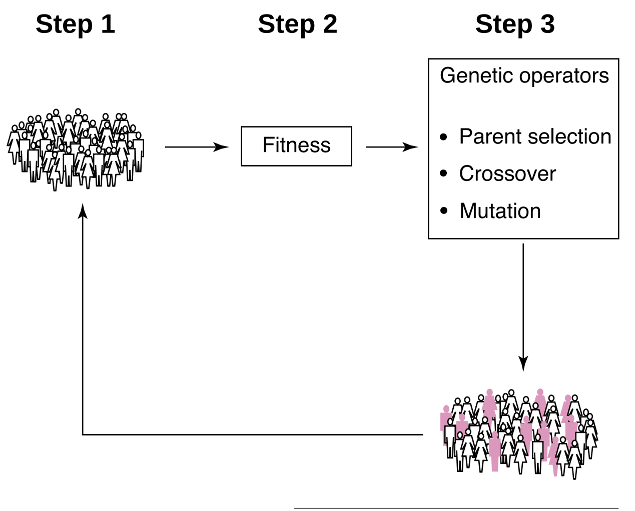
Genetic algorithms (GAs) were introduced by Holland, and mimic nature's evolutionary method of adapting to a changing enviroment[27]. They are stochastic optimization methods and provide a powerful means to perform directed random searches in a large problem space as encountered in chemometrics and drug design. Each individual in a population is represented by a chromosome. After initialization of the first generation (step 1), the fitness of each individual is evaluated by an objective function (step 2). In the reproduction step (step 3), the genetic operators of parent selection, crossover and mutation are applied, thereby providing the first offspring generation. Iteration of steps 2 and 3 is performed until the objective function converges (Fig. 5).

An introduction and overview of the applications of GAs are reported by Venkatasubramanian and Sundaram[28] as well as by Jones[29]. The use of GAs in drug design is reviewed by Judson[30] and Devillers[31]. For a list of the online resources for genetics algorithms, see Box 1.

Applications of GAs in drug design

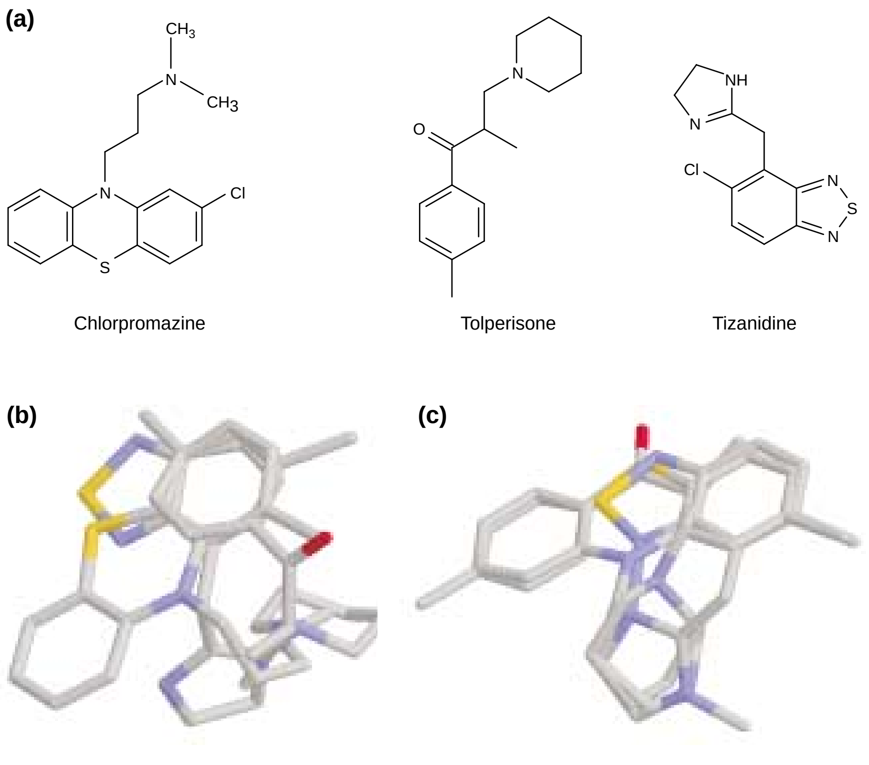
A wide range of studies in QSAR gain advantage from the use of GAs. Hofman *et al.*[32] as well as Turner and Willett[33] improved the predictive value of a QSAR model by variable selection using a GA. Fitness is evaluated by a PLS (partial least squares) cross validation. GA-based variable selection has also proved to be advantageous in comparative molecular field analysis as reported by Kimura *et al.*[34] and Hasegawa *et al.*[35]. The number of field variables was reduced from 1275 to 43, accompanied by an increased predictivity of the model[34].

Pharmacophore perception for receptors with an unknown 3D structure can be carried out by comparing the spatial and electronic requirements of a set of ligands that are known to



Drug Discovery Today: genomics supplement

Figure 5. Schematic view of a genetic algorithm. In the first step, the first population is initialized. The fitness of each individual is then evaluated (step 2). Selection of individuals based on their fitness and application of the genetic operators, crossover and mutation, provides the first offspring generation (step 3). The steps are repeated iteratively.



Drug Discovery Today: genomics supplement

Figure 6. Rigid and flexible multiple alignment of three muscle relaxants (a) mapping 11 atom triples. The RMS (root mean square) value is 2.41 Å for the rigid alignment (b) and 0.81 Å for the flexible alignment (c).

bind to the receptor of interest. Such a comparison is performed by the structural alignment of these ligands. A detailed review on computational methods for the structural alignment of molecules is given by Lemmen and Lengauer[36]. A program called GAMMA (genetic algorithm for multiple molecule alignment; Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, Germany; http://www2.chemie.uni-erlangen.de/research/drugdesign/ga.phtml) enables the flexible alignment of multiple molecules by combining a GA with a Newton optimizer (Fig. 6)[37,38].

Docking of flexible ligands to macromolecules is paramount in structure-based drug design. Some programs that work with a GA also enable automated docking. Such programs include GOLD (genetic optimization for ligand docking; Cambridge CDC, Cambridge, UK; http://www.ccdc.cam.ac.uk/prods/gold)[39], FCEA (family competition evolutionary approach; Bioinformatics Laboratory, Department of Computer Science and Information Engineering, National Taiwan University, Taiwan; http://bioinfo.csie.ntu.edu.tw/english/Research/ea.htm)[40], and AutoDock (automated docking; Molecular Graphics Laboratory, The Scripps Research Institute, La Jolla, CA, USA; http://www.scripps.edu/pub/olson-web/doc/autodock)[41].

Another application of GAs is the automated generation of small organic molecules[42,43]. Douget *et al.*[42] modify a SMILES (simplified molecular input line entry system; Daylight Chemical Information Systems, Mission Viejo, CA, USA; http://www.daylight.com/smiles) string using the lipophilicity, electronic properties and shape-related properties for calculation of the scoring function. Alternatively, Schneider *et al.*[43] assemble drug-derived building blocks and use a similarity measure as fitness function in order to mimic a template. Closely linked to this topic is the generation of combinatorial libraries by means of a GA (Refs 44,45). The number of compounds considered in virtual combinatorial libraries often exceeds the number of compounds that can be synthesized experimentally. A subset of fragments has to be chosen in order to reduce the number of products. Sheridan and Kearsley demonstrate the selection of a subset of amines for the construction of a tripeptoid library with a GA. They use a measure of similarity to a specific tripeptoid target as a scoring function[44].

Combined use of neural networks and GAs

Burden *et al.*[46] suggested more potent dihydrofolate reductase inhibitors by solving the neural network inversion problem for QSAR with a GA. They built a QSAR model based on a dataset of phenyl substituted diaminodihydrotriazines, with a small fully-connected feed-forward/back-propagation neural network, and determined the maximum activity on the structure–activity surface using a GA. Owing to the molecular representation of the compounds (related to the physicochemical properties of the substituents on the phenyl ring), the GA enables prediction of the required molecular properties for higher activity molecules.

Conclusions

Speeding up drug discovery and development is of central interest in all pharmaceutical companies. It has been shown that neural networks and genetic algorithms are powerful tools with a wide range of applications in the field of drug design. Hybrid algorithms that combine GAs and neural networks appear in the literature as promising methods for strengthening the impact of computational methods in drug design. However, the application of neural networks and GAs requires some essential knowledge of these methods before it can be properly employed.

References

1 Rumelhart, D.E. *et al.* (1986) Learning representations by back-propagating errors. *Nature* 323, 533–536

2 Svozil, D. *et al.* (1997) Introduction to multi-layer feed-forward neural networks. *Chemom. Intell. Lab. Syst.* 39, 43–62

3 Zupan, J. and Gasteiger, J. (1999) *Neural Networks in Chemistry and Drug Design* (2nd edn), Wiley

4 Devillers, J. (1996) *Principles of QSAR and Drug Design: Neural Networks in QSAR and Drug Design* (Vol. 2), Academic Press

5 Peterson, K.L. (2000) Artificial Neural Networks and Their Use in Chemistry. In *Reviews in Computational Chemistry* (Lipkowitz, K.B. and Boyd, D.B., eds) (Vol. 16), pp. 53–140, Wiley

6 Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. *Bioorg. Med. Chem. Letters* 8, 11–16

7 Kohonen, T. (2001) *Self-Organizing Maps* (Kohonen, T., ed.) (3rd edn), Springer

8 Hecht-Nielsen, R. (1987) Counterpropagation networks. *Applied Optics* 26, 4979–4984

9 Hecht-Nielsen, R. (1988) Application of counterpropagation networks. *Neural Networks* 1, 131–139

10 Zupan, J. *et al.* (1995) Neural networks with counter-propagation learning strategy used for modelling. *Chemom. Intell. Lab. Syst.* 27, 175–187

11 Anzali, S. *et al.* (1998) The use of self-organizing neural networks in drug design. In *3D QSAR in Drug Design* (Kubinyi, H. *et al.*, eds) (Vol. 2), pp. 273–299, Kluwer

12 Polanski, J. and Walczak, B. (2000) The comparative molecular surface analysis (COMSA): a novel tool for molecular design. *Computers and Chemistry* 24, 615–625

13 Huuskonen, J. *et al.* (1998) Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. *J. Chem. Inf. Comput. Sci.* 38, 450–456

14 Huuskonen, J. (2000) Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *J. Chem. Inf. Comput. Sci.* 40, 773–777

15 Schneider, G. *et al.* (1999) Combining *in vitro* and *in vivo* pharmacokinetic data for prediction of hepatic drug clearance in humans by artificial neural networks and multivariate statistical techniques. *J. Med. Chem.* 42, 5072–5076

16 Jalali-Heravi, M. and Parastar, F. (2000) Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives. *J. Chem. Inf. Comput. Sci.* 40, 147–154

17 Burden, F.R. and Winkler, D.A. (1999) Robust QSAR models using Bayesian regularized neural networks. *J. Med. Chem.* 42, 3183–3187

18 Burden, F.R. *et al.* (2000) Use of automatic relevance determination in QSAR studies using Bayesian neural networks. *J. Chem. Inf. Comput. Sci.* 40, 1423–1430

19 Tominaga, Y. (1999) Comparative study of class data analysis with PCA-LDA, SIMCA, PLS, ANNs and *k*-NN. *Chemom. Intell. Lab. Syst.* 49, 105–115

20 Cai, Y. and Bork, P. (1998) Homology-based gene prediction using neural nets. *Anal. Biochem.* 265, 269–274

21 Uberbacher, E.C. and Mural, R.J. (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. U. S. A.* 88, 11261–11265

22 Jones, D.T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* 287, 797–815

23 Qian, N. and Sejnowski, T.J. (1988) Predicting the secondary structure of globular proteins using neural network models. *J. Mol. Biol.* 202, 568–584

24 Dopazo, J. *et al.* (2001) Methods and approaches in the analysis of gene expression data. *J. Immunol. Meth.* 250, 93–112

25 Schneider, G. (2000) Neural networks are useful tools for drug design. *Neural Networks* 13, 15–16

26 Manallack, D.T. and Livingstone, D.J. (1999) Neural networks in drug discovery: have they lived up their promise? *Eur. J. Med. Chem.* 34, 195–208

27 Holland, J. (1975) *Adaptation in Natural and Artificial Systems*, University of Michigan Press

28 Venkatasubramanian, V. and Sundaram, A. (1998) Genetic algorithms: introduction and applications. In *Encyclopedia of Computational Chemistry* (Schleyer, P.v.R. *et al.*, eds), pp. 1115–1127, John Wiley & Sons

29 Jones, G. (1998) Genetic and evolutionary algorithms. In *Encyclopedia of Computational Chemistry* (Schleyer, P.v.R. *et al.*, eds), pp. 1127–1136, John Wiley & Sons

30 Judson, R. (1997) Genetic algorithms and their use in chemistry. In *Reviews in Computational Chemistry* (Boyd, D.B. and Lipkowitz, K.B., eds) (Vol. 10), pp. 1–73, Wiley

31 Devillers, J. (1996) *Principles of QSAR and Drug Design: Genetic Algorithms in Molecular Modeling* (Vol. 1), Academic Press

32 Hoffman, B.T. *et al.* (2000) 2D QSAR modeling and preliminary database searching for dopamine transporter inhibitors using genetic algorithm variable selection of Molconn Z descriptors. *J. Med. Chem.* 43, 4151–4159

33 Turner, D.B. and Willett, P. (2000) Evaluation of the EVA descriptor for QSAR studies: 3. The use of a genetic algorithm to search for models with enhanced predictive properties (EVA_GA). *J. Comput.-Aided Mol. Design* 14, 1–21

34 Kimura, T. *et al.* (1998) GA strategy for variable selection in QSAR studies: GA-based region selection for CoMFA modeling. *J. Chem. Inf. Comput. Sci.* 38, 276–282

35 Hasegawa, K. *et al.* (1999) GA strategy for variable selection in QSAR studies: application of GA-based region selection to a 3D-QSAR study of acetylcholinesterase inhibitors. *J. Chem. Inf. Comput. Sci.* 39, 112–120

36 Lemmen, C. and Lengauer, T. (2000) Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Design* 14, 215–232

37 Handschuh, S. and Gasteiger, J. (2000) The search for the spatial and electronic requirements of a drug. *J. Mol. Model.* 6, 358–378

38 Handschuh, S. *et al.* (1998) Superposition of three-dimensional chemical structures allowing for conformational flexibility by a hybrid method. *J. Chem. Inf. Comput. Sci.* 38, 220–232

39 Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748

40 Yang, J-M. and Kao, C-Y. (2000) Flexible ligand docking using a robust evolutionary algorithm. *J. Comput. Chem.* 21, 988–998

41 Morris, G.M. *et al.* (1998) Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662

42 Douguet, D. *et al.* (2000) A genetic algorithm for the automated generation of small organic molecules: drug design using an evolutionary algorithm. *J. Comput.-Aided Mol. Design* 14, 449–466

43 Schneider, G. *et al.* (2000) *De novo* design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput.-Aided Mol. Design* 14, 487–494

44 Sheridan, R.P. and Kearsley, S.K. (1995) Using a genetic algorithm to suggest combinatorial libraries. *J. Chem. Inf. Comput. Sci.* 35, 310–320

45 Brown, R.D. and Clark, D.E. (1998) Genetic diversity: applications of evolutionary algorithms to combinatorial library design. *Expert Opin. Ther. Pat.* 8, 1447–1460

46 Burden, F.R. *et al.* (1997) Predicting maximum bioactivity by effective inversion of neural networks using genetic algorithms. *Chemom. Intell. Lab. Syst.* 38, 127–137

Recent articles in the field of genomics from *Drug Discovery Today* and other *Trends* journals include:

Payne, D. *et al.* (2001) Bacterial fatty-acid biosynthesis: a genomics-driven target for antibacterial drug discovery. *Drug Discov. Today* 6, 537–544

Früh, K. *et al.* (2001) Virogenomics: a novel approach to antiviral drug discovery. *Drug Discov. Today* 6, 621–627

Mills, A.A. and Bradley, A. (2001) From mouse to man: generating megabase chromosome rearrangements. *Trends Genet.* 17, 331–339

Richards, R.I. (2001) Fragile and unstable chromosomes in cancer: causes and consequences. *Trends Genet.* 17, 339–345

Trede, N.S. *et al.* (2001) Fishing for lymphoid genes. *Trends Immunol.* 22, 302–307

Southan, C. (2001) A genomic perspective on human proteases as drug targets. *Drug Discov. Today* 6, 681–688

Emahazion, T. *et al.* (2001) SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends Genet.* 17, 407–413