

Drug Design with Artificial Neural Networks

1 OVIDIU IVANCIUC
2 Department of Biochemistry and Molecular Biology,
3 University of Texas Medical Branch, Galveston, USA

4 Article Outline

5 Glossary
6 Definition of the Subject
7 Introduction
8 Perceptron
9 Multilayer Feedforward Artificial Neural Network
10 Radial Basis Function Network
11 Self-Organizing Map
12 Counterpropagation Neural Network
13 Graph Machine Neural Networks
14 Other Neural Networks
15 Future Directions
16 Bibliography

17 Glossary

18 **Artificial neuron** An artificial neuron is a mathematical
19 function that simulates in a simplified form the func-
20 tions of biological neurons. Usually, an artificial neu-
21 ron has four computational functions, namely receives
22 signals through input connections from other neurons
23 or from the environment, sums the input signals, ap-
24 plies a nonlinear functions (transfer function or activa-
25 tion function) to the sum, and sends the result to other
26 neurons or as output from the neural network.

27 **Counterpropagation neural network** The counterprop-
28 agation neural network is a hybrid network that con-
29 sists of a self-organizing map as the hidden layer and
30 an output layer that has as output a computed value for
31 the modeled property. The network implements a su-
32 pervised learning algorithm that converges to a unique
33 solution.

34 Multilayer feedforward artificial neural network

35 A multilayer feedforward (MLF) artificial neural net-
36 work consists of artificial neurons organized in layers.
37 The MLF network has an input layer that receives the
38 structural descriptors for each molecule, an output
39 layer that provides one or more computed properties,
40 and one or more hidden layers situated between the
41 input and the output layers. Each neuron in a hidden
42 layer receives signals from neurons in the preceding
43 layer and sends signals to the neurons in the next layer.

44 **Perceptron** A perceptron is a linear classifier that consists
45 of a layer of input neurons and an output neuron. Each
46 connection between an input neuron and the output
47 neuron has a weight. Depending on the sum of the sig-
48 nals received by the output neuron, its output is +1 or
49 -1.

50 Quantitative structure-activity relationships

51 Quantitative structure-activity relationships (QSAR)
52 represent regression models that define quantita-
53 tive correlations between the chemical structure of
54 molecules and their physical properties (boiling point,
55 melting point, aqueous solubility), chemical properties
56 and reactivities (chromatographic retention, reaction
57 rate), or biological activities (cell growth inhibition,
58 enzyme inhibition, lethal dose). The fundamental
59 hypotheses of QSAR is that similar chemicals have
60 similar properties, and small structural changes result
61 in small changes in property values. The general form
62 of a QSAR equation is $P(i) = f(\mathbf{SD}_i)$, where $P(i)$ is
63 a physical, chemical, or biological property of com-
64 pound i , \mathbf{SD}_i is a vector of structural descriptors of i ,
65 and f is a mathematical function such as linear regres-
66 sion, partial least squares, artificial neural networks, or
67 support vector machines. A QSAR model for a prop-
68 erty P is based on a dataset of chemical compounds
69 with known values for the property P , and a matrix of
70 structural descriptors computed for all chemicals. The
71 learning (training) of the QSAR model is the process
72 of determining the optimum parameters of the re-
73 gression function f . After the training phase, a QSAR
74 model may be used to predict the property P for novel
75 compounds that are not present in the learning set of
76 molecules.

77 **Radial basis function network** The radial basis function
78 (RBF) neural network has three layers, namely an in-
79 put layer, a hidden layer with a non-linear RBF activa-
80 tion function and a linear output layer.

81 **Self-organizing map** A self-organizing map (SOM) is an
82 artificial neural network that uses an unsupervised
83 learning algorithm to project a high dimensional input
84 space into a two dimensional space called a map. The
85 topology of the input space is preserved in SOM, and
86 points that are close to each other in the SOM grid cor-
87 respond to input vectors that are close to each other in
88 the input space. A SOM consists of neurons arranged
89 usually in a rectangular or hexagonal grid. Each neu-
90 ron has a position on the map and a weight vector of
91 the same dimension as the input vectors.

92 **Structural descriptor** A structural descriptor (SD) is
93 a numerical value computed from the chemical struc-
94 ture of a molecule, which is invariant to the number-

Please note that the pagination is not final; in the print version an entry will in general not start on a new page.

ing of the atoms in the molecule. Structural descriptors may be classified as constitutional (counts of molecular fragments, such as rings, functional groups, or atom pairs), topological indices (computed from the molecular graph), geometrical (volume, surface, charged-surface), quantum (atomic charges, energies of molecular orbitals), and molecular field (such as those used in CoMFA, CoMSIA, or CoRSA).

Structure–activity relationships Structure–activity relationships (SAR) represent classification models that can discriminate between sets of chemicals that belong to different classes of biological activities, usually active/inactive towards a certain biological receptor. The general form of a SAR equation is $C(i) = f(\mathbf{SD}_i)$, where $C(i)$ is the activity class of compound i (active/inactive, inhibitor/non-inhibitor, ligand/non-ligand), \mathbf{SD}_i is a vector of structural descriptors of i , and f is a classification function such as k -nearest neighbors, linear discriminant analysis, random trees, random forests, Bayesian networks, artificial neural networks, or support vector machines.

Definition of the Subject

The fundamental hypothesis of the structure-property models is that the structural features of molecules determine the physical, chemical and biological properties of chemical compounds. The first studies that use structure-activity relationships to explain the biological properties of sets of compounds were published by Kopp [74], Crum-Brown and Frazer [18], Meyer [88], and Overton [97]. Modern structure-activity relationships (SAR) and quantitative structure-activity relationships (QSAR) models are based on the Hansch model that predicts a biological property as a statistical correlation with steric, electronic, and hydrophobic indices [27,35,36,37]. The Hansch model shaped the general scene of structure-activity correlations, and almost all subsequent SAR and QSAR models are variations that extend the Hansch model with novel classes of descriptors or with more powerful statistical models, such as partial least squares (PLS), artificial neural networks (ANN), support vector machines (SVM), or other machine learning algorithm. A structural descriptor is a numerical representation of some important molecular features, such as empirical indices (Hammett and Taft substituent constants), physical properties (octanol-water partition coefficient, dipole moment, aqueous solubility), counts of substructures or substituents, graph descriptors [13,54,125], topological indices [4,12,55], connectivity indices [65,66] electrotopological indices [67],

geometrical descriptors (molecular surface and volume), quantum indices (atomic charges, HOMO and LUMO energies) [60,124], and molecular fields (steric, electrostatic, and hydrophobic). SAR represent classification models that are used when the experimental property is a class label (+1/−1), such as substrate/non-substrate, inhibitor/non-inhibitor, ligand/non-ligand, toxic/non-toxic, or carcinogen/non-carcinogen. Classification models are used to screen chemical libraries and to identify compounds that have a desired biological activity, such as inhibitor for a particular enzyme. QSAR represent regression models that are used for experimental properties with continuous values, such as hydrophobicity, aqueous solubility, membrane penetration, lethal concentration, or inhibition constant for an enzyme.

The drug discovery and development process is lengthy, expensive, and has a high attrition rate. The duration of the drug development phases, namely pre-clinical (1 to 5 years), clinical (5 to 11 years), and approval (0.5 to 2 years) [9], puts at 18 years the upper limit for bringing a drug to market, without considering the duration of the drug discovery phase. For 168 drugs approved between 1992 and 2002, the median clinical trial and approval periods were 5.1 and 1.2 years, respectively, with no tendency of decreasing despite exceptional technological advances [63]. Cost data for 68 randomly selected drugs from 10 pharmaceutical companies show that better discovery and screening programs can reduce the total cost per approved drug by up to \$US 242 million [20]. A similar analysis shows that the average out-of-pocket cost per new drug is \$US 403 million (2000 dollars), which increases to \$US 802 after adding the opportunity cost (the cost of pursuing one choice instead of another) [21]. Depending on the therapy or the developing firm, the cost per new drug varies between \$US 500 million to \$US 2,000 million [1]. The attrition rate of the chemical compounds is also impressive, considering that the process starts with millions of compounds tested in high throughput screening and eventually ends with one successful drug on the market. From all compounds that enter the clinical trials, 30% fail due to lack of efficacy, 30% fail in toxicological and safety tests, and only 11% finish the trials [73]. A study considering all 548 new chemical entities approved between 1975 and 1999 found that 45 drugs (8.2%) received one or more black box warnings and 16 (2.9%) were withdrawn from the market [76]. Examples of drugs withdrawn from the market due to adverse reactions include troglitazone (Rezulin) in 2000 [25], cerivastatin (Baycol, Lipobay) in 2001 [28], nefazodone (Serzone) in 2003 [17], pemoline (Cylert) in 1999 (Canada) and 2005 (USA) [42], and rofecoxib (Vioxx) in 2004 [22]. For all difficulties, expenses,

194 and failures that are associated with the drug develop- 243
195 ment process, the reward for a successful drug can be sub- 244
196 stantial. The top 2006 best selling drugs (in \$US billions) 245
197 are lipitor with 14.39 (Pfizer, cholesterol), advair with 246
198 6.13 (GlaxoSmithKline, asthma), plavix with 6.06 (Bristol- 247
199 Myers Squibb, vascular disease), nexium with 5.18 (As- 248
200 traZeneca, acid reflux), norvasc with 4.87 (Pfizer, hyper- 249
201 tension). 250

202 Computer-assisted drug design (CADD) uses compu- 251
203 tational chemistry to increase the chances of finding valu- 252
204 able drug candidates. Among the broad range of compu- 253
205 tational tools used in CADD, artificial neural networks 254
206 (ANN) have a special place due to their abilities to model 255
207 with high accuracies the complex relationships between 256
208 chemical structures and molecular properties. ANN can be 257
209 used both for classification (SAR) and regression (QSAR), 258
210 and they are essential tools in drug discovery cycles to 259
211 identify active compounds in chemical libraries, and to 260
212 optimize a wide range of physico-chemical and biologi- 261
213 cal properties, such as enzyme inhibition, target selectiv- 262
214 ity, or membrane transport. The SAR and QSAR models 263
215 based on ANN may predict with high accuracy properties 264
216 for novel chemicals, even before their chemical synthesis. 265
217 CADD systems based on ANN may increasing the chances 266
218 of bringing a drug to market, when they are integrated into 267
219 the chemical, biological, pharmacological and clinical pro- 268
220 cedures used by the pharmaceutical industry. 269

221 Introduction

222 The simulation of brain mechanisms with artificial neural 272
223 networks was initiated by McCulloch and Pitts [84,101] 273
224 and by [106]. Minsky and Papert investigated the math- 274
225 ematical properties of these simple neural networks in 275
226 their seminal book *Perceptrons* [92]. Although perceptrons 276
227 are able to solve some problems, they fail to learn the 277
228 XOR (exclusive-or) operation. As a result, the research in 278
229 this field was limited until Hopfield demonstrated that by 279
230 adding nonlinear functions to each artificial neuron the 280
231 network could simulate complex behavior [43]. 281

232 Another milestone in ANN research was the discovery 282
233 of the backpropagating algorithm for the training of mul- 283
234 tilayer feedforward artificial neural networks [107,108]. 284
235 Multilayer feedforward (MLF) networks consist of multi- 285
236 ple layers of artificial neurons, all connected in a feedfor- 286
237 ward way, from the input layer to the output layer. MLF 287
238 networks have at least one hidden layer of neurons, and the 288
239 neurons in the hidden and output layers have nonlinear 289
240 activation functions. The hidden neurons and the nonlin- 290
241 ear activation functions are essential in modeling nonlin- 291
242 ear relationships between input variables and the output 292
293

property. MLF networks represent the most widespread 243
ANN type used to model physico-chemical, biological, and 244
toxicological properties of chemical compounds. 245

246 In addition to MLF, other networks used in drug de- 247
248 sign are radial basis function networks, self-organizing 248
249 maps, counter-propagation networks, graph machine neu- 249
250 ral networks, and probabilistic neural networks. Several 250
251 reviews [49,52,144] and books are recommended for a de- 251
252 tailed overview of neural networks and their applications: 252
253 *Neural Computing* by Wasserman [134], *Neural Networks* 253
254 *for Pattern Recognition* by Bishop [10], *Pattern Recogni-* 254
255 *tion and Neural Networks* by Ripley [104], *Neural Net-* 255
256 *works for Chemical Engineers* by Bulsari [15], and *Neu-* 256
257 *ral Networks in Chemistry and Drug Design* by Zupan and 257
Gasteiger [145]. 258

259 The artificial neuron and the perceptron are de- 259
260 scribed in Sect. “Perceptron”. MLF networks are used to 260
261 model physico-chemical properties, such as melting tem- 261
262 perature [75], boiling temperature [33], aqueous solubil- 262
263 ity [46], and *n*-octanol/water partition coefficients [123]. 263
264 Drug design applications of MLF networks include mod- 264
265 els for enzyme inhibition [120], human intestinal absorp- 265
266 tion [136], virtual screening of chemical libraries [102], 266
267 and genotoxicity [127]. Another important application of 267
268 MLF networks is in cancer diagnosis from gene expression 268
269 data [70,119,135]. Section “Multilayer Feedforward Arti- 269
270 ficial Neural Network” is an overview of the most important 270
271 MLF network models in structure-property prediction and 271
in drug design. 272

273 The radial basis function (RBF) neural network be- 273
274 longs to the class of feedforward artificial neural network, 274
275 and it has three layers, namely an input layer, a hidden 275
276 layer with a non-linear RBF activation function and a lin- 276
277 ear output layer [93]. RBF networks give predictive mod- 277
278 els for various properties, such as boiling temperature [81], 278
279 QSAR for enzyme inhibitors [99,100], ligand-target bind- 279
280 ing affinity [142], and aquatic toxicity [87]. SAR and QSAR 280
281 applications of the RBF network are presented in Sect. “Ra- 281
dial Basis Function Network”. 282

283 The self-organizing map (SOM) was developed by Ko- 283
284 honen as an unsupervised neural network that transforms 284
285 a high dimensional input space into a two dimensional 285
286 space [71,72]. SOM maps the objects by trying to preserve 286
287 the topological properties of the input space. To predict 287
288 the class of a new molecule, a vector of structural descrip- 288
289 tors is sent as input for a trained SOM and projected onto 289
290 a neuron from the map. The molecule is classified in the 290
291 same class with the training molecules that are projected 291
292 onto the same neuron. SOM networks cluster chemical 292
293 compounds according to their similarity, and thus can be 293
used to identify groups of molecules with similar proper-

ties [3]. Applications of SOM are found in virtual screening of chemical libraries [105], in modeling the aquatic toxicity of organic chemicals [98], or in structure-activity studies [11]. These and other SOM applications are examined in Sect. “Self-Organizing Map”.

The counterpropagation neural network (CPNN) was developed by Hecht-Nielsen by transforming a self-organizing map into a supervised learning neural network [39,40,41]. The input layer of neurons is connected to the hidden layer that is a self-organizing map, and the excited neuron from SOM sends a signal to the output layer. CPNN may be used for classification or for regression. SAR applications of CPNN include the mutagenicity prediction of aromatic amines [59], and the classification of organic compounds based on their mode of toxic action [117]. Applications of the counterpropagation neural network in SAR and QSAR are presented in Sect. “Counterpropagation Neural Network”.

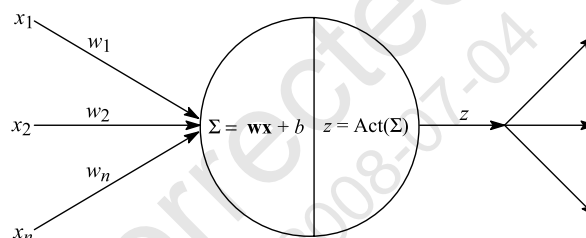
Graph machine neural networks learn SAR and QSAR models directly from the molecular graph [32], by translating the chemical structure into the network topology. For each chemical compound presented to the graph machine, the neural network adopts a structure derived from the molecular graph of the chemical. There are several graph machines used with success in SAR and QSAR, such as recursive neural networks [5,89], the Baskin–Palyulin–Zefirov (BPZ) neural device [7], ChemNet [69], and MolNet [47,50]. The BPZ neural device is constructed by analogy with a biological vision system and contains a sensor field, a set of eyes, and a brain. A chemical structure investigated with the BPZ neural device is represented as a molecular matrix that is superimposed over a sensor field that receives the network input. The information received by the sensor field is transmitted to a set of eyes that transforms it into several MLF networks, and sends signals to the next block, the brain. ChemNet encodes a molecular distance matrix between the input and hidden layers, and has as output an atomic property. In MolNet each atom in a molecule has a corresponding input and hidden neuron, and the connections between the input and hidden layers correspond to the weighted interactions between atoms. The output neuron offers the computed value for a molecular property. In Sect. “Graph Machine Neural Networks” we review structure-activity models computed with graph machine neural networks.

Other ANN used in drug design and molecular property modeling are the Bayesian regularized neural network [31,133], the probabilistic neural network [38,77], and the general regression neural network [141]. Their applications in SAR and QSAR are presented in Sect. “Other Neural Networks”.

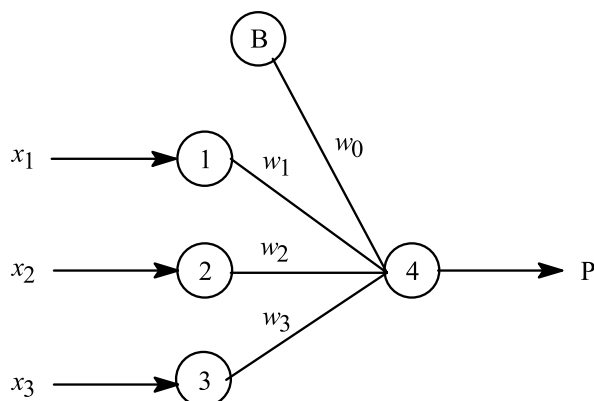
Perceptron

An artificial neuron is a simple model of a biological neuron expressed as a mathematical function. Artificial neurons represent the basic computational units in an artificial neural network. Each artificial neuron has four basic functions (Fig. 1): (a) receives input signals through connections from the neurons in the previous layer or from the environment; (b) makes a summation of the inputs signals; (c) the result of summation is passed to the activation function and transformed with a mathematical function; (d) the result of the activation function is used by a transfer function to produce an output value that is sent to the neurons in the next layer of neurons. The artificial neuron depicted in Fig. 1 receives the input signals x_1, \dots, x_n through connections with the weights w_1, \dots, w_n . The summation of the input signal is computed as the dot product between \mathbf{w} and \mathbf{x} plus a bias parameter b . An activation function Act is applied to the sum Σ to give the output value z . The first artificial neuron was the threshold logic unit proposed by McCulloch and Pitts [84,101]. The activation function was a step function, which has the value is zero for negative argument and one for positive argument. Other important activation functions are: linear, $\text{Act}(z) = z$; sigmoid, $\text{Act}(z) = 1/(1 + e^{-z})$ [43]; hyperbolic tangent, $\text{Act}(z) = \tanh(z)$; symmetric logarithmoid, $\text{Act}(z) = \text{sign}(z) \ln(1 + |z|)$ [15].

A perceptron is the simplest type of ANN, consisting of a layer of input neurons and an output neuron (Fig. 2) [92,106]. The perceptron shown in Fig. 2 has three input neurons (1, 2, 3), a bias neuron B, and an output neuron 4. An input neuron receives an input signal, scales it and then sends it to the output neuron. A simple perceptron is a linear classifier, which cannot learn simple operations, such as the XOR (exclusive-or) operation. A major development was the introduction of nonlinear activation functions that allows ANN to simulate any nonlinear function [43].



Drug Design with Artificial Neural Networks, Figure 1
Structure and functions of an artificial neuron



Drug Design with Artificial Neural Networks, Figure 2
Perceptron structure and signal flow

382 Multilayer Feedforward Artificial Neural Network

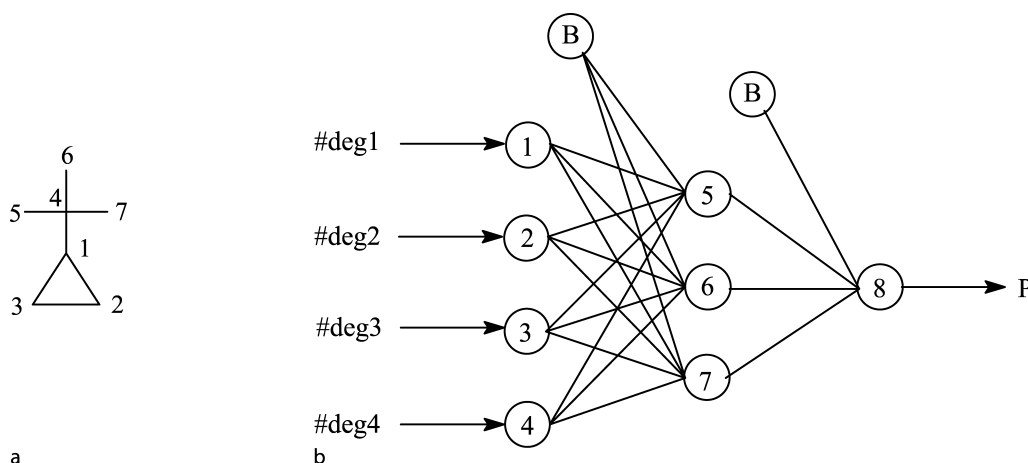
383 A multilayer feedforward artificial neural network, called
384 also a multilayer perceptron (MLP) is a generalization of
385 the simple perceptron obtained by inserting one or more
386 hidden layer of neurons between the input layer and the
387 output layer. Another important characteristic is the use
388 of nonlinear activation functions in the hidden and output
389 layers, although the output neurons can have also a linear
390 activation function. MLF ANN are universal approxima-
391 tors, namely they can approximate any function with arbi-
392 trary precision provided that enough neurons are used in
393 the hidden layer.

394 In SAR and QSAR applications the structural descrip-
395 tors represent the input data received by the input neu-
396 rons. As an example, we consider topological descrip-

397 tors computed from the molecular graph of *t*-butylcy-
398 clopropane (Fig. 3a) that are used as input in an MLF
399 ANN with one hidden layer (Fig. 3b). The input layer is
400 formed by neurons 1–4, the hidden layer contains neurons
401 5–7, and the output layer is represented by neuron 8. The
402 bias neurons B send a signal with the value +1 through
403 weighted connections, and have the role to fine-tune each
404 neuron. The topological descriptors represent the counts
405 of vertex degrees, namely $\text{Deg} = \{3, 2, 1, 1\}$, i. e., there
406 are three atoms with degree 1, two atoms with degree 2,
407 one atom with degree 3, and one atom with degree 4. These
408 numbers are used as input to neurons 1–4, which perform
409 a scaling and then send the values to each neuron from
410 the hidden layer. The output neuron 8 offers the computed
411 value of a property P for *t*-butylcyclopropane.

412 The MLF ANN application to property modeling con-
413 sists of two phases, namely training and prediction. Dur-
414 ing the training (learning) phase the weights of all con-
415 nections are optimized (adjusted) in order to estimate
416 with high precision the investigated molecular property.
417 The most popular training method is the backpropagat-
418 ing algorithm [107,108], a variant of the gradient algo-
419 rithms. More efficient training algorithms are direction-
420 set methods (Powell's method), methods that require the
421 computation of the first derivatives such as conjugate gra-
422 dient methods (Fletcher-Reeves or Polak-Ribiere), Lev-
423 enberg-Marquardt, or quasi-Newton (variable metric)
424 methods (Davidon-Fletcher-Powell or Broyden-Fletcher-
425 Goldfarb-Shanno).

426 MLF ANN can be trained with artificial intelligence
427 algorithms, such as particle swarm optimization, which
was used in a QSAR study of inhibitors of platelet-derived



Drug Design with Artificial Neural Networks, Figure 3

ANN prediction of molecular properties from the molecular structure: topological descriptors computed from a the molecular graph of *t*-butylcyclopropane are used as input in b a multilayer feedforward neural network with one hidden layer

428 growth factor receptor phosphorylation [113]. Such algo-
429 rithms can provide a complete optimization of an MLF
430 ANN, namely they can optimize both the network topol-
431 ogy and the connection weights. The network topology is
432 defined by the number of hidden layers and by the num-
433 ber of neurons in each hidden layer. The complete opti-
434 mization of ANN with particle swarm optimization has
435 fast convergence to the global minimum [85] and avoids
436 the overfitting of the learning dataset of chemicals [112].
437 The selection of the most important descriptors for the
438 ANN input is an important step in obtaining predictive
439 models. Several artificial intelligence algorithms can be
440 used for an efficient feature selection, such as genetic algo-
441 rithm [115,116], particle swarm optimization [2], and ant
442 colony optimization [58].

443 Taskinen and Yliruusi reviewed the MLF ANN appli-
444 cations in predicting several physico-chemical properties
445 of interest in drug development, such as octanol-water
446 partition coefficient, aqueous solubility, boiling tempera-
447 ture, and vapor pressure [121]. The melting temperatures
448 for sulfur-containing organic compounds were predicted
449 from simple molecular graph indices [75]. The melting
450 temperatures of 717 ionic liquids were modeled with MLF
451 ANN, associative neural networks, support vector ma-
452 chines, k -nearest neighbors, and multiple linear regres-
453 sion [126]. The structural descriptors evaluated in these
454 models were electrotopological indices [67] and Dragon
455 descriptors [124]. The moderate accuracy of predictions
456 obtained with all algorithms can be explained by the diffi-
457 culty to encode the solid state characteristics of ionic liq-
458 uids. Hall and Story used electrotopological indices and
459 an MLF ANN to model the boiling temperatures of 298
460 compounds and the critical temperatures of 165 com-
461 pounds [33]. The best results for the boiling temperatures
462 were obtained with a neural network with five hidden neu-
463 rons with a mean absolute error of 3.93 K, whereas a net-
464 work with four hidden neurons gave good predictions for
465 the critical temperatures with a mean absolute error of
466 4.52 K. In a similar study, atom type electrotopological
467 state indices were used to model the boiling temperatures
468 of alkanes, alcohols and chloroalkanes [34].

469 Huuskonen et al. developed MLF ANN models for the
470 aqueous solubility of a diverse set of 734 organic com-
471 pounds [46]. Atom type electrotopological state indices
472 were used as input data in a neural network with five hid-
473 den neurons. Good results were obtained both in train-
474 ing ($s = 0.52$) and prediction ($s = 0.75$), showing that
475 the aqueous solubility for a large and diverse set of or-
476 ganic compounds can be predicted from topological in-
477 dices. Tetko et al. developed a neural network model for
478 the octanol-water partition coefficients of 12,908 organic

479 compounds [123]. The model consists of an ensemble of
480 50 ANN, each ANN having 75 input neurons and 10 neu-
481 rons in the hidden layer. The leave-one-out cross-valida-
482 tion indicate that the model gives good predictions, with
483 $r^2 = 0.95$. The ^{13}C NMR chemical shift can be predicted
484 from simple topological parameters, as shown for a dataset
485 of acyclic alkenes [56,57].

486 To compensate for the lack of toxicological data for
487 aquatic organisms, a number of neural networks QSAR
488 models were proposed for the toxicity of organic chemi-
489 cals against fathead minnow, rainbow trout, bluegill sun-
490 fish, *Daphnia magna*, *Tetrahymena pyriformis*, and *Vib-
491 rio fischeri* [61]. These ANN models can be used to pre-
492 dict the environmental impact of novel chemicals, and to
493 prioritize their experimental evaluation in aquatic toxic-
494 ity assays. Huuskonen tested multiple linear regression
495 and artificial neural networks in QSAR models for toxic-
496 ity of 140 organic chemicals against fathead minnow [45].
497 A neural network trained with 14 electrotopological state
498 indices provided a predictive model for the experimen-
499 tal toxicity. Basak et al. modeled the toxic modes of
500 action for 283 chemicals with ANN and topological in-
501 dices [6]. The toxic compounds were classified as nar-
502 cotics, electrophiles/proelectrophiles, uncouplers of oxida-
503 tive phosphorylation, acetylcholinesterase inhibitors, and
504 neurotoxicants. The modes of action SAR models provide
505 reliable predictions, with rates of correct classification be-
506 tween 65% and 95%. Lower predictions are obtained for
507 classes with a low number of compounds.

508 Dopamine antagonists, serotonin antagonists, and
509 serotonin-dopamine dual antagonists are used as anti-
510 psychotics. Based on a training dataset consisting of
511 1135 dopamine antagonists, 1251 serotonin antagonists,
512 and 386 serotonin-dopamine dual antagonists, Kim et al.
513 tested several machine learning algorithms, including
514 ANN and recursive partitioning [68]. The average classi-
515 fication rate in training was 73.6% and in prediction was
516 69.8%, indicating that these models can be used in virtual
517 screening to identify new active compounds. Siu and Che
518 used multiple linear regression, partial least squares, and
519 ANN to model the α -amino acids affinities for H^+ and for
520 sodium, copper, and silver cations [114]. The cross-vali-
521 dation tests indicate that the best predictions are obtained
522 with the artificial neural network. These QSAR models are
523 useful in elucidating the binding properties of the α -amino
524 acids.

525 Votano et al. used ANN, k -nearest neighbors, and
526 decision forest to model the mutagenicity of 3363 di-
527 verse compounds tested for their Ames genotoxicity [127].
528 The dataset was split into 2963 training compounds and
529 400 prediction compounds, and the SAR models

were developed with 148 topological indices that included electrotopological state indices and molecular connectivity indices. All three classifiers gave good predictions, with a slight advantage for the neural network, as indicated by the area under the receiver operator characteristic (AUROC) curve, namely AUROC = 0.93 for ANN, AUROC = 0.92 for k -nearest neighbors, and AUROC = 0.91 for decision forest.

A good intestinal absorption is desirable property for drugs, but its experimental determination is costly and time-consuming. Wessel et al. used ANN to estimate the percent human intestinal absorption (% HIA) of 86 drugs [136]. A neural network optimized with a genetic algorithm provided the best results, with an error of 9.4% HIA units for training, 19.7% HIA units the cross-validation, and 16.0% HIA units for an external prediction set.

Sutherland et al. evaluated ANN, PLS, genetic function approximation, and genetic PLS in QSAR models for inhibitors of angiotensin converting enzyme, acetylcholinesterase, benzodiazepine receptor, cyclooxygenase-2, dihydrofolate reductase, glycogen phosphorylase B, thermolysin, and thrombin [120]. ANN gave the best predictions for QSAR models obtained with graph indices and geometrical descriptors. Another large scale evaluation of structure-activity models was made for chemicals that target HIV-reverse transcriptase, COX2, dihydrofolate reductase, estrogen receptor, and thrombin [102]. A comparison of several nonlinear modeling algorithms shows that ANN are successful in identifying active compounds from chemical libraries and can be used in virtual screening. Pharmaceutical companies constantly enrich their collections of chemical compounds by acquisitions that increase the structural diversity of molecules available for high throughput screening. Muresan and Sadowski proposed an ANN system to compute an “in-house likeness” score for compound acquisition [94]. The analysis of several datasets show that a set of atom-type counts used as input to the ANN models represents an efficient way of identifying structural patterns that are missing from an in-house collection.

Votano et al. compared multiple linear regression, ANN, k -nearest neighbors, and support vector machines in QSAR models for the human serum protein binding of 1008 chemicals (808 for training and 200 for prediction) [128]. The best predictions ($r^2 = 0.7$) were obtained with ANN models, indicating that these QSAR can assist the drug design process. Katritzky et al. evaluated ANN and multiple linear regression in QSAR models for 277 inhibitors of glycogen synthase kinase-3 [62]. The descriptors were selected from a pool consisting of subgraph counts, topological indices, geometrical parameters and

quantum indices. The QSAR models highlight the structural factors that influence the inhibition of glycogen synthase kinase-3. Inhibitors of the hERG (human ether-á-go-go-related gene) potassium channel can lead to a prolongation of the QT interval that can trigger torsade de pointes, an atypical ventricular tachycardia. Seierstad and Agrafiotis used ANN to model the nonlinear structure-activity relationship for hERG channel inhibitors [110]. The QSAR models with best predictive power are obtained with a feature selection to prevent over-fitting and by aggregation several ANN into an ensemble model to minimize the instability in predicting novel chemicals. The time- and dose-dependent anti-inflammatory in vivo activity of substituted imidazo[1,2-*a*]pyridines was modeled with artificial neural networks [109]. This study shows that ANN have unique properties to study pharmacodynamic and pharmacokinetic properties of drugs and drug-like compounds. Peptides that have a good binding affinity for major histocompatibility complex (MHC) class I molecules are candidates in the development of new vaccines against cancer and viral infections. Filter et al. modeled the peptide binding to MHC class I allele HLA-A*0201 with artificial neural networks [26]. The ANN system can identify peptides that do not correspond to the usual recognition motifs, as shown for several new melanoma-associated peptides.

Artificial neural networks represent the method of choice is cancer prediction and prognosis [19]. Their use improves the accuracy of predicting cancer susceptibility, recurrence and mortality. Molecular data, such as protein biomarkers and microarrays, together with ANN models can improve the understanding of cancer development and progression. The mechanism of action of drugs against 60 malignant cell lines in the National Cancer Institute drug screening program was investigated with ANN [135]. Out of the 141 drugs, the ANN gives incorrect predictions for only 12 of them, showing that the model can be used to guide the screening program. Khan et al. used ANN and gene expression data to identify the class of small, round blue-cell tumors [64]. Although the classification of these tumors difficult in clinical practice, the neural networks identified the relevant genes and correctly classified all samples. ANN trained with gene microarray data can identify with success several types of cancers, such as colon cancer [111] and esophageal cancer [138].

Radial Basis Function Network

An RBF neural network is a feedforward network with three layers, in which the neurons in the hidden layer have a RBF activation function [93]. RBF neural networks are

581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629

universal approximators, that can approximate any continuous function with arbitrary precision if enough neurons are allowed in the hidden layer. An early application of the RBF network in structure-property studies was reported by Lohninger for the boiling temperatures of organic compounds [81]. The training process of a radial basis function artificial neural network (RBF ANN) consists of selecting the network topology, finding the centers and widths of the RBF neurons, and computing the connection weights between the hidden and output layers. Zhou et al. showed that the particle swarm optimization is very efficient in training an RBF neural network [143]. The procedure was tested with a dataset of 40 inhibitors of murine P388 leukemia cells and 70 structural descriptors, and the results indicate that the swarm optimization converges fast to a global optimum. Wan and Harrington compared two algorithms for RBF ANN training, namely K-means clustering and linear averaging [131]. The results obtained for two datasets (polychlorobiphenyl mass spectra and Italian olive oil classification) indicate that the linear averaging method gives better predictions.

QSAR models for glycine/NMDA receptor antagonists, which may treat stroke or seizure, were computed for a dataset consisting of 109 compounds [99]. The set of structural descriptors included topological indices, geometric descriptors, quantum indices, and polar surface parameters. The optimum set of descriptors was selected with a genetic algorithm and an RBF network. SAR models for inhibitors of protein tyrosine phosphatase 1B, which are potential agents for the treatment for type 2 diabetes and obesity, were computed for 128 compounds [100]. The classification models were obtained with RBF networks, linear discriminant analysis, and k -nearest neighbors. A correct classification rate of 85.7% was obtained for an external prediction set, indicating that the SAR model can be used to find inhibitors in virtual libraries. Neuraminidase is a key target treating infections with the influenza virus, and many classes of inhibitors are tested for this enzyme. In order to identify the structural features that influence the inhibition of this enzyme, Lü et al. evaluated RBF networks in QSAR models for 46 neuraminidase inhibitors [82]. Five important descriptors were found with a heuristic search among several hundred indices. Prediction tests show that the RBF network results are comparable with those obtained with a multiple linear model, indicating a possible absence of nonlinear structure-activity relationships for these inhibitors.

Zheng et al. compared RBF networks, support vector machines, and partial least square in QSAR models for the Ah receptor binding affinities of polychlorinated, polybrominated, and polychlorinated-brominated

dibenzo-p-dioxins [142]. The best leave-one-out cross-validation predictions are obtained by the RBF network, with $q^2 = 0.88$. Melagraki et al. used RBF networks to model the toxicity of 39 organic compounds against *Vibrio fischeri* [87]. The descriptors used in the simulation include topological indices, electronegativity parameters, and lipophilicity. The good QSAR predictions indicate that the model can be used to evaluate the toxicity of diverse chemicals. QSAR models based on RBF networks were developed for the toxicity of 221 phenols against *Tetrahymena pyriformis* [86]. Prediction tests performed with cross-validation and with an external set show that the RBF network gives more accurate predictions than multiple linear regression.

Self-Organizing Map

A self-organizing map is an artificial neural network that projects a high-dimensional input space into a low-dimensional (usually two dimensional) output space [71,72]. A SOM is trained with unsupervised learning, meaning that the training molecules do not need a class label or property value. The input vectors are mapped in a two dimensional space by preserving the topological properties of the input space. The output layer in a SOM is a matrix of neurons, with the property that an input vector is projected into a single output neuron. Several similar input vectors may be project into the same output neuron, or into adjacent output neurons that form clusters of similar objects. This SOM property is used to discover clusters in collections of chemical compounds. The chemical similarity of large chemical libraries is easily evaluated by projecting the molecules onto a SOM that auto-organizes similar compounds by preserving their similarity relationships in the descriptor space. SOM are used to screen libraries to identify clusters of structurally similar compounds that may have similar biological properties.

SOM may be used also for classification by assigning class labels to a trained network (Fig. 4). This example shows an 8×8 SOM with labels attached to the output neurons. The class labels are attached after training and represent the majority class for objects projected onto a particular output neuron. The objects are organized in four clusters, two for class 1, one for class 2, and one for class 3. The SOM predictions are straightforward, namely a new object is projected onto an output neuron and takes the label of the output neuron as its predicted class. Single molecules may be mapped onto a SOM network by projecting the points situated on the molecular surface [3]. The Cartesian coordinates of each surface point are used as input for the SOM network that provides a two-dimen-

2	2	2		1	1	1	1
2	2				1	1	1
2	2			1	1	1	1
3	3	3					
3	3		1		1		
3	3	3	1	1	1	1	
			1	1	1		

Drug Design with Artificial Neural Networks, Figure 4
A self-organizing map with three classes

sional display of the molecular surface. To add more structural information on the map, output neurons are colored according to the molecular electrostatic potential of the surface points projected onto them.

SOM networks are very efficient in discovering activity clusters in chemical libraries that contain active compounds for several targets. Terfloth and Gasteiger clustered with SOM a number of 299 compounds representing four activity classes, namely 75 5-hydroxytryptamine 5-HT_{1a}-receptor agonists, 75 histamine H₂-receptor antagonists, 74 monoamine oxidase MAOA inhibitors, and 75 thrombin inhibitors [122]. All classes are generally clustered in distinct regions, but a small number of neurons represent chemicals from different classes. Such degenerate mapping may be solved by increasing the size of the map or by testing other structural descriptors that may discriminate better these activity classes. Drug candidates are routinely screen to determine if they are hERG channel blockers, because such chemicals can cause sudden cardiac death. Roche et al. examined the hERG activity of a chemical library with SOM, PLS, MLF ANN, principal component analysis, and substructure analysis [105]. These classification models offer good predictions for a validation set, with 71% success rate for active compounds and 93% for inactive compounds.

Trypanosoma cruzi is deposited on the skin surface by bugs from the Reduviidae family, and then it infects the human host by penetrating through insect bites, thus causing trypanosomiasis, or Chagas disease. The chronic form of the disease may develop more than 10 years after the infection, and may be fatal. Boiani et al. developed several computational models to evaluate the anti-*Trypanosoma cruzi* activity of *N*-oxide containing heterocycles [11]. A SOM was applied to separate between activity classes and to select an optimum set of descrip-

tors. Other models tested with success for this dataset are *k*-nearest neighbors and decision tree.

QSAR models have limitations for the domain of chemical structures and biological activities for which they give reliable predictions. For example, interpolation gives better predictions than extrapolation, and extrapolations far away from the training structural space or activity space are not reliable. Also, all compounds used to train a QSAR should have the same mechanism of interaction with the biological target. A SOM can easily identify if a prediction compound is similar or not with the training set of chemicals. If the prediction compound is mapped in the SOM regions occupied by the training compounds, then the QSAR prediction has a high confidence level. SOM is also used to select from a large dataset only those chemicals that act through a common mechanism or form a cluster of compounds with similar structures. Gini et al. applied such an approach to separate molecules into homogeneous groups of chemicals with similar biological and structural properties [30]. These groups were then used to compute local QSAR models for the aquatic toxicity of chemicals against *Pimephales promelas* and *Tetrahy-mena pyriformis*. A proper validation of a QSAR model gives vital information regarding its prediction error and validity domain. SOM is a good choice to split a dataset into a training set and a validation (prediction) set. Papa et al. applied this method to develop QSAR models for the toxicity of organic chemicals against *Pimephales promelas* [98].

SOM networks are flexible tools to cluster and visualize large chemical libraries. von Korff and Hilpert applied SOM to cluster a library containing 3000 G-protein-coupled receptor (GPCR) ligands for about 130 receptors and 3000 non-GPCR ligands [147]. The chemical structure was described with fingerprints and topological pharmacophores. Cross-validation tests indicate that SOM can separate GPCR ligands from non-GPCR ligands. A possible application of the SAR model is in the design and identification of focused libraries that target a specific GPCR receptor. The supervised self-organizing map (sSOM) is a SOM variant that uses property labels during the training phase. Xiao et al. compared sSOM with *k*-nearest neighbors, PLS and other linear methods in their ability to model a set of 256 inhibitors of dihydrofolate reductase [137]. The predictions tests show that sSOM is insensitive to noise and provides more accurate predictions than linear QSAR methods.

The experimental data obtained in the National Cancer Institute antitumor drug screening program were investigated with SOM to identify clusters of chemicals having the same biological action [103]. SOM clusters high-

816 light four classes of mechanism of action of drugs, namely
817 mitosis, nucleic acid synthesis, membrane transport and
818 integrity, and phosphatase- and kinase-mediated cell cy-
819 cle regulation. SOM and fuzzy C-means clustering were
820 used by Wang et al. to identify marker genes and to classify
821 tumors based on DNA microarray data [132]. The meth-
822 ods were tested with good results for four datasets, namely
823 leukemia, colon cancer, brain tumors, and National Can-
824 cer Institute cancer cell lines. For each dataset the algo-
825 rithm identifies sets of marker genes with predictive power
826 for tumor classification.

827 Counterpropagation Neural Network

828 Hecht-Nielsen developed the counter-propagation neural
829 network (CPNN) as a supervised learning network that
830 combines a self-organizing map that sends signals to an
831 output layer of neurons [39,40,41]. CPNN applications in
832 chemistry and in structure-activity are reviewed by Zupan
833 et al. [146] and by Vracko [129]. Recent CPNN studies
834 in molecular modeling and structure-activity relationships
835 are reviewed in this section.

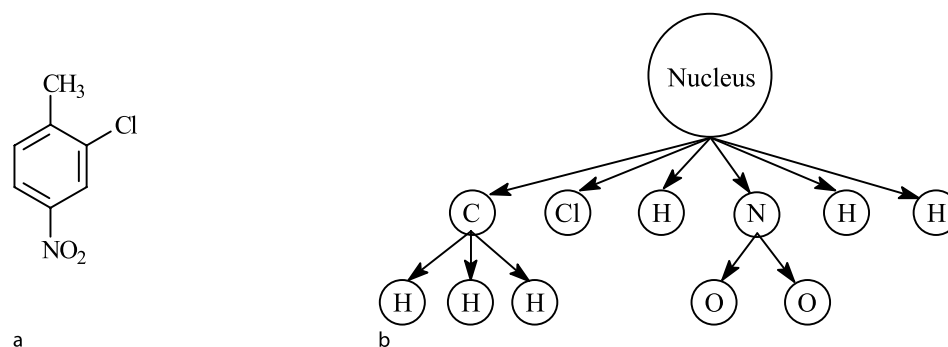
836 Jezierska et al. used CPNN to select an optimum group
837 of descriptors for mutagenicity QSAR models obtained
838 for 95 aromatic and heteroaromatic amines [59]. Start-
839 ing from 275 descriptors, CPNN selected six topologi-
840 cal indices and the LUMO energy as the best combina-
841 tion of input parameters. The reduced set of descriptors
842 give good predictions, with $R_{cv} = 0.85$ obtained in leave-
843 one-out cross-validation. Numerous organic chemicals are
844 environmental pollutants, and a considerable number of
845 studies are dedicated to the computational prediction of
846 their mechanism of aquatic toxicity (MOA). The reli-
847 able prediction of MOA has major applications in se-
848 lecting the appropriate QSAR model, to identify chemi-
849 cals with similar toxicity mechanism, and in extrapolat-
850 ing toxic effects between different species and exposure
851 regimens [53]. Spycher et al. developed MOA classifiers
852 for 115 compounds comprising nine MOA classes and 24
853 descriptors [117]. The four classifiers evaluated, namely
854 CPNN, PLS, logistic regression, and linear discriminant
855 analysis, had an overall correct classification between 52%
856 and 59%, as determined by five-fold cross-validation. Sev-
857 eral MOA classes contain a small number of compounds,
858 which might explain the low prediction rate. In a related
859 study, CPNN and logistic regression classifiers were com-
860 puted for a dataset of 220 phenols with four MOA [118].
861 The rate of correct predictions in five-fold cross-valida-
862 tion was 92%, much higher than in the previous study
863 that considered nine MOA classes. The higher prediction
864 rate may be a result of a larger number of compounds in

each MOA class and of a more homogeneous set of chem- 865
icals. 866

867 CPNN and support vector machines (SVM) were com-
868 pared in a QSAR study for the estrogen receptor binding
869 affinity of 131 compounds [29]. Several methods of feature
870 selection were tested, and the predictions obtained with
871 CPNN and SVM vary significantly with the different sets
872 of selected descriptors. Mazurek et al. investigated CPNN
873 structure-selectivity models for a set of artificial metalloen-
874 zymes [83]. These catalysts perform the enantioselective
875 hydrogenation acetamidoacrylic acid with a high enan-
876 tiomeric excess (%ee). The structural descriptors are com-
877 puted from the geometry of the complexes ligand-metal-
878 loenzyme, and the CPNN is trained to predict the %ee. The
879 predictions obtained for an external test set ($R = 0.953$
880 and $RMS = 16.8\%$ ee) provide convincing evidence for
881 the utility of ligand-metalloenzyme descriptors in mod-
882 eling structure-selectivity relationships. Wagner et al. ap-
883 plied CPNN to structure-activity models for the NF-KB
884 inhibition by 103 sesquiterpene lactones [130]. The activ-
885 ity range for NF-KB inhibition was partitioned in six ac-
886 tivity classes, and the chemical structure was expressed
887 with autocorrelation descriptors based on atomic prop-
888 erties. A ten-fold cross-validation shows that the rate of
889 correct classification is 80.6%, and an external validation
890 with 14 new compounds has 78.6% correct predictions.
891 CPNN and k -nearest neighbors classifiers were applied
892 to the classification of plants from the Asteraceae fam-
893 ily based on a set of sesquiterpene lactones isolated from
894 individual species [44]. The separation of plants in three
895 tribes and seven subtribes is based on structural descrip-
896 tors computed from the molecular structure of sesquiter-
897 pene lactones. This approach can be applied for the au-
898 tomatic identification of plant species based on chemical
899 analysis and machine learning.

900 Graph Machine Neural Networks

901 Usual structure-activity models consider the chemical
902 structure only indirect, through descriptors computed
903 from various representations of molecules. In these mod-
904 els, the topology of neural networks remains constant
905 for all training and prediction molecules. All training
906 molecules, large or small, are represented by a constant
907 length vector of descriptors, thus restricting the use of
908 the chemical structure in models. A different approach
909 is taken in artificial neural networks based on graph ma-
910 chines which learn SAR and QSAR models directly from
911 the molecular graph, by translating the chemical struc-
912 ture into the network topology [32]. In this section we
913 review four graph machine neural networks, namely re-



Drug Design with Artificial Neural Networks, Figure 5

Representation of congeneric chemical compounds in recursive neural networks: a 2-chloro-1-methyl-4-nitrobenzene b represented as a tree

914 cursive neural networks [5,89,90], the Baskin–Palyulin–
915 Zefirov (BPZ) neural device [7], ChemNet [69], and Mol-
916 Net [47,50].

917 Recursive Neural Networks

918 Recursive neural networks (RecNN) map a directed ordered
919 acyclic graph to the real set. All chemical compounds
920 modeled in a RecNN model must be represented as a directed
921 ordered acyclic graph. As an example consider 2-chloro-1-methyl-4-nitrobenzene (Fig. 5a)
922 as a member of series of compounds that have a common
923 skeleton, which in this case is benzene. The molecular
924 structure is transformed into an ordered tree as shown in
925 Fig. 5b. All molecules in the training set are transformed
926 into ordered tree. Using these structured data as input,
927 RecNN reflects the molecular topology as encoded in the
928 ordered tree. The theory and applications of RecNN in
929 SAR and QSAR are presented in several reviews [5,89,90].
930 RecNN were applied to several structure-property models,
931 including models for the glass transition temperature of
932 polymers [23,24], QSAR for the binding affinities of
933 ligands to benzodiazepine receptor [91], and for the
934 Gibbs free energy of solvation in water of organic
935 compounds [8].
936

937 Baskin–Palyulin–Zefirov Neural Device

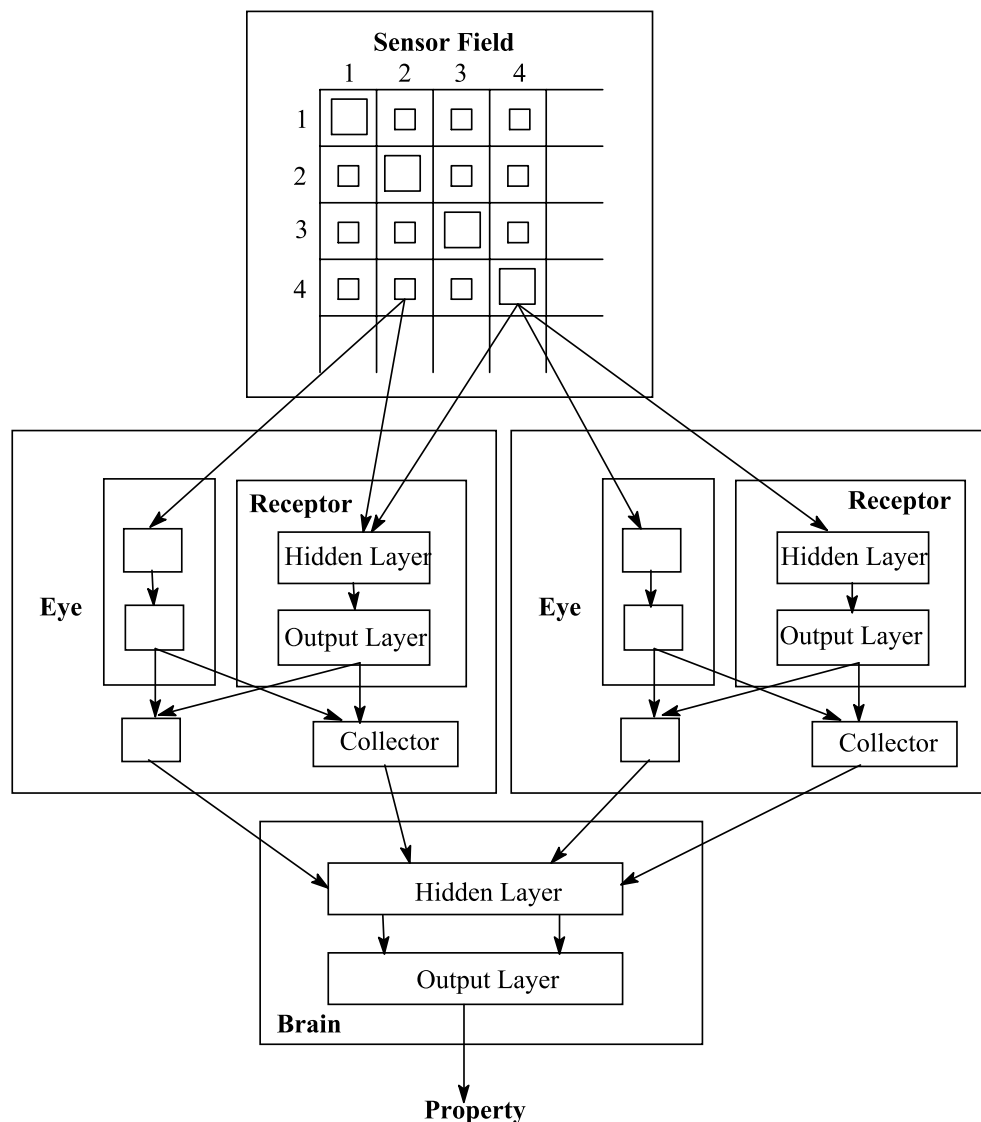
938 The Baskin–Palyulin–Zefirov (BPZ) neural device (Fig. 6)
939 is a computational implementation of a biological vision
940 system and contains a sensor field, a set of eyes, and
941 a brain [7]. A molecule submitted for evaluation to the
942 BPZ neural device is represented as a molecular matrix
943 superimposed over the sensor field. The sensor field
944 receives the network input as a matrix representation of
945 a molecule, and then sends this information to a set of
946 eyes. The structural information is processed in the eyes by several multi-

947 layer feedforward neural networks, and the output signal
948 is sent to the brain. The signals received from the eyes
949 are processed in the brain by an MLF ANN and the output
950 represents the computed property for the molecule sent
951 as input to the sensor field. A set of rules is used to
952 translate a molecule into the structure of the sensor field
953 and of the eyes.

954 A BPZ neural device with a minimal configuration has
955 one receptor in each eye, and can be used as a template
956 to generate a neural device for any molecule presented
957 to the sensor field (Fig. 7). In the minimal configuration,
958 eye 1 receives signals from individual atoms, whereas
959 eye 2 receives from the sensor field signals corresponding
960 to pairs of bonded atoms. The information from the
961 sensor neuron 1 goes to neurons 3 and 4 representing the
962 hidden layer of the receptor R1 from Eye 1. The signal
963 goes then to the output neuron 7 and then to the collector
964 C1 that sends the output to the brain. A similar
965 information flow is similar in Eye 2, with the only
966 difference that the input signal represents information
967 regarding bonds between atoms. Neurons 11 and 12
968 form the hidden layer from the brain, whereas neuron
969 13 is the output neuron that offers the computed
970 molecular property.

970 ChemNet

971 ChemNet is an MLF neural network that translates
972 a molecular structure into the network topology to
973 compute an atomic property [69]. Each molecule is
974 represented by its labeled hydrogen-depleted molecular
975 graph. For each atom i from the molecular graph
976 ChemNet has two corresponding neurons with the same
977 label i , namely one in the input layer and the second
978 one in the hidden layer. The output layer has only
979 one neuron, representing the atom from the molecular
980 graph whose property is computed with ChemNet. The
981 connections between the

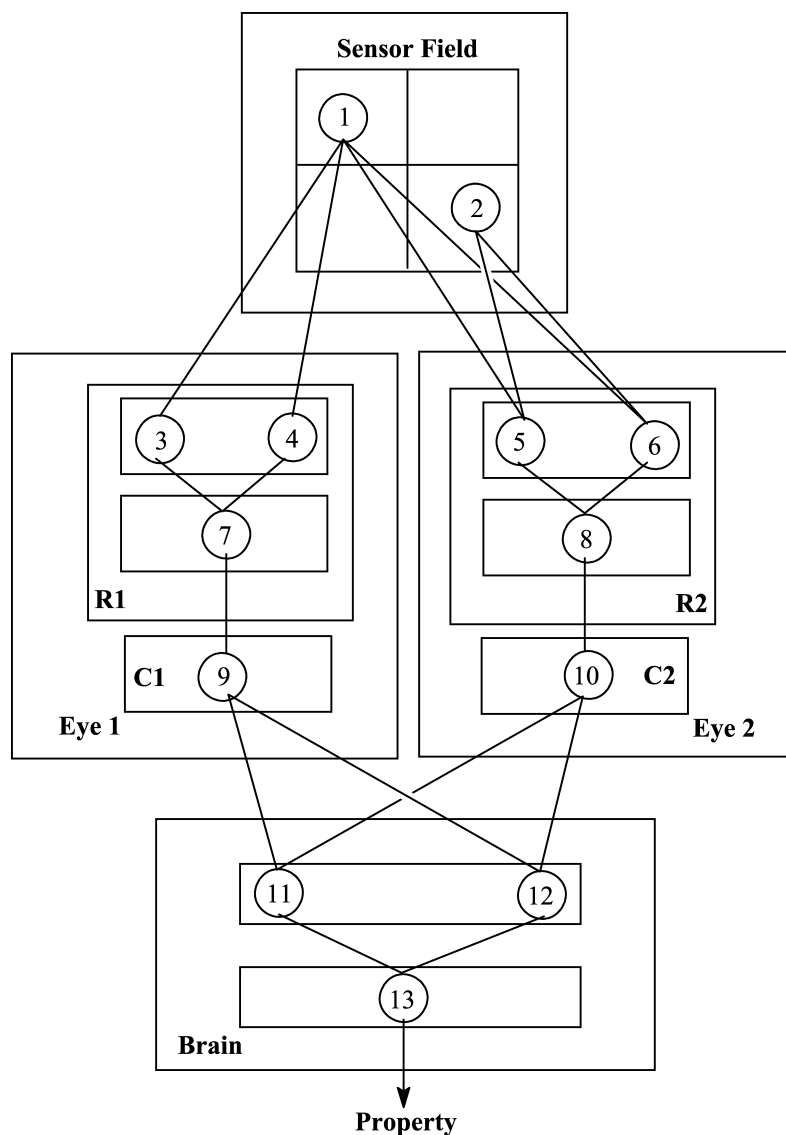


Drug Design with Artificial Neural Networks, Figure 6
The structure of the Baskin-Palyulin-Zefirov neural device

981 input and hidden layers correspond to the bonding rela-
982 tionships between pairs of atoms, i. e. two pairs of atoms
983 separated by the same number of bonds have identical
984 connection weights. The ChemNet structure is demon-
985 strated for 2-methylbutane (Fig. 9a). The distance matrix
986 for the molecular graph of 2-methylbutane is:

$$987 \quad \mathbf{D} = \begin{pmatrix} 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 3 \\ 2 & 1 & 2 & 3 & 0 \end{pmatrix}$$

988 Each neuron from the input and hidden layers of
989 ChemNet corresponds to the atom with the same label
990 from the molecular graph of 2-methylbutane (Fig. 9).
991 Input values represent the number of hydrogen atoms
992 attached to each atom in the molecular graph of 2-methylbu-
993 tane. All pairs of atoms that are separated by the same
994 number of bonds are characterized by input-hidden con-
995 nections with identical weights. The distance matrix de-
996 termines the connections that have the same weight. All
997 ChemNet connections corresponding to pairs of atoms
998 separated by one bond have the same weight (Fig. 9a). The
999 connections between atoms separated by two bonds are

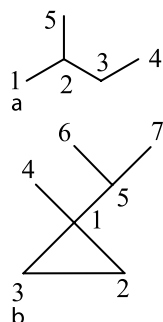


Drug Design with Artificial Neural Networks, Figure 7
A minimal configuration of the Baskin-Palyulin-Zefirov neural device

1000 shown in Fig. 9b, whereas the connections between atoms
1001 separated by three bonds are shown in Fig. 9c. All connec-
1002 tions from the bias neuron to the 5 hidden neurons have
1003 the same weight (Fig. 9d). The ChemNet structure from
1004 Fig. 9 computes a property for atom 1. The connections
1005 between the hidden layer and the output neuron corre-
1006 spond to the bonding relationships between atom 1 and
1007 all other atoms in the molecule, namely atoms situated at
1008 distance 1 (Fig. 9a), distance 2 (Fig. 9b), and distance 3
1009 (Fig. 9c).

MolNet

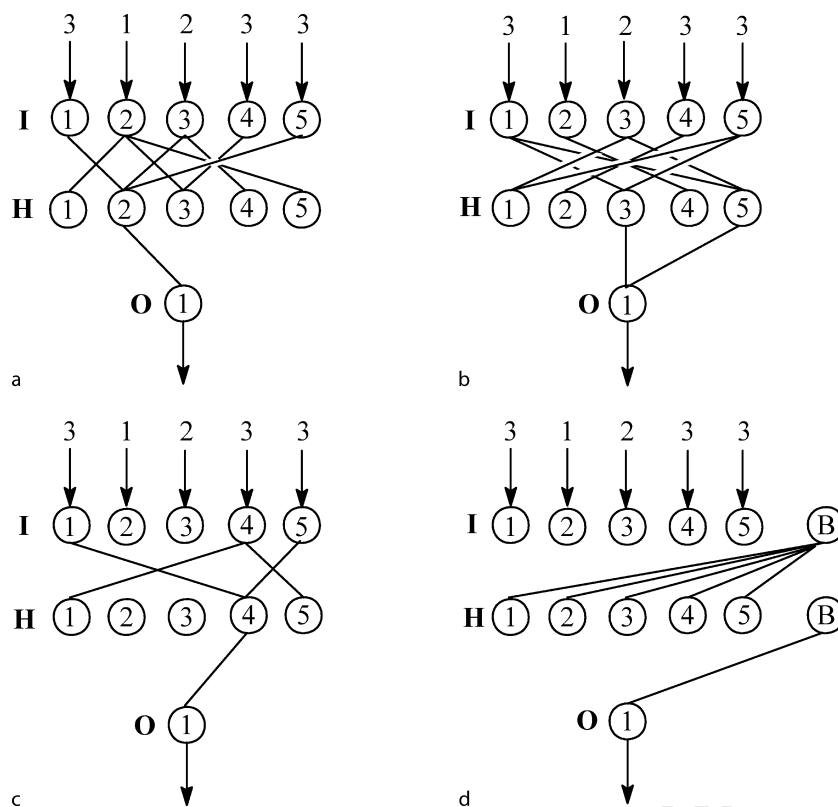
1010 MolNet is a graph machine that maps the distance matrix
1011 in the ANN structure and computes a molecular property
1012 using atomic descriptors as input data [47,48,50,51]. The
1013 number of neurons in the input and hidden layers is equal
1014 to the number of atoms from the molecular graph, and
1015 each atom in a molecule has a corresponding neuron in
1016 the input and hidden layers. The output layer has only one
1017 neuron that provides the computed value for the investi-
1018 gated molecular property. The bonding pattern between
1019 two atoms is defined by the ordered sequence of atom
1020 types and bond types situated on the shortest path between
1021



Drug Design with Artificial Neural Networks, Figure 8
Molecular graphs used to generate ChemNet and MolNet neural networks: a 2-methylbutane; b 1-methyl-1-isopropylcyclopropane

1022 the two atoms. All identical bonding patterns correspond
1023 to network connections with identical weights. An input
1024 neuron i is connected to the hidden neuron i with a con-

1025 nection type that depends on the chemical nature of the
1026 corresponding atom i . All $i - i$ connections correspond-
1027 ing to the same atomic species have identical weights. The
1028 connections from the hidden layer to the output layer (HO
1029 connections) are classified according to the type of atoms
1030 represented by the hidden neurons. The atoms are parti-
1031 tioned into classes according to their atomic number Z ,
1032 the hybridization state and the degree. All hidden neurons
1033 representing atoms of the same type, either in the same
1034 molecule or in different molecules, are connected to the
1035 output neuron by connections with identical weights. The
1036 bias neuron is connected to each hidden neuron with con-
1037 nections (BH connections) partitioned similarly with the
1038 connections between the hidden and output layers, i. e.
1039 according to the atom types as defined above. The bias
1040 neuron is also connected with the output neuron (BO con-
1041 nection). The MolNet structure of 1-methyl-1-isopropyl-
cyclopropane (Fig. 9b) is based the distance matrix com-



Drug Design with Artificial Neural Networks, Figure 9
ChemNet topology for 2-methylbutane (Fig. 8a) that computes a property of atom 1. Each neuron in the input (I) and hidden (H) layers corresponds to the atom with the same label from the molecular graph of 2-methylbutane. The connections between atoms situated at topological distances 1, 2, and 3 are presented in a, b, and c, respectively, whereas the connections from the bias neuron B are presented in d. Input values represent the number of hydrogen atoms attached to each atom in the molecular graph of 2-methylbutane

1042 puted from the molecular graph:

$$1043 \quad \mathbf{D} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 2 & 2 \\ 1 & 0 & 1 & 2 & 2 & 3 & 3 \\ 1 & 1 & 0 & 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 0 & 2 & 3 & 3 \\ 1 & 2 & 2 & 2 & 0 & 1 & 1 \\ 2 & 3 & 3 & 3 & 1 & 0 & 2 \\ 2 & 3 & 3 & 3 & 1 & 2 & 0 \end{pmatrix}$$

1044 The MolNet connections between the input and hidden
1045 layers corresponding to 1-methyl-1-isopropylcyclo-
1046 propane are shown in Fig. 10. The neural network is sep-
1047 arated into four sections, each corresponding to a bond-
1048 ing relationship. Because the molecule considered here has
1049 only carbon atoms and single bonds, all connections depic-
1050 ted in a section from Fig. 10 have identical weights. In-
1051 put values represent vertex degrees, but any atomic index
1052 may be used as input to the network. The network con-
1053 sists of connections between atoms with the same label
1054 (Fig. 10a), and between atoms situated at topological dis-
1055 tances 1 (Fig. 10b), 2 (Fig. 10c), and 3 (Fig. 10d), respec-
1056 tively.

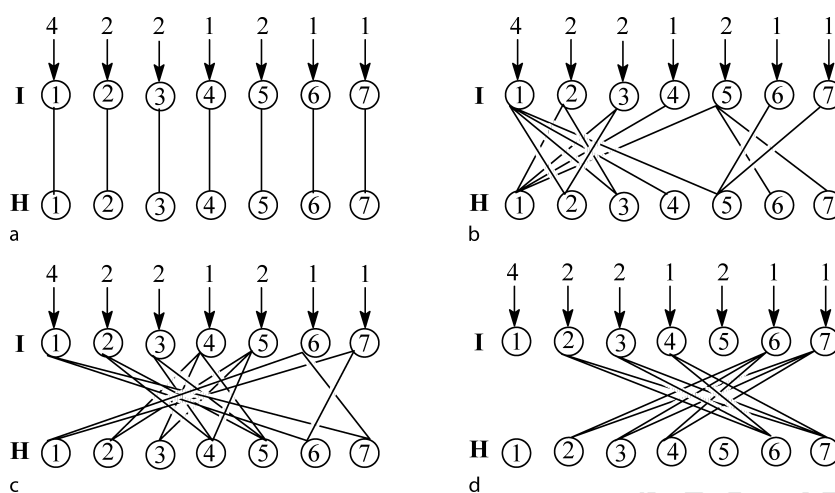
1057 The connections between the hidden and output layers
1058 (HO connections) are determined by the structure of the
1059 molecule presented to MolNet. HO connections in alkanes
1060 are classified according to the degree of the carbon atoms:
1061 hidden neurons representing atoms with identical degrees
1062 are linked to the output neuron with connections hav-
1063 ing identical weights. The connections between the bias

1064 neuron and the neurons in the hidden layer (BH connec-
1065 tions) are classified according to the same rules used for
1066 HO connections. MolNet connections between the hidden
1067 and output layers for 1-methyl-1-isopropylcyclopropane
1068 are presented in Fig. 11.

1069 Other Neural Networks

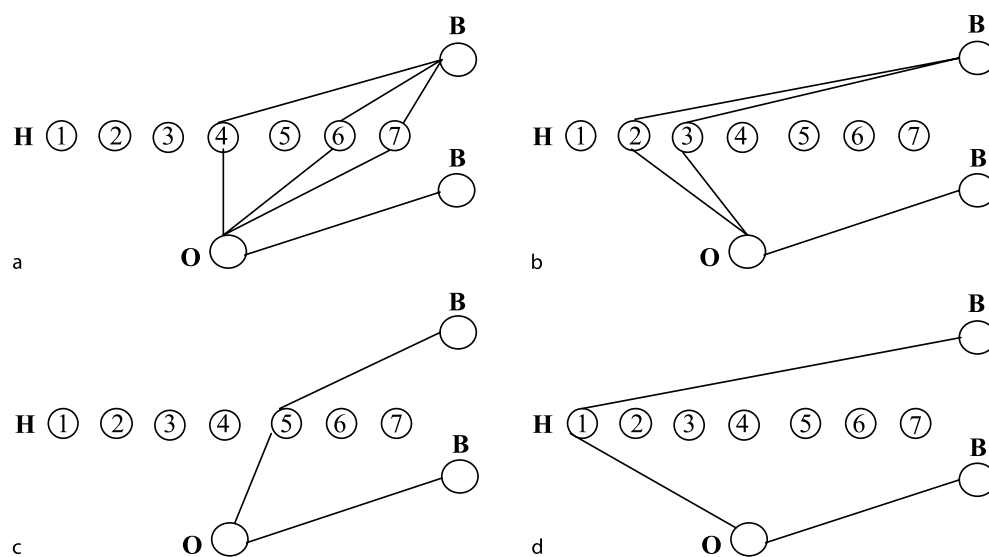
1070 In this section we present an overview of other artifi-
1071 cial neural networks applied in structure-activity models
1072 and in drug design. A Bayesian regularized neural net-
1073 work (BRNN) was evaluated in QSAR for K_m values of
1074 cytochrome P450 3A4 substrates [133]. Correlations with
1075 electrotopological state indices give predictive and robust
1076 models. Bruneau and McElroy applied BRNN to mod-
1077 els for the distribution coefficient $\log D_{7,4}$ computed for
1078 a set of 5000 compounds [14]. The neural model obtained
1079 is stable and can be applied for compounds in different
1080 ionization states. The steady-state volume of distribution
1081 V_{ss} for drugs and drug-like chemicals was computed with
1082 BRNN for a set of 199 compounds (human V_{ss}) and a sec-
1083 ond set of 2086 compounds (rat V_{ss}) [31]. Both models
1084 have good predictions statistics, and can be used to filter
1085 chemicals in the early stages of the drug discovery pro-
1086 cess. Caballero et al. computed a BRNN model for the in-
1087 hibition constant of 46 inhibitors of the cysteine protease
1088 cruzain [16]. The neural network was also used to rank the
1089 importance of the structural descriptors.

1090 Probabilistic neural networks (PNN) and general regression
neural network (GRNN) have similar architec-



Drug Design with Artificial Neural Networks, Figure 10

MolNet connections between the input (I) and hidden (H) layers for 1-methyl-1-isopropylcyclopropane (Fig. 8b); each neuron corresponds to the carbon atom with the same label from the molecular graph of 1-methyl-1-isopropylcyclopropane. The connections between atoms with the same label are presented in a, whereas the connections between atoms situated at topological distances 1, 2, and 3 are presented in b, c, and d, respectively. Input values represent vertex degrees



Drug Design with Artificial Neural Networks, Figure 11

MolNet connections between the hidden (H) and output (O) layers for 1-methyl-1-isopropylcyclopropane (Fig. 8b); the bias neuron is labeled with B. The bias-hidden and hidden-output connections to/from hidden neurons representing atoms with the degree 1, 2, 3, and 4 are presented in a, b, c, and d, respectively

1091 tures, with the difference that PNN perform classification
 1092 whereas GRNN perform regression. Niwa applied PNN to
 1093 identify the biological target from the chemical structure
 1094 of their ligands [96]. The neural network was trained with
 1095 a set of 799 compounds having activities against seven bi-
 1096 ological targets. On average, 90% of the compounds were
 1097 correctly classified. PNN classification models were de-
 1098 veloped for the genotoxic potential of 85 quinolones and
 1099 115 quinolines [38]. The quinolone dataset contains 23
 1100 genotoxic and 62 nongenotoxic compounds, whereas the
 1101 quinoline dataset contains 44 genotoxic and 71 nogeno-
 1102 toxic chemicals. An ensemble of nine PNN models was de-
 1103 veloped for each classification model, and the final class
 1104 attribution (genotoxic/nongenotoxic) was decided by
 1105 a majority vote of the trained classifiers. Simulated an-
 1106 nealing was used to select between three and ten struc-
 1107 tural descriptors for each PNN classifier. The ensemble
 1108 PNN model for quinolone derivatives was able to predict
 1109 correctly 16 of the 23 genotoxic chemicals and 60 of the
 1110 62 nongenotoxic compounds, with an overall accuracy of
 1111 89.4%, an overall accuracy for genotoxic class of 69.6%,
 1112 and an overall accuracy for nongenotoxic class of 96.8%.
 1113 Other structure-activity and drug design applications of
 1114 PNN include the prediction of estrogen receptor ago-
 1115 nists [78], identification of drugs that penetrate the blood-
 1116 brain barrier [79], estimation of the genotoxicity [77], pre-
 1117 diction of P-glycoprotein substrates [140], classification of
 1118 chemicals that are toxic for *Tetrahymena pyriformis* [139],

and identification of factor Xa inhibitors [80]. The general
 regression neural network was applied for the computa-
 tion of the total clearance CL_{tot} of 503 compounds [141].
 The CL_{tot} of a drug characterizes its bioavailability and
 elimination, and thus may be used to determine its dose
 and steady-state concentration. Based on the statistics ob-
 tained for a validation dataset of 105 compounds, the best
 predictions are obtained with the support vector regres-
 sion followed by GRNN. The percent human intestinal
 absorption (%HIA) of 86 drug and drug-like chemicals
 was computed with GRNN and PNN [95]. For an exter-
 nal prediction set, GRNN has a root mean square error
 of 22.8%HIA and PNN has 80% rate of correct classifica-
 tion. The models for human intestinal absorption are
 based on topological indices, thus making the GRNN and
 PNN models strong candidates for screening in large vir-
 tual libraries.

Future Directions

Artificial neural networks have a unique appeal in struc-
 ture-activity predictions because they promise to deliver
 superior modeling power inspired by the functions and
 mechanisms of the human brain. Although the initial ex-
 aggerated expectations never materialized, and artificial
 intelligence is still a faraway dream, the drug design ap-
 plications reviewed here provide compelling evidence for
 the exceptional modeling abilities of artificial neural net-

works. Current applications span all major SAR and QSAR properties relevant to drug design and discovery, and new applications are reported in the literature. Future developments should consider an integrated development of optimally predictive models, in which feature selection, network topology optimization, network training, and validation form a single optimization problem. We anticipate significant improvement in the optimization algorithms used in ANN training, inspired by the remarkable results obtained with particle swarm optimization and ant colony optimization. Novel ANN algorithms and procedures should be adopted from the disciplines of computer science and machine learning. It is expected that the continuous increase in computer power will make feasible applications of ANN ensembles for datasets relevant to the pharmaceutical industry. Significant improvement in SAR and QSAR models may come from graph machine applications that incorporate structured data into the ANN topology.

1164 Bibliography

- 1165 1. Adams CP, Brantner VV (2006) Estimating the cost of new
1166 drug development: is it really 802 million dollars? *Health Aff*
1167 *25*:420–428
- 1168 2. Agrafiotis DK, Cedeño W (2002) Feature selection for struc-
1169 ture-activity correlation using binary particle swarms. *J Med*
1170 *Chem* *45*:1098–1107
- 1171 3. Anzali S, Gasteiger J, Holzgrabe U, Polanski J, Sadowski J,
1172 Teckentrup A, Wagener M (1998) The use of self-organizing
1173 neural networks in drug design. *Perspect Drug Discov Design*
1174 *9*–11:273–299
- 1175 4. Balaban AT, Ivanciuc O (1999) Historical development of
1176 topological indices. In: Devillers J, Balaban AT (eds) *Topologi-
1177 cal Indices and Related Descriptors in QSAR and QSPR*. Gordon
1178 and Breach Science Publishers, Amsterdam, pp 21–57
- 1179 5. Baldi P, Pollastri G (2004) The principled design of large-
1180 scale recursive neural network architectures-DAG-RNNs and
1181 the protein structure prediction problem. *J Mach Learn Res*
1182 *4*:575–602
- 1183 6. Basak SC, Grunwald GD, Host GE, Niemi GJ, Bradbury SP
1184 (1998) A comparative study of molecular similarity, statistical,
1185 and neural methods for predicting toxic modes of action. *En-
1186 viron Toxicol Chem* *17*:1056–1064
- 1187 7. Baskin II, Palyulin VA, Zefirov NS (1997) A neural device for
1188 searching direct correlations between structures and proper-
1189 ties of chemical compounds. *J Chem Inf Comput Sci* *37*:715–
1190 721
- 1191 8. Bernazzani L, Duce C, Micheli A, Mollica V, Sperduti A, Starita
1192 A, Tiné MR (2006) Predicting physical-chemical properties of
1193 compounds from molecular structures by recursive neural
1194 networks. *J Chem Inf Model* *46*:2030–2042
- 1195 9. Berndt ER, Gottschalk AHB, Strobeck MW (2005) Opportu-
1196 nities for improving the drug development process: Results
1197 from a survey of industry and the FDA. National Bureau of
1198 Economic Research Workshop on Innovation Policy and the
1199 Economy, NBER Working Paper No 11425, Washington, DC
- 1200 10. Bishop CM (1996) *Neural Networks for Pattern Recognition*.
1201 Oxford University Press, Oxford, 504 pp
- 1202 11. Boiani M, Cerecetto H, M González, Gasteiger J (2008) Mod-
1203 eling anti-Trypanosoma cruzi activity of *N*-oxide containing
1204 heterocycles, *J Chem Inf Model* *48*:213–219
- 1205 12. Bonchev D (1983) *Information Theoretic Indices for Char-
1206 acterization of Chemical Structure*. Research Studies Press,
1207 Chichester
- 1208 13. Bonchev D, Rouvray DH (eds) (1991) *Chemical Graph The-
1209 ory. Introduction and Fundamentals*. Abacus Press/Gordon &
1210 Breach Science Publishers, New York
- 1211 14. Bruneau P, McElroy NR (2006) $\log D_{7.4}$ Modeling using
1212 Bayesian regularized neural networks. Assessment and cor-
1213 rection of the errors of prediction. *J Chem Inf Model* *46*:1379–
1214 1387
- 1215 15. Bulsari AB (1995) *Neural Networks for Chemical Engineers*. El-
1216 sevier, Amsterdam, 609 pp
- 1217 16. Caballero J, Tundidor-Camba A, Fernández M (2007) Model-
1218 ing of the inhibition constant (K_i) of some cruzain ketone-
1219 based inhibitors using 2D spatial autocorrelation vectors and
1220 data-diverse ensembles of Bayesian-regularized genetic neu-
1221 ral networks. *QSAR Comb Sci* *26*:27–40
- 1222 17. Choi S (2003) Nefazodone (Serzone) withdrawn because of
1223 hepatotoxicity. *Can Med Assoc J* *169*:1187–1187
- 1224 18. Crum-Brown A, Frazer T (1868–1869) On the connection be-
1225 tween chemical constitution and physiological action. Part 1.
1226 On the physiological action of the ammonium bases, derived
1227 from Strychia, Brucia, Thebaia, Codeia, Morphia and Nicotia.
1228 *Trans Royal Soc Edinburgh* *25*:257–274
- 1229 19. Cruz JA, Wishart DS (2006) Applications of machine learning
1230 in cancer prediction and prognosis. *Cancer Inform* *2*:59–78
- 1231 20. DiMasi JA (2002) The value of improving the productivity of
1232 the drug development process: faster times and better deci-
1233 sions. *Pharmacoeconomics* *20*(S3):1–10
- 1234 21. DiMasi JA, Hansen RW, Grabowski HG (2003) The price of in-
1235 novation: new estimates of drug development costs. *J Health*
1236 *Econ* *22*:151–185
- 1237 22. Dieppe PA, Ebrahim S, Martin RM, Jüni P (2004) Lessons from
1238 the withdrawal of rofecoxib. *Br Med J* *329*:867–868
- 1239 23. Duce C, Micheli A, Solaro R, Starita A, Tiné MR (2006) Predic-
1240 tion of chemical-physical properties by neural networks for
1241 structures. *Macromol Symp* *234*:13–19
- 1242 24. Duce C, Micheli A, Starita A, Tiné MR, Solaro R (2006) Predic-
1243 tion of polymer properties from their structure by recursive
1244 neural networks. *Macromol Rapid Commun* *27*:711–715
- 1245 25. Faich GA, Moseley RH (2001) Troglitazone (Rezulin) and hep-
1246 atic injury. *Pharmacoepidemiol Drug Saf* *10*:537–547
- 1247 26. Filter M, Eichler-Mertens M, Bredenbeck A, Losch FO, Sharav
1248 T, Givehchi A, Walden P, Wrede P (2006) A strategy for the
1249 identification of canonical and non-canonical MHC I-binding
1250 epitopes using an ANN-based epitope prediction algorithm.
1251 *QSAR Comb Sci* *25*:350–358
- 1252 27. Fujita T, Iwasa J, Hansch C (1964) A new substituent con-
1253 stant, π , derived from partition coefficients. *J Am Chem Soc*
1254 *86*:5175–5180
- 1255 28. Furberg CD, Pitt B (2001) Withdrawal of cerivastatin from the
1256 world market. *Curr Control Trials Cardivasc Med* *2*:205–207
- 1257 29. Ghafourian T, Cronin MTD (2006) The effect of variable selec-
1258 tion on the non-linear modelling of oestrogen receptor bind-
1259 ing. *QSAR Comb Sci* *25*:824–835

- 1260 30. Gini G, Craciun MV, König C, Benfenati E (2004) Combin- 1321
 1261 ing unsupervised and supervised artificial neural networks to 1322
 1262 predict aquatic toxicity. *J Chem Inf Comput Sci* 44:1897–1902 1323
 1263 31. Gleeson MP, Waters NJ, Paine SW, Davis AM (2006) In silico 1324
 1264 human and rat V_{55} quantitative structure-activity relationship 1325
 1265 models. *J Med Chem* 49:1953–1963 1326
 1266 32. Goulon-Sigwalt-Abram A, Duprat A, Dreyfus G (2005) From 1327
 1267 Hopfield nets to recursive networks to graph machines: Num- 1328
 1268 erical machine learning for structured data. *Theor Comput* 1329
 1269 *Sci* 344:298–334 1330
 1270 33. Hall LH, Story CT (1996) Boiling point and critical tempera- 1331
 1271 ture of a heterogeneous data set: QSAR with atom type elec- 1332
 1272 trotopological state indices using artificial neural networks. 1333
 1273 *J Chem Inf Comput Sci* 36:1004–1014 1334
 1274 34. Hall LH, Story CT (1997) Boiling point of a set of alkanes, 1335
 1275 alcohols and chloroalkanes: QSAR with atom type electro- 1336
 1276 topological state indices using artificial neural networks. *SAR* 1337
 1277 *QSAR Environ Res* 6:139–161 1338
 1278 35. Hansch C (1969) A quantitative approach to biochemical 1339
 1279 structure-activity relationships. *Acc Chem Res* 2:232–239 1340
 1280 36. Hansch C, Fujita T (1964) $\rho - \sigma - \pi$ analysis. A method for 1341
 1281 the correlation of biological activity and chemical structure. 1342
 1282 *J Am Chem Soc* 86:1616–1626 1343
 1283 37. Hansch C, Maloney PP, Fujita T, Muir RM (1962) Correlation of 1344
 1284 biological activity of phenoxyacetic acids with Hammett sub- 1345
 1285 stituent constants and partition coefficients. *Nature* 194:178– 1346
 1286 180 1347
 1287 38. He L, Jurs PC, Kretsoulas C, Custer LL, Durham SK, Pearl GM 1348
 1288 (2005) Probabilistic neural network multiple classifier system 1349
 1289 for predicting the genotoxicity of quinolone and quinoline 1350
 1290 derivatives. *Chem Res Toxicol* 18:428–440 1351
 1291 39. Hecht-Nielsen R (1987) Counterpropagation networks. *Appl* 1352
 1292 *Optics* 26:4979–4984 1353
 1293 40. Hecht-Nielsen R (1988) Applications of counterpropagation 1354
 1294 networks. *Neural Netw* 1:131–139 1355
 1295 41. Hecht-Nielsen R (1990) *Neurocomputing*, Addison-Wesley, 1356
 1296 Reading 1357
 1297 42. Hogan V (2000) Pemoline (Cylert): market withdrawal. *Can* 1358
 1298 *Med Assoc J* 162:106–106 1359
 1299 43. Hopfield JJ (1982) Neural networks and physical systems with 1360
 1300 emergent collective computational abilities. *Proc Natl Acad* 1361
 1301 *Sci USA* 79:2554–2558 1362
 1302 44. Hristozov D, Da Costa FB, Gasteiger J (2007) Sesquiterpene 1363
 1303 lactones-based classification of the family Asteraceae using 1364
 1304 neural networks and k -nearest neighbors. *J Chem Inf Model* 1365
 1305 47:9–19 1366
 1306 45. Huuskonen J (2003) QSAR modeling with the electrotopologi- 1367
 1307 cal state indices: Predicting the toxicity of organic chemicals. 1368
 1308 *Chemosphere* 50:949–953 1369
 1309 46. Huuskonen J, Rantanen J, Livingstone D (2000) Prediction of 1370
 1310 aqueous solubility for a diverse set of organic compounds 1371
 1311 based on atom-type electrotopological state indices. *Eur J* 1372
 1312 *Med Chem* 35:1081–1088 1373
 1313 47. Ivanciuc O (1998) Artificial neural networks applications, Part 1374
 1314 9. MolNet prediction of alkane boiling points. *Rev Roum Chim* 1375
 1315 43:885–894 1376
 1316 48. Ivanciuc O (1999) Artificial neural networks applications. Part 1377
 1317 11. MolNet prediction of alkane densities. *Rev Roum Chim* 1378
 1318 44:619–631 1379
 1319 49. Ivanciuc O (1999) Molecular graph descriptors used in neural 1380
 1320 network models. In: Devillers J, Balaban AT (eds) *Topological* 1381
 Indices and Related Descriptors in QSAR and QSPR. Gordon and Breach Science Publishers, Amsterdam, pp 697–777 1322
 50. Ivanciuc O (1999) The neural network MolNet prediction of 1323
 alkane enthalpies. *Anal Chim Acta* 384:271–284 1324
 51. Ivanciuc O (2000) Artificial neural networks applications. Part 1325
 12. The prediction of alkane heat capacities with the MolNet 1326
 neural network. *Rev Roum Chim* 45:391–403 1327
 52. Ivanciuc O (2001) New neural networks for structure-property 1328
 models. In: Diudea MV (ed) *QSPR/QSAR Studies by Molecular* 1329
Descriptors. Nova Science Publishers, Huntington, pp 213– 1330
 231 1331
 53. Ivanciuc O (2003) Aquatic toxicity prediction for polar and 1332
 nonpolar narcotic pollutants with support vector machines. 1333
Internet Electron J Mol Des 2:195–208 1334
 54. Ivanciuc O (2003) Graph theory in chemistry. In: Gasteiger J 1335
 (ed) *Handbook of Chemoinformatics*, vol 1. Wiley-VCH, Wein- 1336
 heim, pp 103–138 1337
 55. Ivanciuc O (2003) Topological indices. In: Gasteiger J (ed) 1338
Handbook of Chemoinformatics, vol 3. Wiley-VCH, Weinheim, 1339
 pp 981–1003 1340
 56. Ivanciuc O, Rabine J-P, Cabrol DB, Panaye A, Doucet JP (1996) 1341
 ^{13}C NMR chemical shift prediction of sp^2 carbon atoms in 1342
 acyclic alkenes using neural networks. *J Chem Inf Comput Sci* 1343
 36:644–653 1344
 57. Ivanciuc O, Rabine J-P, Cabrol DB, Panaye A, Doucet JP (1997) 1345
 ^{13}C NMR chemical shift prediction of the sp^3 carbon atoms in 1346
 the α position relative to the double bond in acyclic alkenes. 1347
J Chem Inf Comput Sci 37:587–598 1348
 58. Izrailev S, Agrafiotis DK (2002) Variable selection for QSAR by 1349
 artificial ant colony systems. *SAR QSAR Environ Res* 13:417– 1350
 423 1351
 59. Jezierska A, Vračko M, Basak SC (2004) Counter-propaga- 1352
 tion artificial neural networks as a tool for the independent 1353
 variable selection: Structure-mutagenicity study on aromatic 1354
 amines. *Mol Divers* 8:371–377 1355
 60. Jurs P (2003) Quantitative structure-property relationships. 1356
 In: Gasteiger J (ed) *Handbook of Chemoinformatics*, vol 3. 1357
 Wiley-VCH, Weinheim, pp 1314–1335 1358
 61. Kaiser KLE (2003) The use of neural networks in QSARs 1359
 for acute aquatic toxicological endpoints. *J Mol Struct* 1360
 (Theochem) 622:85–95 1361
 62. Katritzky AR, Pacureanu LM, Dobchev DA, Fara DC, Duchow- 1362
 icz PR, Karelson M (2006) QSAR modeling of the inhibition of 1363
 glycogen synthase kinase-3. *Bioorg Med Chem* 14:4987–5002 1364
 63. Keyhani S, Diener-West M, Powe N (2006) Are development 1365
 times for pharmaceuticals increasing or decreasing? *Health* 1366
Aff 25:461–468 1367
 64. Khan J, Wei JS, Ringnér M, Saal LH, Ladanyi M, Westermann 1368
 F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer 1369
 PS (2001) Classification and diagnostic prediction of cancers 1370
 using gene expression profiling and artificial neural networks. 1371
Nature Med 7:673–679 1372
 65. Kier LB, Hall LH (1976) *Molecular Connectivity in Chemistry* 1373
 and Drug Research. Academic Press, New York 1374
 66. Kier LB, Hall LH (1986) *Molecular Connectivity in Structure-* 1375
Activity Analysis. Research Studies Press, Letchworth 1376
 67. Kier LB, Hall LH (1999) *Molecular Structure Description. The* 1377
Electrotopological State. Academic Press, San Diego 1378
 68. Kim H-J, Choo H, Cho YS, Koh HY, No KT, Pae AN (2006) Classi- 1379
 fication of dopamine, serotonin, and dual antagonists by de- 1380
 cision trees. *Bioorg Med Chem* 14:2763–2770 1381

- 1382 69. Kireev DB (1995) ChemNet: A novel neural network based 1442
 1383 method for graph/property mapping. *J Chem Inf Comput Sci* 1443
 1384 35:175–180 1444
- 1385 70. Ko D, Xu W, Windle B (2005) Gene function classification using 1445
 1386 NCI-60 cell line gene expression profiles. *Comput Biol Chem* 1446
 1387 29:412–419 1447
- 1388 71. Kohonen T (1989) *Self-Organization and Associative Memory*, 1448
 1389 3 edn. Springer, Berlin 1449
- 1390 72. Kohonen T (1995) *Self-Organizing Maps*. Springer, Berlin 1450
- 1391 73. Kola I, Landis J (2004) Can the pharmaceutical industry reduce 1451
 1392 attrition rates? *Nat. Rev. Drug Discov* 3:711–715 1452
- 1393 74. Kopp H (1844) Ueber den Zusammenhang zwischen der 1453
 1394 chemischen Constitution und einigen physikalischen Eigen- 1454
 1395 schaften bei flüssigen Verbindungen. *Ann Chem Pharm* 1455
 1396 50:71–144 1456
- 1397 75. Koziol J (2002) Neural network modeling of melting tempera- 1457
 1398 tures for sulfur-containing organic compounds. *Internet Elec- 1458
 1399 tron J Mol Des* 1:80–93 1459
- 1400 76. Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DU, Wolfe 1460
 1401 SN, Bor DH (2002) Timing of new black box warnings and 1461
 1402 withdrawals for prescription medications. *J Am Med Assoc* 1462
 1403 287:2215–2220 1463
- 1404 77. Li H, Ung CY, Yap CW, Xue Y, Li ZR, Cao ZW, Chen YZ (2005) 1464
 1405 Prediction of genotoxicity of chemical compounds by statisti- 1465
 1406 cal learning methods. *Chem Res Toxicol* 18:1071–1080 1466
- 1407 78. Li H, Ung CY, Yap CW, Xue Y, Li ZR, Chen YZ (2006) Prediction 1467
 1408 of estrogen receptor agonists and characterization of asso- 1468
 1409 ciated molecular descriptors by statistical learning methods. 1469
 1410 *J Mol Graph Modell* 25:313–323 1470
- 1411 79. Li H, Yap CW, Ung CY, Xue Y, Cao ZW, Chen YZ (2005) Effect of 1471
 1412 selection of molecular descriptors on the prediction of blood- 1472
 1413 brain barrier penetrating and nonpenetrating agents by statisti- 1473
 1414 cal learning methods. *J Chem Inf Model* 45:1376–1384 1474
- 1415 80. Lin HH, Han LY, Yap CW, Xue Y, Liu XH, Zhu F, Chen YZ (2007) 1475
 1416 Prediction of factor Xa inhibitors by machine learning meth- 1476
 1417 ods. *J Mol Graph Modell* 26:505–518 1477
- 1418 81. Lohninger H (1993) Evaluation of neural networks based on 1478
 1419 radial basis functions and their application to the prediction 1479
 1420 of boiling points from structural parameters. *J Chem Inf Comput 1480
 1421 Sci* 33:736–744 1481
- 1422 82. Lü WJ, Chen YL, Ma WP, Zhang XY, Luan F, Liu MC, Chen XG, 1482
 1423 Hu ZD (2008) QSAR study of neuraminidase inhibitors based 1483
 1424 on heuristic method and radial basis function network. *Eur J 1484
 1425 Med Chem* 43:569–576 1485
- 1426 83. Mazurek S, Ward TR, Novič M (2007) Counter propagation arti- 1486
 1427 ficial neural networks modeling of an enantioselectivity of 1487
 1428 artificial metalloenzymes. *Mol Divers* 11:141–152 1488
- 1429 84. McCulloch WS, Pitts WH (1943) A logical calculus of the ideas 1489
 1430 immanent in nervous activity. *Bull Math Biophys* 7:115–133 1490
- 1431 85. Meissner M, Schmuker M, Schneider G (2006) Optimized Parti- 1491
 1432 cle Swarm Optimization (OPSO) and its application to artificial 1492
 1433 neural network training. *BMC Bioinformatics* 7:125 1493
- 1434 86. Melagraki G, Afantitis A, Makridima K, Sarimveis H, Igglessi- 1494
 1435 Markopoulou O (2006) Prediction of toxicity using a novel RBF 1495
 1436 neural network training methodology. *J Mol Model* 12:297– 1496
 1437 305 1497
- 1438 87. Melagraki G, Afantitis A, Sarimveis H, Igglessi-Markopoulou 1498
 1439 O, Alexandridis A (2006) A novel RBF neural network training 1499
 1440 methodology to predict toxicity to *Vibrio fischeri*. *Mol Divers 1500
 1441 10:213–221* 1501
88. Meyer H (1899) Zur Theorie der Alkoholnarkose, welche 1502
 Eigenschaft der Anästhetica bedingt ihre narkotische 1503
 Wirkung. *Arch Exp Pathol Pharmacol* 42:109–118 1504
89. Micheli A, Portera F, Sperduti A (2005) A preliminary empiri- 1505
 cal comparison of recursive neural networks and tree kernel 1506
 methods on regression tasks for tree structured domains. 1507
Neurocomputing 64:73–92 1508
90. Micheli A, Sperduti A, Starita A (2007) An introduction to re- 1509
 cursive neural networks and kernel methods for cheminform- 1510
 atics. *Curr Pharm Design* 13:1469–1495 1511
91. Micheli A, Sperduti A, Starita A, Bianucci AM (2001) Analysis 1512
 of the internal representations developed by neural networks 1513
 for structures applied to quantitative structure-activity rela- 1514
 tionship studies of benzodiazepines. *J Chem Inf Comput Sci 1515
 41:202–218* 1516
92. Minsky ML, Papert SA (1969) *Perceptrons*. MIT Press, Cam- 1517
 bridge 1518
93. Moody J, Darken CJ (1989) Fast learning in networks of lo- 1519
 cally-tuned processing units. *Neural Comput* 1:281–294 1520
94. Muresan S, Sadowski J (2005) “In-house likeness”: Compari- 1521
 son of large compound collections using artificial neural net- 1522
 works. *J Chem Inf Model* 45:888–893 1523
95. Niwa T (2003) Using general regression and probabilistic neu- 1524
 ral networks to predict human intestinal absorption with 1525
 topological descriptors derived from two-dimensional chemi- 1526
 cal structures. *J Chem Inf Comput Sci* 43:113–119 1527
96. Niwa T (2004) Prediction of biological targets using probabi- 1528
 listic neural networks and atom-type descriptors. *J Med 1529
 Chem* 47:2645–2650 1530
97. Overton CE (1901) *Studien über die Narkose. Zugleich ein 1531
 Beitrag zur Allgemeinen Pharmakologie*. Gustav Fisher Ver- 1532
 lag, Jena 1533
98. Papa E, Villa F, Gramatica P (2005) Statistically validated 1534
 QSARs, based on theoretical descriptors, for modeling 1535
 aquatic toxicity of organic chemicals in *Pimephales promelas* 1536
 (fathead minnow). *J Chem Inf Model* 45:1256–1266 1537
99. Patankar SJ, Jurs PC (2002) Prediction of glycine/NMDA 1538
 receptor antagonist inhibition from molecular structure. 1539
J Chem Inf Comput Sci 42:1053–1068 1540
100. Patankar SJ, Jurs PC (2003) Classification of inhibitors of 1541
 protein tyrosine phosphatase 1B using molecular structure 1542
 based descriptors. *J Chem Inf Comput Sci* 43:885–899 1543
101. Pitts WH, McCulloch WS (1947) How we know universals: The 1544
 perception of auditory and visual forms. *Bull Math Biophys 1545
 9:127–147* 1546
102. Plewczynski D, Spieser SAH, Koch U (2006) Assessing different 1547
 classification methods for virtual screening. *J Chem Inf Model 1548
 46:1098–1106* 1549
103. Rabow AA, Shoemaker RH, Sausville EA, Covell DG (2002) Min- 1550
 ing the National Cancer Institute’s tumor-screening database: 1551
 Identification of compounds with similar cellular activities. 1552
J Med Chem 45:818–840 1553
104. Ripley BD (2008) *Pattern Recognition and Neural Networks*. 1554
 Cambridge University Press, Cambridge, 416 pp 1555
105. Roche O, Trube G, Zuegge J, Pflimlin P, Alanine A, Schneider G 1556
 (2002) A virtual screening method for prediction of the hERG 1557
 potassium channel liability of compound libraries. *Chem Bio 1558
 Chem* 3:455–459 1559
106. Rosenblatt F (1962) *Principles of Neurodynamics: Percep- 1560
 trons and the Theory of Brain Mechanisms*. Spartan Books, 1561
 Washington 1562

- 1503 107. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323:533–536
- 1504 108. Rumelhart DE, McClelland JL (eds) (1986) *Parallel Distributed Processing*. MIT Press, Cambridge, 344 pp
- 1505 109. Saxena AK, Schaper K-J (2006) QSAR analysis of the time- and dose-dependent anti-inflammatory in vivo activity of substituted imidazo[1:2-*a*]pyridines using artificial neural networks. *QSAR Comb Sci* 25:590–597
- 1506 110. Seierstad M, Agrafiotis DK (2006) A QSAR model of hERG binding using a large, diverse, and internally consistent training set. *Chem Biol Drug Des* 67:284–296
- 1507 111. Selaru FM, Xu Y, Yin J, Zou T, Liu TC, Mori Y, Abraham JM, Sato F, Wang S, Twigg C, Oлару A, Shustova V, Leytin A, Hytirogrou P, Shibata D, Harpaz N, Meltzer SJ (2002) Artificial neural networks distinguish among subtypes of neoplastic colorectal lesions. *Gastroenterology* 122:606–613
- 1508 112. Shen Q, Jiang J-H, Jiao C-X, Lin W-Q, Shen G-L, Yu R-Q (2004) Hybridized particle swarm algorithm for adaptive structure training of multilayer feed-forward neural network: QSAR studies of bioactivity of organic compounds. *J Comput Chem* 25:1726–1735
- 1509 113. Shen Q, Shi W-M, Yang X-P, Ye B-X (2006) Particle swarm algorithm trained neural network for QSAR studies of inhibitors of platelet-derived growth factor receptor phosphorylation. *Eur J Pharm Sci* 28:369–376
- 1510 114. Siu F-M, Che C-M (2006) Quantitative structure-activity (affinity) relationship (QSAR) study on protonation and cationization of α -amino acids. *J Phys Chem A* 110:12348–12354
- 1511 115. So S-S, Karplus M (1996) Evolutionary optimization in quantitative structure-activity relationship: An application of genetic neural networks. *J Med Chem* 39:1521–1530
- 1512 116. So S-S, Karplus M (1996) Genetic neural networks for quantitative structure-activity relationships: Improvements and application of benzodiazepine affinity for benzodiazepine/GABA_A receptors. *J Med Chem* 39:5246–5256
- 1513 117. Spycher S, Nendza M, Gasteiger J (2004) Comparison of different classification methods applied to a mode of toxic action data set. *QSAR Comb Sci* 23:779–791
- 1514 118. Spycher S, Pellegrini E, Gasteiger J (2005) Use of structure descriptors to discriminate between modes of toxic action of phenols. *J Chem Inf Model* 45:200–208
- 1515 119. Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S (2005) A comprehensive evaluation of multiclass classification methods for microarray gene expression cancer diagnosis. *Bioinformatics* 21:631–643
- 1516 120. Sutherland JJ, O'Brien LA, Weaver DF (2004) A comparison of methods for modeling quantitative structure-activity relationships. *J Med Chem* 47:5541–5554
- 1517 121. Taskinen J, Yliruusi J (2003) Prediction of physicochemical properties based on neural network modelling. *Adv Drug Deliv Rev* 55:1163–1183
- 1518 122. Terfloth L, Gasteiger J (2001) Neural networks and genetic algorithms in drug design. *Drug Discov Today* 6:S102–S108
- 1519 123. Tetko IV, Tanchuk VY, Villa AEP (2001) Prediction of *n*-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *J Chem Inf Comput Sci* 41:1407–1421
- 1520 124. Todeschini R, Consonni V (2003) Descriptors from molecular geometry. In: Gasteiger J (ed) *Handbook of Chemoinformatics*, vol 3. Wiley-VCH, Weinheim, pp 1004–1033
- 1521 125. Trinajstić N (1992) *Chemical Graph Theory*. CRC Press, Boca Raton
- 1522 126. Varnek A, Kireeva N, Tetko IV, Baskin II, Solov'ev VP (2007) Exhaustive QSPR studies of a large diverse set of ionic liquids: How accurately can we predict melting points? *J Chem Inf Model* 47:1111–1122
- 1523 127. Votano JR, Parham M, Hall LH, Kier LB, Oloff S, Tropsha A, Xie Q, Tong W (2004) Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis* 19:365–377
- 1524 128. Votano JR, Parham M, Hall LM, Hall LH, Kier LB, Oloff S, Tropsha A (2006) QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation. *J Med Chem* 49:7169–7181
- 1525 129. Vracko M (2005) Kohonen artificial neural network and counter propagation neural network in molecular structure-toxicity studies. *Curr Comput-Aided Drug Des* 1:73–78
- 1526 130. Wagner S, Hofmann A, Siedle B, Terfloth L, Merfort I, Gasteiger J (2006) Development of a structural model for NF- κ B inhibition of sesquiterpene lactones using self-organizing neural networks. *J Med Chem* 49:2241–2252
- 1527 131. Wan C, Harrington PB (1999) Self-configuring radial basis function neural networks for chemical pattern recognition. *J Chem Inf Comput Sci* 39:1049–1056
- 1528 132. Wang J, BøTH JI, Myklebost O, Hovig E (2003) Tumor classification and marker gene prediction by feature selection and fuzzy c-means clustering using microarray data. *BMC Bioinformatics* 4:60
- 1529 133. Wang Y-H, Li Y, Li Y-H, Yang S-L, Yang L (2005) Modeling K_m values using electrotopological state: Substrates for cytochrome P450 3A4-mediated metabolism. *Bioorg Med Chem Lett* 15:4076–4084
- 1530 134. Wasserman PD (1989) *Neural Computing*. Van Nostrand Reinhold, New York, 230 pp
- 1531 135. Weinstein JN, Kohn KW, Grever MR, Viswanadhan VN, Rubinstein LV, Monks AP, Scudiero DA, Welch L, Koutsoukos AD, Chiaus A, Paull KD (1992) Neural computing in cancer drug development: Predicting mechanism of action. *Science* 258:447–451
- 1532 136. Wessel MD, Jurs PC, Tolan JW, Muskal SM (1998) Prediction of human intestinal absorption of drug compounds from molecular structure. *J Chem Inf Comput Sci* 38:726–735
- 1533 137. Xiao Y-D, Clauset A, Harris R, Bayram E, Santago P, Schmitt JD (2005) Supervised self-organizing maps in drug discovery, 1. Robust behavior with overdetermined data sets. *J Chem Inf Model* 45:1749–1758
- 1534 138. Xu Y, Selaru FM, Yin J, Zou TT, Shustova V, Mori Y, Sato F, Liu TC, Oлару A, Wang S, Kimos MC, Perry K, Desai K, Greenwald BD, Krasna MJ, Shibata D, Abraham JM, Meltzer SJ (2002) Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. *Cancer Res* 62:3493–3497
- 1535 139. Xue Y, Li H, Ung CY, Yap CW, Chen YZ (2006) Classification of a diverse set of *Tetrahymena pyriformis* toxicity chemical compounds from molecular descriptors by statistical learning methods. *Chem Res Toxicol* 19:1030–1039
- 1536 140. Xue Y, Yap CW, Sun LZ, Cao ZW, Wang JF, Chen YZ (2004) Prediction of P-glycoprotein substrates by a support vector machine approach. *J Chem Inf Comput Sci* 44:1497–1505
- 1537 141. Yap CW, Li ZR, Chen YZ (2006) Quantitative structure-pharmacokinetic relationships for drug clearance by using statistical learning methods. *J Mol Graph Modell* 24:383–395

- 1625 142. Zheng G, Xiao M, Lu XH (2005) QSAR study on the Ah receptor-binding affinities of polyhalogenated dibenzo-*p*-dioxins
1626 using net atomic-charge descriptors and a radial basis neural
1627 network. *Anal Bioanal Chem* 383:810–816
1628
- 1629 143. Zhou Y-P, Jiang J-H, Lin W-Q, Zou H-Y, Wu H-L, Shen G-L, Yu
1630 R-Q (2006) Adaptive configuring of radial basis function net-
1631 work by hybrid particle swarm algorithm for QSAR studies of
1632 organic compounds. *J Chem Inf Model* 46:2494–2501
- 1633 144. Zupan J (2003) Neural networks. In: Gasteiger J (ed) *Hand-*
1634 *book of Chemoinformatics*, vol 3. Wiley-VCH, Weinheim,
1635 pp 1167–1215
- 1636 145. Zupan J, Gasteiger J (1999) *Neural Networks in Chemistry and*
1637 *Drug Design*. Wiley-VCH, Weinheim
- 1638 146. Zupan J, Novič M, Gasteiger J (1995) Neural networks with
1639 counter-propagation learning-strategy used for modeling.
1640 *Chemometrics Intell Lab Syst* 27:175–187
- 1641 147. von Korff M, Hilpert K (2006) Assessing the predictive power
1642 of unsupervised visualization techniques to improve the
1643 identification of GPCR-focused compound libraries. *J Chem*
1644 *Inf Model* 46:1580–1587

Uncorrected Proof
2008-07-04