J Comput Aided Mol Des (2012) 26:1143–1157
DOI 10.1007/s10822-012-9605-7

Composite multi-parameter ranking of real and virtual compounds for design of MC4R agonists: Renaissance of the Free-Wilson methodology

Ingemar Nilsson · Magnus O. Polla

Received: 22 May 2012 / Accepted: 7 September 2012 / Published online: 2 October 2012

Abstract Drug design is a multi-parameter task present in the analysis of experimental data for synthesized compounds and in the prediction of new compounds with desired properties. This article describes the implementation of a binned scoring and composite ranking scheme for 11 experimental parameters that were identified as key drivers in the MC4R project. The composite ranking scheme was implemented in an AstraZeneca tool for analysis of project data, thereby providing an immediate re-ranking as new experimental data was added. The automated ranking also highlighted compounds overlooked by the project team. The successful implementation of a composite ranking on experimental data led to the development of an equivalent virtual score, which was based on Free-Wilson models of the parameters from the experimental ranking. The individual Free-Wilson models showed good to high predictive power with a correlation coefficient between 0.45 and 0.97 based on the external test set. The virtual ranking adds value to the selection of compounds for synthesis but error propagation must be controlled. The experimental ranking approach adds significant value, is parameter independent and can be tuned and applied to any drug discovery project.

Keywords Composite ranking · QSAR · Free-Wilson

Electronic supplementary material The online version of this article (doi:10.1007/s10822-012-9605-7) contains supplementary material, which is available to authorized users.

I. Nilsson (✉) · M. O. Polla
CVGI AstraZeneca R&D Mölndal, 431 83 Mölndal, Sweden
e-mail: Ingemar.nilsson@astrazeneca.com

M. O. Polla
e-mail: Magnus.polla@astrazeneca.com

Introduction

The pharmaceutical industry is currently facing a number of challenges to the ultimate goal of delivering clinically beneficial and profitable drugs to the market. To be successful in the drug discovery process, companies try to decrease the time from lead generation through lead optimization (LO) to clinical development, as well as finding ways to improve the quality of the compounds progressed. The progress heavily depends on the quality of the data but also on decisions taken by the project team during the course of the work. The decision process is multidimensional, where a range of readily accessible in vitro parameters as well as more complex in vivo pharmacokinetic (PK), pharmacodynamic (PD) and safety parameters are analyzed and used in the design of new compounds to ultimately meet the Candidate Drug criteria. Thus, the drug discovery project team needs to, constantly, reflect on incoming data and judge whether that will change the order of ranking of current best compounds and reprioritize project activities accordingly. In the MC4R project, a GPCR obesity target [1], we decided to investigate if an automated "objective" ranking system for a set of physical and in vitro PK and PD parameters could be created. The ranking system was intended to support the project team in decision making and highlighting possible issues. The ranking system has been implemented within IBIS Explore, the in-house database and query tool containing all AZ screening data as well as physicochemical property data [2]. The multi-parameter ranking provides an objective view of all data and most importantly highlights compounds not fully recognized earlier.

The concomitant design and synthesis of new compounds is a complex, time-consuming task, in which the medicinal chemist seeks to balance potency, off-target

interactions, pharmacokinetic properties and toxicity. A critical step in project progression is to select the right compounds to synthesize from an almost infinite number of virtual compounds. Thus, computational models that propose compounds with the right properties are highly desirable. Quantitative structure–activity relationships (QSARs) or quantitative structure–property relationships (QSPRs) approaches analyze the correlation between molecular properties and the corresponding property of interest using multivariate statistics [3]. The QSAR are derived by correlating information about activity or some other property of interest with a molecular descriptor set. This can be 1D (e.g. molecular weight, logP, properties count, or structural descriptors), 2D (e.g. structural keys or hashed fingerprints) or 3D descriptors (CoMFA[4], pharmacophore fingerprints). Linear (e.g. Partial Least Square and Multi Linear Regression) and non linear methods (e.g. Random Forest, Neural Network and Support Vector Machines) have been used extensively to develop QSAR and numerous models have built to explain and predict biological activity [5–10]. The Free-Wilson approach was one of the first mathematical techniques developed for QSAR of a series of chemical analogues [11–13]. The basic idea in Free-Wilson models is that the biological activity of a molecule can be described as the sum of the activity contributions of the specific substructures (parent core and the corresponding substituents (R-groups)). The Free-Wilson method does not require any substituent parameters to be defined; in contrast to Hansch analysis [14, 15] where physicochemical properties are correlated with biological activity values only the activity is needed. The Free-Wilson method makes a number of assumptions that are not necessarily valid for all data sets:

1. Each substituent on the parent structure makes a constant contribution to the activity regardless of the structural variation in the rest of the molecule.
2. The contributions from each substituent are additive.
3. There are no interactions between the parent core and its substituents or between the substituents themselves.

When these conditions are satisfied the Free-Wilson methodology should provide a model with good predictive power, and the quality of the model is mostly set by the accuracy of the modeled response. However, as soon as interactions occur between R-groups, the Free-Wilson model is not able to properly account for the structure–activity relationship or at least result in a model with less good predictions. The Free-Wilson methodology has been used less frequently than the Hansch approach partly as a consequence of the arguments discussed above and partly because it is limited to predict properties of molecules for which the R-groups have been used in the training set. Successful application of the Free-Wilson approach has been reported by several groups. Jorissen et al. [16] reported on the R-group additivity approach in the analysis and design of HIV protease inhibitors. Sciabola et al. [17] have used Free-Wilson methodology in the analysis and prediction of selectivity of a panel of 45 kinase assays. Hoefgen et al. [18] used the Free-Wilson approach to analyze substituent effect on a set of imidazo[1,5-a]pyrido[3,2-e]pyrazines targeting phosphodiesterase 10A (PDE10A). Patel et al. [19] have made a thorough Free-Wilson analysis on combinatorial libraries assayed on several GPCR, ion channel, kinase and P450 targets, where approximately half of the models were successful whereas the remaining had non-additive contributions. Tomic et al. [20] made an interesting use of the Free-Wilson method in a study of binding of transcription factors to DNA. Recently Freeman-Cook et al. [21] reported on the use of Free-Wilson analysis to maximize lipophilic efficiency. A comprehensive list of earlier application of Free-Wilson treatments can be found in reviews by Kubinyi [22, 23]. In this paper we describe the application of the Free-Wilson methodology in the design of MC4R agonists. Strengthened by the very positive outcome of the automated ranking for experimental descriptors we developed a similar virtual multi-parameter scoring scheme based on the Free-Wilson predictions. The virtual multi-parameter ranking scheme was subsequently used to select well designed high quality compounds for synthesis. We strongly believe that the composite ranking based on experimental descriptors that will be presented here can be generally applied to drug discovery projects and add significant value in the decision process. We also believe that composite ranking based on good predictions, as in this case Free-Wilson models, can provide strong support in compound design.

Methods

Experimental data

The experimental data of eleven parameters were taken from IBIS, the AstraZeneca in-house database. These eleven descriptors formed the basis for the progression of compounds to in vivo studies. The parameters were the MC4R EC_{50}, chromatographic $logD_{7.4}$ [24], water solubility at pH 7.4 from DMSO solution (Sol), potency against five different cytochrome P450's, hERG inhibition, selectivity versus MC1R (MC1R K_i/MC4R K_i), intrinsic clearance in human liver microsomes (Cl_{int} HLM) and human hepatocytes (Cl_{int} hHeps), permeability (P_{app}) in caco2 cells at pH 6.5, efflux ratio in caco2 cells at pH 7.4 and human plasma protein binding (hPPB). An aggregated parameter, CYPMIN, is given the value of the highest potency against any of the cytochrome P450's (i.e. 1A2, 3A4, 2C8, 2D6, 2C9, 2C19).

For hERG we also defined an additional parameter as the ratio between hERG inhibition and the plasma concentration needed to obtain a free concentration equal to the MC4R EC_{50}, $hERG_{diff}$ = (hERG IC_{50}/((MC4R EC_{50}/hPPB) × 100)). Similarly a parameter $CYPMIN_{diff}$ was defined as $CYPMIN_{diff}$ = (CYPMIN IC_{50}/((MC4R EC_{50}/hPPB) × 100)). The chromatographic $LogD_{7.4}$ used in this work, although not equivalent to the shake flask $LogD_{7.4}$, is for simplicity further on named $LogD_{74}$. A more detailed description of experimental methods is found in supplementary material.

Structural data

The compound series in the MC4R project, schematically depicted in Fig. 1, have a fairly rigid framework consisting of two rings joined by a short linker. Each ring has two R-groups, which according to calculated 3D models show only modest interactions with each other in neither the predicted biologically active conformation based on a pharmacophore model (Phase) [25] nor in the predicted low energy conformation (Macromodel Low Mode conformational search, OPLS 2005) [26].

To provide a flavor of the structural variation the R-groups are described in terms of simple molecular descriptors in Table 1 and in Fig. 2 are shown a subset of R-groups (R1 to R4). The substituents in R1 are mostly substituted phenyl rings, whereas R2 may include larger variations but with some clear size limitation. R3 is hydrophobic in nature but only a few variations have been tested. R4 allows for greater structural diversity and a hydrogen bond donor is beneficial for MC4 activity.

Although no further details on the dataset are revealed and admittedly some interesting SAR discussions as a consequence are absent, we strongly emphasize that the important message and conclusions in this article are independent of that information. This paper demonstrate the utility of Free-Wilson models in combination with composite ranking that is independent of the disclosure of structural information and the methodology can be used for any dataset where measured activities and Rgroup contribution correlates in an additive manner.

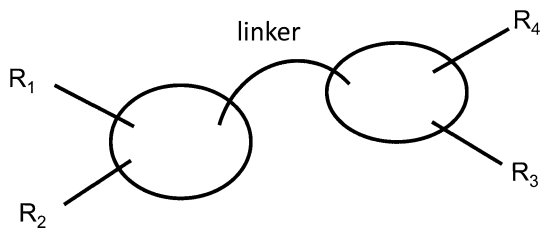

Fig. 1 The schematic scaffold used in the MC4R project

Experimental error

The quality of the theoretical models depends among other things on the quality of the experimental Y-response and the noise it introduce. No comprehensive investigation was made but three experimental variables, $LogD_{74}$, LogCaco and MC4R pEC_{50}, which gave the best, intermediate and lowest quality of the Free-Wilson models, were analyzed in some detail. The standard deviation of the chromatographic $LogD_{74}$ model is <0.1 as determined from repetitive measurement of four control sample. In Fig. 3 are shown the individual MC4R pEC_{50} records for compounds (n = 87) that has been recorded at least twice.

The variation is independent of potency and has a standard deviation of 0.3 for the MC4R pEC_{50} estimated from all compounds measured in triplicate, close to the control sample standard deviation of 0.27 (n = 90). The corresponding standard deviation of the logCaco model, averaged over all compounds measured in triplicate, is 0.26, which is in line with values of the three controls of 0.16, 0.27 and 0.42 (n ~ 240).

Binned experimental data

The ranking scheme was implemented within IBIS Explore, the IBIS query data browser, as custom columns using mathematical expressions, together with the experimental data. The experimental parameters described above were binned from 1 to 5, where 1 is bad and 5 is good. Records registered as non-numeric data with > or < signs were included in the data set by transformation to a slightly larger or smaller numerical values. The binning rules of MC4R EC_{50} in Fig. 4a serve as an illustrative example.

The sum of the binned parameters is divided by the number of measured parameters to create the final

Table 1 Variation in properties of R1-, R2-, R3- and R4-groups as calculated from simple descriptors

	Number of variations	Heavy atoms	Rotatable bonds	Mwt	PSA	NPSA	HBA	HBD	cLogP
R1	10	7–11	0–1	95–164	0–32	81–151	0–2	0–1	−0.3 to (+2.9)
R2	19	1–9	0–2	1–127	0–45	36–159	0–3	0–2	−0.4 to (+1.5)
R3	5	3–7	0–1	44–86	0–	93–135	0	0	1.7 to 2.7
R4	121	1–14	0–6	17–203	18–86	0–206	1–5	0–4	−1.8 to (+1.3)

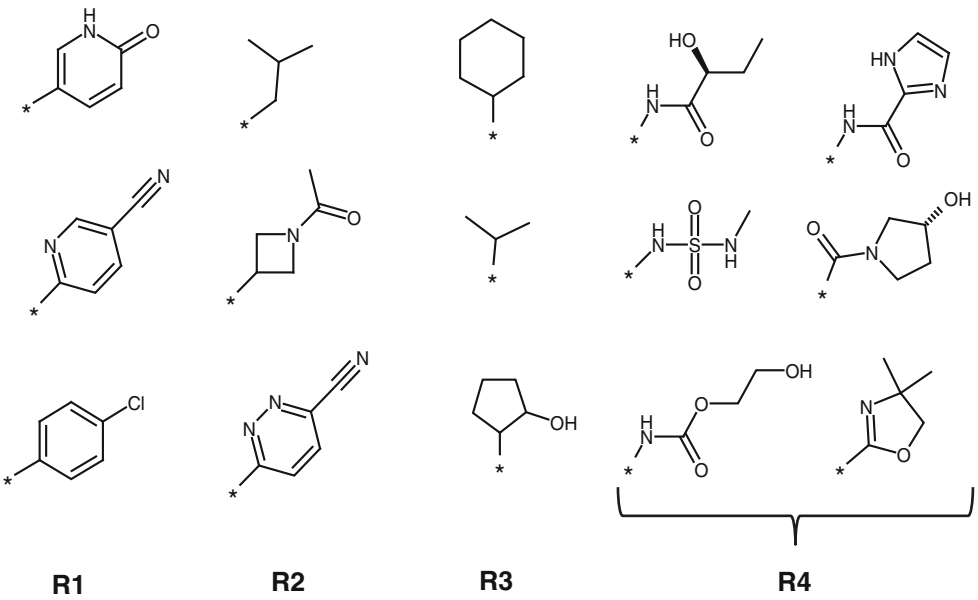

Fig. 2 Subset of R1- to R4-group variation used in the Free-Wilson analysis

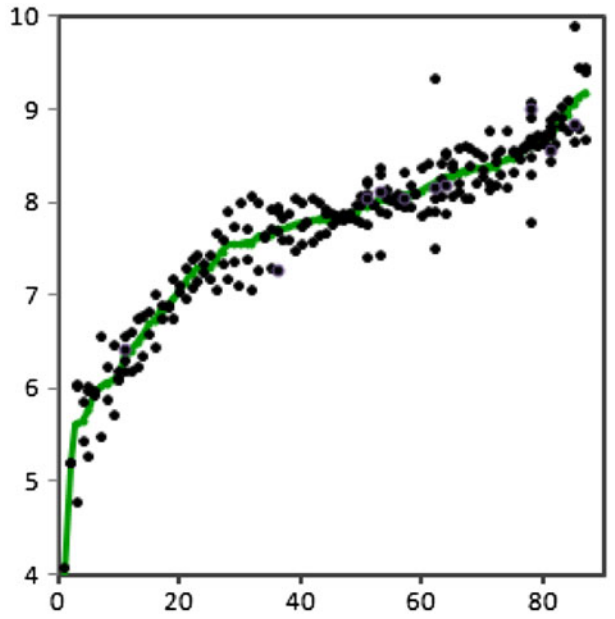

Fig. 3 The variability (Y-axis) in MC4R pEC50 data as a function of MC4R agonist activity. The X-axis corresponds to individual compounds (1-87) with replicate measurements (2-7 times). The mean values are connect by a *green* line

normalized score (NormRank). Thus, the ranking of a compound is independent of how many experimental data points that have been obtained. For convenience the NormRank value was multiplied with 10 and thus the maximum possible score is 50. The final ranking scheme was revised to account for proper handling of negative results. Depending on the degree of "badness" for a parameter a flag was set to OrangeFlag, RedFlag or Showstopper and given a penalty (Fig. 4b). The total penalty which is the sum of OrangeFlag, RedFlag and Showstopper penalties were subtracted from the NormRank values.

Modeling description

The dataset used in this work comprise of 298 synthesized compounds. The training set of 238 compounds for the computational models was randomly picked using the random uniform function in JMP (version 7.0) [27] from the 264 compounds hitherto made at the time the models were built. The remaining 26 compounds were used as an external test set (test set A) in the model building. The additional 34 compounds, a temporal test set (test set B), were synthesized after the models have been built and validated. The quality of the models were consequently assessed by the R^2 of linear regression of the measured versus predicted values for the training set and the predictive power as defined by the Q^2 of the linear regression of the measured versus predicted values of the two external test sets A and B. The general statistical analyses were also carried out using JMP. Visualizations were done using TIBCO Spotfire (version 3.1) [28]. The theoretical molecular descriptors were calculated using Clab, the AstraZeneca in-house descriptor engine. The descriptor set comprises descriptors as cLogP, polar surface area (PSA), hydrogen bond donors (HBD) and MW as well as fingerprints. The Random Forest models were generated using Rule Discovery System (RDS version 2.6.0) [29]. The Partial Least Squares (PLS) analysis was made using SIMCA P + (version 12.0.1.0) [30–33]. The Free-Wilson modeling was carried out using the Fit Model function, the multi-linear regression module in JMP, which in essence is the same as the original Free-Wilson method. The only

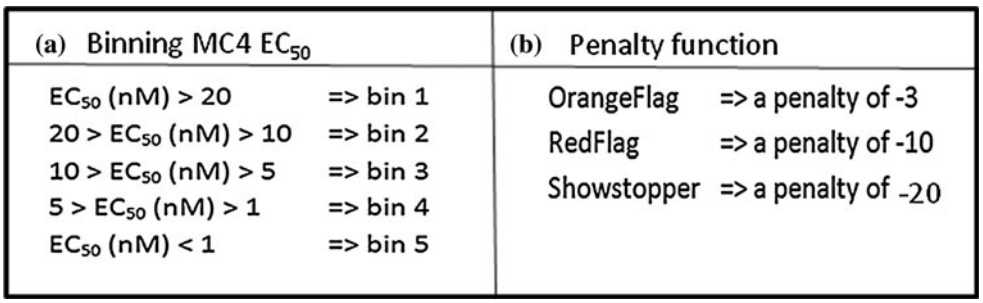

Fig. 4 **a** Example of binning applied to MC4R EC_{50} that was used in the compound scoring. **b** Penalties added to the original scoring function

difference is the different definition of symmetry equation, which results in different intercept and coefficient of the multi-linear regression equation but the predicted values are the same. The R-group descriptors were represented by their canonical smiles strings. Most parameters were transformed to a logarithmic scale to prevent models being dominated/biased by very low or very high values and the additivity concept as described in the classical Free-Wilson method has been frequently applied on logarithmic values [22, 23]. Thus, six models were built on logarithmic values of MC4R EC_{50}, Solubility, Caco2 P_{app}, hERG IC_{50}, CYPMIN IC_{50}, hPPB. Additionally, the $LogD_{74}$ lipophilicity measurement is already on a logarithmic scale. The Cl_{int} HLM, Cl_{int} hHeps and efflux ratio data were kept non-transformed.

Attempts were made to generalize R-groups to make Free-Wilson prediction for R-groups not previously used experimentally. The first method SMARTS Multiplet Generator (SMG), detects six key pharmacophore elements in R-group side chains and reports back their shortest distance in bonds from the attachment point [34]. SMG also generates a 'pharmacophore fingerprint' of the side chain, allowing side chains with similar functionality to be grouped together. This results in a set of parameters that describes the pharmacophore features of the R-group. The second method, ChemMiner, relies on a set of predefined smarts pattern that categorize substituents both in detailed and generalized matches [35]. This collapses R-groups into classes at different levels of specificity, e.g. they can group aryl side chains, meta-substituted aryls and O-linked meta-substituted aryls together depending on the level of "fuzziness" selected. Extended connectivity fingerprints (ECFPs) were applied to the R-groups to evaluate its potential as generalized substituent description in PLS modeling [36, 37].

Results and discussion

A drug discovery project is constantly faced with the challenge to objectively evaluate incoming data. The amount and complexity of data makes it almost impossible to have a current up-to-date and comprehensive picture. Well tuned automated ranking may seem as an unrealistic goal but a step in that direction is very appealing.

Composite multi-parameter ranking

In the MC4R project we decided to investigate the possibility of creating a composite scoring function. The goal was to implement the scoring function within a project query in IBIS Explore, the AZ global system for storing, processing and analyzing physicochemical properties and in vitro and in vivo screening data. Thus, by applying a multi-parameter ranking system we would get an instant picture as soon as new data arrives. To test the concept we decided to create models and scoring functions for eleven of the experimentally measured parameters previously described in the experimental section. In Fig. 5 is shown the IBIS Explore query for a set of project compounds. The data are sorted based on the NormRank value, which is the sum of the binned scores for the individual parameters divided by the number of experimental descriptors seen as the column Sum_Count in the Fig. 5. The final NormRank is also subtracted by a penalty function to properly manage bad data. The penalty function in turn gets its contribution from a set of "badness" flags, OrangeFlag, RedFlag and Showstopper, which are set for the descriptors. These accumulated OrangeFlag, RedFlag and Showstopper flags are also shown as columns in the query table colored green–yellow–red. As seen in Fig. 5, the OrangeFlag column becomes red when a compound accumulates four OrangeFlags. Similarly the RedFlag column is colored red after having obtained more than two flags, whereas the Showstopper column becomes immediately red after having obtained one flag. The query also contains original data as well as binned columns and "badness flags" for individual descriptors in the project spreadsheet view. The question mark before the numerical value arise as a consequence of conversion of out of range measurements to numerical values and are purely flags to indicate that original data contains that prefix.

The approach provides the flexibility to tune parameters as a way to reflect the manual ranking of the project

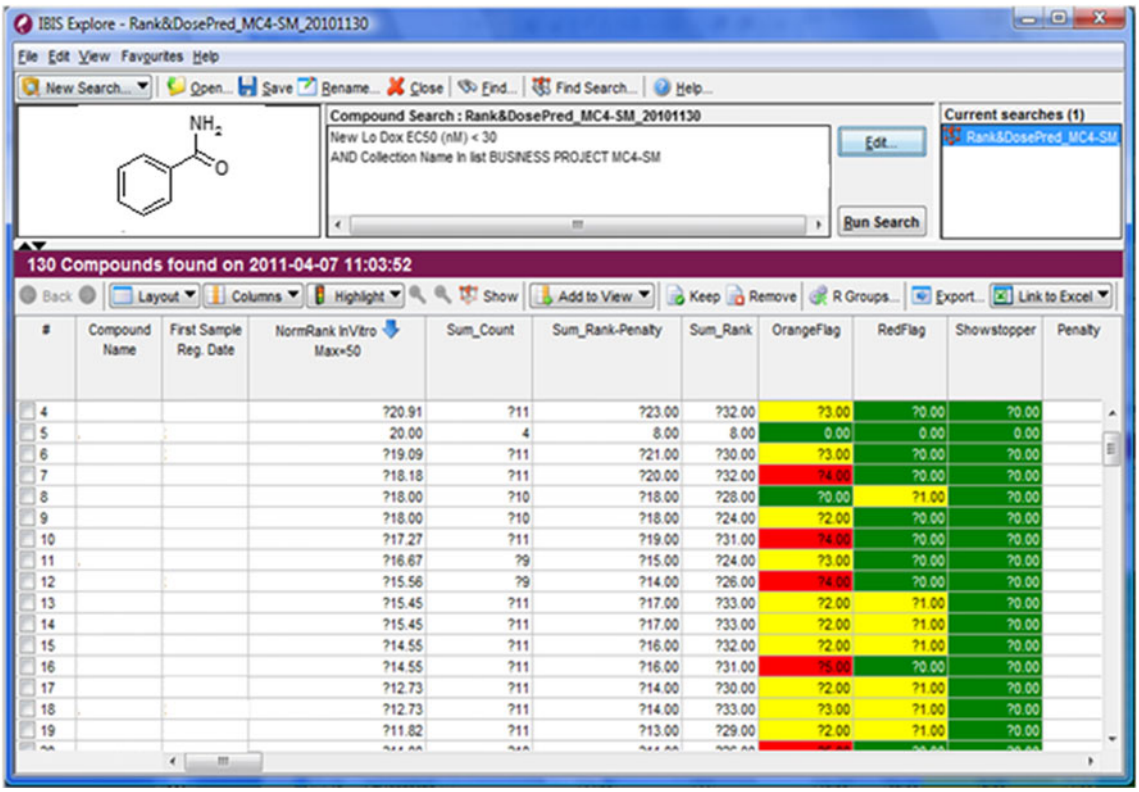

Fig. 5 The IBIS Explore user interface. The compounds are sorted according to the ranking. Number of OrangeFlags, RedFlags and Showstoppers and degree of issue is highlighted by *color code*

members and enables fast analysis of the details behind the ranking. Thus, the NormRank order and the badness flags provide an immediate picture of the status and issues of a compound. Secondly, the additional summary columns of OrangeFlag, RedFlag and Showstopper values, which are presented in conjunction with the NormRank, when color coded as shown in Fig. 5, provide an immediate feedback of the quality of the compounds.

In order to test the relevance of this approach, we applied the revised ranking protocol to a set of compounds selected as candidates for additional in vitro and in vivo studies. In parallel, eight members of the project team created their personal compound ranking. The individual ranking lists were then discussed and merged at a project team meeting. The intention of the process was to select around six compounds for extended studies, out of a set of ~50 pre-filtered compounds. Gratifyingly, when applying the revised NormRank score it ranked compounds in a way similar to the manual rank of the project team members. The first three compounds where top-ranked in the same order in both lists. The remaining three compounds in the team list were scored within the 15 best of the automated ranking list. Interestingly, two of these compounds ended up on the team list more as a result of subjective thinking rather than hard data. Notably, these two compounds were later dropped due to inferior results in follow up studies. However, even more importantly, the automated ranking strongly scored one compound that had been overlooked by the project team and this compound was upgraded to the final list for in vivo studies. An additional strength is that whenever new data appears the ranking table within IBIS Explore is automatically updated. Clearly there is a substantial benefit in this automated ranking as it provides an immediate picture of the most recent data as soon as it is displayed and complements/reinforce the human ranking.

In silico models and virtual ranking

Having an automated composite scoring function in place for real compounds, we decided to investigate the possibility to make a similar composite ranking for virtual compounds. Thus, we wanted to explore if we could build robust models for the experimental parameters discussed above, and use the composite ranking in the selection of compounds for synthesis. As described in the methods section, out of 264 compounds in the original data, 238 compounds were picked randomly and used as training set throughout all model buildings. Remaining 26 compounds comprise test set A. A second temporal test set (set B) of 34 compounds were synthesized and measured after the

models were built and provides a strong validation of the model performance outside the original data set. The quality of all models were assessed by linear regression of the measured versus predicted values for the training set with simple parameters as correlation coefficient R^2 as well as intercept (a) and slope (b) to report deviation from line of unity. Similarly the correlation and predictive power for the external test sets A and B was characterized by the same parameters. The correlation coefficient for the external test sets A and B is for clarity consistently named Q^2 throughout the paper.

QSAR models using molecular descriptors

Initially we focused our modeling efforts using traditional QSAR with molecular descriptors and linear or non-linear modeling techniques. The modeling was based on a set of molecular descriptors generated from Clab and models were built using PLS (Simca) and Random Forest (RDS). The $LogD_{74}$ PLS model based on 62 significant variables, looked quite promising ($R^2 = 0.87$ and $Q^2 = 0.85$ (cross validation as implemented in Simca)) for the training set and the correlation is close to line of unity. The 62 descriptors for the PLS models were manually selected from ~200 Clab descriptors as judged from the importance of the coefficient value and the height of the confidence interval bar as implemented in Simca. The number of descriptors after selection is approximately ¼ of the number of observations in the training set and is within the guidelines of not being a chance correlation [38]. Still, the corresponding correlation for the external test sets A and B, although acceptable ($Q^2 = 0.63$) is significant lower when compared to the training set. Moreover the correlation deviates significantly from the line of unity. There is also a difference in the correlation coefficient for set A ($Q^2 = 0.72$) and set B ($Q^2 = 0.61$) weakly indicating that more recent compounds in set B have slightly different properties from the training set compared to set A. The results for the Random Forest (RF) models, as implemented in RDS, were similar to those of PLS. Although the RF models described more of the variation in data with $R^2 = 0.96$ for the $LogD_{74}$ model, the external predictions were in the same order as the PLS with a Q^2 of 0.68 and 0.63 for the test sets A and B. The Q^2 for the simple $LogD_{74}$ for the external test set may still be regarded as acceptable but also an indication that the molecular descriptor set is inadequate and models for more complex measurements might be inferior. In view of this it was not surprising, that the quality of the PLS models for the Caco2 P_{app} and MC4 pEC_{50} dropped to $R^2 = 0.73$ and $R^2 = 0.58$, respectively. The Q^2 of 0.54 for the external test sets of the Caco2 P_{app} model is acceptable. However, the predictive power in the MC4 pEC_{50} model completely disappeared ($Q^2 = 0.05$). Similarly the Random Forest model for MC4R pEC_{50} gave deceptively better correlation ($R^2 = 0.93$) than PLS, but the predictive power was absent with $Q^2 = 0.12$ for the test sets. Clearly, the Clab molecular descriptor set does not result in predictive QSAR models using established methods. We also made some attempts using Clab descriptors for the R-groups, however, these attempts did not result in viable and predictive models and the approach was dropped.

Free-Wilson modeling

The failures in obtaining robust models using molecular descriptors and PLS/RF turned our attention to the possibility to apply the Free-Wilson methodology in the model building step. Much to our satisfaction the initial Free-Wilson model for the $LogD_{74}$ parameter looked very promising. In Fig. 6a is shown the plot of the experimental versus predicted $LogD_{74}$.The training set is in the plot and all subsequent figures shown as blue unfilled circles. Similarly, in Fig. 6b and subsequent figures the test set A is shown as red filled circles and the temporal test set B as green filled circles. The linear regression of the training set has an R^2 of 0.991 and the regression line coincides with the line of unity (Fig. 6a). In Fig. 6b the prediction for the joint test sets A and B are shown, with an almost as impressive correlation ($Q^2 = 0.972$) but with a slight deviation from line of unity ($a = 0.216$, $b = 0.96$). Notably, as a result of the modeling and design, the temporal set B (green circles) is confined to a much smaller $LogD_{74}$ range.

The model clearly indicates the absence of dominating interactions between the R-groups either through bonds or space. Thus, it seems almost as an ideal case for the Free-Wilson methodology. However, it must be emphasized, that although a good model is obtained for one experimental parameter, devoid of R-group interactions, it does not imply that R-group interaction terms are not present for other observables nor that there is enough data to reliably estimate new parameters. This may be one source of variability of model quality we have observed and together with the uncertainty in the experimental data provides the limit of the models. However, the very encouraging results prompted us to extend the Free-Wilson study. Consequently Free-Wilson models were built for all the experimental project parameters previously discussed.

In Fig. 7 are shown the Free-Wilson correlations of the logarithmic MC4R EC_{50} Caco2 P_{app}, CYPMIN and hERG data. The correlation coefficients for the test set for all these models are in all cases better than 0.9, which in general is better than the other QSAR models presented before. However, the correlation drops considerably for the external test set. Q^2 is 0.41 for the MC4R EC_{50} model for the joint test

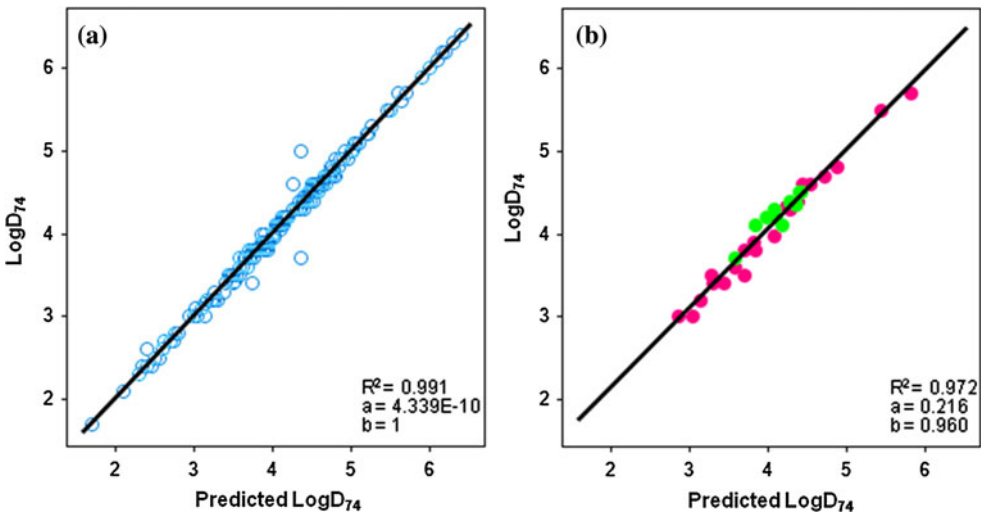

Fig. 6 **a** Free-Wilson model of chromatographic $LogD_{74}$. *Blue unfilled circles* training set. *Red filled circles* external dataset (set A) used when model was prepared. *Green filled circles* prediction on compounds (set B) synthesized after model was prepared. The correlation coefficient R^2 and the coefficients for the straight line fit (a = intercept, b = slope) for the training set are shown in the *lower right corner*. **b** Prediction showing only test set A (*red*) and set B (*green*) and concomitant line statistics in the *lower right corner*

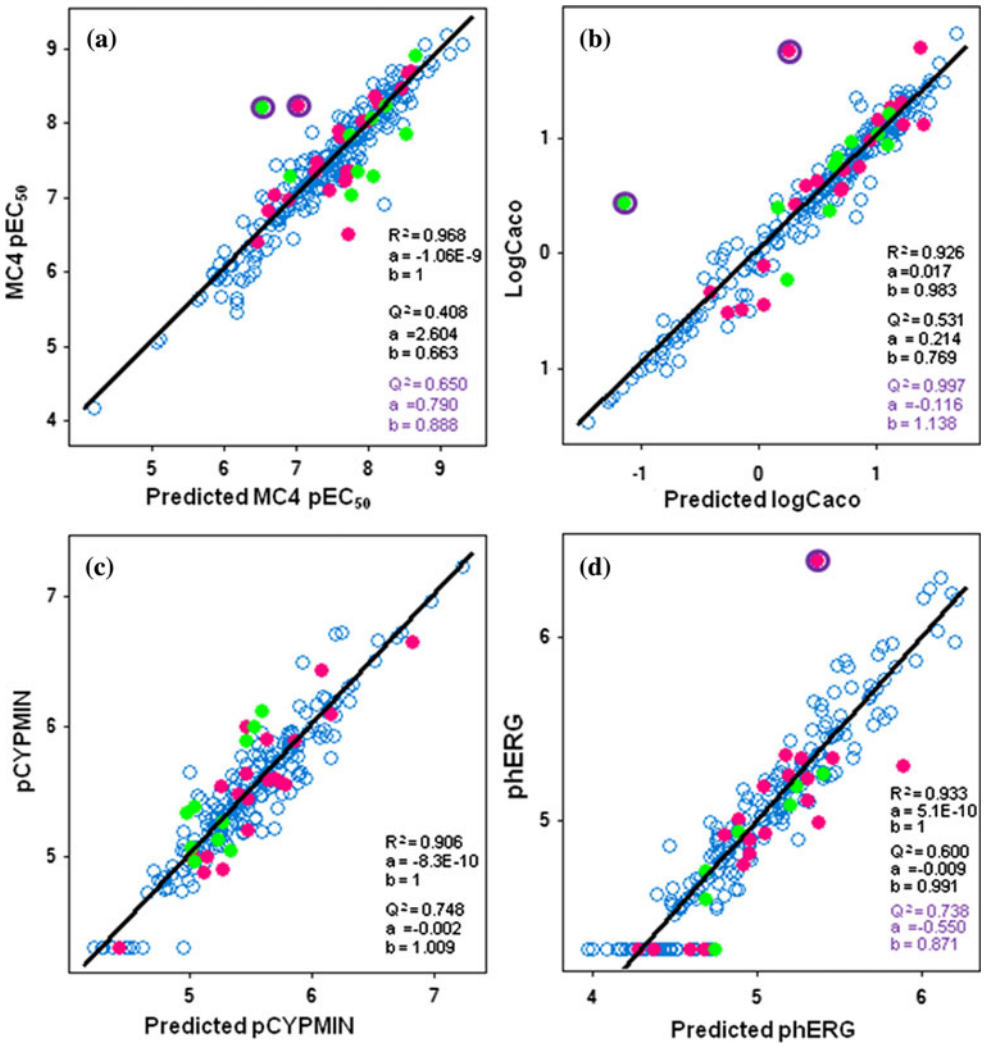

Fig. 7 Free-Wilson models of logarithmic values of **a** MC4R EC_{50} **b** Caco2 P_{app} **c** CYPMIN **d** hERG. *Blue unfilled circles* training set. *Red filled circles* external dataset (set A) used when model was prepared. *Green filled circles* prediction on compounds (set B) synthesized after model was prepared. *Large purple unfilled circles* indicate outliers that were removed in part of the statistical analysis. The correlation coefficient R^2 and the coefficients for the straight line fit (a = intercept, b = slope) for the training set followed by the correlation coefficient Q^2 with all for compounds the test sets A and B and corresponding straight line parameters are shown in *black the lower right corner*. In *purple* are the corresponding statistics when the outliers were removed

set A and temporal set B (Fig. 7a) as compared to 0.97 for the training set. Similarly Q^2 is 0.53 for the joint test set A and B for Caco2 P_{app} whereas R^2 for the training set is 0.93. Removing two outliers in the MC4R EC_{50} test set, as marked by open large purple circles, improves the correlation to 0.65. Admittedly, there is no scientific reason for removing

those records. They have been re-measured in the MC4R EC_{50} assay as well as the other assays and the experimental difference is within the expected range. Interestingly, these two records are also outliers in the Caco2 P_{app} as indicated by similar open purple circles in Fig. 7b and one of the compounds is also an outlier in the hERG model (Fig. 7d). There are no obvious reasons from a structural perspective for the poor predictions on these two compounds. The structural verification MS and NMR and purity analysis are consistent with the proposed structures and other compounds with these Rgroups in other combinations behave as expected. Obviously the Free-Wilson method does not have the necessary information to predict these compounds correctly. The CYPMIN model (Fig. 7c) is fairly robust with an R^2 of 0.91 and a Q^2 of 0.75, whereas the hERG model with an R^2 of 0.93 is slightly less predictive with a Q^2 of 0.60 which upon removal of one of the common outlier improves to a decent Q^2 of 0.73. Evidently models other than $LogD_{74}$ are not as well behaving and the correlation coefficient for the external test set drops considerably compared to the $LogD_{74}$ model. One possible reason for this is larger experimental errors for the other parameters which inevitably will introduce model noise and limit parameter estimate. Our analysis of three experimental descriptors $LogD_{74}$, LogCaco and MC4R pEC_{50} showed some differences in the experimental error. $LogD_{74}$ had a standard deviation of <0.1 and both LogCaco and MC4R pEC_{50} had a mean standard deviation of $\sim$0.3 independent of the absolute activity (For further details see experimental section). However, these differences in experimental errors cannot explain the differences in the quality of the models especially not the difference between LogCaco and MC4R pEC_{50} models, which have similar experimental error. Evidently it is not the only contribution to the decreased predictive power of these models in general and the MC4R pEC_{50} model in particular. The most plausible explanation for the deteriorated models is the existence of specific interactions and interaction terms mediated by the protein–ligand complex and the simple additivity of the R-group contribution in the Free-Wilson approach, which seems perfectly valid for the $LogD_{74}$, does not hold for more complex systems. We have not analyzed experimental error for the other experimental descriptors but we expect the experimental error to be of the same order. Despite the failures to predict some compounds we still regard these models as useful for further prediction purpose and a foundation for a composite virtual scoring function. Similarly the logarithmic hPPB model on logarithmic data exhibits the same behavior with an R^2 of 0.93, which drops to 0.67. Consistently, removal of one of the outliers described earlier also seen in the other models improves the correlation with a Q^2 of 0.81. Thus, in line with previous discussion, the hPPB model is also considered to fulfill our criteria for being used in a composite virtual scoring function. The Free-Wilson models for the non-logarithmic descriptors Cl_{int} hHeps and efflux ratio are shown in Fig. 8a, b respectively. Notwithstanding, the hHeps model for the training set is almost of the same quality ($R^2 = 0.86$) as the logarithmic models described earlier, which is quite encouraging as the experimental data is composed of several mechanisms such as cell membrane penetration and CYP mediated degradation.

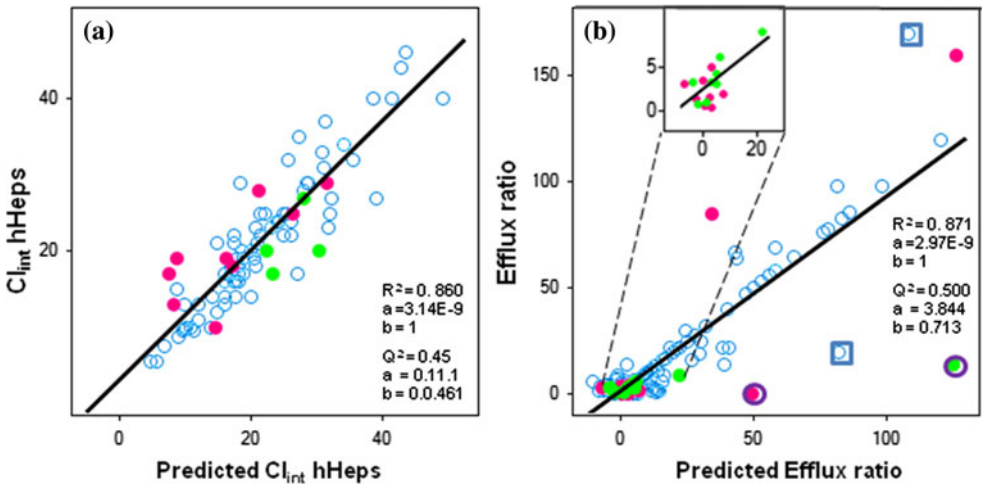

Fig. 8 Free-Wilson models for **a** Cl_{int} hHeps and **b** efflux ratio. *Blue unfilled circles* training set. *Red filled circles* external dataset (set A) used when model was prepared. *Green filled circles* prediction on compounds (set B) synthesized after model was prepared. *Blue rectangles* of training set and *large purple unfilled circles* of the test set A and B indicate outliers that were removed in part of the statistical analysis. The correlation coefficient R^2 and the coefficients for the straight line fit (a = intercept, b = slope) for the training set followed by the correlation coefficient Q^2 for all compounds in test sets A and B and corresponding straight line parameters are shown in *black in the lower right corner*

The correlation coefficient for the Cl_{int} hHeps model drops for the two external test sets to 0.45, but it should be seen in light of the accidental narrow experimental range encompassed by the test set compared to the training set. Consequently the Cl_{int} hHeps model was also considered to be good enough to be included in the virtual multi-parameter ranking described in the following. The Cl_{int} HLM model shows the same behavior as the Cl_{int} hHeps model, with an R^2 of 0.81 for the training set and Q^2 of 0.50 for the test set. It is perhaps surprising that the simpler Cl_{int} HLM model, which does not involve membrane transport, does not perform better than the Cl_{int} hHeps model and we have no obvious explanation for that, even though experimental complexity not necessarily relates to model complexity. The efflux ratio model (Fig. 8b), which inherently has an uncertainty as it is calculated from two independently measured rate constants and consequently have larger experimental uncertainties. Despite this, the model on the training set has an R^2 of 0.87 and the Q^2 for the test sets is 0.50. The outliers marked by open purple circles are the same as reported for the other assays. Two records in the training set stand out, as marked with blue large rectangles in Fig. 8b, one over and one under predicted but there is no obvious explanation for that. In the test set the absolute correlations for the high efflux compounds are low; however, the predicted ranking is correct. It is worth noticing that these training and test compounds share at least one R1 and/or R4-substituent. The most alarming incorrect prediction is the record with the observed value 0.32 which is predicted to be 49. We cannot explain this discrepancy but observe that it contains one of the R2-substituents that cause deviation in this and other models. The two high efflux compounds (85 and 160) in the test set are under predicted (35, 126) but this is not seen as an obstacle for using the model. The expansion in the upper left part of Fig. 8b shows the prediction of the test compounds in the low or medium efflux region. Clearly the prediction is far from perfect ($Q^2 = 0.46$ and line characteristics distant from line of unity), but we still think that the model can be used in our composite virtual ranking. It has in this case been used as a classification where predicted efflux ratio <5 was set to 5 in the binned ranking, between 5 and 30 was set to 3 and >30 was set to 1. Knowing the peculiar behavior of certain R1 and R4 substituents we can keep an eye on virtual compounds where these R-groups are included.

R-group contribution to experimental observables

The additivity inherent in the Free-Wilson methodology provides the possibility to estimate the contribution made by each R-group for a given model. These contributions can be nicely visualized in Bar Chart plots in TIBCO Spotfire. This may be seen as an extension of matched pair analysis [39], were variations on one R-group can be analyzed, while the remaining part of the molecule is the same. In Free-Wilson analysis we will have the average contribution for one R1-substituent in all its combination with R2, R3 and R4 for the given scaffold. In Fig. 9 are shown the normalized contribution plots for MC4 pEC_{50}, logCaco, pCYPMIN and Cl_{int} hHeps for the R1, R2 and R3 groups. One immediately spots the contribution for a given R-group for the experimental variables. The R2-substituent 6-cyanopyridazin-3-yl depicted in Fig. 9 have obviously a strongly positive effect on the MC4R agonistic activity (trace a) as well as the permeability (logCaco, trace b) of the compound compared to similar compounds lacking this substituent. However, it also contributes to the increased inhibition of the CYP 450 metabolic enzymes as seen in trace c and increased metabolism as measure by hHeps (trace d) and is clearly detrimental for the overall properties of the compounds having this substituent. These plots also guide the computational chemist of which R-groups that should be included in the Free-Wilson analysis of virtual libraries described in the following section.

The virtual composite ranking

In previous sections we have elaborated on the composite ranking using experimental descriptors. We will delineate our approach to the analogous composite scoring in the virtual space using the predicted descriptors calculated from the Free-Wilson models. The virtual library generated from all combinations of the current set of 10 R1-, 14 R2-, 5 R3- and 103 R4-groups comprise of 72,100 potential compounds. The set of virtual compounds could have be reduced before enumeration by analyzing the contribution plot discussed above and remove substituents that are detrimental for one or several experimental descriptors. However, one is still faced with a large set of virtual compounds and we sought to identify areas of interest in the virtual compound space and to identify missed opportunities and select the most promising virtual compounds for synthesis. Having an automated composite scoring function in place for real compounds and real data, we therefore turned our attention to a similar composite ranking for virtual compounds. As the Free-Wilson methodology produced acceptable results for a set of the experimental variables, we decided to use the predicted values from the Free-Wilson predictions, based on the previously described models, to the 72 K virtual compounds. A similar binning (1–5) was applied to the predicted values for MC4R EC_{50}, chromatographic $LogD_{74}$, water solubility at pH 7.4 from DMSO solution, potency against five different CYP's, hERG inhibition, Cl_{int} HLM

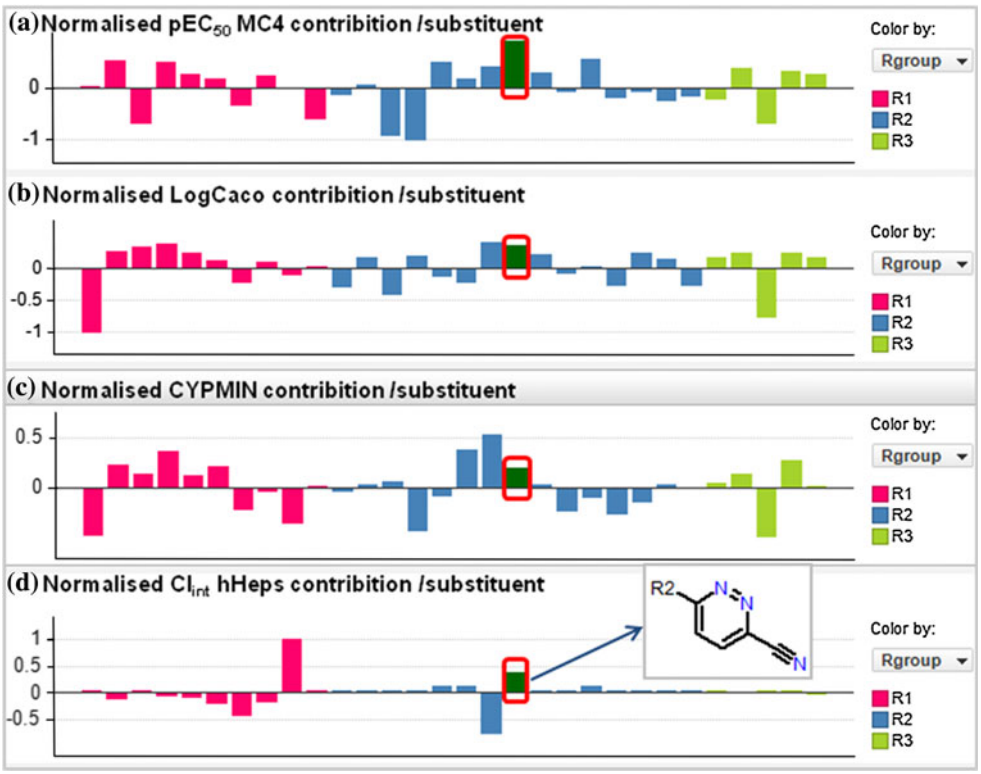

Fig. 9 Normalized R-group contribution plots for selected Free-Wilson models

and Cl_{int} hHeps, Caco2 permeability, efflux ratio in caco2 cells and hPPB. The only difference was that the limits for the upper (5) and lower (1) was adjusted to be slightly more forgiving on false negative predictions. The inherent limitation of the Free-Wilson methodology prevents predictions for virtual compounds where an experimental value for a particular R-group has not been measured. Thus, there will be substantially fewer predictions for descriptors where the model training set is limited. For instance, in this set of compounds the number of predictions for Cl_{int} hHeps was significantly less than those of MC4R EC_{50}. The predicted binned values were then summed up and the sum was divided by the number of Free-Wilson predictions for the virtual compound to create the final score (VirtNormRank). Thus, the virtual ranking is in similarity to the real scoring normalized by the number of predictions obtained. In Fig. 10a the VirtNormRank is plotted against the predicted MC4R pEC_{50} values. In Fig. 10b the ranking applied to the real compounds, that were used as training set for the virtual compound predictions, are shown and the virtual predictions are shown as a grey shadow.

The coloring indicates the number of predictions obtained for each compound. Obviously the experimental data of the training set does not allow Free-Wilson predictions to be made for every single virtual compound. This is a consequence of the screening cascade for the real compounds. A negative experimental value results in a reduced interest from the project to fulfill the tests

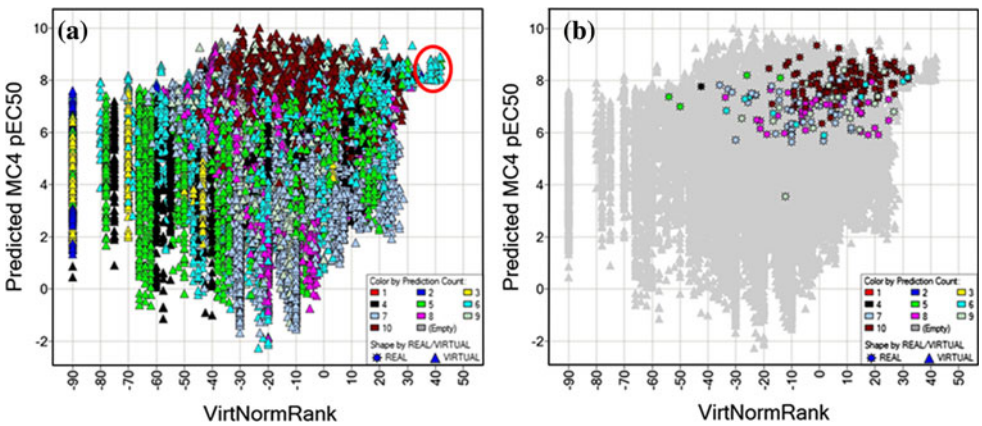

Fig. 10 **a** Score (VirtNormRank) plotted against the predicted MC4R pEC_{50}. **b** Score (VirtNormRank) for the training set plotted against the predicted MC4R pEC_{50}. The *color coding* indicates the number of Free-Wilson predictions the score for virtual compound is based on

requested for a particular compound, which limits the Free-Wilson prediction space. Nevertheless, the in silico multi-parameter scoring for the MC4R project was very encouraging and has been used to select compounds in regions of high score. A set of compounds in the region marked by the red circle in Fig. 10a were, based on the Free-Wilson composite virtual scoring protocol, selected for synthesis. In addition virtual compounds, located close to real compounds that were assessed as interesting, were also selected for synthesis. Figure 11 shows the results from the MC4R pEC_{50} and the Cl_{int} HLM models of 9 compounds that were selected and synthesized on the basis of the virtual ranking. In Table 2 the corresponding statistics for all models are shown.

The MC4R pEC_{50} correlation shown in Fig. 11a has a Q^2 of 0.46 that is of dubious predictive utility but in agreement with the earlier test sets. On the other hand the Cl_{int} HLM correlation (Fig. 11b) is, except for one "outlier" marked with red circle, almost close to perfect with a Q^2 of 0.99 when excluding that record but drops to 0.56 when included. The latter is in accordance with $Q^2 = 0.60$ for the external test set as presented earlier. The intercept (6.35) and slope (1.07) is also good considering the two orders of magnitude of data range. The statistics of the predictions for all the models are summarized in Table 2. Examining Table 1 reveals that Cl_{int} hHeps only has four experimental observations and has no correlation ($Q^2 = 0.1$). The experimental variation is also quite modest (17–26 mL/min/kg) and the predictions fall within a similar range (24–34 mL/min/kg). The logarithmic solubility (pSol) correlation is also low ($Q^2 = 0.35$), but also in this case is the experimental variation quite modest. The remaining models exhibit decent to very good correlations ($Q^2 = 0.66$–0.82). Interestingly the hERG model although with only six measurements is even better than the $LogD_{74}$ model. This could be due to the fact that the data set is small.

It must also be emphasized that the linear curve fitting coefficients (parameter a and b in Table 2) deviates in most cases from the coefficients of the training set models even though the correlation coefficients R^2 and Q^2 are good. Thus, although the ranking may be fairly good the absolute values may deviate significantly. This may have implications on the comparison of the composite real scoring function and the corresponding virtual ranking. The correlation of the composite real and virtual scoring of the nine compounds, which were selected from the virtual ranking are shown in Fig. 12.

The real and virtual ranking in Fig. 12a is simply the sum of the individual bins divided by the number of parameters counted and does not take into account showstoppers, red or orange flag discussed earlier, as they had negligible effect on

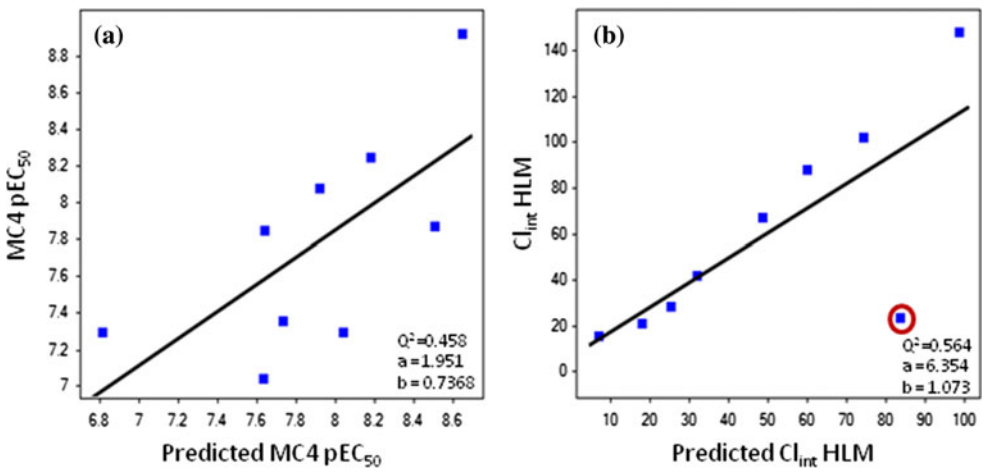

Fig. 11 **a** Predicted versus measured MC4R pEC_{50} for the 9 Free-Wilson selected compounds. **b** Measured versus predicted Cl_{int} HLM for the 9 Free-Wilson selected compounds

Table 2 Statistics for the nine compounds selected based on the basis of Free-Wilson models and virtual ranking

Parameter	LogD74	pEC50 MC4	logCaco	Efflux ratio	phERG	pCYPMIN	loghPPB	Clint HLM	Clint hHEPS	pSol
Q2	0.87	0.46	0.8	0.75	0.92	0.66	0.63	0.57 (0.99)	0.1	0.35
a	0.69	1.95	−0.12	1.45	0.92	−2.1	0.31	6.35	13.8	3.26
b	0.85	0.74	1.05	0.33	0.83	1.43	0.61	1.1	0.26	0.2
RMSE	0.09	0.47	0.21	1.43	0.09	0.26	0.14	31.9 (5.3)	4.9	0.05
Mean	4.25	7.73	0.7	3.85	4.96	5.47	0.96	59.6	21.1	4.1
Observations	9	9	9	8	6	9	9	9	4	9

The parameter a is the intercept and b the slope of the linear regression

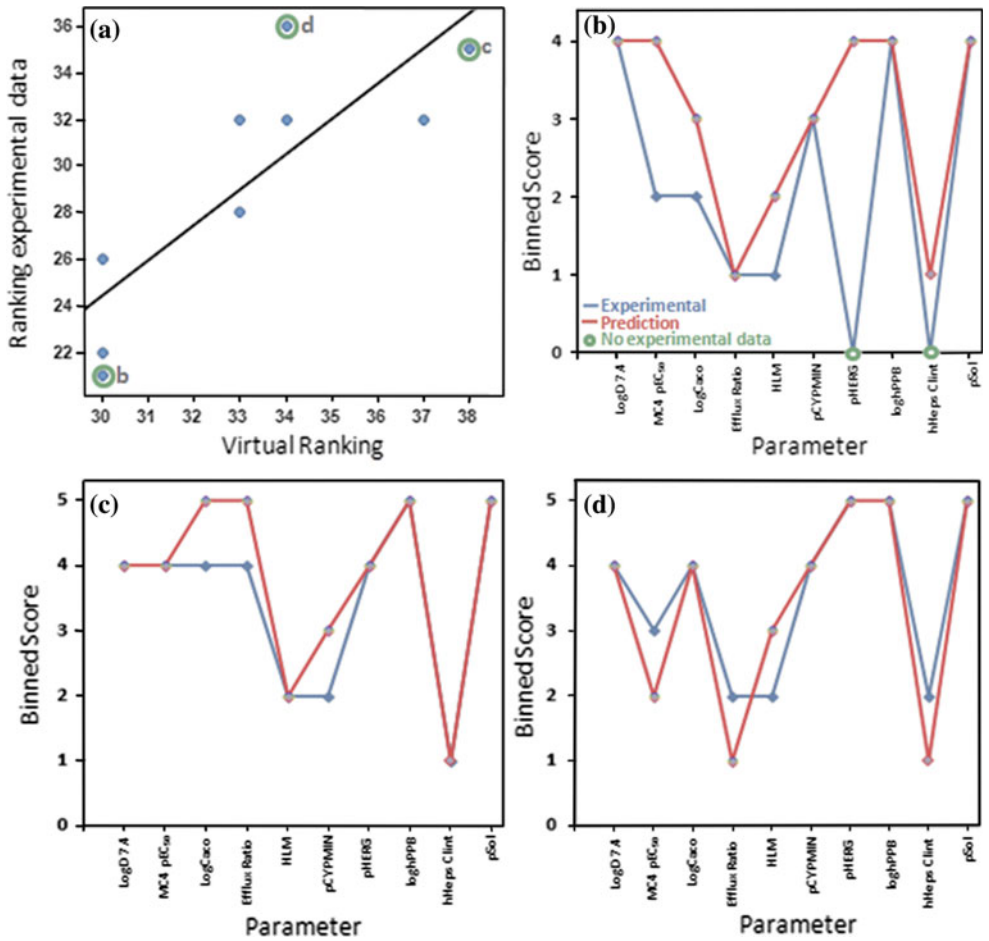

Fig. 12 **a** Real versus virtual ranking for the 9 compounds from second external test set. The *dots* marked **b–d** are the binned score plots in corresponding **b–d** figures. **b–d** Profile charts for 3 of the 9 selected compounds. Binned experimental values are shown as *blue lines* and corresponding calculated values in *red*

the overall correlation and ranking. The overall correlation in Fig. 12a does perhaps not appear that impressive but if one considers the fairly small range of values 26–36 for the real rank and 30–38 for the virtual rank we may probably not expect a better correlation. This impression is reinforced by investigating the profile chart for selected individual compounds (Fig. 12b–d), which shows the binned (1–5) score for the 10 parameters. The binned experimental values are shown as blue lines and the binned calculated values in red. Clearly, the experimental and calculated profiles in Fig. 12c, d, marked with the corresponding letter in Fig. 12a, overlap quite nicely. The virtual profile in 12c is over predicted with one bin in MC4R pEC_{50}, logCaco and pCYPMIN and the remaining parameters are correctly predicted. On the other hand the virtual profile in 10d is under predicted in MC4R pEC_{50}, efflux ratio and Cl_{int} hHeps but still only one bin unit, whereas Cl_{int} HLM is over predicted by one bin unit and the rest are correctly predicted. It is very gratifying to report that these compounds were added to the project shortlist of compounds to be subjected to extended in vivo studies. In Fig. 12b is shown the profile chart of the dot marked by b in Fig. 12a where the real and virtual ranking deviates mostly. Despite this the alignment of the real and virtual profiles in the profile chart are quite good. In one case, MC4R pEC_{50}, the difference is two units with calculation over predicting. In two cases no experimental value has been reported which in the plot is seen as a zero value and marked by a green circle. The profile charts for the remaining six compounds exhibits a similar pattern, where in only one case (efflux ratio) the difference between real (2) and predicted score (5) is three units. In summary we strongly believe that, although substantial work needs to be done to perform the virtual ranking in the profile chart and despite some deficiencies noticed, the approach has got its clear merits.

Future perspectives

The application of the Free-Wilson methodology and composite parameter scoring of virtual compounds is obviously a very appealing in silico support to the drug

discovery projects. Clearly, when the appropriate conditions are fulfilled Free-Wilson will add significant value to a drug discovery project as described in this article. However, also the sun has its dark spots. The Free-Wilson methodology is not applicable outside the box defined by the already used experimental R-groups. Thus, a Free-Wilson model will not be able to predict the contribution from an R-group that has not already been used and measured, which is seen in Fig. 10, where only a subset have a rank based on a predictions for all models. This also implies that Free-Wilson treatment is not applicable until you have synthesized a substantial number of compounds and obtained the associated experimental data. One approach to solve this dilemma would be to create R-group descriptions that introduce some sort of fuzziness or generality. Initial attempts have been performed using both a 2D pharmacophore fingerprint SMG (SMARTS Multiplet Generator) and a fuzzy R-group description. It must be emphasized that this has not been the intended use of these descriptors and in neither of these cases the quality of the model was good enough. Clearly, there is a subtle balance between the predictive power and the generality of a model. The more general/fuzzy the parameters used to create the model, the larger the uncertainty/error in the predictions becomes. We are currently investigating different approaches in this field using signatures and extended fingerprints as fuzzy R-group descriptors and intend to report on these studies in due course.

Conclusions

We have described the development of a composite parameter scoring protocol based on experimental data of compounds in a drug discovery project. The automated ranking was able to rank correctly the top 3 compounds from >300 compounds in a shortlist of 6 compounds produced independently by the project team. Even more important the ranking was able to highlight compounds that for some reason have been neglected by the project team, which lead to further studies on these compounds. The ranking system has been implemented in IBIS Explore, the AstraZeneca in-house database and query tool for project data, thereby providing immediate ranking updates as data arrives and we have clearly shown that a well tuned scoring function can add significant value to the project decisions. The philosophy of the ranking system is general and can be tuned to fit any drug discovery project and include physical, in vitro as well as in vivo parameters and we have recently used it successfully in that context in another project as support for the selection of candidates for further candidate drug profiling.

We have also shown that Free-Wilson methodology, using R-group description rather than traditional molecular descriptors and QSAR, has resulted in models with medium to very good predictive power, as measured for two test sets, for a wide variety of physical and biological responses. Inspired by the successful outcome of the experimental composite ranking, we undertook a similar virtual exercise. We generated the library (72 k) of virtual compounds from the R-groups present in the experimental compounds. The Free-Wilson predictions for 10 experimental variables were all included although some models such as the MC4 pEC_{50} is on the border line to be useful in a composite ranking. A binned score and ranking scheme, as for the real compounds, was generated for the virtual library. Compounds were selected for synthesis and were shown to have activity profile for the measured properties in good agreement with the predictions. Thus, we may conclude that the Free-Wilson models can be very effective in retrospective R-group analysis. Finally, objective ranking based on the Free-Wilson models provides a powerful tool to guide the multi-parameter optimization and clearly highlights virtual compounds to be added to the high priority synthesis queue.

Acknowledgments Thanks to Thomas Leek for providing statistics for the chromatographic $LogD_{74}$ model. Stephan Hjorth is acknowledged for strong support during the development of the models. Alleyn Plowright, Christian Tyrchan, Anders Hogner and Niklas Blomberg are acknowledged for constructive criticism of the manuscript.

References

1. Nozawa D, Chaki S, Nakazato A (2008) Recent advances in the development of melanocortin-4 receptor ligands. Exp Opin Ther Pat 18:403–427
2. AstraZeneca International Bioscience Information System (IBIS) (www.bio-itworld.com/BioIT_Article.aspx?id=78250)
3. Hutter MC (2009) In silico prediction of drug properties. Curr Med Chem 16:189–202
4. Cramer RD III, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967
5. Chadwick A, Hajek M (2004) Learning to improve the decision-making process in research. Drug Discov Today 9:251–257
6. Li H, Yap CW, Xue Y, Li ZR, Ung CY, Han LY, Chen YZ (2005) Statistical learning approach for predicting specific pharmacodynamic, pharmacokinetic, or toxicological properties of pharmaceutical agents. Drug Dev Res 66:245–259
7. Duch W, Swaminathan K, Meller J (2007) Artificial intelligence approaches for rational drug design and discovery. Curr Pharm Des 13:1497–1508
8. Yap CW, Xue Y, Li H, Li ZR, Ung CY, Han LY, Zheng CJ, Cao ZW, Chen YZ (2006) Prediction of compounds with specific pharmacodynamic, pharmacokinetic or toxicological property by statistical learning methods. Mini-Rev Med Chem 6:449–459
9. Yap CW, Li H, Ji ZL, Chen YZ (2007) Regression methods for developing QSAR and QSPR models to predict compounds of

specific pharmacodynamic, pharmacokinetic and toxicological properties. Mini-Rev Med Chem 7:1097–1107
10. Mager DE (2006) Quantitative structure-pharmacokinetic/pharmacodynamic relationships. Adv Drug Deliv Rev 58:1326–1356
11. Free SMJ, Wilson JW (1964) A mathematical contribution to structure-activity studies. J Med Chem 7:395–399
12. Craig PN (1972) Comparison of the Hansch and Free-Wilson approaches to structure-activity correlation. Adv Chem Ser 114:115–129
13. Kubinyi H, Kehrhahn OH (1976) Quantitative structure-activity relationships. 3.1 A comparison of different Free-Wilson models. J Med Chem 19:1040–1049
14. Hansch C, Maloney PP, Fujita T, Muir RM (1962) Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature (London, UK) 194:178–180
15. Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M (1963) The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. J Am Chem Soc 85:2817–2824
16. Jorissen RN, Reddy G, Kiran Kumar S, Ali A, Altman MD, Chellappan S, Anjum SG, Tidor B, Schiffer CA, Rana TM, Gilson MK (2009) Additivity in the analysis and design of HIV protease inhibitors. J Med Chem 52:737–754
17. Sciabola S, Stanton RV, Johnson TL, Xi H (2011) Application of Free-Wilson selectivity analysis for combinatorial library design. Methods Mol Biol 685:91–109
18. Hoefgen N, Stange H, Schindler R, Lankau H, Grunwald C, Langen B, Egerland U, Tremmel P, Pangalos MN, Marquis KL, Hage T, Harrison BL, Malamas MS, Brandon NJ, Kronbach T (2010) Discovery of Imidazo[1,5-a]pyrido[3,2-e]pyrazines as a new class of phosphodiesterase 10A inhibitors. J Med Chem 53:4399–4411
19. Patel Y, Gillet VJ, Howe T, Pastor J, Oyarzabal J, Willett P (2008) Assessment of additive/nonadditive effects in structure-activity relationships: implications for iterative drug design. J Med Chem 51:7552–7562
20. Tomic S, Nilsson L, Wade RC (2000) Nuclear receptor-DNA binding specificity: a COMBINE and Free-Wilson QSAR analysis. J Med Chem 43:1780–1792
21. Freeman-Cook KD, Amor P, Bader S, Buzon LM, Coffey SB, Corbett JW, Dirico KJ, Doran SD, Elliott RL, Esler W, Guzman-Perez A, Henegar KE, Houser JA, Jones CS, Limberakis C, Loomis K, McPherson K, Murdande S, Nelson KL, Phillion D, Pierce BS, Song W, Sugarman E, Tapley S, Tu M, Zhao Z (2012) Maximizing lipophilic efficiency: the use of Free-Wilson analysis in the design of inhibitors of acetyl-CoA carboxylase. J Med Chem 55:935–942
22. Kubinyi H (1988) Free-Wilson analysis. Theory, applications and its relationship to Hansch analysis. Quant Struct Act Relat 7:121–133
23. Kubinyi H (2004) 2D QSAR models: Hansch and Free-Wilson analyses. Comput Med Chem Drug Discov, 539–570
24. Valko K, Bevan C, Reynolds D (1997) Chromatographic hydrophobicity index by fast-gradient RP-HPLC: a high-throughput alternative to log P/log D. Anal Chem 69:2022–2029
25. (2009) Phase version 3.1, Schrodinger, LLC New York, NY
26. (2011) MacroModel version 9.9, Schrodinger, LLC New York, NY
27. JMP Statistical Software (www.jmp.com/software)
28. TIBCO Spotfire (spotfire.tibco.com)
29. Compumine, Rule Discovery System (www.compumine.com)
30. Wold H (1966) Nonlinear estimation by iterative least squares procedures. In: David FN (ed) Research papers in statistics. Wiley, New York, pp 411–444
31. Wold H (1982) Soft modeling: the basic design and some extensions. In: Joreskog KG, Wold H (eds) Systems under indirect observation: causality-structure-prediction. North-Holland, Amsterdam, vol II, Chap. 1, pp 1–54
32. Wold H (1985) Partial least squares. In: Kotz S, Johnson NL (eds) Encyclopedia of statistical sciences. Wiley, New York, vol 6, pp 581–591
33. SIMCA P + (www.umetrics.com)
34. Cosgrove D. AstraZeneca internal development
35. Harrison M. AstraZeneca internal development
36. Rogers D, Brown RD, Hahn M (2005) Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. J Biomol Screen 10:682–686
37. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50:742–754
38. Topliss JG, Edwards RP (1979) Chance factors in studies of quantitative structure-activity relationships. J Med Chem 22:1238–1244
39. Leach AG, Jones HD, Cosgrove DA, Kenny PW, Ruston L, MacFaul P, Wood JM, Colclough N, Law B (2006) Matched molecular pairs as a guide in the optimization of pharmaceutical properties; a study of aqueous solubility, plasma protein binding and oral exposure. J Med Chem 49:6672–6682