J Comput Aided Mol Des (2007) 21:207–221
DOI 10.1007/s10822-006-9098-3

ORIGINAL PAPER

Caco-2 cell permeability modelling: a neural network coupled genetic algorithm approach

Armida Di Fenza · Giuliano Alagona · Caterina Ghio · Riccardo Leonardi · Alessandro Giolitti · Andrea Madami

Received: 4 July 2006 / Accepted: 14 December 2006 / Published online: 30 January 2007

Abstract The ability to cross the intestinal cell membrane is a fundamental prerequisite of a drug compound. However, the experimental measurement of such an important property is a costly and highly time consuming step of the drug development process because it is necessary to synthesize the compound first. Therefore, in silico modelling of intestinal absorption, which can be carried out at very early stages of drug design, is an appealing alternative procedure which is based mainly on multivariate statistical analysis such as partial least squares (PLS) and neural networks (NN). Our implementation of neural network models for the prediction of intestinal absorption is based on the correlation of Caco-2 cell apparent permeability (P_{app}) values, as a measure of intestinal absorption, to the structures of two different data sets of drug candidates. Several molecular descriptors of the compounds were calculated and the optimal subsets were selected using a genetic algorithm; therefore, the method was indicated as Genetic Algorithm–Neural Network (GA-NN). A methodology combining a genetic algorithm search with neural network analysis applied to the modelling of Caco-2 P_{app} has never been presented before, although the two procedures have been already employed separately. Moreover, we provide new Caco-2 cell permeability measurements for more than two hundred compounds. Interestingly, the selected descriptors show to possess physico-chemical connotations which are in excellent accordance with the well known relevant molecular properties involved in the cellular membrane permeation phenomenon: hydrophilicity, hydrogen bonding propensity, hydrophobicity and molecular size. The predictive ability of the models, although rather good for a preliminary study, is somewhat affected by the poor precision of the experimental Caco-2 measurements. Finally, the generalization ability of one model was checked on an external test set not derived from the data sets used to build the models. The result obtained is of interesting practical application and underlines that the successful model construction is strictly dependent on the structural space representation of the data set used for model development.

Keywords Neural network · Genetic algorithm · Caco-2 cell permeability · Intestinal absorption · Oral bioavailability

Electronic supplementary material The online version of this article (doi:10.1007/s10822-006-9098-3) contains supplementary material, which is available to authorized users.

A. Di Fenza (✉) · G. Alagona · C. Ghio
Molecular Modelling Lab, Institute for Physico-Chemical Processes (IPCF) - CNR, Via G. Moruzzi 1, 56124 Pisa, Italy
e-mail: A.DiFenza@ipcf.cnr.it

R. Leonardi
Ion Trading s.r.l., Via S. Martino 52, 56125 Pisa, Italy

A. Giolitti
Menarini Ricerche SpA, Via Sette Santi 3, 50131 Firenze, Italy

A. Madami
Menarini Ricerche SpA, Via Tito Speri 10, 00040 Pomezia, Roma, Italy

Introduction

The prediction of human oral absorption is a major target in the design of new orally administered drugs. Several studies are being made with the aim of finding

optimal computational methods that would allow the substitution of high–cost and time-consuming laboratory experiments. The first approach has been the so-called ''rule of 5'' [1], derived from the analysis of successful drugs, which has led to the essential requisites a compound has to satisfy to exhibit good intestinal absorption. In that study, the physico-chemical properties emerged to characterize the intestinal absorption phenomenon were: lipophilicity, molecular weight and the number of H-bond donors and acceptors. After that, a good number of different attempts to predict quantitatively oral absorption were proposed, which generally make use of data-based approaches [2] such as quantitative structure-property relationship (QSPR) or similarity searches. In fact, it is difficult to separate the contribution of several distinct processes involved in the phenomenon, and characterized by hardly delineated molecular mechanisms (dissolution, membrane passive permeation, membrane active transport and metabolism) to adopt a structure-based approach such as ligand-protein docking and pharmacophore modelling. The data-based modelling approach basically consists of representing, by appropriate molecular descriptors, the chemical structure of the data compounds and then to correlate the target property with the descriptors through multivariate statistical analyses. Multivariate statistics are defined as those methods that examine multiple variables simultaneously [3] looking for either linear or non-linear type of correlations. Despite the robustness of these statistical techniques, major drawbacks are still present for the multivariate oral absorption model building. In particular, the capacity to correlate the chemical structure with the known per cent of absorbed molecule strongly depends on the quality and quantity of the data used to build the model, i.e. the entity of the experimental error (data errors decrease performance in the prediction model and extrapolative prediction is basically difficult) and the amount of measurements. Most importantly, for the successful construction of a good predictive model it is of the utmost importance that the drug-like space occupied by the compounds to be predicted is covered by the training set used in the model development.

Artificial neural networks belong to the non-linear type of multivariate statistical techniques. Their application to the field of QSPR has been proved to be comparable in performance to classical multivariate linear regression and recently the amount of studies reporting neural network models which outperform simple linear regression ones is increasing [4–10]. The attractive properties of neural networks are explained with the inherent non-linearity of their models. From non-linearity derives their parsimonious type of approximation and the ability to handle data that require the inclusion of power terms in order to fit a multiple regression model. Neural networks like multilayer perceptrons allow a better performance to be reached than that attained using other approximation methods, albeit with the same number of parameters [11]. Thus, neural networks are able to get more significant information out of a given amount of data. This property is particularly useful in the case of limited data sets which are the actual case.

Experimental measurements usually performed for screening the oral absorption of new drugs are in vivo oral availability and in vitro cellular permeability assays, e.g. Caco-2 cells. Bioavailability measurements may underestimate the real drug absorption because a fraction of the absorbed drug may not reach the systemic circulation. Actually, after absorption from the gastrointestinal tract, the drug passes directly to the liver via hepatic portal vein where it may be extensively metabolized or excreted before reaching the systemic circulation. In addition, bioavailability measurements have high experimental costs and can be performed only in clinical trial phases. The epithelial Caco-2 cell line, which originally derived from human colorectal carcinoma cells, is widely used as intestinal absorption model since several studies showed sigmoidal relationships between the fraction absorbed in humans and the apparent permeability across Caco-2 cell monolayers [12–15]. Caco-2 permeability measurements are routinely performed by pharmaceutical companies soon after the synthesis that is in an early stage of drug development. This means that many data of this kind are available and even more will be in the next years, which will allow for the construction of large and structurally diverse databases of molecules. This is not the case for bioavailability measurements: drugs who reach the late stages of development are only a fraction of those screened at the first stages and, importantly, are mostly well absorbed. Consequently, databases of molecules with their corresponding bioavailability are smaller as compared to Caco-2 cell databases and strongly biased towards high absorption values. For the aforementioned reasons, Caco-2 cell apparent permeability values are preferable to bioavailability ones for building regressions in order to predict oral absorption.

To the best of our knowledge, the prediction of Caco-2 cell permeability using a neural network (NN) approach has been first performed by Hashida and coworkers in 2002 [16]. They correlated five molecular descriptors (dipole moment, polarizability, sum of charges of nitrogen atoms, oxygen atoms and hydrogen

atoms bonding to nitrogen or oxygen atoms) of 87 compounds with $\log P_{app}$ and demonstrated the superiority of the NN approach with respect to simple and quadratic multiple regressions. They reported a prediction cross-validated correlation coefficient (R) of 0.71, 0.73 and 0.79 for simple multivariate regression, quadratic regression and neural network model, respectively. In the same year, Hashida and coworkers [17], by means of a genetic algorithm-based partial least square (PLS) method applied to the same set of molecules and Caco-2 cell permeability values, reported a prediction with $R = 0.83$, but they excluded 14 molecules, possibly involved in active membrane transport processes, from the original data set [16]. In their study, Hashida and coworkers [17], started from a total of 220 Molconn-Z [18] derived topological descriptors for each compound and used a genetic algorithm applied to PLS analysis to select an optimal subset of 24 descriptors. Despite the good predictive ability of the model developed, no functional discussion about the meaning of the selected optimal descriptors related to Caco-2 cell monolayer permeation was possible, because the Molconn-Z descriptors used, being based on graph theory, are very difficult to interpret.

We present here a genetic algorithm-based neural network approach, its validation and its application to the construction of a model using two different sets of molecules with their relative Caco-2 cell values. The experimental error on these measurements is rather high, especially on the low values, but the positive feature is that the experiments were conducted in the same laboratory for both data sets and with the same procedure. In addition, the two data sets are quite large as compared to those exploited in previously reported studies [16, 17]. Therefore, they constitute a large enough database to ensure fair predictions, made up of Caco-2 cell values free from any variance due to factors such as cell passage number, culture time, type of support and medium, which usually affect Caco-2 cell results from different laboratories or methodologies [19]. Finally, the descriptors used to represent the compound molecular structures are based of the concept of Volsurf [20], a procedure which is specifically designed to produce descriptors related to pharmacokinetic properties. Volsurf extracts the information present in 3D molecular field maps into few quantitative numerical descriptors which have a clear chemical meaning and are easy to understand and to interpret. An additional, important advantage of Volsurf is that it does not need to use complex algorithms to calculate surfaces, volumes and other related descriptors since they are directly obtained from the 3D molecular fields, thus making the calculus of the molecular descriptors very fast and hence suitable for large databases of molecules.

The use of the GA-NN algorithm developed here, coupled to the Volsurf features for the descriptor generation, should constitute a reliable and fast tool in the very first stages of drug design.

Materials and methods

Data sets

Two different sets of pharmacological molecules, Men_1 and Men_2, and their experimentally derived Caco-2 cell values were supplied by Menarini Ricerche SpA. Men_1 gathers 106 NK-2 receptor antagonists [21, 22] (Fig. 1) and Men_2 101 compounds with antitumor activity, whose structures are undisclosed at present. The Men_1_2 data set, constituted by the union of Men_1 and Men_2, was also used. The Caco-2 cell P_{app} is expressed as 10^{-6} cm/s.

Descriptor generation, pre-processing and selection

The 3-D structures of the compounds were generated with CONCORD [23]. The program Volsurf [20, 24, 25] was used with all the nine probes to generate 166 descriptors in total for each compound. Only a single conformer was considered for descriptors generation, in fact Volsurf descriptors were demonstrated to be relatively independent of conformational sampling and averaging [20, 24, 26]. Since the descriptors depend on the ionic state of the considered molecule, the pK_a value and population of the different ionic states for each molecule at the pH value during the experimental procedure (7.4) were calculated with the software ACD/pK_a Batch [27]. In case of molecules with more than a single ionization state, the descriptors for each state have been generated; then, the weighted mean values of the descriptors according to the molar fraction of the corresponding ionic state have been used.

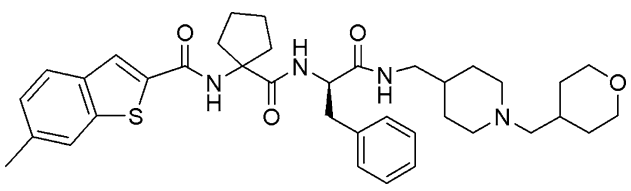

Fig. 1 Structure of MEN15596, the lead compound of the tachykinin NK-2 receptor antagonists belonging to the data set Men_1

The data set Men_2 was the only one entirely composed of neutral molecules.

The total number of descriptors for each molecule was subjected to initial reduction according to redundancy and correlation. In particular, descriptors having more than 80% identical values on all the molecules have been eliminated. One out of two descriptors found to have a Pearson correlation coefficient (R) greater than 0.95 or 0.80 have been randomly eliminated. After these reductions, respectively 124 ($R \geq 0.95$) and 56 ($R \geq 0.80$) descriptors for Men_1 have been obtained; 96 ($R \geq 0.95$) descriptors for Men_1_2 and 94 ($R \geq 0.95$) final descriptors for Men_2. The reduced descriptors were still too many compared to the data they describe, especially when $R \geq 0.95$ is used. Therefore, to alleviate the worst effects of the curse of dimensionality (in many problems, reducing the number of input variables can sometimes lead to improved performance for a given data set, even though information is being discarded [28]) and overfitting, a feature selection by a genetic algorithm has been performed.

Genetic algorithm

Genetic algorithms (GA) are combinatorial optimization schemes, conceptually based on the principles regulating the biological evolutionary mechanisms. In short, with the aim to find the minimum of a function, several solutions are generated, among which the best ones are selected and combined in such a way that the convergence to a point of optimum is reached. In problems like the selections of molecular structure descriptors to be used in the construction of neural network models as well as of other kind of multivariate regressions, the genetic algorithm showed to perform very well [29–39]. The principal steps of the computational GA used in this study are:

1. An initial random population of solutions is generated. Each solution is represented by a combination of descriptors. Each descriptor is indicated with a cardinal number.
2. For each descriptor combination the performance is evaluated using a defined cost function.
3. The combinations are ranked and selected according to their performance values.
4. A new population is created by means of three types of actions on the components of the previous population: elitism (the first n combinations with better performance are transferred in the new population without any modification), crossover (a determined per cent of combinations is used to create new combinations by mutually exchanging a selected part in pairs of randomly selected combinations), mutation (a single combination is subjected to random modifications originating a new combination).

The 2–4 cycles are repeated until a stopping criterion is satisfied. The genetic algorithm here developed presents also the following characteristics:

1. If two identical descriptor combinations are generated in the same population, one is eliminated and, to keep constant the number of combinations in the population, a new one is created according to the rules for the generation of the GA initial population.
2. If a generated combination contains two or more identical descriptors, only one is kept, evaluating the performance on the reduced combination. This feature of the GA allows, in addition, combinations with a smaller number of descriptors with respect to that chosen in the initial settings to be possibly found.
3. The Cost Function (CF) of the GA is obtained averaging out the values of the below defined cost function [29] calculated on twenty different neural network models:

$$\mathrm{CF} = \mathrm{RMSE}_{\mathrm{train}} + 0.4 * |\mathrm{RMSE}_{\mathrm{train}} - \mathrm{RMSE}_{\mathrm{valid}}| \quad (1)$$

where $\mathrm{RMSE}_{\mathrm{train}}$ and $\mathrm{RMSE}_{\mathrm{valid}}$ are the root mean squared errors on the training and validation set (see the Early Stopping Section below), respectively. The use of the mean over twenty RMSE instead of a single RMSE was a necessary strategy to lower the rather high degree of variability exhibited by network RMSE (Fig. 2).

In this study, the computational genetic parameters, common to all the GA-NN runs presented here, were: population size (50 combinations), elite count (4) and mutation rate (0.02). The GA-NN calculation was allowed to run until a plateau in the cost function value was reached (from 100 to 1,000 generations). The crossover fraction is the fraction of individuals in the next generation, other than the elite ones, which are created by crossover, and dictates also the number of individuals created by mutation on the present generation individuals. To show the influence of the reproduction nature on the GA-NN performance, the results relative to an optimized crossover fraction value representing the presence of crossover, i.e. 0.6, and a

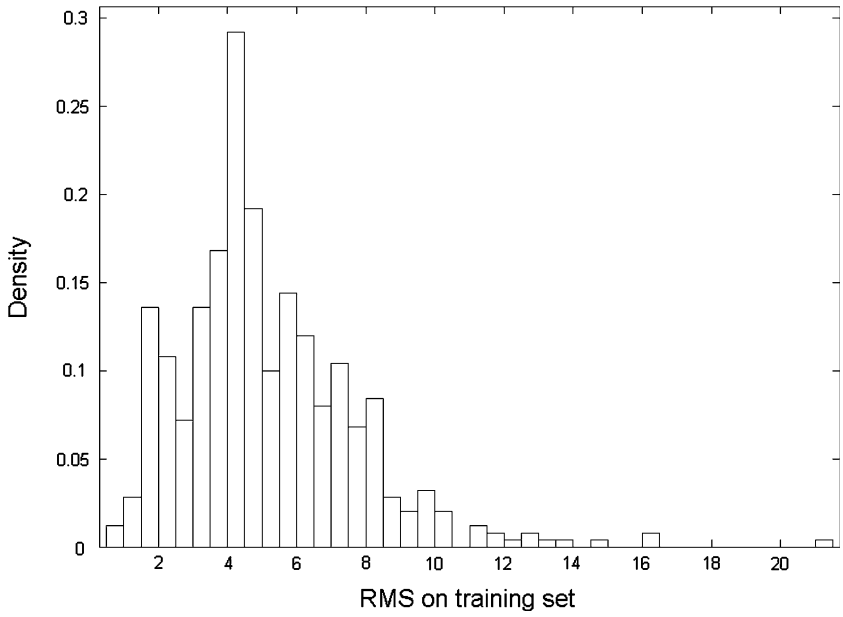

Fig. 2 Distribution of the values of RMSEs for 500 runs of training on the Men_1 data set using model 1 of Table 2

crossover fraction value of 0, indicating the absence of crossover, are reported.

For each set of parameters and conditions, the GA-NN was run three times. The 10 top scoring combinations of each run were tested for the generalization performance (see the Generalization Performance Section below) and the best predictive model, among the 30 generated solutions, was selected. It is worth noting that almost always the top 10 models of a GA-NN run were composed of 3 or 4 different combinations at most, invariably consisting of less than 10 distinct combinations. This is a clear indication that GA-NN convergence had been attained.

Neural network

A neural network was used to build a model based on the mapping of the molecular descriptors with the P_{app} absorption. Neural networks are algorithms whose structure derives from the anatomy of biological neural networks. A neural network model, characterized by its own training determined parameters, represents an implicit non-linear function, where the molecular descriptors are the independent variables, while the target property is the correlated dependent variable. For details of the neural network implementation, readers are referred to the relevant references or standard texts [28, 40]. A feed-forward fully connected multilayer perceptron (MLP), without short connections and with a single layer of hidden neurons, was used. The number of input and hidden layer neurons, 12 and 4, respectively, were chosen in order to keep the network architecture as simple as possible. The input values (descriptor coordinates) and target values (P_{app} measurements) were scaled between 0.0 and 0.90 before feeding the network. The training algorithm employed for the optimization of the network weights was the Levenberg–Marquardt algorithm [41–43]. This algorithm is designed specifically for minimizing a sum-of-square error and appears to be the fastest method for training moderate-sized feed-forward neural networks (up to several hundred weights). It approaches second-order training speed without having to compute the Hessian matrix [44].

Weights and biases were initialized according to the Nguyen–Widrow initializating algorithm [45]. This algorithm chooses values in such a way as to distribute the active region of each neuron in a layer nearly uniformly spaced along the input space of that layer.

Early stopping

The most important way to obtain a generalizating neural network model, able to make good predictions about new inputs, is to avoid the network to learn an exact representation of the training data itself, a phenomenon called overfitting which is one of the most delicate problems of neural networks. In order to limit this effect we trained the neural networks with early stopping (ES) [46]. Accordingly, the compounds in the data set Men_1 were split by a binning method [34] in a training set of 79 molecules, a validation set of 14 molecules and a test set of 13 molecules. The sub division for Men_2 consisted in 77 (training), 12

(validation), 12 (test) and for Men_1_2 in 155 (training), 27 (validation) and 25 (test). As apparent from their names, the training set is used to train the neural network, the validation set to indicate when the training has to stop to prevent data overfitting and the test set is used as an independent set to preliminarily measure the generalization performance of the neural network model. The binning method provides a uniform distribution of the target values (apparent permeabilities) in the three sub-sets necessary for the early stopping procedure. To obtain this, each data set was binned according to the molecule target values and the sub-sets were populated from the bins providing each sub-set with a uniform representation of the dependent variable range (Fig. 3).

Generalization performance evaluation

For each of the Men_1 and Men_2 data sets a final best model was selected comparing the generalization performance values. The generalization performance of a network model is commonly evaluated by measuring the error made by the model in predicting the target values of a data set not used for the network training. The ''test set'' obtained from the binning of the original data set is often used to this aim. Nevertheless, such a test set is somewhat dependent on the way it was extracted from the totality of data, thus its composition is strictly related to the composition of the training and validation sets. For these reasons, in this study the generalization performance was measured with an alternative procedure based on a cross-validation scheme, the leave-one-out (LOO) analysis [47, 48, 49].

The regression coefficients obtained were calculated from a committee of networks [50, 51]. This method consists in training a network N times and considering the output of each of the trained network models, which only differ because were trained to different local minima of the error function. A simple form of committee was used, where the final committee output is the average of the outputs of the N networks composing the committee. Significant improvements in the prediction on new data can be obtained in this way, because the performance of a committee can be better than the performance of the best single network used alone. The averaging over many solutions, in fact, leads to a reduced variance which makes the performance error decrease. In this work, in particular, a committee of 60 neural network models having the same architecture and different initial conditions (different initial weights and biases) was used. For the LOO procedure, due to the high computational cost of this algorithm, the committee was composed of only five neural network models.

The generalization performance of the final best models built with Men_1 and Men_2 was also measured by the ability to predict the apparent permeabilities of a set of molecules taken from the literature [52], which was defined the ''external set''.

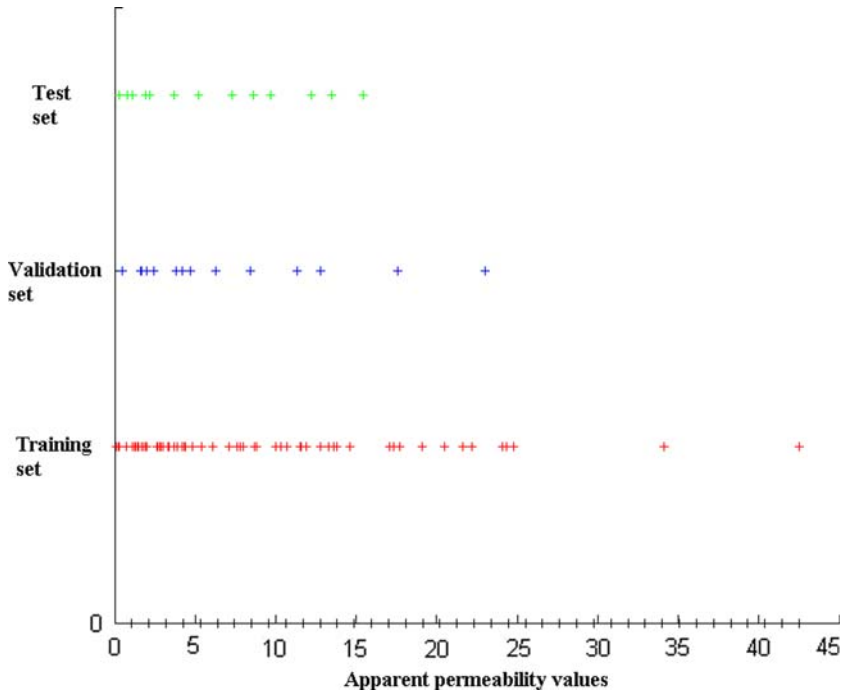

Fig. 3 Example of distribution of the target values (apparent permeabilities) in the three sub-sets obtained with the binning method

Results and discussion

Method validation

The GA-NN method was validated using the data set and the descriptors reported in the study of Guha et al. [34], who correlated the molecular structures of Artemisinin analogues with their biological activities. Both the molecular descriptors (generated by the program ADAPT [53]) and the corresponding activity values were kindly supplied by the authors. The GA-NN was run using the early stopping method as in the work of Guha et al. [34] and the final model was considered as that having the best LOO performance among the models derived from different GA-NN runs. The performance of the best final combination of descriptors obtained with the present method in comparison to that of the Guha et al. [34] combination is reported in Table 1.

The GA-NN training procedure used in Ref. [34] was early stopping and the best model was selected according to the RMS and R^2 values on the ''test'' set, while only subsequently a leave *14%* out regression coefficient was calculated. We also employed ES for training but, as explained in the Method section, we performed the model selection comparing the different model LOO regression coefficients. Interestingly enough, from a perusal of Table 1, it is evident that this procedure provided our model with an R^2 for the leave *14%* out test which is better ($R^2 = 0.73$) than that found in Ref [34] ($R^2 = 0.69$).

In our opinion, the Guha model is seemingly overfitted on the ''test'' set, since its R^2 is rather higher than that measured with the leave *14%* out procedure ($R^2 = 0.88$ and $R^2_{\text{Leave-}14\%\text{-Out}} = 0.69$). On the contrary, our model shows a better consistency between the two generalization tests ($R^2 = 0.79$ and $R^2_{\text{Leave-}14\%\text{-Out}} = 0.73$). This fact enlightens the likely pitfall when using a *hold out* method for model selection: the results can heavily depend on the data set subdivision in training, validation and test set. For this reason, as mentioned in the Method section, in this study the final generalization test was performed by the LOO procedure. Very interestingly, descriptors 6, 8, 10 and 21, that is four descriptors out of ten, are shared between our combination and the Guha one.

The result obtained on the Guha data demonstrated the efficacy and robustness of the GA-NN method developed here and allowed to exploit it on new data sets as described in the following.

Men_1 analysis

Several different runs of the GA-NN were performed on the Men_1 data set with the aim to find the optimal subset of descriptors which could give the best neural network model for the prediction of the Caco-2 cell permeability. The GA-NN was applied to both the data set reduced to 124 descriptors and to that reduced to 56 descriptors. The influence of the crossover fraction value on the genetic algorithm performance was also investigated. Since previous studies showed that the absence of crossover gave an optimal set of solutions [31, 39], two crossover fraction values were compared, i.e., 0 and 0.6.

The best results were obtained using a crossover fraction value of 0.6 (Table 2). A possible explanation for this fact, which is in contrast with previously reported findings [31, 39] can be found in the unfortunate combination of lack of hereditary transmission (when crossover is off) and high variability of the network error function value (for a detailed explanation see Figure S1 in Supporting Information). In addition no differences were found in using 56 or 124 descriptors (data not shown). This is an indication that the molecular descriptors are highly correlated and that the descriptor pool size does not have a noticeably influence on the GA-NN performance. Instead, the GA-NN is especially able to find those non-linear relationships among descriptors as it is evident from the selection of multiple and different descriptor combinations giving the same informative contribution related to the P_{app}.

The values reported in Table 2 also show that, thanks to the early stopping procedure, overfitting phenomena are clearly absent (see also Fig. 4). In fact, the value of R on the validation and test sets are not significantly worse than that on the training set: on the

Table 1 Regression coefficients of the present model as compared to those obtained by Guha et al.[a]

[a] Ref [34]

[b] The neural network architecture used to evaluate the regression coefficients

Descriptors combination	Regression coefficient (R^2)			
	Train	Validation	Test	Leave-*14%*-out
This study (9–5–1)[b] (3, 6, 8, 10, 19, 21, 25, 49, 59)	0.86	0.86	0.79	0.73
Guha et al.[a] (10–5–1)[b] (1, 6, 8, 10, 13, 20, 21, 40, 45, 48)	0.96	0.94	0.88	0.69

Table 2 Pearson correlation coefficients for the best performing descriptor combinations, (No.1 = 4, 6, 8, 12, 52, 55, 61, 74, 116, 117, 119 and No.2 = 3, 6, 15, 50, 52, 55, 63, 74, 93, 103, 111, 115), selected by the GA-NN[a] for data set Men_1

No.	Crossover fraction	*R*_train[b]	*R*_val[b]	*R*_test[b]	LOO (*R*)[b]
1	0.6	0.779 ± 0.025	0.755 ± 0.015	0.771 ± 0.019	0.723 ± 0.006
2	0	0.751 ± 0.016	0.764 ± 0.027	0.827 ± 0.028	0.694 ± 0.003

[a] Parameters of the GA/NN run are: network architecture (12–4–1), mutation rate (0.02), elite count (4), population size (50)

[b] Mean and standard deviation derived from three repetitions

contrary, the test set regression coefficient is even better. This fact is probably due to the particular subdivision of the data set used.

The best descriptor combination selected by the GA-NN identifies a neural network model which is able to make predictions on the P_{app} with a regression coefficient of 0.723. This result is very good considering the rather high experimental error affecting especially the low permeability values used to build the models. The experimental errors were ~100% on absorption values (P_{app}) ≤ 1, ~50% on values between 1 and 10, and ~20% for values above 10. These kinds of measurements are used as a pre-screening in drug development assays. They serve primarily to discriminate among poor absorption ($P_{app} \leq 1$), intermediate/variable absorption ($1 < P_{app} \leq 10$) and optimal absorption ($P_{app} > 10$). The total RMS error on the predictions is 3.97×10^{-6} cm/s, which taken like this does not tell too much. When, instead, the percent RMS error on each of the three intervals of values is calculated and normalized on the intervals, the following prediction RMS errors relative to each considered interval can be found:

$$\begin{aligned} \text{RMS} &= 623.3\% \text{ (for Caco-2 cell permeability} \leq 1) \\ \text{RMS} &= 38.0\% \text{ (for } 1 < \text{Caco-2 cell permeability} \leq 10) \\ \text{RMS} &= 21.2\% \text{ (for Caco-2 cell permeability} > 10) \end{aligned}$$

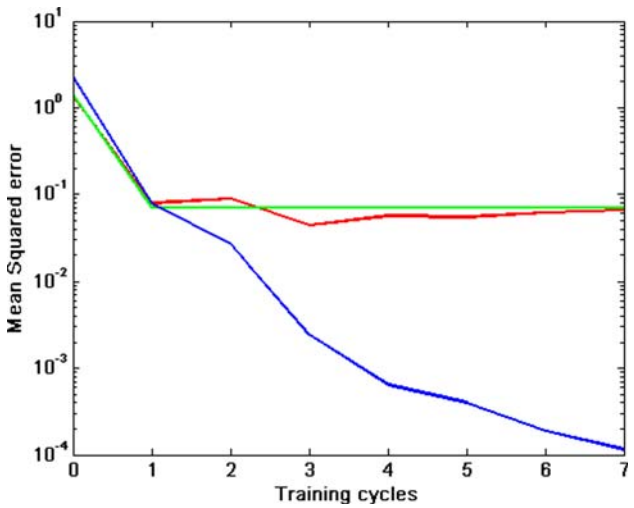

Fig. 4 Plot of the errors on the training (blue), validation (green) and test sets (red) during network training showing the absence of overfitting phenomena: all the errors decrease and there are no permanent increases in those relative to the validation and the test sets

Very interestingly, this trend emerges clearly also from the predicted versus experimental values reported in Table 3 and plotted in Fig. 5. The errors for the P_{app} values >1 are in perfect agreement with the experimental error: actually a prediction error lower than the experimental one would be meaningless. In addition, the high prediction error found for values ≤1 indicates that there is an upper limit for the error on experimental data for feeding neural networks in order to get good predictions. This error can be as high as 50% or even more, but clearly it should be lower than 100%. Despite the average error for calculated permeability values on the three intervals is consistent with the corresponding experimental error, it is worth noting that some values can be considered outliers.

Typical examples are the compounds MEN14766, MEN14982 and MEN15409, whose predicted values differ from the experimental ones, respectively, by –46.6%, –1,141.7% and +1,620% (the last percentage errors are, of course, amplified by the very small permeability values measured for those compounds; see Table 3). These deviations can be explained considering some of the approximations introduced to make the procedure faster. First, since Volsurf descriptors seem to be only marginally influenced by conformational sampling [20, 24, 25], they were generated considering only a single conformer for each compound. However, as the authors claim, ''these results may depend on the set of compounds and their property space'' [24]. Clearly, in the worst case, considering a single conformer model is a rough guess of the in vivo situation, because the biological conformation can be completely different from that used. An analogous consideration can be carried out for the molecular charge. Volsurf descriptors in some cases may depend

Table 3 List of Men_1 molecules and their experimental and predicted Caco-2 cell permeabilities. The predicted values were produced with the LOO test ($R = 0.723$) using model 1 of Table 2

Name	Exptl Permeab.[a]	Predict. Permeab.	Name	Exptl Permeab.[a]	Predict. Permeab.
MEN13719	13.1	12.6	MEN15420	9.1	5.5
MEN13725	15.4	15.5	MEN15528	0.8	2.7
MEN13918	4.2	7.6	MEN15531	0.5	–0.3
MEN13926	2.5	–1.1	MEN15532	0.5	2.7
MEN14103	12.6	10.4	MEN15533	3.4	4.9
MEN14105	10.7	9.3	MEN15536	7.3	11.3
MEN14123	2.7	0.9	MEN15539	0.5	1.6
MEN14180	2.0	3.7	MEN15545	4.2	5.6
MEN14181	3.2	1.2	MEN15546	4.1	5.8
MEN14183	1.2	–0.5	MEN15547	2.2	3.2
MEN14188	2.0	0.9	MEN15555	4.3	7.9
MEN14268	9.0	9.4	MEN15561	2.6	2.3
MEN14330	0.5	–1.1	MEN15562	0.9	6.9
MEN14478	1.0	–1.7	MEN15566	1.5	3.4
MEN14732	18.0	15.5	MEN15568	2.9	4.2
MEN14745	6.0	8.1	MEN15569	4.4	7.9
MEN14766	32.2	17.2	MEN15570	9.1	5.4
MEN14769	6.9	11.6	MEN15571	4.9	4.8
MEN14818	10.6	5.3	MEN15574	4.6	5.1
MEN14848	5.9	7.5	MEN15575	1.1	0.6
MEN14861	12.4	9.4	MEN15577	7.9	5.5
MEN14888	0.5	3.7	MEN15579	4.5	4.8
MEN14889	2.0	2.8	MEN15583	1.0	–6.9
MEN14890	1.4	–1.3	MEN15585	10.2	8.4
MEN14891	0.6	0.0	MEN15586	5.5	6.3
MEN14970	11.7	10.6	MEN15587	8.3	2.6
MEN14982	1.2	–12.5	MEN15588	9.7	10.0
MEN14983	0.5	–0.6	MEN15590	2.2	3.0
MEN14987	3.5	1.1	MEN15591	0.5	4.1
MEN14988	14.0	17.1	MEN15596	5.1	2.0
MEN14993	0.5	3.2	MEN15598	0.8	5.0
MEN15023	13.4	16.3	MEN15599	0.3	1.7
MEN15055	1.3	–0.6	MEN15604	3.6	8.9
MEN15092	13.1	15.3	MEN15605	0.5	5.3
MEN15166	1.6	0.6	MEN15607	0.5	3.6
MEN15168	1.5	7.5	MEN15612	1.8	1.9
MEN15281	0.9	2.3	MEN15716	10.1	4.5
MEN15289	0.5	–0.8	MEN15718	11.6	7.2
MEN15301	9.6	8.3	MEN15719	10.6	7.9
MEN15305	1.4	5.5	MEN15720	2.1	3.8
MEN15309	6.5	4.0	MEN15721	10.4	6.0
MEN15353	8.5	7.1	MEN15728	3.5	2.0
MEN15356	9.6	8.3	MEN15731	2.7	–1.6
MEN15366	3.9	3.7	MEN15734	1.3	4.2
MEN15370	11.5	15.1	MEN15736	0.5	3.5
MEN15384	2.2	–4.9	MEN15738	9.3	8.3
MEN15398	1.6	2.8	MEN15739	7.1	4.0
MEN15400	1.0	2.5	MEN15747	1.9	6.6
MEN15403	8.4	14.7	MEN15749	3.3	5.5
MEN15409	0.5	8.6	MEN15754	0.5	–0.9
MEN15411	0.5	–0.3	MEN15755	0.5	5.5
MEN15417	6.6	5.2	MEN15757	1.3	3.0
MEN15418	0.5	–13.1	MEN15758	2.8	3.5

[a] 10^{-6} cm/s

on the ionic state of the molecule of interest; therefore the populations of the different ionic states for each molecule at pH 7.4 were estimated, considering the weighted averaged descriptor values according to the ionic state molar fraction. Nevertheless, on the whole, the models derived possess a generalization ability comparable to that obtained by Hashida and coworkers ($R = 0.79$) [16], confirming that in silico

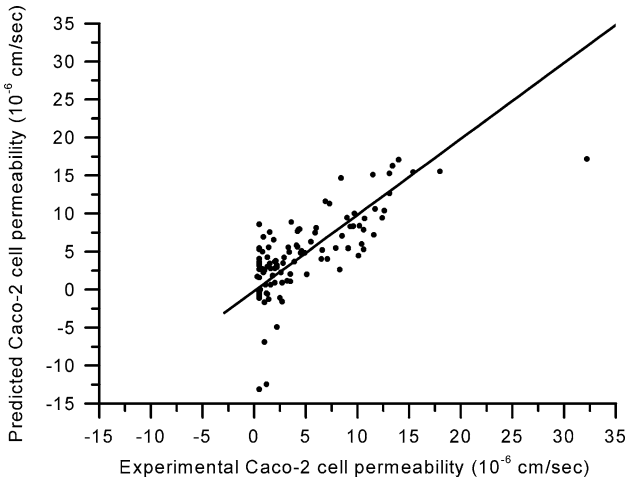

Fig. 5 A plot of experimental versus predicted Caco-2 cell permeability for the Men_1 data set produced from model 1 of Table 2. The predicted values are obtained with the LOO test (R = 0.723)

modelling of Caco-2 cell permeability is a rather difficult task. Definitely the GA-NN procedure can not be blamed, because our GA-NN model developed for the Artemisinin analogues resulted to be very good and outperformed the model derived using an analogous GA-NN technique by Guha et al. on the same data set [34]. It can be argued that the membrane permeation is a very complex event, characterized by the occurrence of several physico-chemical equilibria. Membrane penetration is certainly a more complex phenomenon than the action exerted by Artemisinin analogues in the Plasmodium Falciparum parasite, in the sense that this last event is highly specific and strictly reliant on the presence in Artemisinin analogues of a definite pharmacophore group. Thus, modelling of a multi-mechanism process is more complicated than other QSPR studies where the property analyzed is characterized by a single mechanism. Lipinski commented on the need to deconvolute multiple mechanisms: ''Computational models for multi-mechanism assays (such as ADME) typically get worse as more data is accumulated. By contrast, computation models for single mechanism assays (biological receptor affinity) typically get better as more data is accumulated. The reason is that as more and structurally more diverse compounds are screened in a multi-mechanism system more data is obtained on more mechanisms, and the noise level for each individual mechanistic component rises.''[54] This comment gives also an explanation for the unseen improvement which should have derived from the present use of large sets of data with respect to previous studies [16, 17].

Men_1_2 and Men_2 analysis

The GA-NN was applied to Men_1_2 and Men_2 to check for any possible dependence of the procedure upon the data set analyzed. A crossover fraction of 0.6 was used, since the results obtained with the Men_1 data set demonstrated better performance of the GA-NN with this parameter value. The original number of descriptors of both data sets was reduced eliminating descriptors linearly correlated according to a value of $R \geq 0.95$.

The results in Table 4 indicate that, although the LOO regression coefficient of Men_2 is slightly higher (0.752) than the regression coefficients calculated for Men_1 and Men_1_2 (0.723 and 0.726, respectively), these three values are very close to each other. The reason is that all the data sets considered have equivalent and high experimental errors. Furthermore, importantly, both the Men_1 and Men_2 data sets, despite constituted by molecules which are analogues of two different lead compounds, thus encompassing different regions of the structural space, constitute a similarly limited sampling of it. In addition, as shown below (Fig. 6), the regions of the structural space covered by the two data sets are rather close to each other.

External set prediction

Most of the databases available for the study of problems like the one addressed here, i.e. the construction of a QSPR model, consist of compounds characterized by analogous structures. Thus, the assessment of the model generalization ability, even based on the LOO test, does only give a measure of the capacity of the model to make predictions inside the structural space it was developed on. Nevertheless, to be of use in a real application, a model should be as much generalizating as possible. Therefore, since the final models (model 1 and 3, Table 4) in this study were developed from data sets of compounds structurally related (Men_1 and Men_2, respectively), an analysis of their potential applicability to a real problem was simulated by evaluating their ability to make reliable predictions of the P_{app} relative to an external data set composed of 50 structurally diverse compounds taken from the literature [52]. The procedure consisted in training the network using the model specified descriptors (model 1 and 3, Table 4) on the corresponding data set Men_1 or Men_2 which the models were developed on. Once the two neural networks were trained and their parameters (weights and biases) were determined,

Table 4 Comparison of the best performing descriptor combinations of Men_1, Men_1_2 and Men_2 selected by the GA-NN[a]

No.	Data set	Model	LOO regression coefficient (R)[b]
1	Men_1	4 6 8 12 52 55 61 74 116 117 119	0.723 ± 0.006
2	Men_1_2	1 8 9 35 40 42 44 75 85 88 96	0.726 ± 0.009
3	Men_2	1 7 20 35 38 43 58 64 67 80 81 83	0.752 ± 0.002

[a] Parameters common to all the runs are: network architecture (12–4–1), crossover fraction (0.6), mutation rate (0.02), elite count (4), population size (50)

[b] Mean and standard deviation derived from three repetitions of the LOO regression coefficient calculus

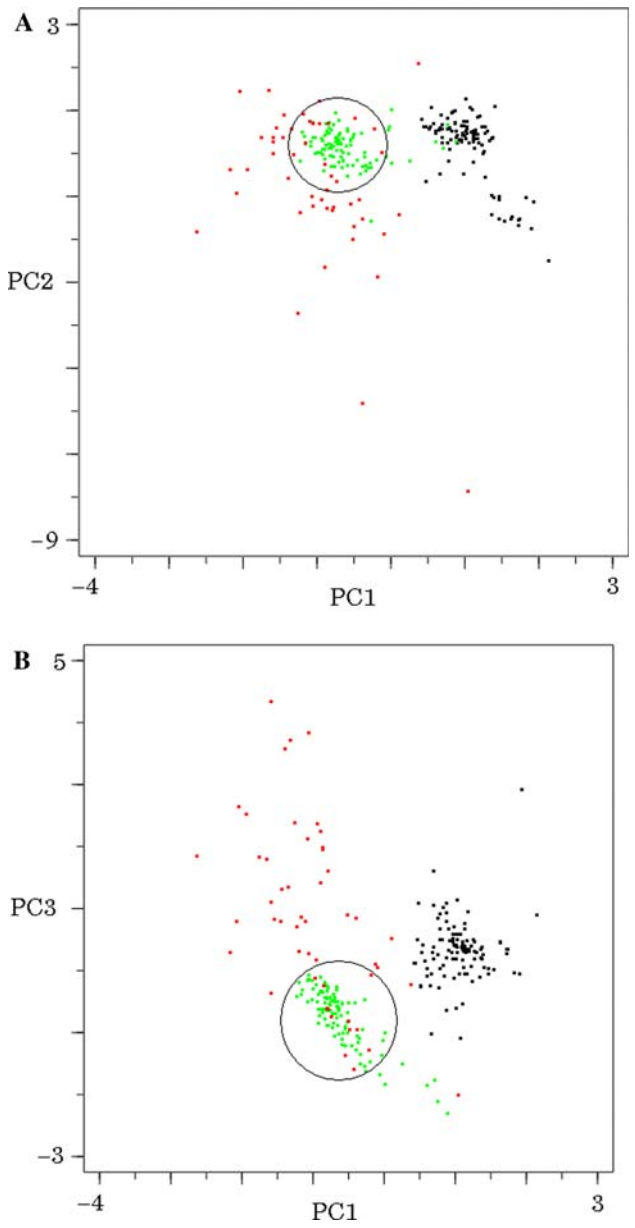

Fig. 6 Plots of the first two, PC1 and PC2, (**A**) and of the first and third, PC1 and PC3, (**B**) principal components for the ensemble of compounds of Men_1 (black), Men_2 (green) and the external set (red).

these were used as network function coefficients to calculate the P_{app} of the external set compounds.

Since the two models were both developed from a restricted structural space and with a rather high experimental error on the corresponding P_{app} values, quantitatively appealing results were not expected. However, the correlation coefficients between predicted and experimental P_{app} values of the external set corresponding to the two models were acceptable even if contrasting. Actually, the Men_1 model (1 in Table 4) produced $R = 0.40$, while the Men_2 model (3 in Table 4) produced $R = 0.61$.

The difference in predictive ability exhibited by the two models and the overall rather low R values are definitely explained by the different structural composition of the data sets. To demonstrate this, the analysis of the principal components (PCA) on the Volsurf descriptors of the data set constituted by the union of Men_1, Men_2 and the external set (Fig. 6) was carried out. The new descriptors obtained with PCA are a linear combination of the Volsurf ones and are linearly independent. Importantly, they are constructed and ordered in such a way that most of the information (technically, most of the explained variance) is contained in the first descriptors (the principal components or PCs).

In this case, the first three components contained respectively the 34%, 16% and 10% (that is the 60%) of the total variance. Thus, 2-D plots of the data using the first (PC1) and the second (PC2) or the third (PC3) components were built (Fig. 6 A, B), showing the relative location of the compounds belonging to the three data sets (Men_1, Men_2 and external set) in the obtained 2-D reduced structural space.

It is evident from both Fig. 6A and 6B that the Men_1 and Men_2 compounds are confined to small close regions of the diagram, confirming the existence of structure similarity inside each data set. On the contrary, the external set covers a wider region of the chart, indicating a higher structural diversity of its compounds. Moreover, it interestingly appears that a

significant part of the compounds of the external set occupy a region of the plot (circled area) common to that populated by some of the Men_2 molecules; whereas, points corresponding to Men_1 compounds (black) are never superimposed to those corresponding to the external set molecules (green). This indicates that the external set molecules are structurally more similar to those of Men_2 than to the Men_1 ones. Thus explaining the higher predictive ability displayed on the external set compounds of the model developed with Men_2 than that produced with Men_1.

The rather low regression coefficients found (0.61 and 0.40), thus, underline the limits of a data-based statistical approach, confirming that the predictive power of models, developed according to this procedure, is strictly connected to the structural space sampling of the compounds used for the training with respect to those to be predicted. Despite this, the results obtained here are still of fair practical applicability for a qualitative estimate, i.e. in terms of ''high/intermediate/poor'' answers, of the P_{app} given the molecular structure. In addition, better previsions on new molecules could be obtained building a GA-NN model with the compounds from Men_1 and Men_2 along with those from the ''external'' set, i.e., training the neural network on molecules which sample larger structural and property spaces.

Finally, it is interesting that the plot of the calculated versus experimental values of the external set (Fig. 7) shows the same trend as in Fig. 5: the error made by the neural network follows the pattern of the experimental error, i.e. it is higher on the lower values.

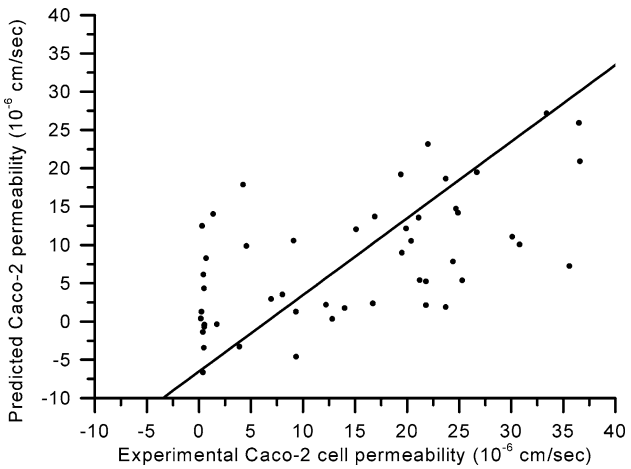

Fig. 7 A plot of experimental versus predicted Caco-2 cell permeability of an external data set produced from model 3 of Table 4

Physico-chemical meaning of the best descriptors selected by the algorithm

The qualitative meaning of the best descriptors selected using Men_1, Men_1_2 and Men_2 is shown in Table 5. The 166 Volsurf descriptors generated are divided in 160 GRID descriptors (depending on the GRID force-field [55, 56]) and six additional descriptors not depending on GRID maps. The GRID descriptors are indicated with the symbol for that particular descriptor, followed by the specific probe molecule or group used to characterize the polar or hydrophobic potential interaction sites. As anticipated above, the selection of different descriptor combinations, exhibiting similar predictive ability, at each different GA-NN run on the same data set, demonstrated that the descriptors generated with Volsurf are highly correlated. In other words, most of them share a great part of the carried information, at least with respect to the property under scrutiny (P_{app}). The same consideration can be made when descriptors selected using different data sets are compared. Thus, it is not surprising that the best descriptors selected by the GA-NN on the three data sets are different. However, below and in Table 5 it is pointed out that they correspond to descriptors having a similar physico-chemical meaning. Given these assumptions, it is remarkable that two descriptors, namely descriptors 78 and 93, are present in each of the three models (Tables 4 and 5) and indicate, respectively, W1 O- (hydrophilic regions) and Iw5 OH2 (integy moments). Before discussing in depth the meaning of these descriptors, it is advisable to examine also the other descriptors selected for each data set reported in Table 5, and how they are qualitatively and quantitatively related.

First of all it is clear from Table 5 that the selected descriptors are prevalently hydrophilic with a ratio of around 9 over 11 or 12 total descriptors, constant in each model, with respect to hydrophobic (1 or 3:11 or 12) and molecular shape (1:11 or 12). This is not unexpected since Volsurf descriptors are hydrophilic for more than 70%. Nevertheless, in our opinion, the choice of this kind of descriptors has been only partially influenced by this fact. The selected descriptor combinations possessed a good predictive ability indeed. If it is true that the selection was predetermined by the high probability to create an initial population with mainly hydrophilic descriptors, it is also true that low performance combinations were discarded during the GA-NN runs. Therefore the predominance of hydrophilic descriptors in the selected combinations is to be ascribed only to their strong correlation to the apparent permeability. Whereas, according to the

Table 5 Physico-chemical meaning of the best descriptors selected by the GA-NN

Data set Men_1	Data set Men_1_2	Data set Men_2	Descriptors class
(7) W6 OH2	(80) W3 O-	(37) W8 O	Hydrophilic regions
(72) W3 N3+	(70) W1 N3+	(72) W3 N3+	
(78) W1 O-	(78) W1 O-	(78) W1 O-	
–	(136) HB8 O	(148) HB4 N1	Hydrogen bonding
(158) HB6 N:=	(154) HB2 N:=	(149) HB5 N1	
(160) HB8 N:=	(151) HB7 N1	(151) HB7 N1	
(10) BV11 OH2	–	–	Hydrophilic best volumes
(12) BV31 OH2	(15) BV32 OH2	(15) BV32 OH2	
(93) Iw5 OH2	(93) Iw5 OH2	(93) Iw5 OH2	Hydrophilic Integy moments
(108) D12 OH2	–	–	Hydrophilic local int. energy min. dist.
(16) D1 DRY	(16) D1 DRY	–	Hydrophobic regions
–	–	(123) D13 DRY	Hydrophobic local int. energy min. dist.
–	(166) Log P	–	Water/octanol partition coeff.
–	–	(111) ID1 DRY	Hydrophobic integy moments
–	–	(120) Emin2 DRY	Hydrophobic local int. energy min.
(162) MW	–	–	Molecular Weights
–	(1) V OH2	(1) V OH2	Molecular volume

models, hydrophobicity has a minor, although still significant, role. Three 'hydrophilic regions' descriptors are present in each model. Among them the above mentioned W1 O- (which expresses the molecular envelope which is accessible to and attracts the anionic phenolate oxygen atom) is identical in all three models and must, therefore, have a high physico-chemical content related to apparent permeability. This is confirmed by the fact that W1 O- was calculated at a level of interaction energy (W1) which accounts for polarisability and dispersion forces, and polarity/polarisability is a molecular property which has a well known important role in the phenomenon of membrane permeation. The other descriptor which is constant in the three models is Iw5 OH2, which belongs to the 'INTEraction enerGY (integy) moments' class. It defines the unbalance between the centre of mass of a molecule and the position of the hydrophilic regions around it. Depending on the value of this descriptor, it is possible to distinguish between molecules having their hydrated regions clearly concentrated in just one part of the molecular surface, and those with polar groups either close to the centre of mass or at the opposite ends of the molecule. This descriptor can be an indication that the distribution of the polar groups on the molecular surface is relevant as well to endow a drug with the ability to penetrate cell membranes.

Hydrogen bond (HB) descriptors are those describing hydrogen bonding capability, another key molecular property for membrane diffusion. The amount of HB descriptors is rather constant and significant in each model: three descriptors in models Men_1_2 and Men_2, and two in model Men_1 (Table 5). The hydrophobic descriptor typology is rather diverse inside each model and these descriptors are not numerically equally distributed: only one descriptor (D1 DRY) in the Men_1 model, two in the Men_1_2 and up to three in the Men_2 model. It is very interesting that one of the hydrophobicity descriptors selected is Log *P*, whose importance in the intestinal absorption is well known and has been already postulated in the Lipinsky ''rule of five''. According to this rule and many subsequent studies, another important feature involved in oral absorption is the molecular weight. Our GA-NN selected separately the Molecular Weight and Molecular Volume descriptors, respectively, for one data set and for the other two, thus confirming the importance of these quantities in drug oral absorption and the ability of the GA-NN procedure to select out of a large descriptor pool only a small set of them which are important in the description of the membrane permeability phenomenon.

Conclusions

A GA-NN procedure to model Caco-2 apparent permeability, a property which is correlated with oral absorption, has been developed. The final performance of the selected models was evaluated by the leave-one-out (LOO) procedure that provides a better assessment of the model generalization capability. A GA-NN directly based on the LOO procedure would be too computationally expensive. The models derived from the application of the GA-NN on two independent data sets are characterized by good and similar LOO regression coefficients ($R = 0.723$ and $R = 0.752$), despite the rather high experimental error affecting the measurements of Caco-2 permeability used for their

development and the approximations (use of a single molecular conformation, error on the estimate of the different ionic state molar fractions) introduced to make the procedure very fast. The fact that the GA-NN selection method performs equally well on data sets consisting of compounds covering close regions of the structural space and affected by the same experimental error confirms that the results obtained were not fortuitous.

The models obtained were also tested on an external set of molecules with rather good results, if the already mentioned experimental error and the limited structural space encompassed by the molecules employed for the model construction are considered. In this case, as expected, strong data set composition dependence for the prediction has emerged. Therefore our study underlines that the prediction of Caco-2 cell permeability is actually more limited by the scarce availability of large and varied databases of molecules with known permeability values than by the existence of valid and robust regression techniques.

The physico-chemical meaning of the structural descriptors constituting the models developed emphasizes the importance of four main features correlating with Papp such as hydrophilicity and hydrogen bond propensity, hydrophobicity and their function Log *P*, besides molecular weight or volume.

As a final conclusion, the GA-NN procedure developed here to model oral absorption can be a valid tool in early drug design research and its application to real problems can be actualized and optimized by using large data sets with enough structural space diversity and measures of the target property with low enough experimental errors.

Acknowledgment A. D. F. is grateful to Menarini Ricerche SpA for a post-doctoral fellowship allowing her to carry out research activity at IPCF, and to Dr. Giovanni Barcaro (IPCF, CNR, Pisa), Dr. Roberto Bartolini (ILC, CNR, Pisa) and Dr. Marco Galimberti (IPCF, CNR, Pisa) for several helpful discussions. We thank Dr. Rose-Marie Catalioto (Menarini Ricerche S.p.A, Firenze) who performed the Caco-2 P_{app} measurements and Dr. Antonio Triolo (Menarini Ricerche S.p.A., Firenze) for MS analysis.

References

1. Lipinsky CA, Lombardo F, Dominy BW, Feeney P (1997) Adv Drug Deliv Rev 23:3
2. Yamashita F, Hashida M (2004) Drug Metab Pharmacokin 19:327
3. Motulsky H, (1995) Intuitive biostatistics, Oxford University Press, New York
4. Song XH, Yu RQ (1993) Chemom Intell Lab Syst 19:101
5. Hirst JD, King RD, Sternberg MJE (1994) Comp Aided Mol Design 8:405
6. Breindl A, Beck B, Clark T (1997) J Mol Model 3:142
7. Goll ES, Jurs PC (1999) J Chem Inf Comput Sci 39:974
8. Patankar SJ, Jurs PC (2000) J Chem Inf Comput Sci 40:706
9. Yang L, Wang P, Jiang Y, Chen J (2005) J Chem Inf Model 45:1804
10. Wessel MD, Jurs PC (1995) J Chem Inf Model 35:841
11. Duprat AF, Huynh T, Dreyfus G (1998) J Chem Inf Comput Sci 38:586
12. Hildalgo IJ, Raub TJ, Borchardt RT (1989) Gastroenterology 96:736
13. Artursson P (1990) J Pharm Sci 79:476
14. Artursson P, Karlsson J (1991) Biochem Biophys Res Commun 175:880
15. Delie F, Rubas WA (1997) Crit Rev Ther Drug Carrier Syst 14:221
16. Fujiwara S, Yamashita F, Hashida M (2002) Int J Pharm 237:95
17. Yamashita F, Wanchana S, Hashida M (2002) J Pharm Sci 91:2230
18. Molconn-Z software, Hall Associated Consulting, Quincy, MA
19. Artursson P, Palm K, Luthman K (1996) Adv Drug Deliv Rev 22:67
20. Cruciani G, Crivori P, Carrupt PA, Testa B (2000) THEOCHEM -J Mol Struc 503:17
21. Sisto A, Caciagli V, Altamura M, Giolitti A, Fedi V, Guidi A, Giannotti D, Harmat N, Nannicini R, Pasqui F, Maggi CA, WO03037916, MENARINI RICERCHE SPA and inventors, 2003
22. Fattori D, Porcelloni M, D'Andrea P, Rossi C, Altamura M, Maggi CA, WO2004094412, MENARINI RICERCHE SPA and inventors, 2004
23. Pearlman RS, Concord distributed by Tripos Inc., St. Louis, Missouri, 63144, USA
24. Crivori P, Cruciani G, Carrupt PA, Testa B (2000) J Med Chem 43:2204
25. Cruciani G, Pastor M, Clementi S (2000) In: Gundertofte K, Jørgensen FS (eds) Molecular modeling and prediction of bioactivity. Springer, Berlin, pp 73–82
26. Guba W, Cruciani G (2000) In: Guntertofte K, Jørgensen FS (eds) Molecular modeling and prediction of bioactivity. Springer, Berlin, pp 89–94
27. ACD/pKa Batch, Advanced Chemistry Development, Inc.: Toronto ON, Canada, www.acdlabs.com, 2003
28. Bishop CM, (ed) (1995) Neural networks for pattern recognition. Oxford University Press Inc., New York
29. Wessel MD, Jurs PC, Tolan JW, Muskal SM (1998) J Chem Inf Comput Sci 38:726
30. Agatonovic-Kustrin S, Beresford R, Yusof APM (2001) J Pharm Biomed Anal 25:227
31. So SS, Karplus M (1996) J Med Chem 39:1521
32. So SS, Karplus M (1996) J Med Chem 39:5246
33. Yasri A, Hartsough D (2001) J Chem Inf Comput Sci 41:1218
34. Guha R, Jurs PC (2004) J Chem Inf Comput Sci 44:1440
35. Marini F, Roncaglioni A, Novič M (2005) J Chem Inf Model 45:1507
36. Rogers D, Hopfinger AJ (1994) J Chem Inf Comput Sci 34:854
37. Leardi R (2001) J Chemometr 15:559
38. Lavine BK, Davidson CE (2003) J Chem Inf Comput Sci 43:1890
39. Luke BT (1994) J Chem Inf Comput Sci 34:1279
40. Zupan J, Gasteiger J (eds) (1999) Neural networks in chemistry and drug design Wiley-VCH, Weinheim
41. Levenberg K (1944) Quart Appl Math 2:164
42. Marquardt D (1963) SIAM J Appl Math 11:431

43. Gill PE, Murray W (1978) SIAM J Numer Anal 15:977
44. Hagan MT, Menhaj M (1994) IEEE Transactions on Neural Networks 5:989
45. Nguyen D, Widrow B (1990) Proc Int Joint Conference Neural Networks 3:21
46. Nelson MC, Illingworth WT (eds) (1991) A practical guide to neural nets, Addison-Wesley, Reading, MA USA
47. Stone M (1974) J R Statist Soc B 36:111
48. Stone M (1978) Math Operationsforsch Statist Ser Statistics 9:127
49. Wahaba G, Wold S (1975) Commun Statist 4:1
50. Perrone MP, Cooper LN (1993) In: Mammone RJ (ed) Artificial neural network for speech and vision. Chapman & Hall, London, pp 126–142
51. Perrone MP (1993) In: Mozer MC et al (eds) Proceedings Connectionist Models Summer School, Lawrence Erlbaum, Hillsdale NJ, pp 364–371
52. Yazdanian M, Glynn SL, Wright JL, Hawi A (1998) Pharm Res 15:1490
53. Stuper AJ, Brugger WE, Jurs PC (eds) (1979) Computer-assisted studies of chemical structure and biological function. Wiley-Interscience, New York
54. Lipinski CA (2000) J Pharmacol Toxicol 44:235
55. Goodford PJ (1985) J Med Chem 28:849
56. GRID 22, Molecular Discovery Ltd., http://www.moldiscovery.com