Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
International Journal of Pharmaceutics 309 (2006) 109–114
INTERNATIONAL JOURNAL OF PHARMACEUTICS
www.elsevier.com/locate/ijpharm

A simplified model to predict *P*-glycoprotein interacting drugs from 3D molecular interaction field

Xiao-Mei Zhuang, Jun-Hai Xiao, Jin-Tong Li *, Zhen-Qing Zhang, Jin-Xiu Ruan

Beijing Institute of Pharmacology and Toxicology, Beijing, China 100850

Received 18 June 2005; received in revised form 4 November 2005; accepted 13 November 2005
Available online 20 December 2005

Abstract

A new two components partial least squares discriminant analysis (PLS) model for the prediction of *P*-glycoprotein-associated ATPase activity of drugs by using VolSurf compute theoretical molecular descriptors derived from 3D molecular interaction field was reported in the present study. By using 27 diverse drugs from literature, two models were constructed (R^2 = 0.9003, 0.8150; Q^2 = 0.7165, 0.7630) in this paper, which were similar to models that utilized MolSurf parametrization (R^2 = 0.7760, 0.7180; Q^2 = 0.7420, 0.6950) by using 22 drugs reported in the same literature. The results investigated VolSurf software was superior to MolSurf in its simplicity. Properties associated with the volume, polarizability, and hydrogen bond could have important impact on the *P*-glycoprotein-associated ATPase activity.

Keywords: *P*-glycoprotein; VolSurf; QSAR; Molecular interaction field

1. Introduction

P-glycoprotein (*P*-gp), a 170 Kda glycoprotein, is a member of a highly conserved superfamily of ATP-binding cassette (ABC) transport protein, and shares extensive similarity with numerous bacterial, yeast, insect and other mammalian ABC transport proteins (Higgins, 1992). MDR1 gene encodes *P*-gp in humans (Ambudkar et al., 1999; Germann, 1996). High levels of *P*-gp expression have been observed in the endothelial cells of the blood–brain barrier, certain cells of the adrenal gland, liver, pancreas, kidney, colon, jejunum, digestive tract and cells of the lumen surface of the gravid uterus secretory epithelium and in many cancer cells as well. *P*-gp can extrude a range of structurally diverse, toxic xenobiotic compounds from cells (Schinkel, 1997), therefore the broad distribution of *P*-gp not only causes a major problem in the failure of cancer chemotherapy, but also involves ADME properties of drugs, especially in the intestine absorption and tissue distribution in the body. Because of this strategical location, modulation of *P*-gp activity and/or expression at these cellular sites may affect the pharmacokinetic parameters of drugs that are *P*-gp substrates, leading to modified bioavailability and possible adverse drug reactions (Romiti et al., 2004). Knowledge of the factors that determine substrate specificity is crucial for successful drug targeting and rational design of new drugs. It is accepted, however, that interaction of compounds with *P*-gp is a complex process and currently the details of its mechanism of action are still the subjects of controversy (Stouch and Gudmundsson, 2002). Evaluation of such factors is critically important to understand the whole scheme of interaction between *P*-gp and drugs. Many attempts have been made to find early assessment of *P*-gp substrates or inhibitors. The proved several screening assays could help identify the subject of substrates and inhibitors. For example, the cytotoxicity IC_{50} endpoint is one of the evaluated methods. The activity of the reversal agent is generally expressed as a fold reversion that also is usually called the MDR ratio (Dhainaut et al., 1996). Another popular approach is based on the increased accumulation of photo-affinity analogs of anti-tumor agents (Beck and Qian, 1992) or fluorescent compounds (Kessel et al., 1991), which interact with other *P*-gp modulators inside the cell. Transport studies using Caco-2 cell line that expresses *P*-gp have also been used to screen *P*-gp substrates and inhibitors (Burton et al., 1993). Besides these experimental techniques, computational approaches have also been developed to predict *P*-gp interacting drugs because the experimental determination is laborious, expensive, and time-consuming, and requires a sufficient quantity of pure compound. Therefore, there is a considerable demand

* Corresponding author at: Taiping Road 27, Haidian District, Beijing, China. Tel.: +86 10 6687 4610; fax: +86 10 6821 1656.
E-mail address: Leejohnton@yahoo.com (J.-T. Li).

0378-5173/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2005.11.009

for fast and reliable computational methods to assess *P*-gp interactions at an early stage of drug discovery. Unfortunately, so far a truly general conclusive QSAR model has not been found for either substrate or inhibitory activities. Österberg and Norinder had reported a theoretical calculation to predict *P*-gp interaction using MolSurf parametrization and PLS statistics (Österberg and Norinder, 2000). The investigated results explained that MolSurf descriptors could predict *P*-gp associated ATPase activity of drugs on certain extends. However, this method is more complex and the computational requirements are prohibitive for medium-sized data sets.

Recently, a novel method named VolSurf has been developed by Cruciani's group (Cruciani et al., 2000a). VolSurf is an automatic procedure to convert 3D molecular field into physicochemical properties relevant to molecular descriptors and has proven its efficacy and simplicity of usage. The basic concept of VolSurf is to compress the information presented in 3D grid maps into a few quantitative numerical descriptors which are easy to understand and interpret. The principal advantage of these descriptors is that they do not require structural superimposition for a 3D-QSAR analysis, as is usually required when working with grid-field variables (Kubinyi, 1997), and their numerical values are related to conformations submitted to computation. To our best knowledge, no attempt has been made to use descriptors derived from VolSurf to build *P*-gp associated ATPase activities predictive model. In the present paper, we reported the use of VolSurf and PLS statistics for modeling the structure–activity relationship between the ATPase activities and structurally diverse *P*-gp substrates by using not only the 22 drugs introduced in Österberg's paper but also additional five drugs.

2. Computational procedures

2.1. Overview of building predictive model approach

The overall procedures contained the following five major steps:

(1) Collection compounds with *P*-gp associated ATPase activities from literature.
(2) The three-dimensional structure of the compounds was constructed using the Concord program, and the resulting conformations were refined by energy minimization with Tripos force field as implemented in sybyl 6.91 (SYBYL Version 6.91).
(3) The compounds were submitted to multivariate characterization based on their interaction energy with chemical probes. Then we used the GRID program (Goodford, 1985; Bobbyer et al., 1989) to calculate the 3D molecular interaction field.
(4) Molecular descriptors were calculated using the VolSurf program.
(5) Chemo metric tool PLS was used to correlate the data and build a *P*-gp interaction model.

It should be noted that the VolSurf program could perform steps 3–5 automatically.

2.2. Dataset

Log $1/k_1$ data for 27 compounds were compiled from the literature (Litman et al., 1997). Drugs chosen were sets of calmodulin antagonists, steroids, hydrophobic cations, chemotherapeutic substrates of *P*-gp and some other drugs with lower affinity for *P*-gp. We followed the same approach as Österberg's, taking $\log 1/k_1$ as the response variable (Österberg and Norinder, 2000), where $1/k_1$ is the reciprocal of Michaelis constant, k_m, which is directly proportional to affinity, and $\log 1/k_1$, is directly proportional to the free energy of interaction between ligand and receptor. (The chemical structures have been omitted.)

2.3. Calculation of VolSurf descriptor variables

The molecular descriptors were derived from the VolSurf/GRID program. The interaction fields with a water probe (OH), a hydrophobic probe (DRY) and a carbonyl probe (O) were calculated all around the target molecules. O represents a hydrogen bond acceptor probe that offers complementary information in comparison with the water probe, which informs on all the possible hydrogen bond centers without regard to their donor or acceptor characteristics. As a result, VolSurf generated the 72 descriptors were omitted because a detailed explanation of the VolSurf methodology is given everywhere (Cruciani et al., 2000b). Then we used exclude individual variables command to select the active descriptors. The result showed that 55 descriptors were active in the model.

2.4. Statistical analysis

The relationship between the experimental reported $\log 1/k_1$ values and the computed VolSurf descriptors was determined using partial least squares (PLS), which allows quantitative relationship to be established among multiple variables (Wold et al., 1993). The number of significant latent variables and the quality of the models were determined by using the leave-one-out cross-validation procedure (LOO-CV). In such a procedure, each compound is removed once from the dataset, and the remaining compounds are used to develop a new model, with which the compounds left are then predicted.

2.5. Training set selection

We used the same method as the maximin approach and selected the same 14 molecules reported by Österberg as the training set (Österberg and Norinder, 2000, Marengo and Todeschini, 1992).

3. Results and discussion

3.1. P-gp associated ATPase activity data selection

ATPase activity is pre-requisite for *P*-gp to transport substrate and both nucleotide binding domains (NBD's) of *P*-gp must hydrolyze nucleotides for the transport to occur (Stouch and Gudmundsson, 2002). The stimulation/inhibition of *P*-gp

ATPase activity in membranes obtained from cells that express *P*-gp could be detected. Assays are on the hypothesis that drug-induced ATP hydrolysis reflects transport by the transporter. It has been proposed that substrates could be characterized based on their kinetic parameters (Litman et al., 1997; Scarborough, 1995). Therefore, determination of *P*-gp ATPase activity modulations by various drugs is a means of obtaining qualitative and quantitative data describing the interaction between these drugs and *P*-gp (Garrigues et al., 2002).

3.2. The model 1 with the 26 molecules as the training set

Tamoxifen was excluded from the statistical analysis by using exclude objects command. Exclude manual was utilized to remove outliers, split the dataset in a training and a prediction set. Although it was difficult to explain why the calculated pK_1 value of tamoxifen was so smaller than the experimental value, tamoxifen has a very large, spread-out and flexible conjugated π-system that may engage in fairly effective interactions with *P*-gp, which was not correctly described by the currently computed VolSurf descriptors. The relationship between the 3D structure and the *P*-gp substrates of the dataset consisting of 26 compounds and 55 active descriptors was studied in the preliminary investigation of model 1. Two significant principal components were found by LOO-CV technique. The two components explained about 81% of the total variance of the matrix.

3.3. The model 2 with the 14 molecules as the training set

Based on the preliminary investigation, we carried out model 2 with 14 molecules from the 27 molecules and the same 55 descriptors. Two significant latent variables emerged from the PLS model and LOO-CV procedure. These components accounted for about 90% of the total variance of the matrix.

The results of the PLS statistical properties containing the two models are summarized in Table 1. The results suggested that the VolSurf model is similar to MolSurf model. And the plots of experimental versus calculated ATPase-associated activity of the two models are shown in Figs. 1 and 2.

3.4. External prediction of the model 2

In order to validate the predictability of the second model, we selected the remaining 13 molecules as the test set. The results for the 13 molecules and 27 data set compounds with experimental, calculated and predicted *P*-gp ATPase activity by VolSurf procedure are shown in Table 2.

Table 1
PLS statistical properties of the two VolSurf and MolSurf models

Model	Training set	R^2	Q^2	SDEC	SDEP
1[a]	26 Compounds	0.8150	0.7630	0.2682	0.4284
1[b]	21 Compounds	0.7180	0.6950	0.4520	0.4700
2[a]	14 Compounds	0.9003	0.7165	0.3452	0.4693
2[b]	14 Compounds	0.7760	0.7420	0.3900	0.4190

Note: SDEC, standard error of recalculation in fitting; SDEP, standard error of predicting in the prediction phase.
[a] Models were built by VolSurf.
[b] Models were built by MolSurf.

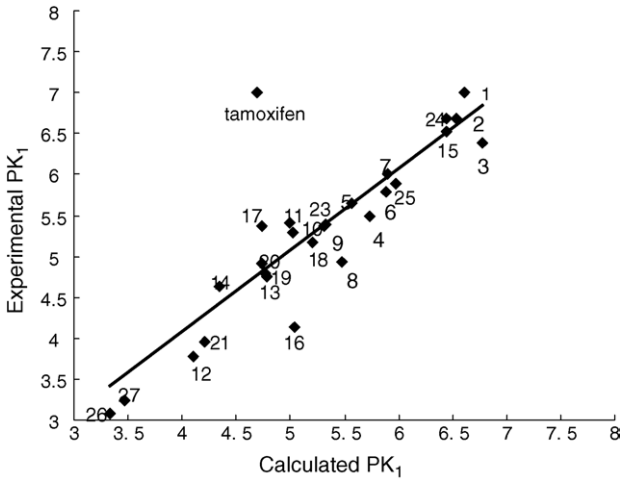

Fig. 1. Relationship between experimental and calculated ATPase activity (PLS model 1).

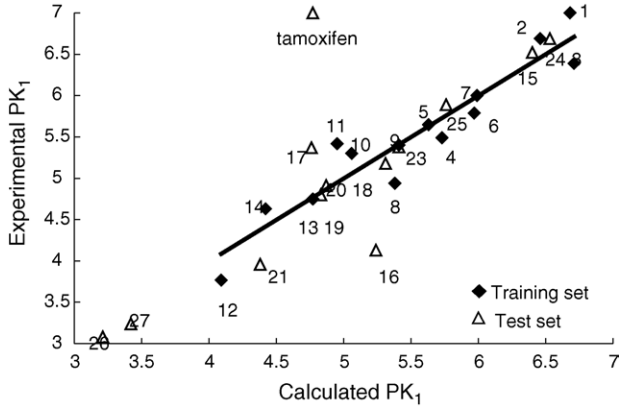

Fig. 2. Relationship between experimental and calculated ATPase activity (PLS model 2).

3.5. The interpretation of VolSurf models

The coefficient plot of model 1 shows the contribution of the 55 active descriptors (Fig. 3). The vertical bars represent the

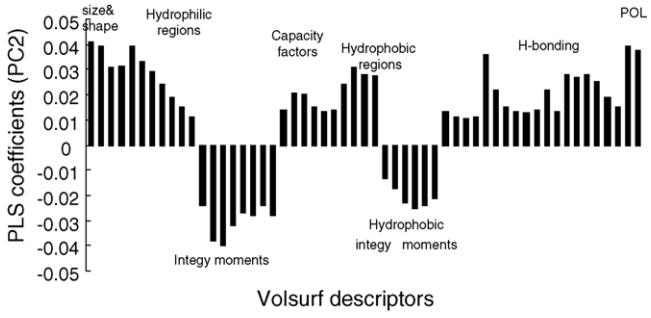

Fig. 3. PLS coefficient plot of the global model 1 for the correlation of 55 active VolSurf descriptors with *P*-gp interacting drugs.

Table 2
Experimental, calculated and predicted *P*-gp ATPase activity

No.	Compound	Experimental[a] activity	Model 1, calculated activity		Model 2			
			VolSurf[c]	MolSurf[b]	Calculated activity		Predicted activity	
					VolSurf[c]	MolSurf[b]	VolSurf[d]	MolSurf[e]
1	Reserpine	7.00	6.61	6.88	6.68	6.96		
2	Epirubicin	6.69	6.54	6.60	6.46	6.74		
3	Dipyridamole	6.39	6.78	6.16	6.71	6.18		
4	Amiodarone	5.49	5.74	5.79	5.73	5.88		
5	Terfenadine	5.65	5.57	5.81	5.63	5.81		
6	D-Verapamil	5.79	5.88	5.62	5.97	5.69		
7	Pimozide	6.00	5.9	5.57	5.99	5.53		
8	Fluphenazine	4.94	5.48	5.30	5.38	5.27		
9	Spironolactone	5.40	5.33	5.00	5.41	5.31		
10	Quinidine	5.30	5.02	4.93	5.06	4.93		
11	Mefloquine	5.42	4.99	4.95	4.95	4.88		
12	*S*-Propranolol	3.77	4.10	4.79	4.09	4.81		
13	Progesterone	4.75	4.78	4.73	4.77	4.80		
14	Promethazine	4.63	4.35	4.38	4.42	4.41		
15	Daunomycin	6.52	6.45	6.52			6.40	6.67
16	Diltiazem	4.13	5.04	5.39			5.24	5.48
17	*S*-Propafenone	5.37	4.73	5.18			4.76	5.25
18	Trifluoperazine	5.18	5.21	4.76			5.31	4.78
19	Trifluopromazine	4.80	4.77	4.56			4.83	4.58
20	Chlorpromazine	4.91	4.74	4.51			4.87	4.53
21	Amitriptyline	3.96	4.21	4.46			4.38	4.49
22	Tamoxifen	7.00	4.69	5.01			4.77	5.04
23	Fucidin	5.38	5.31	–			5.41	–
24	Vincristine	6.69	6.45	–			6.53	–
25	Vinblastine	5.89	5.97	–			5.76	–
26	Colchicine	3.08	3.33	–			3.21	–
27	Methotrexate	3.24	3.47	–			3.42	–

[a] Experimental pK_1 values.
[b] Calculated/fitted pK_1 values for the training set by MolSurf procedure.
[c] Calculated/fitted pK_1 values for the training set by VolSurf procedure.
[d] Predicted pK_1 values for the test set by VolSurf procedure.
[e] Predicted pK_1 values for the test set by MolSurf procedure.

contribution of each single descriptor with a short bar displaying a minor contribution and a long bar a major one. Therefore, we could draw the following conclusions:

(1) The size, shape and volume descriptors have a significant impact on promoting ATPase activity. This is consistent with the observations by Litman and Österberg, both of whom had found an analogical relationship between ATPase activity and the Van der Waals surface area (Litman et al., 1997; Österberg and Norinder, 2000). This phenomenon has been explained as an indication that the binding between modulators and *P*-gp takes place across a wide interaction surface on the protein, other than at a peculiar binding site. It has been proved that *P*-gp is a multisite protein (Orlowski and Garrigos, 1999), where there were clearly at least two binding sites. For example, verapamil and progesterone are known to be bound to different sites, and these binding sites were unequal. However, the performance at one site is contingent on the other being unoccupied, and transport is also sometimes mitigated when the other site is occupied (Wang et al., 2000). This explanation was verified by theoretical calculation of *P*-gp-interacting drugs using MolSurf parametrization, which has been further testified by our VolSurf model.

(2) Descriptors of POL (polarisability) and hydrogen bond show another marked influence on promoting ATPase activity. Hydrogen bond describes the H-bonding capacity of a molecule target, as obtained with a polar probe. The water probe presents an optimal ability to donate and accept hydrogen bonds to and from the target. POL is an estimate of the average molecule polarisability, calculated according to the additive method of Miller, and the correlation between this method and the polarisability calculated with VolSurf is very good. This result is also consistent with the investigation of Österberg et al., who found that factors related to hydrogen bond, such as strengths and intermolecular hydrogen bond, have a major impact on the ATPase activity. From these information, a primary conclusion can be drawn that hydrogen bond capacity is a detrimental factor which may affect the absorption of drugs because it is difficult for drugs with many hydrogen bonds to cross bio-membranes passively, and to make the situation even worse, the likelihood of them being subjected to a *P*-gp-related efflux mechanism is also enhanced.

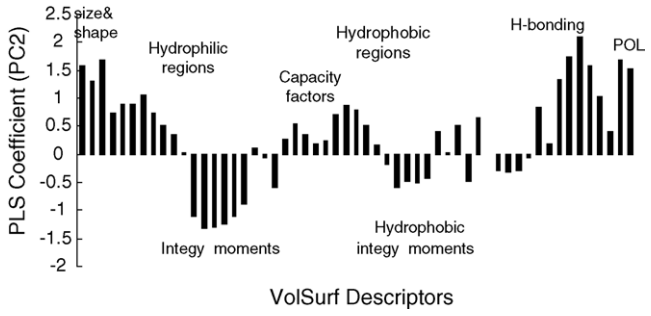

Fig. 4. VolSurf scaled descriptors of reserpine.

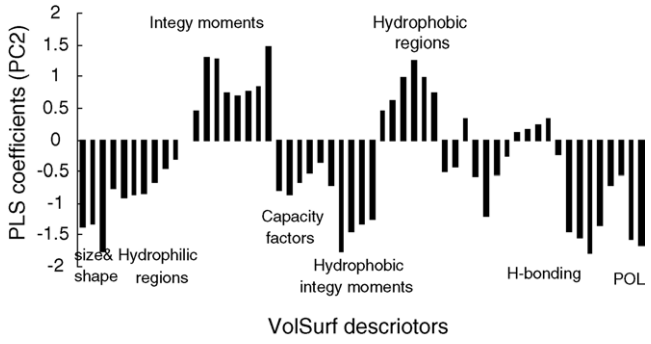

Fig. 5. VolSurf scaled descriptors of methotrexate.

(3) The coefficient plot also shows that not only such descriptors of polarity as hydrophilic regions (W1–W7) and capacity factors (CW1–CW5), but also the descriptors of hydrophobic interactions (D4–D8) are directly correlated to *P*-gp ATPase activity. Although the later one's positive contribution appears somewhat smaller than the former one. This is in line with Seelig's conclusion that the importance of lipophilicity and amphiphilicity might be due as much to membrane partitioning as to protein binding (Seelig, 1998).

3.6. The application to two compounds in the dataset

At last, we discussed in detail about two compounds to exemplify the interpretative value of the VolSurf model. These compounds represent different properties of *P*-gp interaction. Reserpine has high ATPase activity while methotrexate has low ATPase activity. Interestingly, the scaled value of the two compounds also showed the different trends, as shown in Figs. 4 and 5. Descriptors such as size, shape and volume of reserpine are larger than those of methotrexate; polarisability (POL) and hydrogen bond descriptors of methotrexate are smaller than those of reserpine. These results are in agreement with our earlier analysis of the PLS coefficient plot (Fig. 3).

4. Conclusion

The models constructed in this paper for the prediction of *P*-gp interacting drugs involves physiological point of view and consists with the insights from many other QSAR approaches and molecule models to the full extent. We root out that VolSurf descriptors depicting size, shape and volume properties have significant impact on promoting *P*-gp ATPase activity, as well as the hydrogen bond. At the same time, the descriptors describing hydrophilic and hydrophobic properties also facilitate the *P*-gp ATPase activity, which is in agreement with the accepted view that many *P*-gp substrates are amphiphilic molecules. It should be noted, however, that the interaction of compounds with *P*-gp is a complicated process, and so far there is not a very robust assay to probe the interaction. Another drawback of the present models is that it has been constructed/validated on small data sets, which may lead to bias. Our next goal to be achieved is to involve more compounds for improving the model.

Acknowledgments

We are grateful to professor Song Li and his laboratory for their help and support in VolSurf software. The project was supported by the National High Technology Research and Development Program of China (No. 2003AA2Z347B), and by the National Natural Science Foundation of China (No. 30200347).

References

Ambudkar, S.V., Dey, S., Hrycyna, C.A., Ramachandra, M., Pastan, I., Gottesman, M.M., 1999. Biochemical, cellular and pharmacological aspects of the multidrug transporter. Annu. Rev. Pharmacol. Toxicol. 39, 361–398.

Beck, W.T., Qian, X.D., 1992. Photoaffinity substrates for *P*-glycoprotein. Biochem. Pharmacol. 43, 89–93.

Bobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M., 1989. New hydrogen-bond potential for use in determining energically favourable binding sites of molecules of known structure. J. Med. Chem. 32, 1083–1094.

Burton, P.S., Conradi, R.A., Hilgers, A.R., Ho, N.F., 1993. Evidence for a polarized efflux system for peptides in the apical membrane of Caco-2 cells. Biochem. Biophys. Res. Commun. 190, 760–766.

Cruciani, G., Crivori, P., Carrupt, P.A., Testa, B., 2000a. Molecule fields in quantitative structure-permeation relationships: the VolSurf approach. J. Mol. Struct. Theochem. 503, 17–30.

Cruciani, G., Pastor, M., Guba, W., 2000b. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. Eur. J. Pharm. Sci. 11, S29–S39.

Dhainaut, A., Regnier, G., Tizot, A., Pierre, A., Leonce, S., Guilbaud, N., Krans-Berthier, L., Atassi, G., 1996. New purines and purine analogs as modulators of multidrug resistance. J. Med. Chem. 39, 4099–4108.

Garrigues, A., Nugier, J., Orlowski, S., Ezan, E., 2002. A high-throughput screening microplate test for the interaction of drugs with *P*-glycoprotein. Anal. Biochem. 305, 106–114.

Germann, U.A., 1996. *P*-glycoprotein: a mediator of multidrug resistance in tumor cells. Eur. J. Cancer 32A, 927–944.

Goodford, P.J., 1985. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. J. Med. Chem. 28, 849–857.

Higgins, C.F., 1992. ABC transporters-from microorganisms to man. Annu. Rev. Cell Biol. 8, 67–113.

Kessel, D., Beck, W.T., Kukuruga, D., Schulz, V., 1991. Characterization of multidrug resistance by fluorescent dyes. Cancer Res. 51, 4665–4670.

Kubinyi, H., 1997. QSAR and 3D QSAR in drug design. Part 1. methodology. Drug Disc. Today 2, 457–467.

Litman, T., Zeuthen, T., Skovsgaard, T., Stein, W.D., 1997. Structure-activity relationships of *P*-glycoprotein interacting drugs: kinetic characterization of their effects on ATPase activity. Biochim. Biophys. Acta 1361, 159–168.

Marengo, E., Todeschini, R., 1992. A new algorithm for optimal, distance-based experimental design. Chemomet. Intel. Lab. Syst. 16, 37–44.

Orlowski, S., Garrigos, M., 1999. Multiple recognition of various amphiphilic molecules by the multidrug resistance *P*-glycoprotein: molecular mecha-

nisms and pharmacological consequences coming from functional interactions between various drugs. Anticancer Res. 19, 3109–3123.

Österberg, T., Norinder, U., 2000. Theoretical calculation and prediction of *P*-glycoprotein-interacting drugs using MolSurf parametrization PLS statistics. Eur. J. Pharm. Sci. 10, 295–303.

Romiti, N., Tramonti, G., Donati, A., Chieli, E., 2004. Effects of grapefruit juice on the multidrug transporter *P*-glycoprotein in the human proximal tubular cell line HK-2. Life Sci. 76, 293–302.

Scarborough, G.A., 1995. Drug-stimulated ATPase activity of the human *P*-glycoprotein. J. Bioenerg. Biomembr. 27, 37–41.

Schinkel, A.H., 1997. The physiological function of drug-transporting *P*-glycoproteins. Semin. Cancer Biol. 8, 161–170.

Seelig, A., 1998. A general pattern for substrate recognition by *P*-glycoprotein. Eur. J. Biochem. 251, 252–261.

Stouch, T.R., Gudmundsson, O., 2002. Progress in understanding the structure-activity relationships of *P*-glycoprotein. Adv. Drug Deliv. Rev. 54, 315–328.

SYBYL, Molecular Modeling Software, Version 6.91, Tripos Inc., December 2002.

Wang, E.J., Casciano, C.N., Clement, R.P., Johnson, W.W., 2000. Two transport binding sites of *P*-glycoprotein are unequal yet contingent: initial rate kinetic analysis by ATP hydrolysis demonstrates intersite dependence. Biochim. Biophys. Acta 1481, 63–74.

Wold, S., Johansson, E., Cocchi, M., 1993. PLS-partial least-squares projections to latent structures. In: Kubinyi, H. (Ed.), 3D QSAR in Drug Design: Theory, Methods and Applications. ESCOM Science Publisher BV, Lei den, pp. 523–550.