Optimizing the Performance of In Silico ADMET General Models According to Local Requirements: MARS Approach. Solubility Estimations As Case Study

Julen Oyarzabal,*[,||,†] Joaquin Pastor,[||,‡] and Trevor J. Howe[§]

Departments of Molecular Informatics and Medicinal Chemistry, Johnson & Johnson Pharmaceutical Research and Development, Jarama 75, 45007 Toledo, Spain and Department of Molecular Informatics, Johnson & Johnson Pharmaceutical Research and Development, Turnhoutseweg 30, 2340 Beerse, Belgium

Received August 17, 2009

The quality of in vitro data used to build in silico absorption, distribution, metabolism, and toxicity (ADMET) models is, in many cases, inconsistent. The paucity of data from single laboratory sources has led to the mixing of data sets with varying experimental conditions and to the coverage of restricted chemical space in models which are purported to be of general applicability. In order to overcome these shortcomings, a method, Metropolis/Monte Carlo adaptive ranking simulation (MARS) has been developed. This aims to estimate "optimal flexible threshold points" in order to achieve better correlation between any in silico ADMET model and any discrete qualitative experimental data. The MARS method covers three key factors: the predictive model, the experimental procedure for the assay, and the chemical series or scaffold. When large and general solubility data sets (>650 compounds) are analyzed against commercially available in silico models, using MARS, an improvement in κ statistics up to 16.2% is obtained. When particular chemical series are addressed, improvements up to 46.0% are seen on κ statistics. This coefficient then allows an investigation into the effectiveness of a classifier by assessing the improvement over chance. These improvements in ranking estimations allow more predictive decision-making for virtual libraries.

INTRODUCTION

Classical strategies in the drug discovery process now focus on optimizing the ADMET properties of compounds simultaneously with the inherent pharmacodynamic profile of the biological target. Strategies which ignore ADMET often deliver drug development candidates that are not sustainable in development phases. Compounds with suboptimal pharmacokinetic profiles often end in failure (or at least experience a deceleration) during the development process.[1−3] Failure in ADMET remains a primary cause in the high attrition rate for drug candidates. Out of approximately 250 compounds entering preclinical development, only five are clinically tested, and only one is ultimately approved by the FDA.[4] In many of these failures, ADMET factors[5−10] are cited as a cause.

Most of the pharmaceutical industry incorporates ADMET optimization very early in the drug discovery process,[9,11] adopting the "fail early, fail fast"concept[2,12] to decrease attrition rates.[13−15] In filtering out compounds with poor biopharmaceutical properties in early stage candidates, a battery of in vitro ADMET screens are currently implemented and applied[16] with varying degrees of certainty and success. The utility of in silico ADMET tools, therefore, needs to be demonstrated against in vitro data in the same range as the compounds and the assays used to classify them.

Many ADMET assays provide qualitative rather than quantitative values, and so some predictive classification methods have been developed[12,14,17−22] to rank virtual,[23] proposed synthetic candidates. These issues of data quality appear to encourage the development of "threshold" over "continuum" methods. Along the past decade, quantitative models for the prediction of ADMET properties have been developed and validated, so threshold methods could be derived from them. Some reported models handle predefined general and static threshold points[12,18−20,22] based on previously reported experimental data to classify compounds. To get the best performance from such predictive models, in a situation where a correlation between in silico estimations and in vitro experimental data is sought, a Metropolis/Monte Carlo adaptive ranking simulation (MARS) has been developed. This explores the ranking space of an ADMET property to determine optimal threshold point(s), which will describe the optimal performance for an in silico estimation against discrete experimental data. Using this approach, optimal and flexible threshold points are determined dependent on the chemical series under analysis, the experimental procedure, and/or the in silico model. Thus, an informed reading of the estimations contained in commercial generalized ADMET models is obtained.

This approach has been already utilized to evaluate different commercial models predicting solubility, which plays a key role in absorption, against Johnson and Johnson's (J & J's) in-house high-throughput (HT) solubility experimental data set.[24]

* Corresponding author. Telephone: +34 91 7328000. E-mail, joyarzabal@cnio.es.

|| Current address: Spanish National Cancer Research Centre (CNIO), Experimental Therapeutics Programme, Melchor Fernandez Almagro 3, 28029 Madrid, Spain.

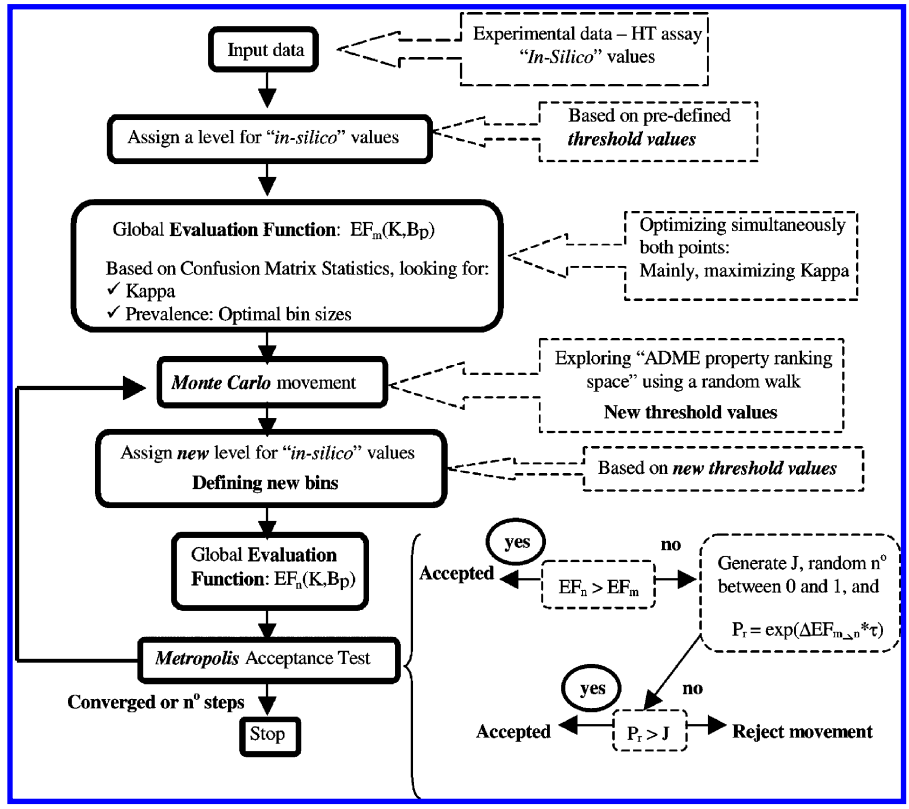
† Department of Molecular Informatics, Johnson & Johnson Pharmaceutical Research and Development.

‡ Department of Medicinal Chemistry, Johnson & Johnson Pharmaceutical Research and Development.

§ Department of Molecular Informatics, Johnson & Johnson Pharmaceutical Research and Development.

10.1021/ci900308u CCC: $40.75
Published on Web 11/17/2009

Scheme 1. General Flowchart for the MARS Code



METHODS

The fundamental precept of the MARS approach is to achieve the best performance for an in silico model through optimal binning. The first step is the development of a probability statistics method to define the threshold values for optimal binning in a flexible manner. This is a function of the given ADMET continuum model, the experimental procedure and the chemical series. The ADMET property ranking space is, thus, explored by a Metropolis algorithm (acceptance test) on Monte Carlo movements (a stochastic technique) using a random walk. The degree of accuracy is determined by comparing the number of binned compounds in each level using the tandem in silico model and MARS against experimental assays. Consequently, relevant statistical measurements, overall accuracy and κ parameters (derived from the corresponding confusion matrix) are obtained during the performance analysis. The κ coefficient can then be defined as the main value for the evaluation function in the Metropolis algorithm. Scheme 1 summarizes the flowchart of the MARS encoded in Perl.

Confusion Matrix Statistics. The confusion matrix, sometimes called the error matrix, is almost universally adopted as the standard error report in supervised classification. The matrix is a symmetrical array of numbers, which express the number of classified compounds in the assigned category relative to the actual category from in vitro experimental assays. The simplest, and most widely used, measure of classification prediction accuracy, known as overall accuracy, is counting the proportion of cases that fall along the diagonal, the number of correctly classified cases, divided by the weight of total cases. This is shown in Table 1.

Alternatively, Cohen has proposed an additional measure, the κ coefficient,[25] κ, for measuring agreement in confusion

Table 1. Confusion Matrix[a]

in silico model	predicted		
experimental	class 1	class 2	total
class 1	a	b	a + b
class 2	c	d	c + d
total	a + c	b + d	a + b + c + d

[a] Where, a is true positive, b is false negative, c is false positive, and d is true negative. Overall accuracy is (a + d)/(a + b + c + d).

matrices. The principle of this statistic is to measure the difference between the agreement of the working process and the agreement of chance in the classification process. The κ statistics can be interpreted as a measure of agreement that exists beyond the degree expected by chance alone. This coefficient makes full use of the information contained in the confusion matrix, not only the diagonalization (overall accuracy) but also those entries containing the number of false positives and negatives. The κ coefficient is the optimal measurement to determine if a classifier predicts better than chance.[26]

The κ coefficient is defined as

$$\kappa = (P_o - P_e)/(1 - P_e)$$

Subject to the numbers of observed (P_o) and expected (P_e) compounds, the percentage of their difference is computed as the coefficient. Observed compounds are represented by accurate classified compounds, which are from the normalized sum of diagonal elements (*ii*). Expected compounds are represented by the designated number of compounds in the correct classification, which are computed from normalized values of the row (*i*) and column (*j*) totals in the confusion

matrix. The mathematical representation of P_o and P_e are given by the following equation:

$$P_o = \sum_{i=1}^{t} P_{ii}$$

$$P_e = \sum_{i=1}^{t} P_{i+}P_{j+}$$

For the computational purposes, the κ coefficient is determined as

$$\kappa = \frac{\sum_{i=1}^{t} P_{ii} - \sum_{i=1}^{t} P_{i+}P_{j+}}{N^2 - \sum_{i=1}^{t} P_{i+}P_{j+}}$$

where, t is the number of rows in the confusion matrix, and N is the total number of data points in the matrix (e.g., compounds).

A higher value indicates that the result is in better classification than the random assignment. For example, κ with a value 0.56 can be interpreted as a "classifier", achieving an accuracy of 56% better than the chance assignment of compounds to categories. In this paper, κ coefficient values are reported as percentages (% κ) in order to highlight the improvement over chance, i.e., % $\kappa = \kappa \times 100$.

The κ coefficient, therefore, utilizes all the information contained in the confusion matrix and provides a meaningful measurement for exploration of property "ranking space". In developing QSAR/QSPR models, κ will be maximized in order to improve the overall accuracy of the in silico predictive power.[27,20−22]

Metropolis Monte Carlo. Depending on the number of bins required to correlate quantitative estimations with qualitative determinations, a number of starting threshold values are defined to set up the simulation. At each iteration of the Monte Carlo simulation, new threshold values are generated by making a random change to the starting threshold value movements in both directions, positive and negative, using a random number generator. A unique random number is generated for each threshold value under study. Therefore, in the case where there is more than one threshold value, each one of them will move in an independent manner. The size of move at each iteration is governed by the maximum possible displacement in any direction, δr_{max}. This is an adjustable parameter, the value of which is usually chosen so that approximately 50% of the trial moves are accepted. If the maximum displacement is too small, then many moves will be accepted, but the "ranking space" will only be explored very slowly. A large simulation would then be needed to be sure that the entire space has been properly explored. Too large a value of δr_{max} will mean that many trial moves will be rejected because they lead to unfavorable overlaps. The maximum displacement is adjusted automatically "on the fly" to achieve an acceptance ratio between 40 and 60%. The main goal when using Metropolis Monte Carlo simulations is to achieve a proper exploration of the ranking space and to find out the optimum threshold values which maximize κ.

Evaluation Function, EF(κ, B_p). The Metropolis algorithm generates a Markov chain of states. Therefore, the outcome of each trial depends only upon the preceding trial and not upon any previous trials, where each trial belongs to a finite set of possible outcomes. The probability of accepting a trial move from m, the preceding state, to n depends on the probability ratio of the two states m and n. In the Metropolis method, a new threshold value n, obtained from the random walk in the ranking space, is accepted if its probability function is better than that of the original state m. In this case, if the evaluation function of state n is higher than that of m. Consequently, the new threshold value is retained as the starting point for the next iteration. If the new threshold value obtained after an additional movement drives the probability function to a worse value than its predecessor, then the probability ratio of these two final states is compared to a random number between 0 and 1.

$$EF(\kappa, B_p) = \alpha(\kappa \times 100) + (\chi \times B_p)$$

where, κ is the κ coefficient from the confusion matrix statistics, α is 0.9, χ is 0.1, and B_p is

$$B_p = \sum_{i=1}^{n} w_i$$

If the probability ratio, P_r, is greater than the random number, then the new threshold value is accepted. If not, then it is rejected, and the initial threshold value is retained for the next move. If the probability function of the new state n, P_n, is very close to that of the old state m, P_m, then the probability ratio will be very close to 1, and so the move is likely to be accepted.

This probability ratio is obtained after applying the corresponding probability functions (P_m and P_n) for the two states m and n:

$$P_r = P_n/P_m = \exp(\Delta EF_{m\to n}(\kappa, B_p) \times \tau(T))$$

$$\text{where,} \qquad P_n = \exp(EF_n(\kappa_n, B_{pn}) \times \tau(T))$$

$$P_m = \exp(EF_m(\kappa_m, B_{pm}) \times \tau(T))$$

$$\tau(T) = 375/T(^\circ K)$$

$$\Delta EF_{m\to n}(\kappa, B_p) = EF_n(\kappa_n, B_{pn}) - EF_m(\kappa_m, B_{pm})$$

This acceptance condition can be more concisely written, thus:

$$\text{rand}(0,1) \le \exp(\Delta EF(\kappa, B_p) \times \tau(T))$$

As is described, the principal goal to achieve during this simulation is to maximize the κ statistics. Hence, the evaluation function is mainly κ dependent, although another parameter, bin population dependence (B_p), is also considered, as shown above.

Bin population (prevalence) plays an important role in the optimal binning task. For example, unequal group sizes after new bins have been defined can influence the scores for many of the classifier methods, such as κ.[26] Data sets considered in this report do not have any extreme bin populations, all

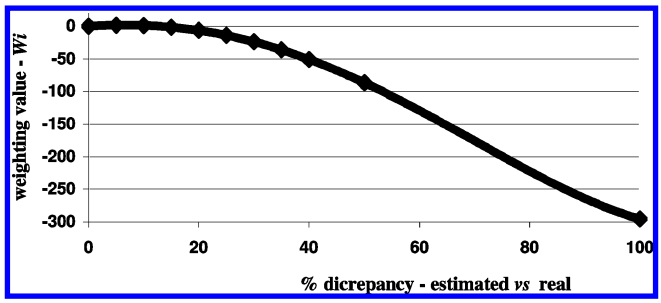


Figure 1. Polynomial function applied as a weighting factor to the evaluation function.

prevalence values are bigger than 0.1 and lower than 0.9. Therefore, an additional function, B_p, has been introduced in this equation, with a moderate influence (10%) so designed to keep estimated bin populations as close as possible to the real experimental data population. Where prevalence values are acceptable, κ is maximized.

This function, B_p, is based on the sum of independent weighting factors. There are as many weighting factors, w_i, as bins are required, and these are calculated "on the fly" from a polynomial function:

$$w_i = (4 \times 10^{-6}) \times x^4 - (4 \times 10^{-4}) \times x^3 - (0.0345 \times x^2) + (0.4937 \times x) + 0.129$$

where, x is the percentage of discrepancy between the real population at a particular bin and the estimated population of that bin.

Depending on how far these estimated bin populations are from a real data set, the weighting factors have negative (or no) influence on the final evaluation function as shown in Figure 1.

In most cases, this Monte Carlo simulation comprises of an equilibration phase followed by a production phase. During equilibration, the bin populations are monitored until they achieve stable values around the real experimental data. After this is achieved the production phase can commence.

The $EF(\kappa, B_p)$, is slightly different during the equilibration phase than during production phase, since the early stages are mainly focused on having optimal group sizes. Therefore:

$$EF(\kappa, B_p) = \alpha(\kappa \times 100) + (\chi \times B_p)$$

At the equilibration phase, both α and χ are 0.5, but at production phase, α and χ are 0.9 and 0.1, respectively.

In order to avoid underflow and overflow during the simulation, those movements involving large changes in the evaluation function sign, positive or negative, will be directly rejected without any further analysis, as shown in the equation below:

$$|\Delta EF_{m \to n}(\kappa, B_p)| > \upsilon(T)$$

$$\text{where,} \qquad \upsilon(T) = T(^{\circ}\text{K})/33$$

This is a classical Monte Carlo simulation which samples from the canonical ensemble. Therefore, the number of compounds (N), the temperature (T), and the volume (V) are fixed. The simulation will stop after a certain number of predefined steps or after convergence is reached. Convergence is defined as a fluctuation in the value of the evaluation function lower than 5 per 1 000 during 1 000 consecutive steps.

Table 2. Commercial Models Utilized for This Analysis

model	implemented in:
Accelrys[12]	accord for Excel v.5.2. (Accelrys)[37]
Syracuse[38,39]	accord for Excel v.5.2. (Accelrys)[37]
ACD	ACD/solubility DB v.6.0. (ACD)[40]
S+Sw	ADMET predictor v.1.1.0. (Simulation Plus) [41a]
MHSw[39]	ADMET predictor v.1.1.0. (Simulation Plus) [41a]
QPmmff[42b]	QikProp v.2.2. (Schrödinger)[43]

[a] Experimental melting points were not used as input for these estimations. [b] Three-dimensional structures were generated using CORINA[44] and were minimized using MMFF.[45−49]

DATA. The main aim of MARS is to explore the ranking space of an ADMET property under analysis and to find the optimal threshold point(s). By using the optimized threshold points in an in silico model, it is expected to achieve the best performance for that model. To validate this novel approach, solubility has been selected, as ADMET property case study, due to one of the key elements of compound characterization during the whole drug discovery and development process: from hit identification and lead optimization to formulation.[28] Thus, during these last years, many efforts are still being devoted to not only developing new in silico models[29−31] but also to validate their utility, as predictive tools, through the solubility challenge.[32,33] Solubility plays a key role in the absorption phase of the pharmacokinetic profile. If the solubility and the rate of dissolution are too low, an orally administrated drug will mostly be excreted without passage into the cardiovascular system.[34] The Food and Drug Administration (FDA) uses solubility together with passive permeability in their biopharmaceutics classification system (BCS).[35] This is used as a guide for the in vivo bioavailability and bioequivalence studies for immediate-release solid oral dosage forms[36] of drug substances.

Several commercial in silico solubility models are readily available and can be evaluated against measured kinetic solubility; in this case, a set of solubility data were used from in-house chemistry and from a single experimental protocol based on turbidity.[24] Thermodynamic solubility is less frequently used for in silico model preparation and validation, since extensive diverse data sets are not readily available. The assay is time-consuming and requires relatively large amounts of pure material not normally prepared in current early phase drug discovery programs. Kinetic solubility measures the propensity for a compound to stay in solution upon dilution, creating supersaturated solutions. This can be measured in a high-throughput mode with little sample.

Table 2 shows the commercial models used for comparison in this analysis, these models were compared with in-house in vitro data from a nephelometry assay.[24] In this analysis, 1 625 experimental data points are considered. The in vitro assay is run at two different starting concentrations, 5 and 20 mM based in dimethylsulfoxide (DMSO) stock solutions. At 5 mM, 968 experimental values were available, and at 20 mM, there were 657 data points.

Chemical space together with the experimental conditions are the two main variables affecting the scope of general in

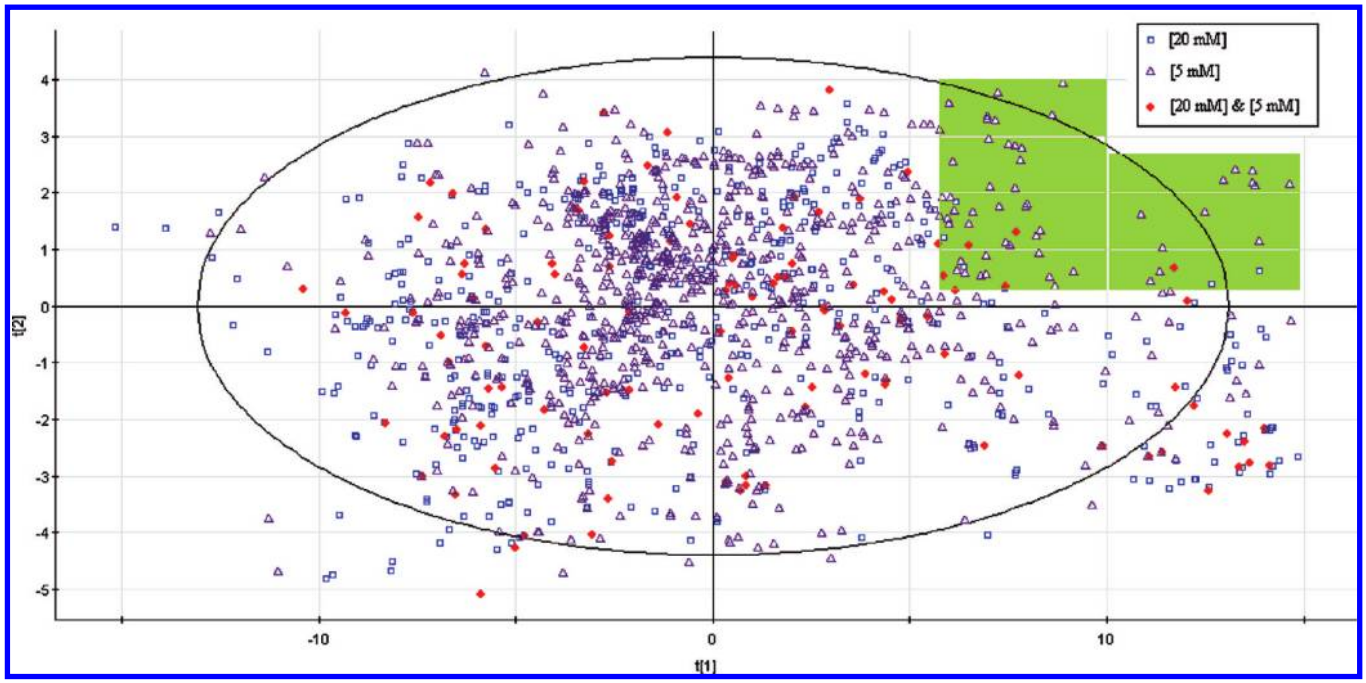


Figure 2. Chemical space covered by those two data sets under analysis and the overlapping subset.

silico models. Thus, both variables are explicitly considered in this analysis. There is, however, little information available on chemical space and experimental conditions for the training sets utilized to build the commercially available general models. Therefore, the retrieved data sets are focused also to analyze the impact of these key points on the overall performance. These data sets come from two different experimental conditions. In order to asses their respective coverage of chemical space, a set of 101 descriptors was calculated for all compounds in each data set, using Cerius[2] 4.9,[50] which includes descriptors of structural, topological,[51−53] and electrotopological indices known as E-state keys.[54] The dimensionality was reduced by the principal component analysis (PCA), using SIMCA-P+ 10.0,[55] to give just two components which account for 85.8% of the variance. Both data sets were treated in this process, and so comparison of the same scoring is a valid and straightforward process as Figure 2 shows. Both chemical spaces are, approximately, covering the same space; although the 20 mM data set is not fully populating the area corresponding to the $t^1 > 6$ and $t^2 > 0$ scores, colored in green.

Of these data, 118 compounds have been assayed at both concentrations. These compounds are represented in the Figure 2 as red dots, and they give a good coverage of the chemical space represented by the two data sets of which they are a common subset. Discrepancies in the experimental responses, 5 mM vs 20 mM, are present in 26.3% of these 118 compounds. This is very useful to evaluate the impact of experimental conditions on models performance since chemical space, the additional variable, is kept constant. Therefore, these compounds are considered as additional independent data sets.

Two chemical series[56−60] enclosed in these data sets, scaffolds I and II, (Figure 3) will be explicitly analyzed to illustrate this approach in a typical drug discovery based approach based on a compound class.

Solubility Classification Thresholds. The predicted solubility of the 1 625 compounds has been calculated from the

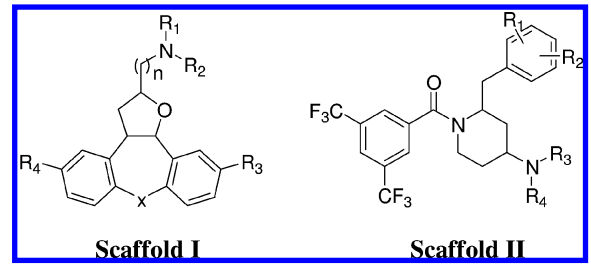


Figure 3. Chemical series under analysis.

in silico models described in Table 2 and ranked into different bins/classes. The binning of the data is the key question, what are the threshold points which define a compound as soluble and insoluble or as having low solubility?

According to the solubility range definition described in Table 3, there is good agreement between solubility classification for compounds assayed by a HT approach, using DMSO stock solutions and nephelometry, and by a HPLC method;[61] thus, the same criteria are utilized to define the experimental classification used here since assay conditions are similar.[24] The fixed threshold values to evaluate the performance of these in silico models versus the J & J HT solubility data are, thus, also taken from Table 3: $x_1 = -4$ and $x_2 = -5$. Logarithmic value of the aqueous solubility (S, mol/L) for the upper limit (x_1), which are considered to be soluble compounds, is −4 and the corresponding value for the lowest limit of solubility (x_2), insoluble compounds, is −5.

If ranking threshold values were not assessed for each solubility experimental method, then solubility range definitions reported in literature might be useful to play that role. There is a consensus for the upper limit (x_1), as it is reported in Table 4. It is popularly argued that the minimum required solubility for a drug having an average dose (1 mg/kg) and average intestinal permeability is 52 μg/mL. In molar terms, that value is equivalent to a log S of −4, for a compound having a molecular weight of 500 Da;[8,15] therefore, our upper

Table 3. Compounds Are Ranked Based on the Experimental Solubility Values Obtained by This High-Throughput Solubility Method[61]

solubility range definition	solubility value (μg/mL)	log solubility value, Log S (mol/L)[a]
sparingly	<10	< -4.7
partially soluble	10−100	> -4.7 and < -3.7
soluble	>100	> -3.7

[a] For a compound having a molecular weight near to 500 Da.

Table 4. Minimum Required Solubility Values to Consider a Compound As Soluble

solubility value (μg/mL)	log solubility value, Log S (mol/L)[a]
52[8,15]	−4
100[9,18,62]	−3.7
100[b,63,64]	−3.7
50	−4[12]

[a] For a compound having a molecular weight near to 500 Da. [b] At any pH from 1 to 8.

Table 5. Lowest Limit for Solubility Values to Consider a Compound As Insoluble

solubility value (μg/mL)	log solubility value, Log S (mol/L)[a]
10[61]	−4.7
1[65]	−5.7
0.5	−6[12]

[a] For a compound having a molecular weight near to 500 Da.

limit (x_1) is set up as −4. But defining the lowest limit for solubility (x_2) shows some discrepancies in literature, as seen in Table 5.

Thus, based on these two possible sets of fixed threshold points: (i) $x_1 = -4$ and $x_2 = -5$, optimal for J & J experimental conditions, and (ii) $x_1 = -4$ and $x_2 = -6$, our set of 1 625 compounds are classified according to their predicted solubility, log S (mol/L).

RESULTS AND DISCUSSION

Applying Fixed Threshold Points. Each of the two data sets (5 mM and 20 mM) are analyzed, and different statistics (% κ coefficient and overall accuracy) from the obtained confusion matrices are derived and shown in Table 6.

Three important observations can be made here. First, depending on the starting concentration, the same model has different performances, % κ coefficients, and consequently different % overall accuracy. This may indicate that the in silico models have been developed using training sets, which experimental conditions might not be identical in all cases — some performing better with a starting concentration of 5 mM and others with a starting concentration of 20 mM. Consequently, if it were the case, some models should perform better than others for given experimental conditions. But due to chemical space not being identical for the two data sets, this is an additional variable. Thus, analyzing those 118 compounds assayed at both concentrations, well spread in chemical space (fixed threshold points, $x_1 = -4$ and $x_2 = -5$, Table 7) show there are cases where there are significant discrepancies in the performance of the model. The Δ % κ >10; therefore, some models clearly perform better at certain experimental conditions than at others. Accelrys is better at 5 mM, and QPmmff is better at 20 mM. This would in turn suggest that their training sets might be enriched with data coming from particular experimental conditions.

Second, each in silico model provides different performance for the same data set, same experimental conditions, and same chemical space. This suggests that training sets used to develop these models cover different chemical spaces or/and come from different experimental conditions.

Third, as Figure 4 shows, depending on the selected fixed threshold points different performances are achieved. In general, better results are obtained if those values coming from closer experimental conditions to our in-house assays are utilized.

Applying Flexible Threshold Points to Large Data Sets. The three observations made from the previous analysis suggests that by using fixed threshold points the in silico models are not being utilized in the optimal manner. Due to these models are not applied in their appropriate context, which is unknown, they underperform. Through this work, by applying MARS, optimal flexible threshold points will be defined depending on the chemical series under analysis, the experimental procedure and/or the in silico model. Therefore, the performance of these general models is maximized according to the requirements. Then, the comparison among those in silico models under evaluation is proper since all of them are considered in their best scenario, i.e., optimized performance, against solubility data sets (Figure 5).

Table 8 shows that by using MARS and defining flexible threshold points (x_1 and x_2), each models performance is improved, not only in the percentage of κ (up to 16.2% at 5 mM and up to 14.1% at 20 mM) but also, as it was assumed, in the percentage of overall accuracy (up to 10.1% at 5 mM and up to 13.1% at 20 mM) to a significant level, comparing these results against those obtained after applying optimal fixed threshold values ($x_1 = -4.00$ and $x_2 = -5.00$), Table 6. If these results are compared after using MARS against those obtained after applying the other set of fixed threshold points ($x_1 = -4.00$ and $x_2 = -6.00$), there are very good improvements (Figures 6 and 7).

These results suggest that each model benefits from personalized flexible threshold points, depending on experimental conditions. This may be due to each model has been developed using different training data and our assay provide different results at different experimental conditions. For example, as Table 8 shows, at 5 mM, x_1 goes from −3.6 (ACD model) to −5.6 (MHSw model) to get the optimal binning/classification for each model. Furthermore, if the same model is inspected at the two different concentrations, threshold values have to vary to get the optimal answer at both concentrations, i.e. at 5 mM the ACD model has as threshold values $x_1 = -3.60$ and $x_2 = -4.39$ and at 20 mM the values go to $x_1 = -2.95$ and $x_2 = -3.95$.

An interesting case may be exemplified from Table 8 at 20 mM. An estimated solubility value (log S) of −3.96 would

Table 6. Statistics for the Overall Accuracy and κ[a]

	% over. acc.	% κ	x_1, log S(mol/L)	x_2, log S (mol/L)	% over. acc.	% κ	x_1, log S (mol/L)	x_2, log S (mol/L)
[5 mM] − 968 compounds								
Accelrys	59.3	35.2	−4	−5	54.5	33.7	−4	−6
Syracuse	58.1	33.0	−4	−5	55.5	34	−4	−6
ACD	55.7	29.5	−4	−5	49.1	23.4	−4	−6
S+Sw	63.0	43.5	−4	−5	52.5	34.5	−4	−6
MHSw	50.4	19.5	−4	−5	46.5	22.4	−4	−6
QPmmff	60.3	36.3	−4	−5	57.3	36.1	−4	−6
[20 mM] − 657 compounds								
Accelrys	62.2	34.9	−4	−5	45.8	24.0	−4	−6
Syracuse	67.4	37.5	−4	−5	60	33	−4	−6
ACD	61.6	39.6	−4	−5	40	18.5	−4	−6
S+Sw	70.9	48.7	−4	−5	42.5	28	−4	−6
MHSw	62.7	24.2	−4	−5	53.9	22.6	−4	−6
QPmmff	58.9	31.5	−4	−5	50.5	25.7	−4	−6

[a] Statistics were obtained from the corresponding confusion matrices after analyzing both data sets. Two different sets of fixed threshold points, x_1 and x_2, have been used to classify compounds.

Table 7. % κ Coefficients Obtained from Fixed Threshold Points[a]

118 compounds	Accelrys[b]	Syracuse[b]	ACD[b]	S+Sw[b]	MHSw[b]	QPmmff[b]
5 mM	29.77	19.74	36.65	27.18	16.46	24.94
20 mM	16.92	23.68	38.77	36.18	14.08	38.31

[a] $x_1 = -4$, $x_2 = -5$. [b] % κ coefficient.

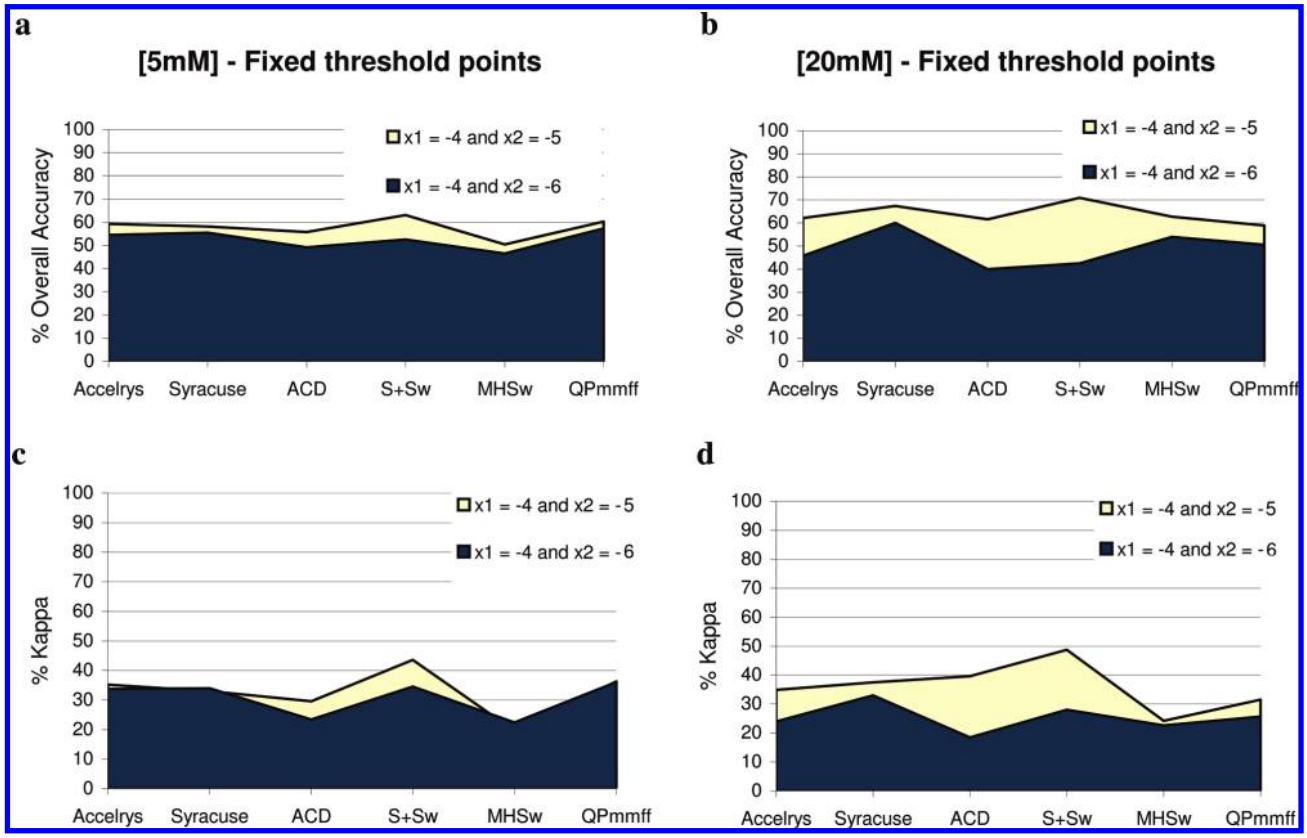


Figure 4. Comparison of statistical measurements obtained from the corresponding confusion matrices after analyzing data coming from HT solubility assay for 968 compounds at 5 mM and 657 compounds at 20 mM vs in-silico models.

mean a potentially soluble compound for S+Sw model, but for ACD, the same log S value would involve an insoluble compound. The ACD model has as threshold values $x_1 = -2.95$ and $x_2 = -3.95$, and S+Sw model has as threshold values $x_1 = -3.96$ and $x_2 = -4.83$; both cases show good performances and optimal % κ coefficients.

By using MARS, the optimal performance for each in silico general solubility model is achieved according to customer requirements, where it is applied (chemical space and experimental conditions), and its own circumstances (training set). Then, in this context, models are properly compared (Figures 6 and 7).

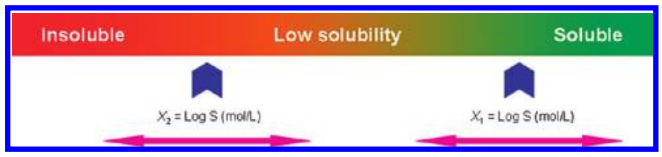


Figure 5. MARS approach explores the ADMET property ranking space in order to find out those threshold points that allow optimal performance for each in silico model.

Table 8. Performance after Applying MARS and Defining Flexible Threshold Points[a]

	% over. acc.	% κ	x_1, log S (mol/L)	x_2, log S (mol/L)
5 mM - 968 compounds				
Accelrys	59.8	35.4	−4.4	−5.28
Syracuse	62.8	39.8	−5.24	−6.22
ACD	57.2	30.4	−3.6	−4.39
S+Sw	69.4	49.5	−4.55	−4.92
MHSw	60.5	35.7	−5.6	−6.3
QPmmff	65.3	42.8	−4.92	−5.67
20 mM - 657 compounds				
Accelrys	66.5	38.3	−3.91	−4.65
Syracuse	68.5	41	−4.35	−5.24
ACD	74.7	53.7	−2.95	−3.95
S+Sw	72.9	50.1	−3.96	−4.83
MHSw	57.5	25.8	−5.04	−5.85
QPmmff	59.8	31.9	−4.09	−4.87

[a] Flexible threshold points (x_1 and x_2) were used to get an optimal binning, resulting statistics (overall accuracy and κ) obtained from the corresponding confusion matrices are shown.

If this analysis had been done without using MARS approach, it would have driven to wrong conclusions. There is a remarkable example at 20 mM: if MARS is not utilized, then the ACD predictive model has a moderate performance, comparing with the rest of the models, but if MARS is utilized, then ACD becomes the best model.

Previous analyses and discussion around the 118 compounds assayed at both concentrations, using fixed threshold points, suggested that model training sets might come from different experimental conditions; therefore, depending on the users experimental conditions, their performance might not be optimal. This issue is overcome by using optimized threshold points; maximizing models performance according to user assay requirements and to the different underlaying experimental conditions for models training sets, as Tables 9 and 10 describe.

The effect of assay conditions is clearly illustrated in Table 9 where optimal threshold points, for the same model, are quite different depending on experimental conditions. Therefore, the meaning of the estimated solubility values will depend not only on the model under consideration but also on the users experimental conditions.

Figure 8 shows that models performance has improved considerably, % κ coefficient up to 19.95% at 5 mM and up to 22.03% at 20 mM, comparing Tables 10 vs 7. As it is expected, some models perform better at certain experimental conditions than at others. Accelrys is better at 5 mM than at 20 mM, but QPmmff and ACD are better at 20 mM.

Applying Flexible Threshold Points to Chemical Series. This evaluation has been done for a large set of experimental data where different chemical structures are considered simultaneously. This could be one of the reasons why overall accuracy hardly reaches 70%, and κ is regularly below 50%, as Figures 6 and 7 show. On a daily basis, project-based analysis in drug discovery, particular chemical series are the data sets where these estimations are applied; therefore these percentages could improve. In order to validate the MARS approach on a "real case", project-based analysis and to evaluate in silico models performance on a practical scenario, two different chemical series,[56−60] scaffolds I and II are analyzed in this report, as described above in Figure 3. In this particular case, just two in silico models have been considered for this analysis: S+Sw (pH-dependent model) implemented in the ADMET Predictor package and Accelrys (pH-independent model) implemented in Accord for Excel.

In total, solubility data for 48 derivatives of scaffold I at 5 mM as starting concentration and for 30 additional derivatives at 20 mM are available. From scaffold II, experimental data for 68 compounds are available at 20 mM.

Once it is checked that models performance can be optimized according to users and models training set experimental conditions, the aim is evaluating the impact of the training sets chemical space on models performance. Thus, scaffold I data sets, at both starting concentrations, are considered. A similar exercise was done comparing those 118 compounds assayed at 5 mM and 20 mM, but in this case, the chemical space is not well spread and is not covered, so only one chemical series is considered. Therefore, the chemical space's influence should be noticeable, even considering experimental conditions effects, if it is not covered in the training set.

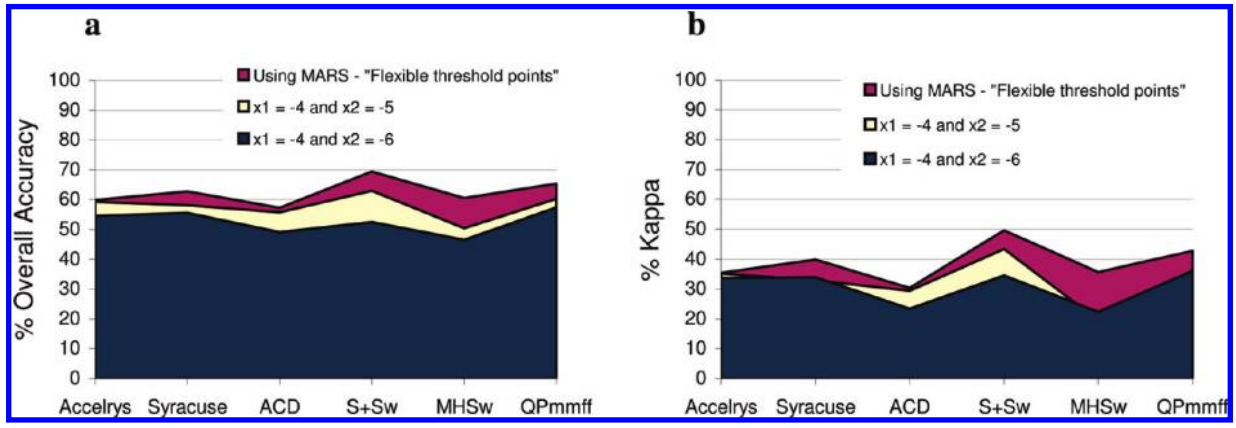


Figure 6. Comparison of statistical measurements obtained from the corresponding confusion matrices after analyzing data coming from HT solubility assay, 968 compounds, at 5 mM vs in-silico models.

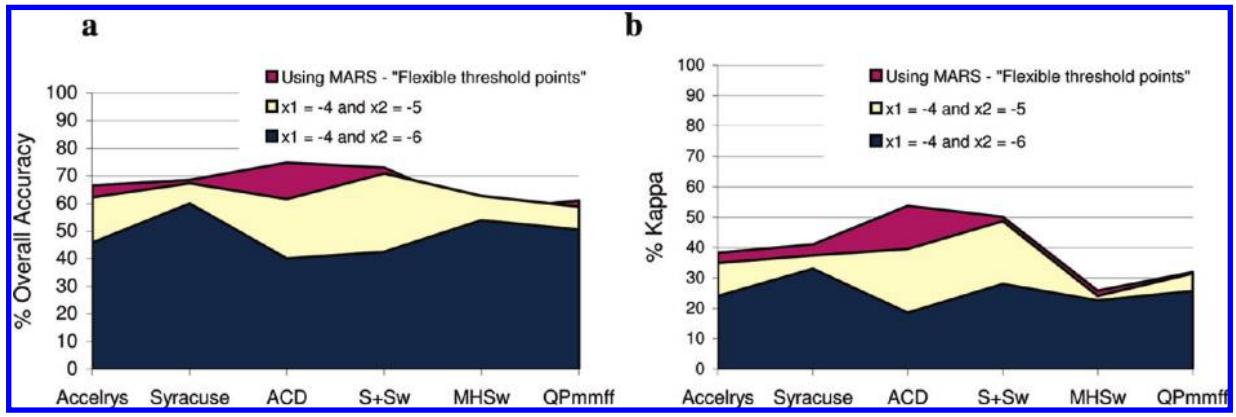


Figure 7. Comparison of statistical measurements obtained from the corresponding confusion matrices after analyzing data coming from HT solubility assay, 657 compounds, at 20 mM vs in-silico models.

Table 9. Optimized Threshold Points for Each in Silico Model Depending on Experimental Conditions[a]

118 compounds	Accelrys		Syracuse		ACD		S+Sw		MHSw		QPmmff	
	x_1[b]	x_2[b]	x_1[b]	x_2[b]	x_1[b]	x_2[b]	x_1[b]	x_2[b]	x_1[b]	x_2[b]	x_1[b]	x_2[b]
5 mM	−4.05	−4.89	−5.99	−6.33	−4.19	−4.87	−4.73	−4.90	−5.53	−6.29	−4.76	−5.33
20 mM	−3.05	−4.56	−5.17	−6.39	−3.26	−4.45	−4.81	−5.53	−5.05	−6.34	−3.60	−5.62

[a] Using the same set of compounds at both starting concentrations. [b] Flexible threshold points: x_1 and x_2 as log S (mol/L).

Table 10. % κ Coefficients Obtained after Running MARS

118 compounds	Accelrys[a]	Syracuse[a]	ACD[a]	S+Sw[a]	MHSw[a]	QPmmff[a]
5 mM	31.74	31.05	40.00	47.51	29.87	33.67
20 mM	21.74	36.28	60.8	42.53	28.99	45.81

[a] % κ coefficient.

Scaffold I Analysis. When fixed threshold points are applied at both starting concentrations, the models performance is very poor, independent of the assay conditions, hardly reaching κ coefficient values around 10%. But, after applying MARS and defining optimal threshold points, outstanding results have been achieved, and improvements in % κ coefficient go from 25.58 to 46.00%, see Table 11 and Figure 9a and b.

Experimental conditions play their role, but, as it is suggested by comparing these data with the above-reported analyses about the overlapping subset, there is clearly an additional factor. This point stresses the models poor performance since in most of the cases κ coefficient is quite poor (lower than 10% when fixed threshold values are applied), and there is only an additional variable, the chemical space in the training set. Therefore, these results suggest this chemical series is poorly covered in these model training sets, and a proper definition in the threshold points is necessary in order to get the maximum models performance — the right meaning of threshold points for this chemical series under these experimental conditions. For example, according to the optimal threshold points, reported in Table 11, a virtual compound with an estimated log S (mol/L) equal to −5.10 would be soluble, however, if a general fixed threshold point were applied (No MARS), then this virtual compound would be insoluble and, therefore, discarded.

Scaffold II Analysis. Comparing scaffold I at 20 mM with scaffold II at the same starting concentration provides very interesting information, since just one variable is present, the covered chemical space for models training sets. Results shown in Table 12 and Figure 10 clearly suggest that this chemical series, scaffold II, is better covered, at least in Accelrys model's training set, than in scaffold I.

This conclusion is in agreement with previous analysis for scaffold I, where data suggested that the chemical series was poorly covered in models training sets.

This is a good example where the role of these flexible threshold points is clearly defined. For the same model and

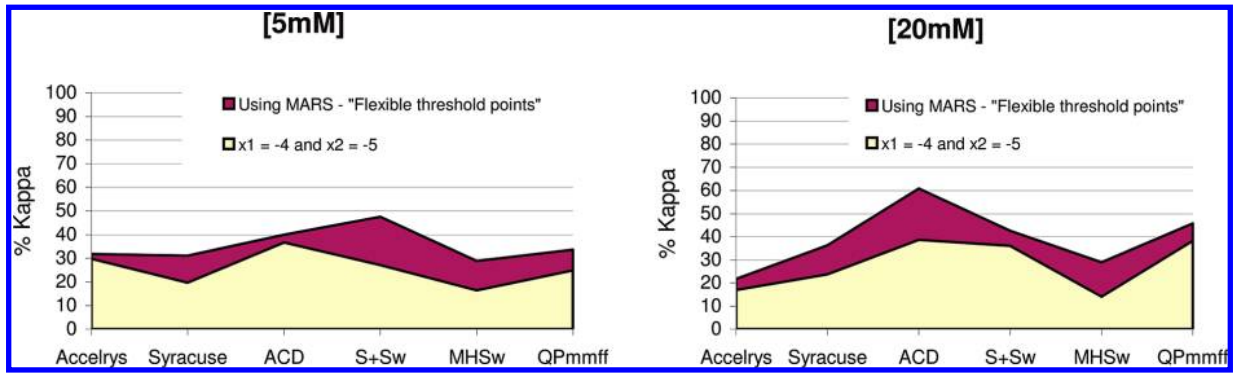


Figure 8. By using MARS models, performance is optimized, improving % κ coefficients, according to user assay requirements and to the different underlaying experimental conditions for models training sets.

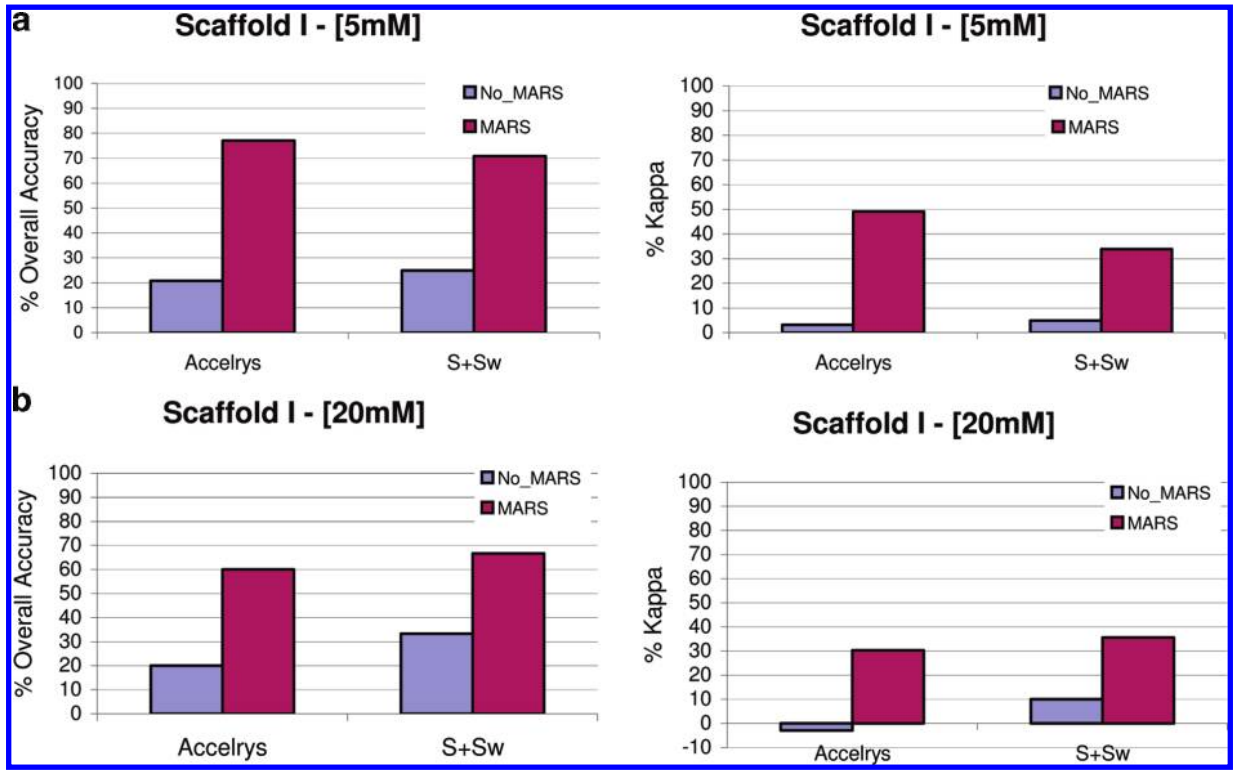


Figure 9. As statistics have shown us, there is a clear improvement in the performance of these in silico models after applying the MARS approach to this chemical series, scaffold I.

Table 11. Performance of In Silico Models after Applying MARS Approach to Scaffold I

	% over. acc.	% κ	x_1, log S (mol/L)	x_2, log S (mol/L)
5 mM − 48 compounds				
Accelrys, no MARS	20.83	3.13	−4.00	−5.00
Accelrys, MARS	77.08	49.13	−5.82	−6.03
S+Sw, no MARS	25.00	4.95	−4.00	−5.00
S+Sw, MARS	70.83	33.86	−5.17	−6.35
20 mM − 30 compounds				
Accelrys, no MARS	20.0	−3.0	−4.00	−5.00
Accelrys, MARS	60.0	30.23	−5.52	−5.92
S+Sw, no MARS	33.33	10.04	−4.00	−5.00
S+Sw, MARS	66.67	35.62	−5.11	−5.79

Table 12. Performance of These In Silico Models after Applying MARS Approach to Scaffold II

20 mM − 68 compounds	% over. acc.	% κ	x_1, log S (mol/L)	x_2, log S (mol/L)
Accelrys, no MARS	72.06	20.83	−4.00	−5.00
Accelrys, MARS	76.47	27.08	−4.06	−4.76
S+Sw, no MARS	75.00	6.77	−4.00	−5.00
S+Sw, MARS	79.41	42.58	−4.30	−5.20

experimental conditions, threshold points have different meanings, depending on each chemical series. Scaffold I at 20 mM, according to the Accelrys model, a virtual compound with a log S (mol/L) value equal to −5.00 would be soluble. But for scaffold II at same starting concentration, according to the Accelrys model, a virtual compound with that estimated solubility value (−5.00) would be insoluble.

Method Validation. Applying Flexible Threshold Points to Test Sets. After reporting the development of this new approach, MARS, its application to different sets of HT experimental data has pointed out that it provides a real improvement in the predictive performance of in silico models. But it is necessary to go one step further in this analysis, and we should identify what is the minimal number of experimental data points we need to get accurate enough flexible threshold points to classify new/virtual compounds, project-based analysis, with an optimal prediction. This is our ultimate goal.

The chemical series, scaffolds I (5 mM) and II (20 mM), utilized in the previous analysis will be used again, except scaffold I at 20 mM, since its data set is quite reduced. The first step is selecting the training set; hence, from each chemical series a set of diverse compounds is selected, but these selected compounds must fit a couple of requirements: first, diverse in the chemical space and second, but not less important, diverse in the response space (i.e., insoluble, soluble, and low solubility compounds) in order to be representative enough.

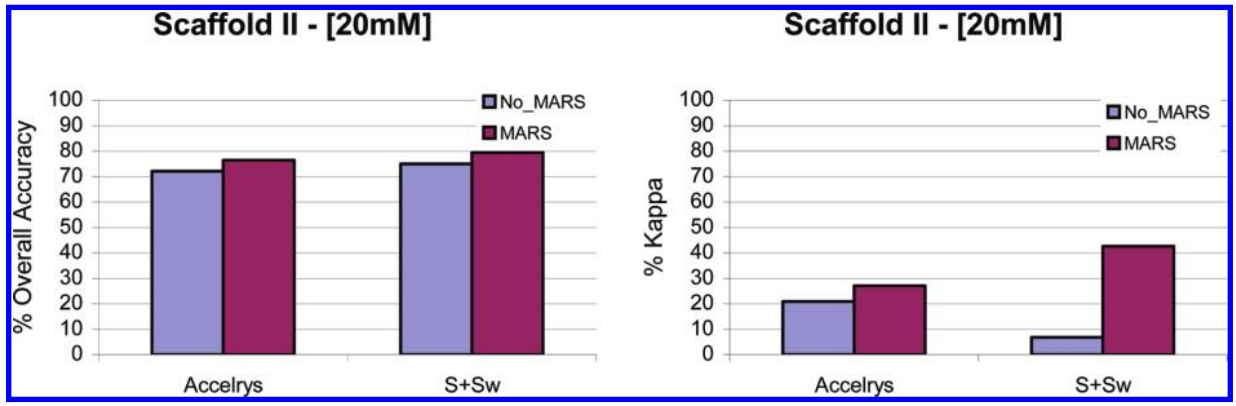


Figure 10. Graphs showing the improvement in the performance of these in silico models after applying the MARS approach to this chemical series is quite poor, except for % κ coefficient when S+Sw model is analyzed.

Table 13. Defined Optimal Threshold Points for Scaffold I[a]

training set - 20 compounds	x_1, log S (mol/L)	x_2, log S (mol/L)
Accelrys	−5.70	−6.36
S+Sw	−5.08	−5.29

Test Set - 28 compounds	% over. acc.	% κ	x_1, log S (mol/L)	x_2, log S (mol/L)
Accelrys, no MARS	17.86	3.30	−4.00	−5.00
Accelrys, MARS	71.43	35.26	−5.70	−6.36
S+Sw, no MARS	25.00	6.67	−4.00	−5.00
S+Sw, MARS	67.86	20.25	−5.08	−5.29

[a] Using a small training set (20 diverse compounds), we are able to classify properly those compounds placed at the test set.

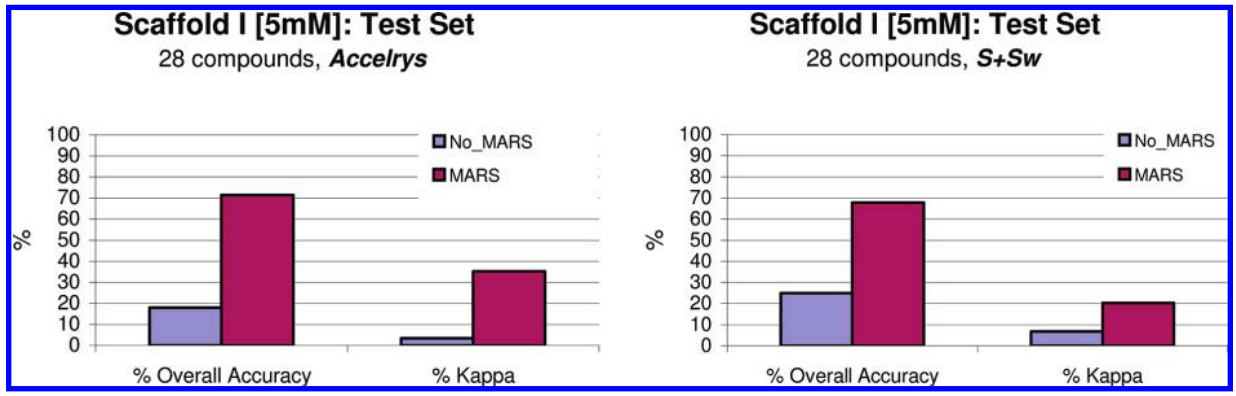


Figure 11. Scaffold I test sets.

To select those compounds fitting the first requirement, diversity in the chemical space, a set of 101 descriptors was calculated for all compounds in each chemical series, using Cerius^2 4.9 package,[50] which included descriptors describing structural, topological,[51−53] and electrotopological indices known as E-state keys.[54] Thus, these descriptors are used to characterize these chemical series within a 101-dimensional chemical space. The dimensionality was reduced by principal component analysis (PCA) down to two components that account for 90.4% of the variance. Once the chemical space was set, a diverse set of compounds was selected using the MaxMin algorithm,[66] distance-based function, implemented within the CombiChem module of Cerius^2. Simultaneously, these selected compounds cover the whole range of response space, including soluble, insoluble, and low solubility compounds. The training set to calculate the optimal flexible threshold for each chemical series/scaffold is defined following the above-described procedure.

Once the training set is defined, the MARS approach is applied to find out which are the optimal threshold points for this chemical series. For the first chemical series, scaffold I, a training a set of 20 compounds, has been utilized that, after being analyzed by MARS, suggests as optimal threshold points: $x_1 = -5.70$ and $x_2 = -6.36$ for the Accelrys predictive model and $x_1 = -5.08$ and $x_2 = -5.29$ for the S+Sw model. Then, the rest of compounds, 28 in this case, are the corresponding test set where those predefined threshold points are applied in order to evaluate the predictive performance of the novel combination, the MARS approach, and the in silico models. Statistics for these 28 compounds using Accelrys as the source of estimations are 71.43% as the overall accuracy and κ is 35.26% and using S+Sw are 67.86% as the overall accuracy and κ is slightly lower at 20.25%. There is an outstanding improvement in models performance when flexible threshold points, coming from the training set, have been applied to classify compounds in this test set, as Table 13 and Figure 11 show.

For the second chemical structure, scaffold II, 21 compounds have been defined as a training set that, after being analyzed by MARS, suggests as threshold points: $x_1 = -2.30$

Table 14. Defined Optimal Threshold Points for Scaffold II[a]

training set − 21 compounds	x_1, log S (mol/L)	x_2, log S (mol/L)
Accelrys	−2.3	−5.10
S+Sw	−3.49	−5.25

test set − 47 compounds	% over. acc.	% κ	x_1, log S (mol/L)	x_2, log S (mol/L)
Accelrys, no MARS	76.60	29.56	−4.00	−5.00
Accelrys, MARS	72.34	21.36	−2.30	−5.10
S+Sw, no MARS	76.60	13.11	−4.00	−5.00
S+Sw, MARS	80.85	53.36	−3.49	−5.25

[a] using a small training set (21 diverse compounds), we are able to classify properly those 47 compounds placed at the test set.

and $x_2 = -5.10$ when Accelrys is utilized as predictive the model and $x_1 = -3.49$ and $x_2 = -5.25$ for the S+Sw model, as Table 14 shows. The rest of compounds, 47 in this case, are the corresponding test set. After analyzing this test set, considering those previously reported personalized threshold points, we have obtained the following results: using Accord, 72.34% is the overall accuracy and 21.36% is κ; in this case no improvement is achieved comparing with default fixed threshold points. The simulation may probably have reached a local minima for the κ value at those coordinates ($x_1 = -2.30$ and $x_2 = -5.10$) whose barrier is not overcome by Monte Carlo movements; therefore, in this particular case, the optimal answer is not obtained. Whereas utilizing S+Sw, both statistical measurements are improved, obtaining results such as 80.85% for the overall accuracy and 53.36% is κ. The increment for κ coefficient, using optimal threshold points, is really outstanding in this case: 40.25. Therefore, this last "classifier", when the S+Sw model is applied, achieves an accuracy that is 53.36% better than the chance assignment of compounds to categories. These results are shown in the Table 14 and Figure 12.

Finally, we should highlight that the MARS approach is a learning method. This report has shown that by using 20 compounds as an initial training set, we obtain good predictive results. Furthermore, taking into account that we are working in an iterative process, these threshold points can be refined once new experimental information is obtained, just by running the MARS approach again, improving our predictive power. This information is confirmed by current in-house ongoing projects. The computation time is quite acceptable for a training set of 100 compounds. If we run MARS until completion for 100 000 steps, then it will take around 3 min on an Octane, SGI workstation.

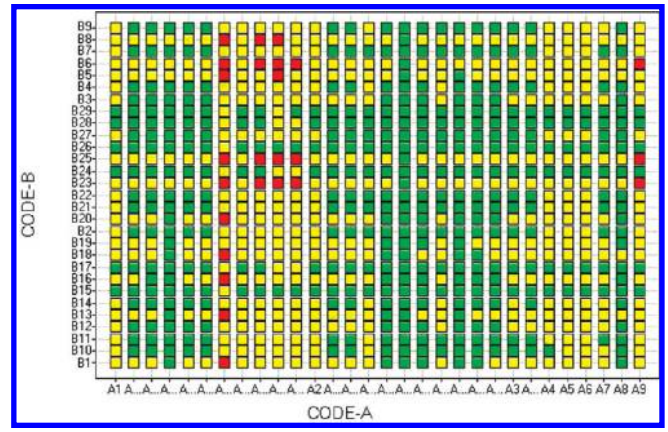


Figure 13. Aiding decision making, based on those results obtained from a predictive optimal binning and a combination of in silico estimations and from the MARS approach, we are able to prioritize the synthesis of those compounds with an optimal profile, color-coded. Green (soluble compounds), yellow (compounds with low solubility), and red (insoluble compounds). Code-B and code-A involve two different substitutions at positions R1 and R2 for a scaffold bearing two diversity points.

The above-described analyses demonstrate that the predictive performance of in silico models, properly optimized with the novel MARS approach through a personalized definition of threshold points (depending on model under analysis, experimental conditions, and chemical series), vs HT experimental data is good enough to be used in the compound and the library design. In fact, this combination of MARS and predictive in silico models, which initially requires a relative small amount of high-throughput experimental data (around 20 diverse compounds, in terms of both chemical and response spaces) as the training set, has been utilized to predict solubility of several in-house virtual libraries. Due to we obtain compounds classification, as HT assay, is quite easy to visualize which library has the optimal profiling (higher percentage of soluble compounds) and which are those selected compounds to be synthesized or at least, to prioritize their synthesis; color-coded flagging compounds (Figure 13). Therefore, this combination of computational approaches could be utilized as a guide to select compounds in drug discovery, aiding decision making, as well as kinetic solubility ranking, used as a guide to thermodynamic solubility but with a great advantage, we can have this prioritization done even before the synthesis is considered with higher efficiency, optimizing resources.

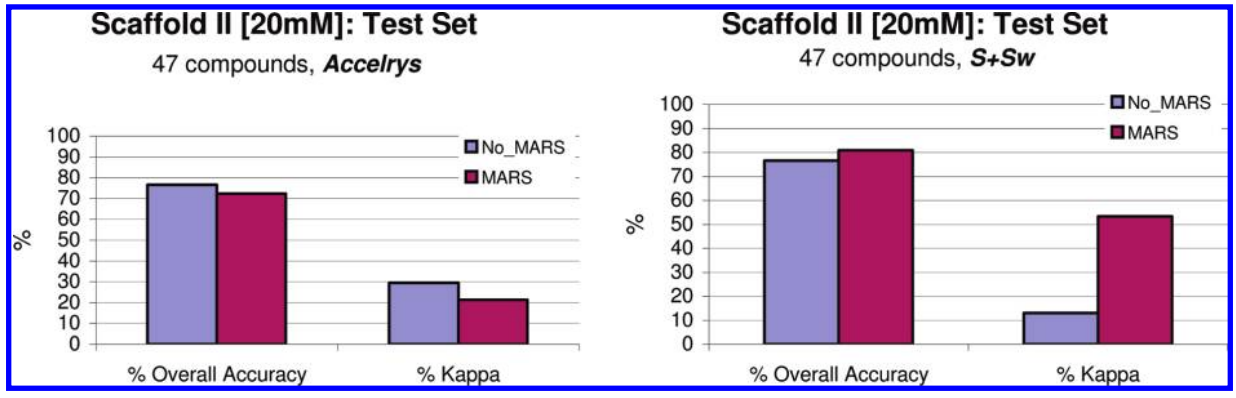


Figure 12. Scaffold II test sets.

CONCLUSION

A novel approach, Metropolis/Monte Carlo adaptive ranking simulation (MARS), has been developed to optimize the performance of in silico absorption, distribution, metabolism, and toxicity (ADMET) predictive general models vs qualitative high-throughput experimental data for local use, trying to overcome the lack of consistency on models training data sets and to overcome their limited or/and publicly unknown coverage of the chemical space. We are focused on these kinds of experiments since in drug discovery we deal with large numbers of compounds, small sample sizes and short time lines; thus, discovery assays must be high-throughput, conservative in sample use, inexpensive, and rapid. In fact, the main source for experimental solubility data in drug discovery is kinetic solubility measured with a high-throughput system.

MARS estimates, which are optimal flexible threshold points, to get the best correlation between estimations, predictive models, and high-throughput experimental measurements, and therefore, provides the best performance for each in silico model, taking into account three factors that have been shown as key: the predictive model under analysis, the user's and training sets experimental conditions, and the user's and training sets chemical space (models application scope). Therefore, an informed reading of the estimations, by generalized ADMET models, is achieved.

We have demonstrated that MARS optimizes, to a considerable degree, the performance for each in silico model vs the large high-throughput (HT) data set. Furthermore, we have tested and validated that the combination of MARS, defining optimal threshold points, with the predictive in silico models drives us to pretty good ranking predictions when we are dealing with a particular chemical series.

These two facts have practical applications on a daily basis work: (i) Proper evaluation, and therefore, assessment of ADMET commercial software packages since they will be fine-tuned to your proprietary chemical space and your own experimental conditions. Thus, evaluated within your current application scope. (ii) Project based, since this allows the use of the MARS approach (thresholds values) together with the predictive models in the profiling of new/virtual libraries of compounds. By color coding the estimated ranks, we are able to identify optimal structures from large virtual libraries and to discard undesired virtual compounds.

In summary, we are able to deal with in silico predictions from an interesting perspective: optimal performance of general models considering the local "environmental" conditions surrounding each particular project; thus, general models become optimized local models to be used under certain requirements, the three key factors. Then, these predictions are handled in an easy manner, as HT experimental data, just through a simple color coding on estimated classes, flagging compounds, that aids decision making, which prioritize/guide the synthesis of certain compounds.

Defining the applicability domain of models (chemical space), dealing with the quality of the experimental data utilized in training/test sets and experimental conditions, assessing models prediction accuracy, and the use of local models instead of global models is nothing new,[67,31] and although we are trying to make progresses, as it has been reported in recent reviews,[68,69] there is clearly still a need to circumvent these issues around ADMET predictions.

Currently, this general and novel approach, MARS, is being used, together with the corresponding in silico models, for the analysis of different ADMET properties (i.e., BBB, etc.), as well as for other properties. In addition, we are currently working on alternative algorithms to explore the ADMET property ranking space and to avoid local minima, an issue previously highlighted (scaffold II test set). Results from these studies will be published in the future.

ACKNOWLEDGMENT

We thank Marc Noppe for running the ADMET Predictor as well as Jef Peeters and Claire Mackie for useful discussions on this approach and on different solubility issues. Finally we want to thank Antonio Gómez for his support.

REFERENCES AND NOTES

(1) Venkatesh, S.; Lipper, R. A. Role of the development scientist in compound lead selection and optimization. *J. Pharm. Sci.* **2000**, *89*, 145–154.
(2) Clark, D. E. Prediction of Intestinal Absorption and Blood-Brain Barrier Penetration by Computational Methods. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 477–496.
(3) Clark, D. E. In Silico Prediction of Blood-Brain Barrier Permeation. *Drug Discov. Today* **2003**, *20*, 927–933.
(4) Pharmaceutical Industry 2009 Profile; Pharmaceutical Research and Manufacturers of America: Washington, DC; http://www.phrma.org/files/PhRMA%202009%20Profile%20FINAL.pdf. Accessed October 8, 2009.
(5) Prentis, R. A.; Lis, Y.; Walker, S. R. Pharmaceutical Innovation by the Seven UK-owned Pharmaceutical Companies (1964−1985. *Br. J. Clin. Pharmacol.* **1988**, *25*, 387–396.
(6) Kennedy, T. Managing the drug discovery/development interface. *Drug Discov. Today* **1997**, *2*, 436–444.
(7) Kubinyi, H. Drug research: myths, hype and reality. *Nat. Rev. Drug Discov.* **2003**, *2*, 665–668.
(8) Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2001**, *44*, 235–249.
(9) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **1997**, *23*, 3–25.
(10) DiMasi, J. A. Success Rates for New Drugs Entering Clinical Testing in the United States. *Clin. Pharmacol. Ther.* **1995**, *58*, 1–14.
(11) Smith, D. A.; van de Waterbeemd, H. Pharmacokinetics and metabolism in early drug discovery. *Curr. Opin. Chem. Biol.* **1999**, *3*, 373–378.
(12) Cheng, A; Merz, K. M., Jr. Prediction of Aqueous Solubility of a Diverse Set of Compounds Using Quantitative Structure-Property Relationships. *J. Med. Chem.* **2003**, *46*, 3572–3580.
(13) Lesney, M. S. Assaying ADMET. *Mod. Drug Discov.* **2004**, *7*, 30–34.
(14) Susnow, R. G.; Dixon, S. L. Use of Robust Classification Techniques for the Prediction of Human Cytochrome P450 Inhibition. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1308–1315.
(15) Lombardo, F.; Gifford, E.; Shalaeva, M. Y. In Silico ADMET Prediction: Data, Models, Facts and Myths. *Mini Rev. Med. Chem.* **2003**, *3*, 861–875.
(16) Van de Waterbeemd, H.; Gifford, E. ADMET *In Silico* Modelling: Towards Prediction Paradise. *Nature Rev. Drug Disc.* **2003**, *2*, 192–204.
(17) Egan, W. J.; Merz, K. M., Jr.; Baldwin, J. J. Prediction of Drug Absorption Using Multivariate Statistics. *J. Med. Chem.* **2000**, *43*, 3867–3877.
(18) Manallack, D. T.; Tehan, B. G.; Gancia, E.; Hudson, B. D.; Ford, M. G.; Livingston, D. J.; Whitley, D. C.; Pitt, W. R. A Consensus Neural Netwok-Based Technique for Discriminating Soluble and Poorly Soluble Compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 674–679.
(19) Jensen, B. F.; Refsgaard, H. H. F.; Bro, R.; Brockhoff, P. B. Classification of Membrane Permeability of Drug Candidates: A Methodological Investigation. *QSAR Comb. Sci.* **2005**, *24*, 449–457.
(20) Chohan, K. K.; Stuart, W. P.; Mistry, J.; Barton, P.; Davis, A. M. A Rapid Computational Filter for Cytochrome P450 1A2 Inhibition

Potential of Compound Libraries. *J. Med. Chem.* **2005**, *48*, 5154–5161.
(21) O'Brien, S. E.; de Groot, M. J. Greater Than the Sum of Its Parts: Combining Models of Useful ADMET Prediction. *J. Med. Chem.* **2005**, *48*, 1287–1291.
(22) Refsgaard, H. H. F.; Jensen, B. F.; Brockhoff, P. B.; Padkjaer, S. B.; Guldbrandt, M.; Christensen, M. C. In Silico Prediction of Membrane Permeability from Calculated Molecular Parameters. *J. Med. Chem.* **2005**, *48*, 805–811.
(23) Lehn, J. M. Dynamic combinatorial chemistry and virtual combinatorial libraries. *Chem.—Eur. J.* **1999**, *5*, 2455–2463.
(24) Our high-throughput system ranks compounds according to their kinetic solubility. Ranking is done into three classes: Class 1, insoluble compounds; Class 2, compounds with low solubility and Class 3, soluble compounds. The ranking is based on turbidity measurements. DMSO-stock solutions are diluted with a single aqueous buffered solvent in three consecutive steps. For every dilution turbidity is measured by a nephelometer. Based on this turbidity measurement the ranking is performed into the three classes.
(25) Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46.
(26) Fielding, A. H.; Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Env. Conserv.* **1997**, *24*, 38–49.
(27) Oyarzabal, J.; Pastor, J. MARS Approach for Drug Discovery: Aiding Decision-Making in In-Silico ADME. In *QSAR and Molecular Modelling in Rational Design of Bioactive Molecules; Proceedings of the 15th European Symposium on QSAR and Molecular Modelling*, Istanbul, Turkey, 2004; Aki, E., Yalçin, I. Eds.; CADD&D Society in Turkey Publishing, 2005, pp. 219—220.
(28) Alsenz, J.; Kansy, M. High throughput solubility measurement in drug discovery and development. *Adv. Drug. Del. Rev.* **2007**, *59*, 546–547.
(29) Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O. Random Forest Models To Predict Aqueous Solubility. *J. Chem. Inf. Model.* **2007**, *47*, 150–158.
(30) Zhou, D.; Alelyunas, Y.; Liu, R. Score of Extended Connectivity Fingerprint as Descriptors in QSPR Study of Melting Point and Aqueous Solubility. *J. Chem. Inf. Model.* **2008**, *48*, 981–987.
(31) Davis, A. M.; Bruneau, P. In Silico Prediction of Solubility. In *Drug Bioavailability. Methods and Principles in Medicinal Chemistry*; 1st ed.; van de Waterbeemd, H., Testa, B., Eds.; Wiley-VCH: Weinheim, Germany, 2009, Vol. 40, pp. 53—68.
(32) Llinas, A.; Glen, R. C.; Goodman, J. M. Solubility Challenge: Can You Predict Solubilities of 32 Molecules Using a database of 100 Reliable Measurements. *J. Chem. Inf. Model.* **2008**, *48*, 1289–1303.
(33) Hopfinger, A. J.; Esposito, E. X.; Llinas, A.; Glen, R. C.; Goodman, J. M. Findings of the Challenge To Predict Aqueous Solubility. *J. Chem. Inf. Model.* **2009**, *49*, 1–5.
(34) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Del. Rev.* **2002**, *54*, 355–366.
(35) Waiver of in vivo bioavailability and bioequivalence studies for immediate-release solid oral dosage forms based on a biopharmaceutics classification system. U. S. Department of Health and Human Services, Food and Drug Administration; http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm128219.htm. Accessed August 3, 2009.
(36) Oprea, T. I.; Zamora, I.; Ungell, A.-L. Pharmacokinetically Based Mapping Device for Chemical Space Navigation. *J. Comb. Chem.* **2002**, *4*, 258–266.
(37) *DS Accord for Excel*, Version 5.2; Accelrys, Inc.: San Diego, CA, 2004.
(38) Meylan, W. M.; Howard, P. H. Atom/Fragment Contribution Method for Estimating Octanol-Water Partition Coefficients. *J. Pharm. Sci.* **1995**, *84*, 83–92.
(39) Meylan, W. M.; Howard, P. H.; Boethling, R. S. Improved method for estimating water solubility from octanol/water partition coefficient. *Environ. Toxicol. Chem.* **1996**, *15*, 100–106.
(40) *ACD/Solubility DB*, Version 6.0; Advanced Chemistry Development, Inc.: Toronto, Canada, 2004.
(41) *ADMET Predictor*, Version 1.1.0; Simulations Plus, Inc.: Lancaster, CA, 2004.
(42) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from Monte Carlo simulations. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155–1158.
(43) *QikProp*, Version 2.2; Schrödinger: Portland, OR, 2004.
(44) *CORINA*, Version 3.0; Molecular Networks GmbH: Erlangen, Germany, 2004.
(45) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
(46) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.
(47) Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553–586.
(48) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587–615.
(49) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616–641.
(50) *Cerius*^{2}, Version 4.9; Accelrys, Inc.: San Diego, CA, 2004.
(51) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Structure-Activity Analysis*, 1st ed; John Wiley & Sons: New York, 1986; pp 13—89.
(52) Katrizky, A. R.; Gordeea, E. V. Traditional Topological Indices vs Electronical, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR Research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835–857.
(53) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
(54) Hall, L. H; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039–1045.
(55) *SIMCA-P+*, Version 10.0; Umetrics AB: Umeå, Sweden, 2004.
(56) Andres-Gil, J. I.; Fernandez-Gadea, F. J.; Gil-Lopetegui, P.; Diaz-Martinez, A. Preparation of dibenzocycloheptafuranalkanamines and analogs as 5-HT antagonists; Patent WO9919317 A1. 28pp., 1999.
(57) Janssens, F. E.; Sommen, F. M.; Surleraux, D. L. N. G.; Leenaerts, J. E.; Van Roosbroeck Y. E. M. Preparation of 1-(1,2-disubstituted piperidinyl)-4-substituted piperazine derivatives as substance-P antagonists; Patent WO9716440 A1. 69pp., 1997.
(58) Janssens, F. E.; Sommen, F. M.; Surleraux, D. L. N. G.; Van Roosbroeck, Y. E. M. Preparation of aroyl 4-piperidinopiperidides and analogs as tachykinin receptor antagonists; Patent WO9724324 A1. 48pp., 1997.
(59) Janssens, F. E.; Sommen, F. M.; Surlerraux, D. L. N. G.; Leenaerts, J. E. N-acyl-2-substituted-4-(benzimidazolyl- or imidazopyridinyl)piperidines as tachykinin antagonists; Patent WO9724350 A1. 50pp., 1997.
(60) Dugovic, C. J.; Janssens, F. E. Use of substance P antagonists for influencing the circadian timing system. Patent WO2001030348 A1. 37pp., 2001.
(61) Bevan, C. D.; Lloyd, R. S. A High-Throughput Screening Method for the Determination of Aqueous Drug Solubility Using Laser Nephelometry in Microtiter Plates. *Anal. Chem.* **2000**, *72*, 1781–1787.
(62) Suzuki, T. Development of An Automatic Estimation System for Both the Partition Coefficient and Aqueous Solubility. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 149–166.
(63) Hoerter, D.; Dressman, J. B. Influence of physicochemical properties on dissolution of drugs in the gastrointestinal tract. *Adv. Drug Del. Rev.* **1997**, *25*, 3–14.
(64) Blasko, A.; Leahy-Dios, A.; Nelson, W. O.; Austin, S. A.; Killion, R. B.; Visor, G. C.; Massey, I. J. Revisiting the Solubility Concept of Pharmaceutical Compounds. *Monatsh.Chem.* **2001**, *132*, 789–798.
(65) Baurin, N.; Richardson, C.; Chen, I.; Foloppe, N.; Potter, A.; Jordan, A.; Roughley, S.; Parratt, M.; Greaney, P.; Morley, D.; Hubbard, R. E. Drug-like Annotation and Duplicate Analysis of a 23-Supplier Chemical Database Totalling 2.7 Million Compounds. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 643–651.
(66) Snarey, M.; Terret, N. K.; Willet, P.; Wilton, D. J. Comparison of Algorithms for Dissimilarity-Based Compound Selection. *J. Mol. Graph. Model.* **1997**, *15*, 372–385.
(67) Stouch, T. R.; Kenyon, J. R.; Johnson, S. R.; Chen, X.-Q.; Doweyko, A.; Li, Y. In silico ADMET/Tox: why models fail. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 83–92.
(68) Balakin, K. V.; Savchuk, N. P.; Tetko, I. V. In Silico Approaches to Prediction of Aqueous and DMSO Solubility of Drug-Like Compounds: Trends, Problems and Solutions. *Curr. Med. Chem.* **2006**, *13*, 223–241.
(69) Tetko, I. V; Bruneau, P.; Mewes, H.-W.; Rohrer, D. C.; Poda, G. I. Can we estimate the accuracy of ADME-Tox predictions. *Drug Discov. Today* **2006**, *11*, 700–707.

CI900308U