Three-Dimensional Quantitative Structure-Permeability Relationship Analysis for a Series of Inhibitors of Rhinovirus Replication

Sean Ekins,*,§ Gregory L. Durst,§ Robert E. Stratford,§ David A. Thorner,‡ Richard Lewis,‡ Richard J. Loncharich,† and James H. Wikel§

Lilly Research Laboratories, Eli Lilly and Co., Lilly Corporate Center, Indianapolis, Indiana 46285, Lilly Research Centre, Sunninghill Road, Windlesham, GU20 6PH, U.K., and Chemical Research Technologies, Eli Lilly and Co., Lilly, 2001 West Main Street, Greenfield, Indiana 46140

Received July 20, 2001

Multiple three-dimensional quantitative structure−activity relationship (3D-QSAR) approaches were applied to predicting passive Caco-2 permeability for a series of 28 inhibitors of rhinovirus replication. Catalyst, genetic function approximation (GFA) with MS-WHIM descriptors, CoMFA, and VolSurf were all used for generating 3D-quantitative structure permeability relationships utilizing a training set of 19 molecules. Each of these approaches was then compared using a test set of nine molecules not present in the training set. Statistical parameters for the test set predictions (r^2 and leave-one-out q^2) were used to compare the models. It was found that the Catalyst pharmacophore model was the most predictive (test set of predicted versus observed permeability, $r^2 = 0.94$). This model consisted of a hydrogen bond acceptor, hydrogen bond donor, and ring aromatic feature with a training set correlation of $r^2 = 0.83$. The CoMFA model consisted of three components with an r^2 value of 0.96 and produced good predictions for the test set ($r^2 = 0.84$). VolSurf resulted in an r^2 value of 0.76 and good predictions for the test set ($r^2 = 0.83$). Test set predictions with GFA/WHIM descriptors ($r^2 = 0.46$) were inferior when compared with the Catalyst, CoMFA, and VolSurf model predictions in this evaluation. In summary it would appear that the 3D techniques have considerable value in predicting passive permeability for a congeneric series of molecules, representing a valuable asset for drug discovery.

Optimizing absorption, distribution, metabolism, and excretion (ADME) properties and minimizing toxicity of new molecular entities is of principal importance to the pharmaceutical industry. The reason for this focus is that the failure of a candidate in clinical evaluation because of one of these properties represents one of the many concerns linked to the increasing cost of drug development. The prediction of drug absorption is recognized as a major challenge to the discovery and development of promising orally efficacious new chemical entities.[1] Caco-2 cells are widely used as an in vitro system to predict human absorption and in this regard are still undergoing development.[2−4] However they possess several limitations including long preparation time, very slow absorption times compared to human intestine and large interlaboratory differences in quantitative results.[2] In light of the limitations in throughput with these systems, the development of higher throughput experimental and computational tools for the reliable prediction of human absorption continues.[4,5] The permeability of molecules via passive diffusion through cellular membranes is presently thought to be dependent upon molecular properties such as lipophilicity, hydrogen bonding, size, and charge.[6] Many computational approaches have taken account of some or all of these properties. Recent successes in computational modeling of absorption include the use of simulation models[1] as well as the incorporation of physicochemical parameters (as well as polar surface area) into models that use multivariate regression analysis.[6−18] These computational modeling advances provide promise that they can be used routinely to provide increased and reliable oral absorption screening capacity. Such models also provide an opportunity for greater insight into the molecular determinants of membrane transport. An additional computational approach that we believe can be applied to Caco-2 data is three-dimensional quantitative−structure activity relationship (3D-QSAR) approaches which have been used in defining important chemical features for cytochrome P450 substrates and inhibitors.[19−23] Application of these 3D-QSAR computational methods to generate 3D-quantitative permeability relationships (3D-QSPR) will further advance structure-transport relationship studies earlier in the discovery process concomitant with optimizing other ADME properties.[24]

The objective of the present study was to evaluate multiple computational approaches for the estimation of passive intestinal absorption by applying them to Caco-2 permeability coefficient data for a series of structurally related 2-aminobenzimidazoles possessing potent and broad-spectrum inhibition of rhinovirus and enterovirus replication.[25] In a recent report on these compounds, the coordinated use of in vitro techniques such as bioavailability surrogates was shown to be valuable in explaining their poor oral bioavailability in rats and in overcoming this problem while maintaining their potency.[25] Using a data set of known passively absorbed

* Corresponding author phone: (317)433-5387; fax: (317)433-0311; e-mail: ekins_sean@lilly.com.
† Chemical Research Technologies.
‡ Lilly Research Centre.
§ Lilly Research Laboratories.

10.1021/ci010330i CCC: $20.00
Published on Web 09/19/2001

and structurally similar molecules represents a foundation for future 3D-QSAR studies that will include actively transported and more structurally diverse molecules. In the present study a comparison of various commercially available 3D-QSAR techniques was made including Catalyst, comparative molecular field analysis (CoMFA), VolSurf, and genetic function approximation (GFA) using 3D molecular surface weighted holistic invariant molecular (MS-WHIM) descriptors. This paper describes how passive absorption can be predicted using commercially available 3D-QSAR approaches to provide a readily interpretable in silico prediction of in vitro Caco-2 permeability.

EXPERIMENTAL SECTION

Materials. 2-Aminobenzimidazoles were synthesized and characterized at the Lilly Research Laboratories as previously reported.[25] Culture medium components and reagents for cell culture were obtained from GIBCO Life Technologies, Inc. (Grand Island, NY). Fetal bovine serum for Caco-2 culture was obtained from Hyclone Corp. (Logan, CT). Sulforhodamine 101, used as a marker for Caco-2 monolayer integrity, was from Molecular Probes (Eugene, OR). All other reagents were purchased from Sigma Chemical Co. (St. Louis, MO) or Fisher Scientific, Inc. (Fairlawn, NJ).

Caco-2 Permeability. Caco-2 cells were obtained from the Memorial Sloan-Kettering Cancer Center and were cultured at 37 °C in a humidified atmosphere of 5% CO_2 in air. They were grown in DMEM:F12 (3:1) media supplemented with 5% fetal bovine serum and 50 μg/mL tobramycin. Monolayers at 75−90% confluency were either subcultured on a weekly basis using a 1:10 split ratio or seeded onto Millicell-PCF polycarbonate inserts (30 mm diameter, 0.4 μm pore size; Millipore Corp., Bedford, MA) at a density of 600 000 cells per filter. The culture medium was replaced with fresh medium every other day. For transport studies, cells from passages 43−56 were used at 21−30 days postseeding. Measurement of transepithelial electrical resistance was made immediately prior to an experiment using a Millicell-ERS system (Millipore Corp.). Cells with a net resistance less than 300 $\Omega \cdot cm^2$ were not used. Measurement of transepithelial flux of the various compounds was made at 37 °C using a side-by-side diffusion chamber described previously.[26] Transport medium consisted of 25 mM HEPES (pH 7.4) or 25 mM MES (pH 6.0) as buffering agents along with 125 mM NaCl, 5.2 mM KCl, 1.2 mM $CaCl_2$, 1.2 mM $MgCl_2$, and 10 mM glucose. To initiate an experiment, a stock solution of an inhibitor in DMSO was added to either the apical or basal side of the monolayer. The final concentration of DMSO was 1.4% (v/v). Samples (200 μL) were taken from the contralateral compartment at various times up to 120 min and replaced with an equal volume of fresh transport buffer. For a given compound, rate of transport under initial rate conditions (<10% transport) was determined by linear regression analysis and divided by initial donor concentration and the filter surface area to calculate the corresponding permeability coefficient. Experiments were conducted in duplicate or triplicate under limited lighting conditions. At the end of an experiment, monolayer integrity was evaluated according to the percentage transport of sulforhodamine 101 in 30 min. Monolayers with greater than 0.2% transport in this time period were not included in the analysis of permeability (there were never fewer than two monolayers used in the estimation of a given permeability coefficient). Analysis of the various compounds was by HPLC-UV and permeability coefficients (P_{app}) were generated as previously described.[25]

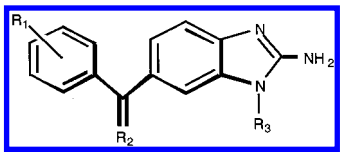
Molecular Modeling. The molecular modeling studies were carried out using Silicon Graphics Octane and O2 workstations. Molecular structures (Table 1) were represented as SMILES (Simplified Molecular Input Line Entry System) string format.[27] Initially 19 of the molecules (Table 1) were selected at random (with proprietary software) as a training set, and the remaining nine were used as a test set (a means of assessing the model produced). These same sets of molecules were used along with the permeability data expressed as $-\log P_{app}$ (except for Catalyst which used $1/P_{app}$) to generate models as described below.

Modeling with Catalyst. Briefly, models were constructed using Catalyst version 4.5 (Molecular Simulations, San Diego, CA) after importing the molecular structures in the SMILES format and using the permeability data expressed as $1/P_{app}$. The 3-D molecular structures were generated as previously described for CYPs.[19−23] Catalyst's "best" conformational search method was used for each compound with the number of conformers limited to 255 with an energy range of 20 kcal/mol. Ten hypotheses were generated using these conformers for each of the molecules and the transformed permeability values, after selection of the following features for the molecules: hydrogen bond donor, hydrogen bond acceptor, hydrophobic, negative ionizable, and ring aromatic. After assessing all 10 hypotheses generated, the lowest energy cost hypothesis was considered the best as this possessed features representative of all the hypotheses and had the lowest total cost.

The goodness of the 3D-QSPR correlation between the estimated and observed permeability values was estimated by means of an r^2 value. Statistical significance of the retrieved hypothesis was verified by permuting (randomizing) the response variable 10 times, i.e., the permeabilities of the training set compounds were mixed a number of times (so that each value was no longer assigned to the original molecule) and the Catalyst hypothesis generation procedure was repeated. The total energy cost of the generated pharmacophores can be calculated from the deviation between the estimated permeability and the observed permeability, combined with the complexity of the hypothesis (i.e. the number of pharmacophore features). A null hypothesis can also be calculated which presumes that there is no relationship in the data, and the experimental permeabilities are normally distributed about their mean. Hence, the greater the difference between the energy cost of the generated hypothesis and the energy cost of the null hypothesis, the less likely it is that the hypothesis reflects a chance correlation.

Validation of the Catalyst Model. The test set contained nine molecules with permeability values not included in the initial training sets as described previously. These test set molecules were fit by the fast-fit algorithm to the Catalyst model in order to predict a permeability value as previously described for CYPs.[19−23] Fast fit refers to the method of finding the optimum fit of the molecule to the hypothesis among all the conformers generated without performing an energy minimization.

Table 1. Structure and Permeability of Vinyl Carboxamide 2-Aminobenzimidazoles



compd no.	R1	R2	R3	structural class	permeability (cm/s × 10^5)[a]
LY122772	all H	oxime	isopropyl sulfonamide	sulfonamide	13.6 ± 2.58
LY153186[b]	all H	carboxamide	isopropyl sulfonamide	sulfonamide	4.4 ± 0.14
LY341904	4-F	carboxamide	isopropyl sulfonamide	sulfonamide	4.7 ± 1.16
LY341908	3-F	carboxamide	isopropyl sulfonamide	sulfonamide	7.3 ± 1.19
LY354400[b]	2,5 di-F	carboxamide	isopropyl sulfonamide	sulfonamide	3.5 ± 0.28
LY355081[b]	2,5 di-F	*N*-methyl carboxamide	isopropyl sulfonamide	sulfonamide	5.5 ± 0.13
LY357132	all H	*N*-methyl carboxamide	isopropyl sulfonamide	sulfonamide	6.3 ± 0.43
LY357822	2,3 di-F	*N*-methyl carboxamide	isopropyl sulfonamide	sulfonamide	6.1 ± 0.37
LY362546	3-F	*N*-methyl carboxamide	*n*-propyl sulfonamide	sulfonamide	6.4 ± 0.26
LY362683	2,3 di-F	carboxamide	isopropyl sulfonamide	sulfonamide	3.7 ± 0.05
LY366092[b]	naphthyl	*N*-methyl carboxamide	isopropyl sulfonamide	sulfonamide	9.3 ± 0.16
LY366347	2,3,4 tri-F	*N*-methyl carboxamide	isopropyl sulfonamide	sulfonamide	7.3 ± 0.32
LY366349[b]	3-F, 4-MeO	carboxamide	isopropyl sulfonamide	sulfonamide	3.3 ± 0.12
LY366572	3,5 di-F	carboxamide	isopropyl sulfonamide	sulfonamide	5.8 ± 0.67
LY366659[b]	2,3,5,6-F	carboxamide	isopropyl sulfonamide	sulfonamide	6.3 ± 0.40
LY366799	2-F	carboxamide	isopropyl sulfonamide	sulfonamide	2.9 ± 0.00
LY366856	2,3,4 tri-F	carboxamide	isopropyl sulfonamide	sulfonamide	4.7 ± 0.09
LY368227[b]	3,4 di-F	carboxamide	isopropyl sulfonamide	sulfonamide	4.4 ± 0.19
LY368228[b]	2,4 di-F	carboxamide	isopropyl sulfonamide	sulfonamide	4.3 ± 0.07
LY368766	2,3 di-F	*N*-ethyl carboxamide	isopropyl sulfonamide	sulfonamide	8.6 ± 0.09
LY354030	3-F	*N*-methyl carboxamide	N,N di-Me sulfonylurea	sulfonylurea	8.8 ± 2.12
LY353462	3-F	carboxamide	N,N di-Me sulfonylurea	sulfonylurea	6.7 ± 1.20
LY362898	3-F	*N*-methyl carboxamide	morpholino sulfonylurea	sulfonylurea	4.1 ± 0.08
LY359353	3-F	*N*-methyl carboxamide	pyrrolidino sulfonylurea	sulfonylurea	8.0 ± 0.44
LY368288	3-F	carboxamide	morpholino sulfonylurea	sulfonylurea	1.8 ± 0.14
LY366094[b]	2,3 di-F	*N*-methyl carboxamide	isopropyl	*N*-alkyl	0.6 ± 0.01
LY366853	2,3 di-F	carboxamide	isopropyl	*N*-alkyl	0.5 ± 0.05
LY368177	2-F	*N*-methyl carboxamide	cyclopentyl	*N*-alkyl	0.7 ± 0.01

[a] Results were obtained in the presence of a pH gradient (apical pH 6.0, basolateral pH 7.4) and are expressed as mean ± one standard deviation; $n = 2-3$ filters. [b] Molecules used for the test set.

Modeling with CoMFA. The 3D grid based CoMFA[28] technique in SYBYL (vers. 6.5, Tripos Inc., St. Louis, MO) was used to build a model relating the steric and electrostatic fields of the molecules to the permeability data. Two different alignments of the compounds were explored to determine which yields the best model. The CoMFA method is very sensitive to alignment of the training set molecules and it can affect whether a significant model is built. In this case, since we are not dealing with a binding site, a reasonable, low energy conformation and atom-to-atom alignment may be sufficient to get a predictive model. However specific conformations and subsequent alignment cannot be ruled out a priori. The first alignment uses the "rms fit" function of SYBYL to overlay the common benzimidazole and amino nitrogens of all the molecules. Lilly compound LY122772 (enviroxime) was used as a template. All structures were generated with CONCORD 3.2.1 (Tripos Inc., St. Louis, MO) and their Mulliken electronic charges calculated using MOPAC-93, (J. J. P. Stewart, Fujitsu Corp.) with the following keywords: AM1 MULLIK MMOK NOINTER.

The second alignment considered was the "permeability pharmacophore" superimposition proposed by the Catalyst analysis. A structure definition file (SDF) of all the superimposed molecules was exported from Catalyst. This file was read directly into SYBYL preserving the Catalyst geometries. The molecules were checked for correct atom and bond typing, then charges were obtained using a 1-SCF calculation with MOPAC-93. Compounds with poor permeability were not aligned with the rest of the set. For a CoMFA analysis these would have to be aligned to some extent, so they were adjusted manually by rigidly superimposing the benzimidazole rings to the rest of the set.

Standard CoMFA. A default 2.0 Å grid spacing with a positively charged sp^3 carbon probe atom was used to calculate the electrostatic and steric interaction energies. No minimum sigma was used to filter the data. Standard CoMFA scaling was performed on all data matrices. Leave-one-out cross-validated PLS analysis was performed to determine the number of components using the PRESS (predictive residual sum of squares) and q^2 statistics as guidelines. The optimum number of components was determined by a minimum PRESS value and maximum q^2. The q^2 statistic is defined as

$$q^2 = 1 - \text{PRESS/SSD}$$

where PRESS = sum of the squared deviations of the actual and predicted dependent variables and SSD = sum of the squared deviations of each dependent variable from the mean of all dependent variables.

PLS analyses with no cross-validation were subsequently performed for the determined number of components to obtain a final model. This work was performed for both alignments in exactly the same manner.

Examining the Predictive Ability of the CoMFA Models. The nine test molecules (*vide supra)* were used to

assess the predictive power of the CoMFA models. Each molecule was built and aligned manually to the model. The permeability for each molecule was predicted with the final model and compared to the observed values.

Modeling Using GFA MS-WHIM. Each molecule was coded as a SMILES[28] and atomic three-dimensional coordinates with Gasteiger-Huckel charges were generated by CONCORD 3.2.1 (Tripos Inc., St. Louis, MO) and an internally written charge lookup program. MS-WHIM descriptors were computed using the program EL3DMD (Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN) as previously described for CYPs.[19,20,22,23]

MS-WHIM descriptors are a set of statistical parameters that contain information about the structure of the molecules in terms of size, shape, symmetry, and distribution of molecular surface point coordinates after "weighted" centering and principal component analysis (PCA). The following weights were applied: (i) unweighted, (ii) positive, and (iii) negative electrostatic potential, (iv) hydrogen bonding acceptor and (v) donor capacity, and (vi) hydrophobicity, which yield a total of 102 descriptors (17 for each weight[8,29]). These descriptors were then imported into Cerius[2] version 4.0 (Molecular Simulations, San Diego, CA), and the GFA was used to construct an equation to relate these descriptors to the log transformed Caco-2 permeability values.

Validation of GFA-MS-WHIM Models. The Cerius[2] software enabled leave-one-out cross validation of the model to be carried out. External validation of the model was performed by predicting the log permeability for the nine molecule test set using the MS-WHIM descriptor values as inputs for the equation.

Modeling with VolSurf. The 19 structures represented as SMILES strings were converted into 3D using Corina (Oxford Molecular Group, Oxford, U.K.). These structures were then imported into VolSurf ([30,31] VolSurf v.2.0, Multivariate Infometrics Analysis, Perugia, Italy 1999). Molecular fields were computed around each structure, using GRID ([32,33] GRID v.18, Molecular Discovery Ltd. Oxford, U.K.) and the water, dry, carbonyl and amide NH probes, which describe solvent, hydrophobic, and hydrogen-bonding effects, respectively. The GRID resolution was 0.5 Å. The permeability data were expressed as described in the other methods. Models were constructed using PLS, allowing up to 10 components. The goodness of the structure−activity correlation can be represented by r^2. However, the correlation increases with the number of components, so the q^2 statistic was used instead to choose the best model. The q^2 values were obtained by both the leave-one-out and leave-two-out procedures as recommended by the authors of VolSurf. In addition, where q^2 values are similar, the more parsimonious model should be chosen.[30]

Validation of the Volsurf Model. The fields around the nine test set structures were computed as described for the training set and their activity estimated using the best PLS model.

RESULTS

In vitro Caco-2 permeability data was generated for 28 molecules (Table 1). A lack of asymmetry in the apical to basolateral versus basolateral to apical permeability of compounds from each of the three structural classes indicated that transport was via a passive mechanism.[25]

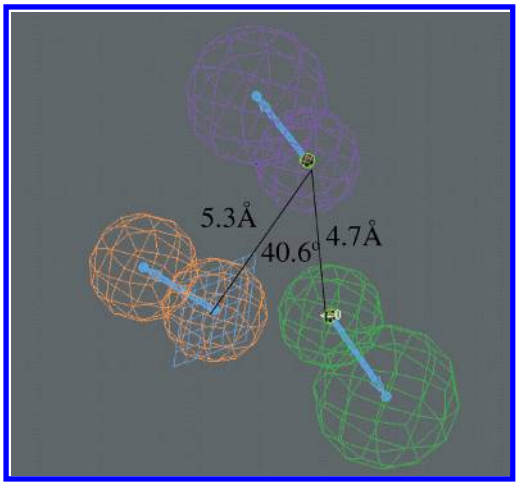


Figure 1. Catalyst pharmacophore for training set showing hydrogen bond acceptor with a vector in the direction of a putative hydrogen bond donor (green), hydrogen bond donor with a vector in the direction of a putative hydrogen bond acceptor (purple), and ring aromatic feature (orange). Distances shown are in Å, and the angle between features is in degrees.

Preliminary studies (results not shown) were carried out using multiple linear regression approaches to relate permeability to selected empirical and calculated molecular descriptors for these compounds (e.g., log D, log P, number of hydrogen bond acceptors, number of hydrogen bond donors, hydrogen bond total, and molecular volume). The best correlations with permeability ($r^2 > 0.9$) were observed with log D and hydrogen bond acceptors suggesting their importance in this congeneric series. Following this we utilized numerous 3D-QSPR approaches to predict Caco-2 permeability.

The first 3D-QSPR model examined was a Catalyst pharmacophore obtained from the permeability data for 19 training set molecules. This consisted of a hydrogen bond acceptor, a hydrogen bond donor, and a ring aromatic feature (Figure 1). This pharmacophore demonstrated an excellent correlation of observed versus estimated permeability ($r^2 = 0.83$). The pharmacophore also demonstrated an excellent overlay for 18 of 19 molecules, while the least permeable molecule (LY366853) was incorrectly aligned to the pharmacophore (Figure 2, blue structure showing the 2-aminobenzimidazole at the top). The best pharmacophore hypothesis was selected and used for prediction of the test set which gave an r^2 for observed and predicted permeability of 0.94. To attempt to show the significance of this model we tried permuting the training set data 10 times and then generating new hypotheses. This resulted in a lower mean r^2 value of 0.41 and models with fewer features than obtained in the model described previously. Ideally this value should have approached zero.

The CoMFA modeling produced a significant model using the initial manual alignment. Figure 3 shows the superimposition rule and the dimensions of the CoMFA grid ($17 \times 23 \times 18$ Å) used for the CoMFA analysis. Three components were found to be optimum for the training set of 19 molecules yielding a PRESS = 0.34 and a $q^2 = 0.30$. The subsequent model built with three components had an $r^2 = 0.96$ and standard error of estimate = 0.08.

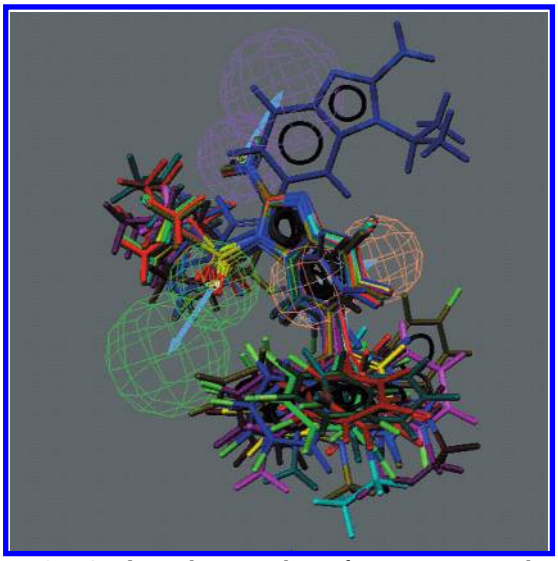


Figure 2. Catalyst pharmacophore for training set showing hydrogen bond acceptor with a vector in the direction of a putative hydrogen bond donor (green), hydrogen bond donor with a vector in the direction of a putative hydrogen bond acceptor (purple), and ring aromatic feature (orange). All training set molecules are overlapped showing the least permeable molecule LY366853 poorly fits to the pharmacophore.

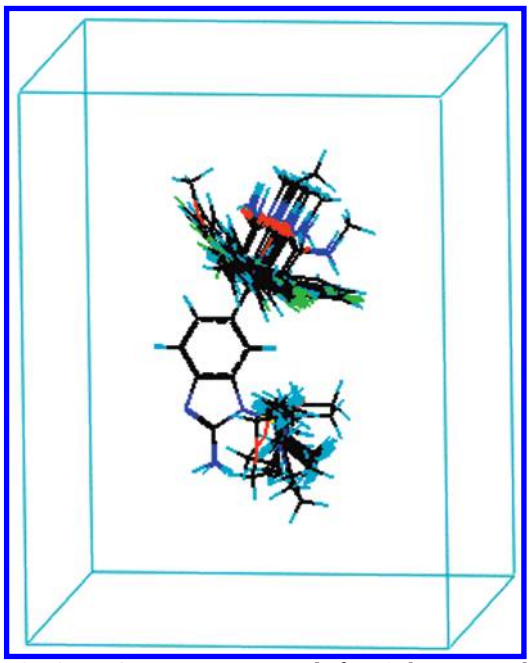


Figure 3. CoMFA superimposition rule for 19 rhinovirus inhibitor compounds.

The test set of molecules was predicted by this CoMFA model with an $r^2 = 0.84$ for observed versus predicted permeability. The standard deviation X coefficients were contoured around compound LY122772 to show the most important fields that explain permeability. The steric field map (Figure 4) shows green areas where steric bulk contributes favorably to the permeability and yellow areas where steric bulk decreases permeability. The electrostatic map (Figure 5) indicates where increasing relative negative charge (red) increases permeability and increasing relative positive charge (blue) increases permeability.

An attempt to use the Catalyst generated pharmacophore alignment instead of the manual overlap for all of the molecules used in the CoMFA analysis did not yield a

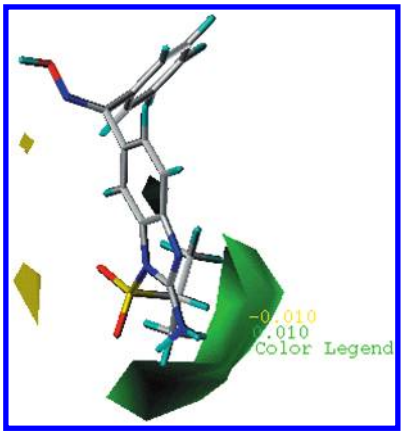


Figure 4. CoMFA steric map contoured around LY122772. Yellow is contoured at −0.01, and green areas are contoured at 0.01.

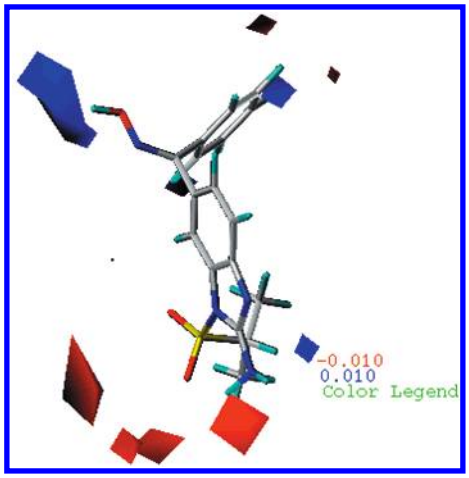


Figure 5. CoMFA electrostatic map contoured around LY122772. Red is contoured at −0.01, and blue is contoured at 0.01.

statistically significant model when utilized for CoMFA model building (based on the PRESS and q^2 statistics).

The Cerius[2] GFA-MS-WHIM model resulted in the following equation using the most significant descriptors:

$$\log \text{permeability} = 5.4407 - (1.24929 * \text{MWE3W1}) - (5.14386 * \text{MWH3W2}) + (0.533754 * \text{MWH3W3}) + (40.5602 * \text{MWP2W5}).$$

MWE3W1 is an unweighted shape descriptor, MWH3W2 is positive electrostatic potential, MWH3W3 is negative electrostatic potential, and MWP2W5 is a hydrophobicity term.[8,29] This model resulted in an r^2 value of observed versus predicted permeability of 0.98 and cross validated r^2 (q^2) value of 0.95. This GFA model was quite reproducible in that repetition resulted in an essentially similar equation and statistics. The equation above was used to predict the permeability of the test set. The r^2 of observed and predicted permeability was 0.46 for the test set.

The VolSurf PLS model chosen had three components, with an r^2 of 0.76 and q^2 (leave one out) of 0.54 and a q^2 (leave two out) of 0.46. The descriptors with the largest coefficients correspond to the volumes of interaction with the amide NH probe and to a lesser extent the water probe, followed by the hydrogen-bond interaction energies for the amide NH and the carbonyl probe (Figure 6). The DRY probe descriptors make virtually no contribution. The amide NH probe map around compound LY122772 (Figure 7)

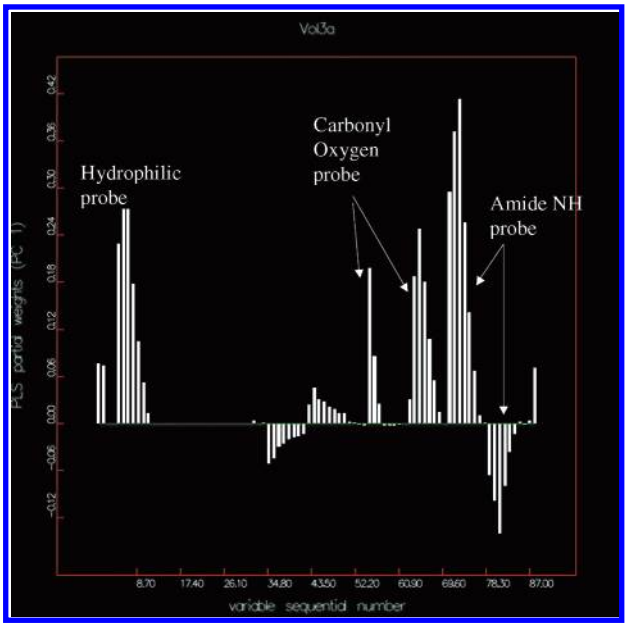


Figure 6. PLS weightings profile 1, with the most influential probe coefficients identified for VolSurf.

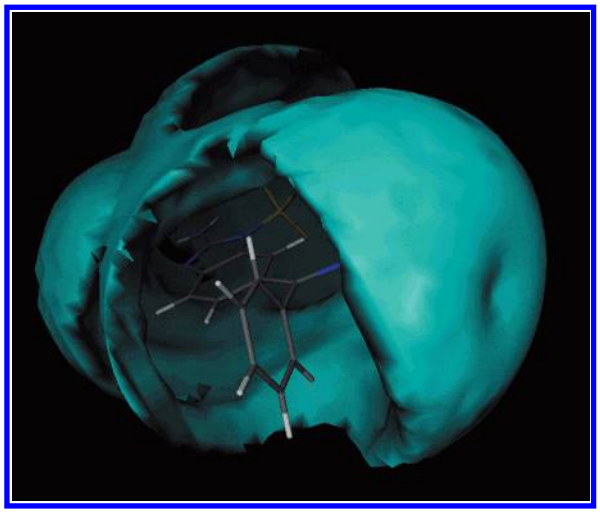


Figure 7. Lilly compound LY122772 with the amide NH field contoured at −1.0 kcal/mol with VolSurf.

contoured at −1.0 kcal/mol gives a visual picture of the magnitude of the key amide NH volume descriptor. This model resulted in an $r^2 = 0.83$ for the test set observed versus predicted permeability values. The model average for all four computational methods with this test set resulted in an r^2 of 0.84 for the observed versus predicted permeability values.

DISCUSSION AND CONCLUSION

The discovery and optimization of new drugs is becoming increasingly reliant upon the combination of experimental and computational approaches. In silico strategies have been widely applied to the prediction of absorption as commonly assessed as permeation across Caco-2 cells. These approaches offer some promise in drug design for optimizing this ADME property. Previously we have investigated the application of in vitro oral bioavailability assays in the design of orally active inhibitors of rhinovirus replication as a method to optimize potency and bioavailabity in parallel.[25] During this process, permeability coefficients across Caco-2 cells were generated, which raised the question as to whether these

Table 2. Test Set Predictions Obtained with Each Computational Approach and a Model Average

compd no.	log obsd permeability	log predicted permeability for each model				
		catalyst	Co-MFA	GFA-MS-WHIM	VolSurf	model av[a]
LY153186	0.64	0.68	0.77	0.72	0.72	0.72
LY354400	0.54	0.77	0.75	0.55	0.62	0.67
LY355081	0.74	0.72	0.77	0.76	0.76	0.765
LY366092	0.97	0.92	0.77	0.12	1.08	0.72
LY366094	−0.22	−0.20	−0.16	−0.60	−0.15	−0.28
LY366349	0.52	0.60	0.76	0.85	0.85	0.76
LY366659	0.80	0.89	0.78	0.82	0.69	0.79
LY368227	0.64	0.74	0.77	0.82	0.61	0.73
LY368228	0.63	0.70	0.75	0.52	0.61	0.64

[a] Model average = average of predicted permeability values for each molecule.

experimental data could be used to build computational structure activity (in this case activity relates to a molecule's permeability) relationships. Previously, regression analysis between absorption and polar molecular surface descriptors[9−11,13−16,34] or other 2D or 3D parameters[6,17,18,34−36] have been used to predict the absorption of well-behaved molecules with some success. Even though the 3D arrangement of physiochemical properties may not be crucial for model generation, none of these previously published methods had provided a 3D visualization of physiochemically relevant parameters necessary for the absorption of a structurally related or diverse molecule data set until recently.[7] These authors predicted hydrogen bonding capacity using GRID calculations for a previously published data set[6] and were able to produce maps of hydrogen bond interactions accounting for directional and electronic properties.[7] As Caco-2 cells represent a useful and widely used measure of human drug absorption in vivo,[37,38] it would appear that data generated in vitro would represent useful information that could be computationally modeled and further readily evaluated experimentally. This would go some way toward answering the question of whether absorption can be predicted from structure alone.[39]

In this study we have used a number of commercially available 3D computational approaches (Catalyst, CoMFA, Cerius[2] GFA, and VolSurf) as well as other (MS-WHIM) methods to model the Caco-2 permeability data for passively absorbed inhibitors of rhinovirus replication to aid future optimization of these molecules. We took the approach of randomly splitting the 28 molecules into a training set of 19 molecules and test set of nine molecules. Thus far it would appear that Catalyst produces very predictive models in terms of the excellent relationships of observed versus predicted permeability for the test set in terms of the r^2 and q^2 values obtained (Table 2). Both CoMFA and VolSurf were also very successful but somewhat less so than Catalyst at providing predictions as determined with the test set. The least successful approach used in this present study with rhinovirus inhibitors was Cerius[2] GFA with MS-WHIM descriptors, which generated poorer test set r^2 and q^2 values (Table 2).

This study presents the first occasion where computational approaches that yield either graphical (Catalyst, CoMFA, and VolSurf) or nongraphical (GFA-MS-WHIM) models have been compared for predicting permeability of a congeneric series of molecules. The advantage of using Catalyst is that manual alignment of molecules is not required, as multiple

conformers are automatically generated and aligned for each molecule. With Catalyst one can clearly visualize structural fragments or functional groups likely to result in increased permeability. In both Catalyst and CoMFA models the aminobenzimidazole cores are aligned similarly and very clearly suggest pharmacophore features or areas where structural modification may increase the permeability parameter. The steric CoMFA map indicates that increased steric bulk at the R3 position on the benzimidazole is favorable to permeability; however, the substituents that varied at this location were also rather polar in nature. This was reflected in the electrostatic map where important negative regions were contoured around the sulfonyl groups. It is important to note that the Catalyst model placed a hydrogen bond acceptor feature in this same location. Thus varying the electronic environment around the sulfonyl groups appears to have an effect on the strength of the hydrogen bonds formed by the sulfonyl oxygens. This may in turn affect permeability. The agreement of both models around this feature lends credence to the ability of the two different modeling techniques to highlight the same important areas of molecules that influence Caco-2 permeability. This is of interest as different conformers are likely to have been used by Catalyst and CoMFA. Such 3D information obtained from these two models can in turn be fed back into the design of new compounds to suggest that other large polar groups with hydrogen bond acceptors at the R3 position might be favorable to increase permeability.

Using Catalyst to generate a permeability pharmacophore alignment as an input for CoMFA did not improve upon manual alignment for this latter software. This is possibly because Catalyst compares general molecular features in 3D, (e.g., H-bond acceptor or donor), whereas CoMFA is more sensitive to specific atomic changes in partial charge and/or steric potential, which is more variable. CoMFA fields are very sensitive to alignment of the molecules, and the Catalyst alignment may not be optimal for capturing the changes in the fields from molecule to molecule on a 2 Å grid in this particular case.

The VolSurf method was designed to take into account the 3D structures of molecules without requiring a detailed alignment that is unlike the fitting that occurs with both Catalyst and CoMFA. The interpretation of the VolSurf model can be done both visually and in terms of the descriptors used. The key terms seem to be the extent of the polar surface described by the amide NH probe and the strength of the hydrogen-bonding interactions. Graphical inspection of the plots indicates that the key groups in the molecules are the sulfonyl oxygens and the carboxamide/oxime function (Figure 6). As seen in the Catalyst model, varying the electronic environment around either of these groups will change the volume of interaction (a less polar function will shrink the volume) and the strength of interaction with the amide NH probe. In this respect, the VolSurf model is consistent with the Catalyst and CoMFA models. Interestingly the descriptors derived from the dry probe were not very important in the model, implying that addition of further lipophilic groups would not change the permeability. This may reflect the need for more structural diversity in the present training set.

Although these graphical techniques were very predictive for a structurally similar data set, it would appear that the nongraphical/descriptor based approach of GFA-MS-WHIM was less predictive even though it appeared internally consistent. Such an observation is surprising in light of previous studies where MS-WHIM[8,29] was used to model a more structurally diverse Caco-2 derived permeability data set. These observations would suggest the value of assessing multiple approaches for future use in permeability prediction as it appears that some methods do better than others for any particular data set. It would be most valuable if we could attain a single prediction with optimal characteristics of both highly predictive and visually interpretable models of absorption. Catalyst, CoMFA, and VolSurf achieve this goal in this case for a congeneric series of antivirals. Another alternative is to take a model average or consensus prediction (Table 2) for each test set molecule. This is a technique used previously with other types of data.[40−44] The strength of using a model average is that different methods have varying predictabilities for the same molecules. If one method does poorly on one molecule, another method may do much better, improving the overall predictability of the property of interest. In our case this model average approach provided good overall predictions and appears to have value as a way to summarize data from multiple approaches, representing a way to unify models for present and future use.

The current study represents an approach that builds on the knowledge that polar surface area descriptors[45,46] and hydrogen bond donors and acceptors are important predictors of absorption.[6,12] Specifically, the Catalyst model picked up the requirement for hydrogen bond acceptor and donor features (Figure 2), and the Cerius[2] GFA was able to build models that utilized hydrophobicity and other descriptors using MS-WHIM. This study also provides perhaps the most extensive comparison of commercially available computational methods for absorption prediction to date. In addition, the present Catalyst and CoMFA comparison suggests that the automated alignment feature within Catalyst may or may not be useful at deriving a likely molecular orientation for field analysis, as determined from the test-set predictions. Volsurf and CoMFA produced comparable results suggesting the alignment independent and dependent methods respectively were equivalent in this case.

Our computational models have immense value in design of rhinovirus inhibitors with improved permeability which can be verified in vitro using Caco-2 cell permeability as an indicator of human absorption. The next logical step is to produce computational models of Caco-2 data that will successfully predict larger data sets containing more diverse structures than a single series of molecules which is passively absorbed, as used in this case. However, the current study has shown the utility of commercially available 3D approaches that may suggest that conformation can have implications for permeability. These future models will also need to be able to predict active uptake or efflux of molecules. Catalyst or VolSurf will likely be favorably applied to this task as they have automated alignment or are alignment independent, respectively. These two techniques are both economical and fast. At some point, models generated for permeability, metabolic lability, drug−drug interactions, and clearance will all need to be integrated as a single ADME screening filter. Such a unified ADME model may be feasible in the future and might necessitate using multiple modeling tools such as those described in the present

study. Such a multivariate analysis incorporating both bioactivity and ADME data is required to make decisions early in the lead optimization paradigm, in terms of the best molecules to test in vitro and in vivo and hopefully increase the efficiency of ADME simultaneously.

ACKNOWLEDGMENT

The authors gratefully acknowledge the helpful advice of our many colleagues in The Departments of Drug Disposition, Computational Chemistry, and Molecular Structure Research and in particular the constructive comments of Dr. Jon A. Erickson and Dr. Steven A. Wrighton. The authors wish to extend their sincere thanks to Dr Gabrielle Cruciani of the University of Perugia, Italy for initially supplying a copy of VolSurf and Dr. Peter Goodford of the University of Oxford, England for supplying an evaluation copy of GRID. The authors also gratefully acknowledge Dr. Gianpaolo Bravi (GlaxoWellcome, Stevenage, U.K.) for implementing MS-WHIM while at Eli Lilly and Co.

REFERENCES AND NOTES

(1) Grass, G. M. Simulation models to predict oral drug absorption from in vitro data. *Adv Drug. Del. Rev.* **1997**, *23*, 199−219.
(2) Barthe, L.; Woodley, J.; Houin, G. Gastrointestinal absorption of drugs: methods and studies. *Fundam. Clin. Pharmacol.* **1999**, *13*, 154−168.
(3) Irvine, J. D.; Takahashi, L.; Lockhart, K.; Cheong, J.; Tolan, J. W.; Selick, H. E.; Grove, J. R. MDCK (Madin-Darby Canine Kidney) cells: a tool for membrane permeability screening. *J. Pharm. Sci.* **1999**, *88*, 28−33.
(4) Kramer, S. D. Absorption prediction from physicochemical parameters. *Pharm. Sci. Technol. Today.* **1999**, *2*, 373−380.
(5) Egan, W. J.; Merz, K. M. J.; Baldwin, J. J. Prediction drug absorption using multivariate statistics. *J. Med. Chem.* **2000**, *43*, 3867−3877.
(6) van de Waterbeemd, H.; Camenenisch, G.; Folkers, G.; Raevsky, O. A. Estimation of Caco-2 cell permeability using calculated molecular descriptors. *Quant. Struct. Act. Relat.* **1996**, *15*, 480−490.
(7) Segarra, V.; Lopez, M.; Ryder, H.; Palacios, J. M. Prediction of drug permeability based on GRID calculations. *Quant. Struct. Act. Relat.* **1999**, *18*, 474−481.
(8) Bravi, G.; Wikel, J. H. Application of MS-WHIM Descriptors: 3. Prediction of molecular properties. *Quant. Struct. Act. Relat.* **2000**, *19*, 39−49.
(9) Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. prediction of intestinal absorption. *J. Pharm. Sci.* **1999**, *88*, 807−814.
(10) Kelder, J.; Grootenhuis, P. D. J.; Bayada, D. M.; Delbressine, L. P. C.; Ploemen, J.-P. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* **1999**, *16*, 1514−1519.
(11) Krarup, L. H.; Christensen, I. T.; Hovgaard, L.; Frokjaer, S. Predicting drug absorption from molecular surface properties based on molecular dynamics simulations. *Pharm. Res.* **1998**, *15*, 972−978.
(12) 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.
(13) Norinder, U.; Osterberg, T.; Artursson, P. Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. *Eur. J. Pharm. Sci.* **1999**, *8*, 49−56.
(14) Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Artursson, P. Correlation of drug absorption with molecular surface properties. *J. Pharm. Sci.* **1996**, *85*, 32−39.
(15) Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **1997**, *14*, 568−571.
(16) Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Beigi, F.; Lundahl, P.; Artursson, P. Evaluation of dynamic polar molecular surface area as a predictor of drug absorption: Comparison with other computational and experimental predictors. *J. Med. Chem.* **1998**, *41*, 5382−5392.
(17) Sugawara, M.; Takekuma, Y.; Yamada, H.; Kobayashi, M.; Iseki, K.; Miyazaki, K. A general approach for the prediction of the intestinal absorption of drugs: regresssion analysis using the physiochemical properties and drug-membrane electrostatic interaction. *J. Pharm. Sci.* **1998**, *87*, 960−966.
(18) Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of human intestinal absorption of drug compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 726−735.
(19) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. *Pharmacogenetics* **1999**, *9*, 477−489.
(20) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. Three-dimensional-quantitative structure activity relationship (3D-QSAR) analyses of inhibitors for CYP3A4. *J. Pharmacol. Exp. Thera.* **1999**, *290*, 429−438.
(21) Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. Three-dimensional quantitative structure activty relationship (3D-QSAR) analysis of CYP3A4 substrates. *J. Pharmacol. Exp. Thera.* **1999**, *291*, 424−433.
(22) Ekins, S.; Bravi, G.; Ring, B. J.; Gillespie, T. A.; Gillespie, J. S.; VandenBranden, M.; Wrighton, S. A.; Wikel, J. H. Three-dimensional-quantitative structure activity relationship (3D-QSAR) analyses of substrates for CYP2B6. *J. Pharmacol. Exp. Thera.* **1999**, *288*, 21−29.
(23) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug. Met. Dispos.* **2000**, *28*, 994−1002.
(24) Ekins, S.; Ring, B. J.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. *Predicting Drug-drug interactions in silico using pharmacophores: a paradigm for the next millenium*; Guner, O. F., Ed.; University International Line: San Diego, 2000; pp 269−299.
(25) Stratford, R. E. J.; Clay, M. P.; Heinz, B. A.; Kuhfeld, M. T.; Osborne, S. J.; Philips, D. L.; Sweetana, S. A.; Tebbe, M. J.; Vasudevan, V.; Zornes, L. L.; Lindstrom, T. D. Application of oral bioavailability surrogates in the design of orally active inhibitors of rhinovirus replication. *J. Pharm. Sci.* **1999**, *88*, 747−753.
(26) Kuhfeld, M. T.; Stratford, R. E. In vitro measurement of drug transport using a new diffusion chamber compatible with millicell culture supports: performance with Caco-2 monolayers. *Int. J. Pharm.* **1996**, *133*, 47−58.
(27) Weininger, D. SMILES 1. Introduction and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
(28) Cramer, R. D., III.; Patterson, D. E.; Bunce, J. D. 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(29) Bravi, G.; Wikel, J. H. Application of MS-WHIM Descriptors: 1. Introduction of new molecular surface properties and 2. Prediction of binding affinity data. *Quant. Struct. Act. Relat.* **2000**, *19*, 29−38.
(30) Cruciani, G.; Pastor, M.; Guba, W. Volsurf: a new tool for the pharmokinetic optimization of lead compounds. *Eur. J. Pharm. Chem.* **2000**, *11 suppl 2*, S29−S39.
(31) Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. Predicting blood-brain barrier permeation from the three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204−2216.
(32) Goodford, P. J. Computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(33) Bobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M. New hydrogen-bond potential for use in determining energetically favorable binding sites of molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083−1094.
(34) Stenberg, P.; Luthman, K.; Ellens, H.; Lee, C. P.; Smith, P. L.; Lago, A.; Elliot, J. D.; Artursson, P. Prediction of the intestinal absorption of endothelin receptor antagonists using three theoretical methods of increasing complexity. *Pharm. Res.* **1999**, *16*, 1520−1526.
(35) Breitkreutz, J. Prediction of intestinal drug absorption properties by three-dimensional solubility parameters. *Pharm. Res.* **1998**, *15*, 1370−1375.
(36) Ghuloum, A. M.; Sage, C. R.; Jain, A. N. Molecular haskeys: a novel methods for molecular characterisation and its application for predicting important pharmaceutical properties of molecules. *J. Med. Chem.* **1999**, *42*, 1739−1748.
(37) Wils, P.; Warnery, A.; Phung-Ba, V.; Scherman, D. Differentiated intestinal epithelial cell lines as in vitro models for predicting the intestinal absorption of drugs. *Cell. Biol. Toxicol.* **1994**, *10*, 393−397.
(38) Walter, E.; Janich, S.; Roessler, B. J.; Hilfinger, J. M.; Amidon, G. L. HT29-MTX/Caco-2 cocultures as an in vitro model for the intestinal epithelium: in vitro-in vivo correlation with permeability data from rats and humans. *J. Pharm. Sci.* **1996**, *85*, 1070−1076.
(39) Curatolo, W. Physical chemical properties of oral drug candidates in the discovery and exploratory development settings. *P.S.I.T.* **1998**, *1*, 387−393.
(40) Rogers, D.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure−activity relationships and quan-

titative structure property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.
(41) So, S.-S.; Karplus, M. Genetic neural networks for quantitative structure−activity relationships: improvements and application of benzodiazepine affinity for benzodiazepine/GABAA receptors. *J. Med. Chem.* **1996**, *39*, 5246−5256.
(42) So, S.-S.; Karplus, M. A comparative study of ligand−receptor complex binding affinity prediction methods based on glycogen phosphorylase inhibitors. *J. Comput. Aided. Mol. Des.* **1999**, *13*, 243−258.
(43) Hou, T. J.; Wang, J. M.; Liao, N.; Xu, X. J. Applications of genetic algorithms on the structure−activity relationship analysis of some cinnamamides. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 775−781.
(44) Ajay On better generalization by combining two or more models: a quantitative structure−activity relationship example using neural networks. *Chem. Intel. Lab Sys.* **1994**, *24*, 19−30.
(45) Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, 43, 3714−3717.
(46) Stenberg, P. Computational Models for the prediction of intestinal membrane permeability. Ph.D. Thesis, Uppsala University, Uppsala, Sweden, 2001.

CI010330I