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