Designing Molecules with Specific Properties from Intercommunicating Hybrid Systems†

James Devillers

CTIS, 21 rue de la Bannière, 69003 Lyon, France

Received February 22, 1996⊗

The concept that computers can be creative in designing new molecules is now being actively explored. Our work has been aimed at developing an intercommunicating hybrid system using a genetic algorithm and a backpropagation neural network model for solving the general problem of designing molecules with specific properties. A case study dealing with biodegradation modeling is presented. The usefulness of the intercommunicating hybrid systems in QSAR and drug design is emphasized.

1. INTRODUCTION

Humans are hybrid information processing devices. Indeed, our behaviors and actions are governed by a combination of genetic information encoded in our chromosomes and knowledge acquired through a complex learning process. This type of hybrid information processing is now tentatively replicated in computers. Thus, for example, artificial neural networks (ANNs) are inspired by the functionality of biological neurons in the brain.[1] Genetic algorithms[2] (GAs) are also biologically inspired and based on the principles of the Darwinian evolution. These techniques constitute valuable tools for solving specific tasks in drug design and in (Quantitative) Structure−Activity Relationship ((Q)SAR) and Quantitative Structure−Property Relationship (QSPR) studies.[3,4] However, it is now well admitted that certain complex problems in molecular modeling cannot be solved by a single technique alone. Each approach presents particular properties that make them suited for specific applications and not for others. Thus, for example, ANNs are powerful tools for recognizing patterns, but they present some difficulties for explaining how they reach their decisions. Fuzzy logic systems,[5,6] which exhibit a wide flexibility for handling imprecise "linguistic" concepts, are good at explaining their decisions, but they cannot automatically acquire the rules they use to make those decisions.[7] GAs are very efficient for solving complex optimization problems; however, there are many parameters to be tuned and often considerable experimentation is needed before obtaining the optimal solution(s). The strengths and weaknesses of these different intelligent[8] techniques have been the central driving force behind the creation of hybrid systems where two or more techniques are combined in order to overcome the limitations of each approach and create a synergy between them in order to provide valuable modeling tools.

According to Goonatilake and Khebbal,[7] hybrid systems can be differentiated on the basis of their functionality. In function-replacing hybrids, a principal function of a given technique is replaced by another intelligent technique. Here, the aim is to take an approach presenting weaknesses in a particular property and combine it with a technique that has strengths in the same property. Thus, for example, GAs have been shown to be very effective at function optimization, efficiently searching large and complex spaces to find nearly global optima. Under these conditions, GAs have been largely employed for designing and/or training ANNs.[9−14] The second class of hybrid systems corresponds to the polymorphic hybrids which use a single processing architecture to achieve the functionality of different intelligent processing techniques. The cellular encoding methodology[15] which allows synthesis of ANNs belongs to this category of hybrid systems. The last class of hybrid systems corresponds to the intercommunicating hybrids which are independent, self-contained, processing modules that exchange information and perform separate functions in order to generate optimal solutions.[7,16] Thus, if a problem can be subdivided into distinct processing tasks, then different independent modules can be used to solve the parts of the problem at which they are the best. They work in synergy for performing a particular task. The different modules can process subtasks either in a sequential manner or in a competitive/cooperative framework. Intercommunicating hybrids, presenting different levels of complexity, are penetrating the field of QSAR and drug design.[17−19] Therefore, the aim of this paper is to present the usefulness of this type of hybrid system through a practical example dealing with the design of molecules with a specific biodegradability.

2. AN INTERCOMMUNICATING HYBRID SYSTEM FOR MODELING BIODEGRADATION

2.1. Description. Sometimes, a slight modification in the structure of an organic compound can considerably change its biological activity. It is particularly true for complex biological activities such as biodegradation. The knowledge of these (Quantitative) Structure−Biodegradability Relationships ((Q)SBRs) is economically very important for optimizing the design and synthesis of new molecules. Recently, the backpropagation neural network (BNN) has been presented as a powerful nonlinear statistical tool for modeling biodegradation due to its ability to learn and generalize from complex and noisy biodegradation data.[20,21] The powerful modeling potency of a BNN can be used in conjunction with the optimization ability of a GA in order to produce a hybrid system able to automatically propose molecular structures for the design of molecules presenting various degrees of biodegradability. In this context, an intercommunicating hybrid system was designed for modeling biodegradation. It is basically constituted of

† Invited paper presented in the Symposium "Frontiers in Mathematical Chemistry. Novel Directions in QSAR" (M. Randić and P. G. Mezey, Presiding) at PACIFICHEM 1995 (December 17−22, 1995, Honolulu, Hawaii).

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1996.

S0095-2338(96)00022-4 CCC: $12.00

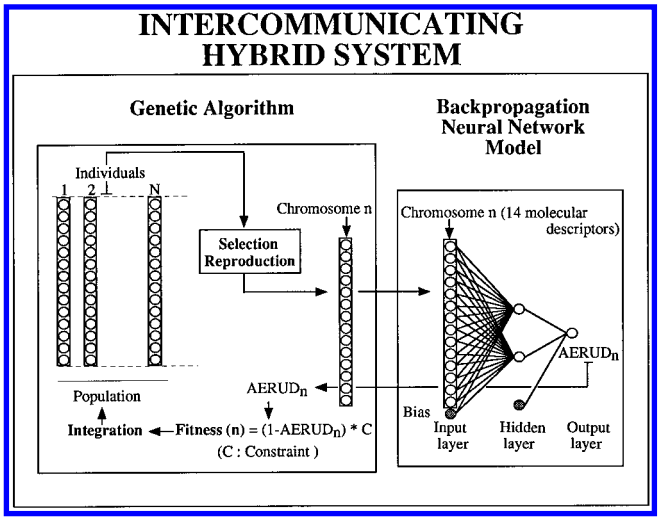

Figure 1. Intercommunicating hybrid system for designing biodegradable molecules.

a GA for searching combinations of structural fragments among a set of descriptors in order to facilitate the design of biodegradable molecules. The biodegradability of the candidate molecules is estimated from a three-layer BNN model. The elaboration of this system has been described in a recent paper.[17] The hybrid system is summarized in Figure 1, and the key steps in the design of the BNN model and GA are recalled below.

The BNN model was derived from a training set of 38 molecules presenting a high degree of structural heterogeneity. The aerobic ultimate degradation in receiving waters (AERUD) of these molecules was used as endpoint to estimate their biodegradability.[22] The generalization ability of the BNN was estimated from a testing set of 49 molecules. Examples of molecules belonging to the training and testing sets are displayed in Figure 2. The molecules were described by means of 13 structural descriptors (Table 1) known to influence the environmental fate of chemicals.[20,21,23−28] They were Boolean descriptors. It is obvious that other descriptors could have been selected since numerous structural features influencing the biodegradation of organic molecules have been identified in the literature.[23,24] However, a trial and error procedure revealed that this set of descriptors (Table 1) offered the best compromise between an optimal description of our training set and valuable generalization perspectives with the BNN model. Last, they were sufficiently general to describe a wide variety of structures as required for GA experiments. In addition to the 13 Boolean descriptors, four classes of molecular weight were defined (Table 1) not only to account for size effects but also to allow the description of all the molecules. Molecular descriptors (except descriptor no. 10) and the AERUD values were scaled by means of eq (1).

$$x'_i = [a(x_i - x_{\min})/(x_{\max} - x_{\min})] + b \qquad (1)$$

In eq (1), x'_i and x_i were the scaled and original values, respectively. For the molecular descriptors, $x_{\min}$ and $x_{\max}$ were the minimum and maximum values found in the different columns, respectively. For descriptor no. 10, the four classes were encoded 0.2, 0.4, 0.6, and 0.8, respectively. For the outputs, $x_{\min}$ and $x_{\max}$ corresponded to the minimum and maximum AERUD values (i.e., 1 and 4) which could be calculated from the procedure proposed by Boethling *et al.*[22] In order to have scaled data ranging between 0.05 and 0.95, the values of a and b equaled 0.9 and 0.05, respectively. Numerous trials were performed with the STATQSAR package[29] to determine the optimal set of parameters for the BNN (i.e., number of hidden neurons, learning rate (η), momentum term (α), number of learning cycles). With a 14/2/1 BNN ($\alpha = 0.9$ and $\eta = 0.4$), we succeeded in correctly modeling all the chemicals belonging to the training set. Indeed, all the residuals were inferior to 0.3 (absolute value). The convergence was obtained within 2500 cycles. The selected BNN model was able to correctly predict the biodegradability of 91.8% (i.e., 45/49) of the test chemicals. The numerous runs performed in our study have shown that our results were highly reproducible. Due to the structural heterogeneity of the chemicals belonging to the training and testing sets, we assume that the selected BNN model can be used to estimate the AERUD values corresponding to the candidate molecules generated by the GA.

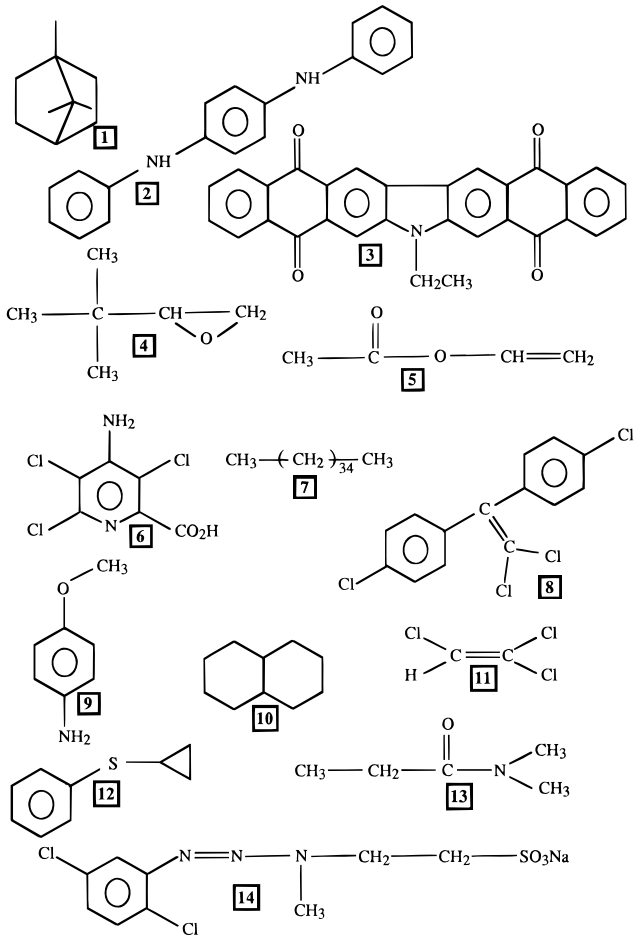

Figure 2. Structure of some molecules included in the training (no. 1−7) and the testing (no. 8−14) sets.

Table 1. Molecular Descriptors

no.	descriptor	no.	descriptor
1	heterocycle N	10	molecular weight (g·mol^{-1})
2	C=O		<100
3	>2 chlorine atoms		100−200
4	fused rings		200−300
5	only C, H, O, N		≥300
6	NO_2	11	C−O: ether + OH II and III
7	≥2 cycles	12	amines + X=N (X = C or N)
8	epoxide	13	conjugated C=O
9	primary or aromatic OH	14	Z=Z except NO_2 (Z = any heteroatom)

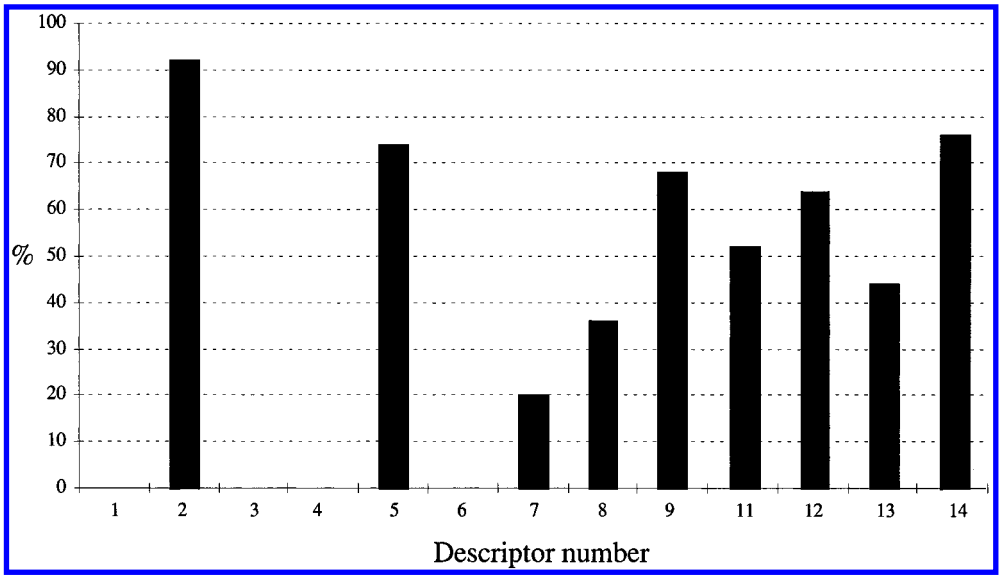

Figure 3. Influence of the 13 Boolean descriptors on the biodegradability of the molecules. See Table 1 for correspondence between the numbers and the descriptors. Due to the fact that the molecular weight (descriptor no. 10 in Table 1) was encoded by means of four classes, it was not represented on this figure.

The basic requirements for applying a GA are that a possible solution to the problem can be encoded and that a given solution can be evaluated quantitatively by means of a fitness function. Initially, a random population of individuals is created. Each individual is represented by a chromosome constituted of genes encoding the studied information. Evaluation of the fitness of each individual allows the selection of candidates for mating. Thus, if the fitness of an offspring is sufficiently high, it replaces a less fit member of the population and becomes a parent. The manipulation of the chromosomes by means of crossovers and mutations allows to increase the overall fitness of the population and therefore to obtain an optimal solution for the problem at hand. In our GA, each chromosome consisted of 14 genes corresponding to the selected molecular descriptors (Table 1). The fitness function was based on the estimation of an AERUD value. However, the objective of the search was to minimize the AERUD values instead of maximizing them since our goal was to select structural fragments increasing the biodegradability of the molecules. As a result, it was necessary to perform a mathematical transformation of the fitness function (see ref 2, p 76). It is obvious that due to the different nature of the molecular descriptors, the reproduction−selection process of the GA could lead to impossible combinations. To overcome this problem, different constraints were added by means of a penalty method.[2] Thus, a penalty was assigned to the fitness for designed molecules which violated the defined constraints. As a result, they were removed from the mating population. For illustrative purposes, two examples of constraints are given below.

−When a molecule contains more than two chlorine atoms (descriptor no. 3), it cannot be only constituted of C, H, O, and N (descriptor no. 5). Thus, if #3 = 1 => #5 = 0 and if #5 = 1 => #3 = 0.

−When a molecule contains fused rings (descriptor no. 4), it obligatorily contains at least two cycles (descriptor no. 7). Thus, if #4 = 1 => #7 = 1 (the converse is not true).

The GA analysis was performed with the STATQSAR package.[29] The GA configuration was obtained by a trial and error procedure. We found that using a population of 100 was a reasonable compromise between having enough sampling over combinatorial space and not taking too much computer time. The crossover (Pcross) and mutation (Pmut) probabilities equaled 0.8 and 0.1, respectively. A 50% population permutation was selected (elitist strategy). Selection was optimized from the prescale, scale, and scalepop routines of Goldberg[2] (p 79). Last, the GA was stopped after 300 generations.

2.2. Analysis of the Results. With our intercommunicating hybrid system (Figure 1), it is possible to propose candidate molecules having low AERUD values. These molecules present particular structures (Figure 3). Thus, on the basis of numerous runs, among the best 50 candidates, it is interesting to note, for example, that biodegradable molecules preferentially contain a C═O group (descriptor no. 2) and/or a primary or aromatic OH (descriptor no. 9). At the opposite, these molecules do not include a nitrogen heterocycle (descriptor no. 1) and/or more than two chlorine atoms (descriptor no. 3) (Figure 3). It must be pointed out that these results are not surprising. Indeed, for example, it is well-known that the presence of OH, CHO, and COOH groups on an aromatic ring facilitates biodegradation, as compared with the NO_2 group and halogens. In addition, the rate of biodegradation usually decreases with an increasing number of substituents on the aromatic ring.[30] Our results clearly show that the proposed intercommunicating hybrid system (Figure 1) can be used for designing molecules with a particular biodegradability since it always proposes logical solutions. It is interesting to note that during the simulations, the hybrid system regularly proposes candidate molecules having the same coding as biodegradable molecules belonging to the training and testing sets.

In the chemical industry, the synthesis of new molecules is often under the dependence of structural constraints. Therefore, it appeared interesting to test the limits of the intercommunicating hybrid system (Figure 1) when structural conditions are voluntarily introduced. In practice, this type of constraints requires to firstly modify the initialization process of the population since the selected descriptor(s) is

Table 2. Number of Biodegradable Candidates Proposed by the GA under Constraints

constrained descriptor	no. of candidates with AERUD <2.5	constrained descriptor	no. of candidates with AERUD <2.5
1, 3, 4, 6, or 7	50/50	3, 6	1/50
2, 5, 9, 12, or 14	50/50	4, 6	5/50
8, 11, or 13	50/50[a]	1, 2	13/50
1, 3	0/50	2, 3	9/50
1, 4	1/50	2, 4	16/50
1, 6	1/50	2, 6	21/50
3, 4	1/50		

[a] Same results are obtained when the presence or absence of these descriptors is imposed.

Table 3. Variation of GA Parameters[a]

parameter	min.	max.	increment	no. of runs
population	10	100	10	500
generation	50	500	50	500
crossover	0.1	0.9	0.1	450
mutation	0.01	0.09	0.01	450
	0.1	0.9	0.1	450

[a] The parameters were varied one at a time. During this process, the values of the other GA parameters were those selected in our hybrid system.

(are) not randomly chosen. In addition, it is also necessary to control the operations of mutations and crossovers in order to avoid the loss of the selected descriptor(s).

In a first series of experiments, the presence of structural fragments decreasing the biodegradability of the molecules (i.e., 1, 3, 4, 6, 7) was imposed. The selection of one of these fragments in the molecule cannot prevent the design of biodegradable molecules by the hybrid system. Indeed, in all cases, it is possible to propose candidate molecules having AERUD values <2.5 (Table 2). In the same way, the absence of structural features increasing the biodegradability of the molecules (i.e., 2, 5, 9, 12, 14) does not also prevent the proposal of biodegradable candidates. When the presence of two different fragments is imposed, different results can be obtained (Table 2). Thus, the presence of two of the most unfavorable structural fragments (i.e., 1, 3, 4, 6) hardly prevents the design of biodegradable molecules. However, the combination of the presence of one of these unfavorable descriptors with the absence of one of the structural features increasing the biodegradability of the molecules (i.e., 2, 5, 9, 12, 14) provides interesting results. This is exemplified with descriptor no. 2 in Table 2.

One limitation of GAs is that the setting of parameters (e.g., population size, crossover and mutation probabilities) is problem dependent and represents a time-consuming process requiring experience. Under these conditions, we have tried to evaluate the influence of the GA parameters (Table 3) on the responses of the intercommunicating hybrid system (Figure 1). Two series of experiments were conducted. The former dealt with the use of GA without imposing structural conditions (Table 4). In the latter, the presence of descriptors no. 1 and 7 (Table 1) was imposed (Table 5).

When no structural constraints are imposed (Table 4), it is worth noting that with a population size varying from 10 to 100, in most cases, the GA finds the very best chromosome (i.e., AERUD value = 1.39). However, when the population size decreases, the number of generations necessary to find it increases. When the maximal number of generations varies between 50 and 500, the GA always discovers the very best chromosome. However, at least 300 generations are necessary to almost always obtain the best 50 chromosomes. As regards Pcross, from 0.1 to 0.9, the GA always finds the very best chromosome and for all the tested values, the GA proposes the best 50 chromosomes in 98% of the trials. For Pmut, optimal values range between 0.04 and ~0.2 (Table 4).

Table 4. Influence of the GA Parameters When No Structural Constraints Are Imposed

parameter		summary results
population	10−100	The GA always finds the best chromosome[a] except in 1 case/50 when the population equals 10.
generation	50−500	The GA always finds the best chromosome.
	50	The GA does not find the best 50 chromosomes except in 1 case/50.
	≥300	The GA always finds the best 50 chromosomes except in 1 case/250.
crossover	0.1−0.9	The GA always finds the best chromosome. In 9 cases/450, the GA does not find the best 50 chromosomes.
mutation	0.01−0.09	The GA always finds the best chromosome.
	0.1−0.4	In 1 case/200, the GA does not find the best chromosome.
	0.5−0.9	In 189 cases/250, the GA does not find the best chromosome.
	≤0.03	In 10 cases/150, the GA does not find the best 50 chromosomes.
	0.04−0.09	In 1 case/300, the GA does not find the best 50 chromosomes.
	0.1	In 1 case/50, the GA does not find the best 50 chromosomes.
	0.2	In 4 cases/50, the GA does not find the best 50 chromosomes.
	0.3	In 27 cases/50, the GA does not find the best 50 chromosomes.
	≥0.4	The GA never finds the best 50 chromosomes.

[a] AERUD value = 1.39.

When descriptors no. 1 and 7 are fixed (Table 5), the best chromosome has an AERUD value of 1.78. The GA finds 21 chromosomes with an AERUD value inferior to 2.5. As for the preceding series of experiments, when the population size decreases, the number of generations necessary to find the best chromosome increases. When the maximal number of generations varies between 50 and 500, the GA always discovers the very best chromosome. As in the previous series of experiments, at least 300 generations are necessary to ensure that the best 21 chromosomes are found in a large majority of the trials. For a Pcross ranging from 0.1 to 0.9, the GA always finds the very best chromosome. For all the tested values, the GA proposes the best 21 chromosomes in 96% of the trials. For Pmut, optimal values range between 0.1 and 0.4 (Table 5).

These results clearly show the difficulty to select GA parameters especially when structural constraints are imposed.

3. DISCUSSION

Our results support the idea that an intercommunicating hybrid system based on the combined use of a GA and a

Table 5. Influence of the GA Parameters When the Presence of Descriptors Nos. 1 and 7 Is Imposed

parameter		summary results
population	10−100	The GA always finds the best chromosome[a] except in 19 cases/50 when the population equals 10 and in 4 cases/50 when the population equals 20.
generation	50−500	The GA always finds the best chromosome.
	50	The GA does not find the best 21 chromosomes[b] except in 5 cases/50.
	≥300	The GA finds the best 21 chromosomes in 241 cases/250.
crossover	0.1−0.9	The GA always finds the best chromosome. In 18 cases/450, the GA does not find the best 21 chromosomes.
mutation	0.01−0.09	The GA always finds the best chromosome.
	0.1−0.5	The GA always finds the best chromosome.
	0.6−0.9	In 43 cases/200, the GA does not find the best chromosome.
	0.01−0.09	In 74 cases/450, the GA does not find the best 21 chromosomes.
	0.1−0.4	In 8 cases/200, the GA does not find the best 21 chromosomes.
	0.5	In 22 cases/50, the GA does not find the best 21 chromosomes.
	0.6	In 49 cases/50, the GA does not find the best 21 chromosomes.
	≥0.7	The GA never finds the best 21 chromosomes.

[a] AERUD value = 1.78. [b] When all GA parameters are fixed at the values selected in our hybrid system, the GA finds 21 chromosomes with an AERUD value inferior to 2.5.

BNN model can be used to propose candidate molecules having a specific biodegradability. More generally, from this example, it is obvious that the synergy between GAs and BNNs offers a rich field of investigations in QSAR and drug design. Thus, for example, recently, we have used this type of hybridization for the proposition of candidate molecules having particular pharmacological activities but presenting a reduced toxicity.[31] In these studies, we exploit the heuristic potency of BNNs for modeling complex biological activities and the ability of GAs for proposing different optimal solutions which can be ranked. However, it is obvious that other types of hybridization can be considered. Thus, Table 6 provides a comparison of four intelligent techniques with respect to five useful properties in QSAR and drug design. From Table 6, it is obvious that intercommunicating hybrid systems can offer solutions to hard modeling problems that are not solvable by an individual technique alone. The strategy consists in selecting methods, as the modeling problems dictate, in order to obtain the best level of synergy between them.

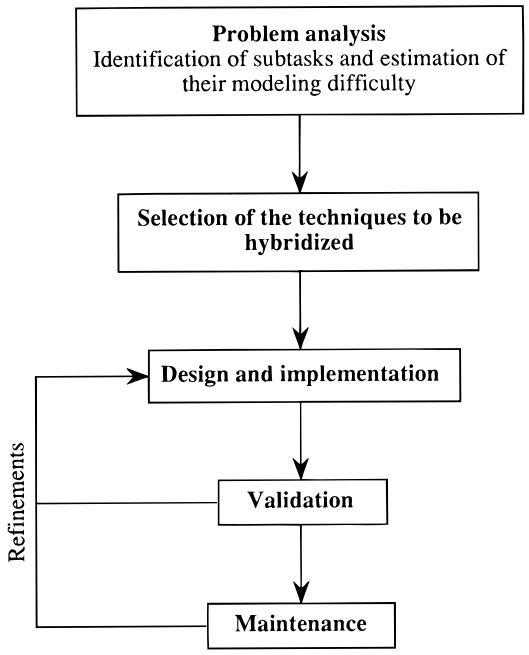

Figure 4. Intercommunicating hybrid system development cycle in QSAR and drug design.

It is important to mention that the notion of intercommunicating hybrid system covers not only the mixing and matching of different intelligent techniques but also the combination of these methods with classical modeling tools (e.g., PLS). In the same way, intercommunicating hybrid systems also deal with the integration of intelligent techniques with conventional computing systems such as spreadsheets and databases. Indeed, to be useful in QSAR and drug design, an intercommunicating hybrid system must be able to extract and use information from a variety of sources. In addition, the results produced by this hybrid system must be easily communicated to other systems for further processing. For these reasons, it is vital to have development environments (e.g., object-oriented programming) providing "the glue" to join together the different techniques and approaches which are hybridized.

In the next few years, intercommunicating hybrid systems will undoubtedly play an increasingly important role in QSAR and drug design. Indeed, they provide powerful alternatives to currently established data analysis when we have to take into account different modeling constraints as it is generally the case in the industry. However, as stressed above, to be accepted within the drug discovery departments of pharmaceutical and agrochemical companies, the intercommunicating hybrids must be able to interact and communicate with a variety of different information systems and existing applications used in these organizations. Furthermore, the other practical problem associated with the development of intercommunicating hybrid systems within private companies is an "educational" one. Indeed, to

Table 6. Comparison of Intelligent Techniques (ITs) and Their Use in QSAR and Drug Design (DD) (Adapted from Goonatilake[16])

ITs	learning	flexibility	adaptation	explanation	discovery	QSAR/DD
BNNs[a]	*****	*****	*****	*	**	****
GAs[b]	*****	****	****	***	*****	**
FSs[c]	*	*****	*	***	*	*
ESs[d]	*	*	*	*****	*	*

[a] BNNs = backpropagation neural networks. [b] GAs = genetic algorithms. [c] FSs = fuzzy systems. [d] ESs = expert systems.

develop successful and powerful hybrid systems, one needs to be aware of the detailed workings of different intelligent techniques as opposed to just knowing a single method. This can pose problems since, basically, there are currently only relatively few personnel in the drug discovery departments of the pharmaceutical and agrochemical companies. In addition, in practice, it is generally difficult to find a person presenting a sufficient training and experience in the development and application of more than one intelligent technique. This is problematic since the construction of an intercommunicating hybrid system requires to strictly follow different steps (Figure 4) for reducing development times and costs. This explains the increasing number of collaborations between chemical industry and consulting companies specialized in the design of hybrid systems. This type of hybridization also provides synergetic results!

REFERENCES AND NOTES

(1) Eberhart, R. C.; Dobbins, R. W. *Neural Network PC Tools: A Practical Guide;* Academic Press: San Diego, 1990.

(2) Goldberg, D. E. *Genetic Algorithms in Search, Optimization & Machine Learning*; Addison-Wesley Publishing Company: Reading, 1989.

(3) Devillers, J. *Neural Networks in QSAR and Drug Design*; Academic Press: London, 1996.

(4) Devillers, J. *Genetic Algorithms in Molecular Modeling*; Academic Press: London, 1996.

(5) Dubois, D.; Prade, H.; Yager, R. R. *Readings in Fuzzy Sets for Intelligent Systems*; Morgan Kaufmann Publishers: San Mateo, 1993.

(6) McNeill, F. M.; Thro, E. *Fuzzy Logic: A Practical Approach;* Academic Press Professional: Cambridge, 1994.

(7) Goonatilake, S.; Khebbal, S. In *Intelligent Hybrid Systems*; Goonatilake, S., Khebbal, S., Eds.; John Wiley & Sons: Chichester, 1995; Chapter 1.

(8) The adjective "intelligent" is a misnomer. No one has ever claimed that a technique or a system possesses a kind of intelligence. However, this adjective is commonly used with neural networks, genetic algorithms, fuzzy logic, and expert systems (see for example the cover of *IEEE Expert* of June 1995). It was easier to keep this terminology in the present paper.

(9) Harp, S. A.; Samad, T.; Guha, A. In *Proceedings of the Third International Conference on Genetic Algorithms;* Schaffer, J. D., Ed.; Morgan Kaufmann Publishers: San Mateo, 1989; p 360.

(10) Miller, G. F.; Todd, P. M.; Hegde, S. U. In *Proceedings of the Third International Conference on Genetic Algorithms;* Schaffer, J. D., Ed.; Morgan Kaufmann Publishers: San Mateo, 1989; p 379.

(11) Jones, A. J. Genetic algorithms and their applications to the design of neural networks. *Neural Comput. Applic.* **1993**, *1*, 32−45.

(12) Shamir, N.; Saad, D.; Marom, E. Using the functional behavior of neurons for genetic recombination in neural nets training. *Complex Syst.* **1993**, *7*, 445−467.

(13) Montana, D. J. In *Intelligent Hybrid Systems*; Goonatilake, S., Khebbal, S., Eds.; John Wiley & Sons: Chichester, 1995; Chapter 5.

(14) Roberts, S. G.; Turega, M. In *Artificial Neural Nets and Genetic Algorithms. Proceedings of the International Conference, Alès, France, 1995;* Pearson, D. W., Steele, N. C., Albrecht, R. F., Eds.; Springer-Verlag: Wien, 1995; p 96.

(15) Gruau, F. In *Intelligent Hybrid Systems*; Goonatilake, S., Khebbal, S., Eds.; John Wiley & Sons: Chichester, 1995; Chapter 13.

(16) Goonatilake, S. In *Intelligent Systems for Finance and Business*; Goonatilake, S., Treleaven, P., Eds.; John Wiley & Sons: Chichester, 1995; Chapter 1.

(17) Devillers, J.; Putavy, C. In *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic Press: London, 1996; Chapter 12.

(18) van Helden, S. P.; Hamersma, H.; van Geerestein, V. J. In *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic Press: London, 1996; Chapter 7.

(19) Venkatasubramanian, V.; Sundaram, A.; Chan, K.; Caruthers, J. M. In *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic Press: London, 1996; Chapter 11.

(20) Cambon, B.; Devillers, J. New trends in structure-biodegradability relationships. *Quant. Struct.-Act. Relat.* **1993**, *12*, 49−56.

(21) Devillers, J. Neural modelling of the biodegradability of benzene derivatives. *SAR QSAR Environ. Res.* **1993**, *1*, 161−167.

(22) Boethling, R. S.; Gregg, B.; Frederick, R.; Gabel, N. W.; Campbell, S. E.; Sabljic, A. Expert systems survey on biodegradation of xenobiotic chemicals. *Ecotoxicol. Environ. Safety* **1989**, *18*, 252−267.

(23) Kuenemann, P.; Vasseur, P.; Devillers, J. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht, 1990; Chapter 19.

(24) Niemi, G. J.; Veith, G. D.; Regal, R. R.; Vaishnav, D. D. Structural features associated with degradable and persistent chemicals. *Environ. Toxicol. Chem.* **1987**, *6*, 515−527.

(25) Howard, P. H.; Boethling, R. S.; Stiteler, W.; Meylan, W.; Beauman, J. Development of a predictive model for biodegradability based on BIODEG, the evaluated biodegradation data base. *Sci. Total Environ.* **1991**, *109/110*, 635−541.

(26) Howard, P. H.; Boethling, R. S.; Stiteler, W. M.; Meylan, W. M.; Hueber, A. E.; Beauman, J. A.; Larosche, M. E. Predictive model for aerobic biodegradability developed from a file of evaluated biodegradation data. *Environ. Toxicol. Chem.* **1992**, *11*, 593−603.

(27) Domine, D.; Devillers, J.; Chastrette, M.; Karcher, W. Estimating pesticide field half-lives from a backpropagation neural network. *SAR QSAR Environ. Res.* **1993**, *1*, 211−219.

(28) Boethling, R. S.; Howard, P. H.; Meylan, W.; Stiteler, W.; Beauman, J.; Tirado, N. Group contribution method for predicting probability and rate of aerobic biodegradation. *Environ. Sci. Technol.* **1994**, *28*, 459−465.

(29) STATQSAR Package (1994). CTIS, Lyon, France.

(30) Pitter, P.; Chudoba, J. *Biodegradability of Organic Substances in the Aquatic Environment;* CRC Press: Boca Raton, 1990.

(31) This intercommunicating hybrid system was designed in the frame of a contract with a pharmaceutical company. It was constituted of two backpropagation neural network (BNN) models and one genetic algorithm. One BNN model was dedicated to the prediction of the pharmacological activity described by fuzzy coding. The other BNN model allowed the prediction of the *in vitro* toxicity of the candidate molecules.

CI960022Y