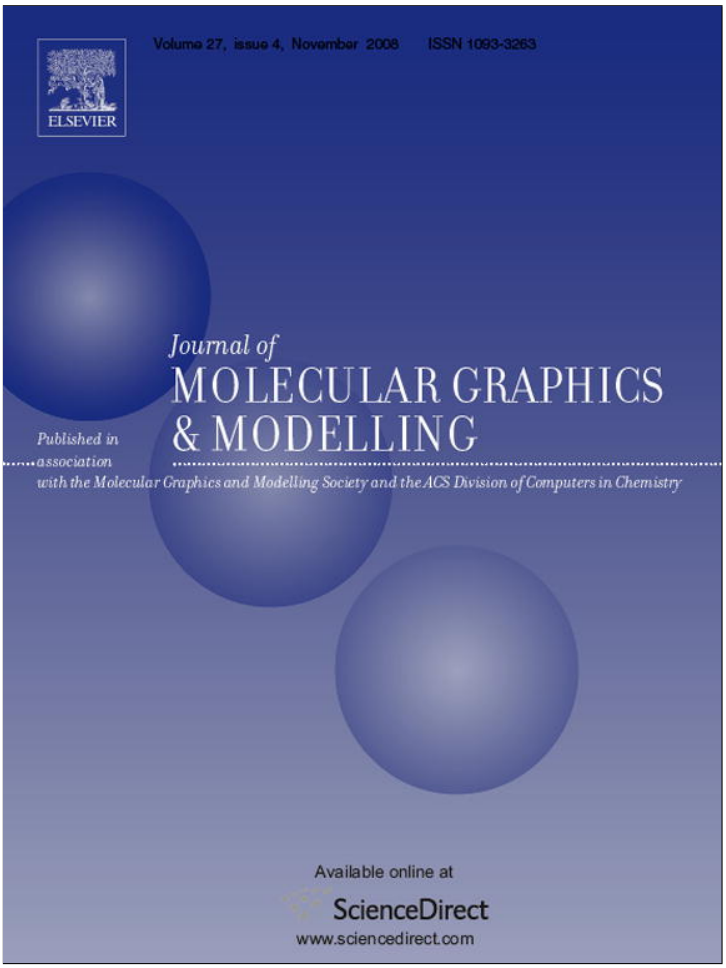


Journal of Molecular Graphics and Modelling 27 (2008) 439–451

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Development of predictive *in silico* model for cyclosporine- and aureobasidin-based P-glycoprotein inhibitors employing receptor surface analysis

Hiba M. Zalloum, Mutasem O. Taha *

Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman 11942, Jordan

ARTICLE INFO

Article history:
Received 26 May 2008
Received in revised form 28 July 2008
Accepted 29 July 2008
Available online 13 August 2008

Keywords:
Cyclosporines
Aureobasidins
P-glycoproteins
Receptor surface analysis
3D-QSAR

ABSTRACT

P-glycoprotein (Pgp) is implicated in multiple drug resistance (MDR) exhibited by several types of cancer against a multitude of anticancer chemotherapeutic agents. This problem prompted several research groups to search for effective P-gp inhibitors. Cyclosporine A (CsA), aureobasidin A (AbA) and related analogues were reported to possess potent inhibitory actions against Pgp. In this work we employed receptor surface analysis (RSA) to construct two satisfactory receptor surface models (RSMs) for cyclosporine- and aureobasidin-based Pgp inhibitors. These pseudoreceptors were combined to achieve satisfactory three-dimensional quantitative structure activity relationship (3D-QSAR) for 68 different cyclosporine and aureobasidin derivatives. Upon validation against an external set of 16 randomly selected Pgp inhibitors, the optimal 3D-QSAR was found to be self-consistent and predictive ($r^2_{LOO} = 0.673$, $r^2_{PRESS} = 0.600$). The resulting 3D-QSAR was employed to probe the structural factors that control the inhibitory activities of cyclosporine and aureobasidin analogues against Pgp.

1. Introduction

Chemotherapy, which is considered the main approach for the treatment of malignancies such as leukemias and lymphomas, has significantly improved during the last decade largely due to the introduction of effective drug combinations and treatment schemes. Unfortunately, many cancers develop resistance to multiple chemotherapeutic agents, which seems to be mainly mediated by P-glycoprotein (Pgp) [1–3].

Pgp is a large membrane-spanning protein belonging to the ATP-binding cassette superfamily (ABC family) of transporters. It is encoded by the *MDR1* gene and acts as energy-dependent efflux pump [4–6]. Pgp is composed of 1280 amino acids arranged into two homologous and symmetrical halves, each comprising a trasmembrane region of six α-helices and a cytosolic domain [7,8]. Pgp is expressed in many normal tissues such as the intestines, liver, kidneys, lungs, endothelia of the brain, testis, and placenta, which is consistent with its role as a natural detoxification system [9,10]. It has pronounced effects on the pharmacokinetics and pharmacodynamics of many drugs [5,10–14].

The fact that Pgp is over-expressed in many cancer cells [11] prompted many research groups to search for effective inhibitors for this glycoprotein. These efforts culminated in the discovery of a large number of structurally and functionally unrelated compounds capable of inhibiting Pgp partially or completely [15,16]. However, since the three-dimensional (3D) structure of Pgp is not well characterized [17–19]. most drug discovery pursuits focused on ligand-based approaches, such as quantitative structure–activity relationship (QSAR) models [17,20,21] and pharmacophore analyses [22], to identify Pgp substrates [17,20–23] and inhibitors [23,24].

The cyclic peptides cyclosporine A (CsA) and aureobasidin A (AbA) were reported to interfere with ABC transporters in fungal and mammalian cells, particularly Pgp [25–28]. Structure–activity relationship analyses suggested common inhibitory binding site for these inhibitors and that their activities can be improved by modifying some of their amino acid residues [29–32]. However, CsAs and AbAs are large and chemically complex molecules, which seems to have discouraged attempts to model their inhibitory structure–activity relationships. Accordingly, absence of published three-dimensional QSAR analyses specific to this group of Pgp inhibitors [23,33–38] combined with the lack of useful Pgp 3D structural information prompted us to model a diverse set of cyclosporins and aureobasidins in order to probe their affinity interactions with their corresponding putative Pgp binding site. However, the sheer size and chemical complexity of these inhibitors restricted our modeling options. For example, cyclosporins and aureobasidins are too large for effective comparative

* Corresponding author. Tel.: +962 6 5355000x23305; fax: +962 6 5339649.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

1093-3263/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.07.005

field analysis (CoMFA), i.e. the number of field points that need to be generated is prohibitively large to be managed by statistical analysis (nearly 26,000 field points). Therefore, we were obliged to employ receptor surface analysis (RSA) [39–41] to achieve predictive QSAR model(s).

RSA relies on energy and steric descriptors gathered at the molecular surfaces of the training compounds, which restrict the number of descriptors involved in statistical analysis (i.e., compared to CoMFA). RSA uses explicit surfaces from compounds of varying activities to characterize the shape of the active site, providing clues on how ligands may be modified to improve activity [39–41]. RSA is a useful tool in situations when the 3D structure of the receptor is unknown, as it helps to build a hypothetical model of the receptor site, known as receptor surface model (RSM). RSM differs from pharmacophore models in that the former tries to capture essential information about the receptor, while the latter captures information about the commonality of compounds that bind to a receptor [39–41]. An RSM embodies essential information about the hypothetical receptor site in the form of 3D surface with associated properties such as hydrophobicity, partial charge, electrostatic (**ELE**) potential, van der Waals (**VDW**) potential, and hydrogen bonding propensity. The surface of the RSM is divided into triangles, whereby the points at the triangles' heads store these properties as associated scalar values [41,42]. These descriptors can be used for 3D-QSAR studies [39–42], whereby the variable selection technique genetic partial least squares (G/PLS), based on genetic function approximation (GFA) and partial least squares (PLS), is usually employed to derive the best possible 3D-QSAR statistical model(s) for each RSM [43–46].

RSA has gained recent interest as a valid tool in the area of computer-aided molecular design and discovery. For example, RSA was used to build receptor models for $GABA_A$ modulators [47], Angiotensin II Receptor Antagonists [48], agonists for tyramine receptor in *Plodia interpunctella* [49], and CCR5 Antagonists [50].

In this study, 84 different cyclosporine- and aureobasidin-based Pgp inhibitors were aligned together and employed to construct several RSMs. The surface descriptors of the best models were combined with a variety of two-dimensional (2D) molecular descriptors. Subsequently, this collection was searched employing G/PLS to find the best possible combination of descriptors capable of accessing self-consistent and predictive QSAR model. The successful RSMs were subsequently used to probe the structural features within cyclosporines and aureobasidines responsible for Pgp inhibition.

2. Methods

2.1. Hardware and software

Pgp inhibitors were sketched in two-dimensional format using ChemDraw Ultra (version 6.0) from Cambridge Soft Corp. (Cambridge, MA, www.cambridgesoft.com) installed on a PC.

Receptor surface analysis and genetic partial least squares statistical analyses studies were performed using CERIUS2 suite of programs (version 4.10) from Accelrys Inc. (San Diego, CA, www.accelrys.com) installed on a Silicon Graphics Octane2 desktop workstation equipped with a 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system.

2.2. Dataset

A set of 84 Pgp inhibitors belonging to cyclosporines (**1**–**59**, Table 1 and Fig. 1) [30] and aureobasidins (**60**–**84**, Fig. 1 and Table 1) [31] was gathered for modeling. The *in vitro* bioactivities of the inhibitors are expressed as the concentration of the test compound that inhibited Pgp activity by 50% (IC_{50}) in Pgp expressing multidrug resistant leukemia CEM cells. The logarithm of measured IC_{50} (μM) values were used in QSAR modeling, thus correlating the data linearly to the free energy change. None of the training compounds contained any ionizable groups, therefore they were modeled neutral.

A training subset was selected from the collected inhibitors for QSAR modeling. However, since it is essential to assess the predictive power of the resulting QSAR models on external set of inhibitors, some of the inhibitors (16 compounds, *ca.* 20%) were employed as external test subset for validating the resulting QSARs. The test molecules were selected as follows: The inhibitors were ranked according to their IC_{50} values, then every fifth compound was selected for the test subsets starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set. The following compounds were selected to validate the resulting QSAR models: **2**, **4**, **5**, **9**, **12**, **15**, **18**, **23**, **24**, **39**, **40**, **53**, **61**, **70**, **75** and **77** (Fig. 1 and Table 1).

2.3. Preparation of the dataset

The 2D structures of compounds **31** and **81** (the most potent representatives of cyclosporines and aureobasidins, respectively) were sketched using ChemDraw Ultra and saved in MDL mol-file format. Subsequently, the two structures were imported into CERIUS2, converted into corresponding standard 3D structures, assigned Gasteiger partial atomic charges [51,52] and energy-minimized to the closest local minima. The resulting structures were utilized as starting conformers for conformational sampling employing Boltzmann jump without setting a particular energy cutoff and by retaining up to 150 unique minimized conformers. The following settings were employed: torsion angle jump = 10°, number of perturbations per jump sequence = 50, temperature = 5000 K [53].

The resulting conformational ensembles (i.e., of **31** and **81**) were employed to align the two structures tightly via FIELDFIT module of CERIUS2 as follows: Initially, the two molecules were grossly pre-aligned based on their electrostatic moments. Subsequently, they were atom-aligned using consensus flexible alignment strategy via common steric and electrostatic fields. The steric field contribution in the alignment was set to 75%, while the electrostatic contribution was set to 25%. The molecular fields were mapped by oxygen atom probe. This particular steric-to-electrostatic field ratio was selected after several preliminary trials and based on the molecular strain energy upon fitting the aligned structures into the optimal receptor models (see below: Section 2.8). The alignment was subsequently fine tuned using RMS atoms match tethering close similar functional groups among the two compounds followed by another round of FIELDFIT alignment based on common steric and electrostatic fields employing the same aforementioned parameters [54].

Thereafter, the aligned lead structures (**31** and **81**) were used as templates to build the 3D structures of the remaining inhibitors of each corresponding group, i.e., **1**–**59** (CsA group) and **60**–**84** (AbA group), respectively. The structures were constructed by appropriately mutating **31** or **81** followed by energy-minimizing the modified fragments. Subsequently, a final round of FIELDFIT alignment was performed whereby the two aligned lead compounds (**31** and **81**) were considered rigid targets and the remaining 82 inhibitors were flexibly fitted against them using their common steric and electrostatic fields. The steric field contribution in the alignment was set to 75%, while the electrostatic contribution was set to 25%. The molecular fields were mapped by oxygen atom probe.

Table 1
The structures of Pgp inhibitors utilized in modeling[a]

Compound	Altered residue[b]	IC_{50} (μM)	Compound	Altered residue[a]	IC_{50} (μM)
1	–	3.40	**44**	Thr^{2} Val^{11}	19.00
2	D-$MeVal^{11}$	15.50	**45**	Bmt^{1} Thr^{2}	17.00
3	$MeGly^{3}$ Leu^{5} D-Hiv^{8}	1.60	**46**	Thr^{2}-Val^{4}	100.00
4	$MeAlle^{11}$	3.40	**47**	Thr^{2}-Leu^{10}	14.10
5	$Melle^{11}$	1.50	**48**	Bmt^{1} Val^{2}	2.10
6	$MeAla^{11}$	16.10	**49**	Val^{2}-Leu^{10}	8.80
7	Deoxy-$MeBmt^{1}$	2.20	**50**	Nva^{2}-Leu^{9}	100.00
8	$MeAoa^{1}$	4.00	**51**	Nva^{2} Leu^{10}	8.80
9	$MeLeu^{1}$	8.00	**52**	Gly^{3} D-Ser^{8}	30.00
10	Me-cyclohexyl-Ala^{1}	8.10	**53**	Thr^{2} Leu^{5} Leu^{10}	6.00
11	Me-cyclized-Bmt^{1}	13.10	**54**	Thr^{2} D-Hiv^{8} Leu^{10}	2.10
12	8′-OH-$MeBmt^{1}$	50.00	**55**	Thr^{2} Leu^{5} Ala^{10}	56.00
13	Ala^{2}	4.70	**56**	8'-OH-$MeBmt^{1}$ $MeGly^{3}$ Leu^{5} D-Hiv^{8}	15.80
14	Val^{2}	2.00	**57**	Gly^{3} Leu^{5} D-Hiv^{8}	4.00
15	Nva^{2}	1.90	**58**	$MeGly^{3}$ Ile^{5} D-Hiv^{8}	1.60
16	Thr^{2}	5.10	**59**	$MeGly^{3}$ Leu^{4} Ile^{5} D-Hiv^{8}	4.30
17	D-$MePhe^{3}$	1.50	**60**	–	2.27
18	L-Pro^{3}	5.20	**61**	D-Hiv^{1}	1.10
19	D-Pro^{3}	3.50	**62**	L-Cha^{3}	5.88
20	$Melle^{4}$	6.40	**63**	β-$HOMePhe^{4}$	2.15
21	γ-OH-$MeLeu^{4}$	30.00	**64**	D-$MeAla^{4}$	4.79
22	$MeVal^{4}$	11.10	**65**	D-$MePhe^{4}$	2.73
23	$MePhe^{4}$	2.40	**66**	L-Cha^{3} D-$MeAla^{4}$	2.91
24	$MeAla^{6}$	4.60	**67**	L-Cha^{3} D-$MeSer^{4}$	5.16
25	Abu^{7}	50.00	**68**	L-Ser^{3} D-$MeAla^{4}$	31.00
26	D-Lys^{8}	130	**69**	L-Tyr^{3} D-$MeAla^{4}$	5.77
27	D-Ser^{8}	42.00	**70**	L-$Tyr(Chm)^{3}$ D-$MeAla^{4}$	2.11
28	Deoxy-$MeBmt^{1}$ Val^{2}	2.70	**71**	D-Pro^{5}	2.73
29	$MeLeu^{1}$-Val^{2}	7.00	**72**	$NH(CH_2)_7CO$-^{3,4,5}	7.42
30	$MeLeu^{1}$-Nva^{2}	5.60	**73**	Val^{6}	1.29
31	Nva^{2}-Nva^{5}	1.30	**74**	Val^{7}	7.55
32	Nva^{2}-Leu^{5}	2.00	**75**	D-$MeVal^{7}$	2.73
33	Thr^{2}-Ile^{5}	8.40	**76**	L-Hiv^{7}	1.66
34	Val^{11}	3.40	**77**	L-Glu^{8}	8.96
35	Bmt^{1}	2.30	**78**	$MeVal^{9}$	2.68
36	Leu^{4}	100	**79**	γ-$HOMeVal^{9}$	1.09
37	Ile^{4}	15.00	**80**	Sar^{9}	0.89
38	Val^{4}	54.00	**81**	2,3-Dehydro-$MeVal^{9}$	0.18
39	Leu^{6}	23.00	**82**	2,3-Epoxy-$MeVal^{9}$	1.55
40	Leu^{9}	100.00	**83**	D-β-$HOMeVal^{9}$	0.39
41	Leu^{10}	5.70	**84**	β-$HOMePhe^{4}$ $MeVal^{9}$	1.36
42	Leu^{6} Leu^{10}	100.00			
43	Ala^{2} Leu^{10}	3.00			

[a] The corresponding scaffolds are in Fig. 1.
[b] *Amino acid abbreviations*: Abu, R-aminobutyric acid; alle, allo-isoleucine; Hiv, 2-hydroxy-isovaleric acid; Mealle, *N*-methyl allo-isoleucine; MeAoa, *N*-methylamino octanoic acid; MeBmt, *N*-methyl-4-butenyl-4-methyl-threonine; MeGly, *N*-methylglycine (sarcosine, Sar); MeLeu, *N*-methylleucine; MePhe, *N*-methylphenylalanine; MeVal, *N*-methylvaline; Melle, *N*-methyl isolleucine; MeAla, *N*-methyl alanine; *N*-Me, *N*-methylated; *N*-desMe, *N*-desmethylated; Hmp, 2-hydroxy-3-methylpentanoic acid; Phe, Phenylalanine; Cha, cyclohexylalanine; Chm, cyclohexylmethyl; Nva, norvaline.

2.4. Generation of different physicochemical descriptors

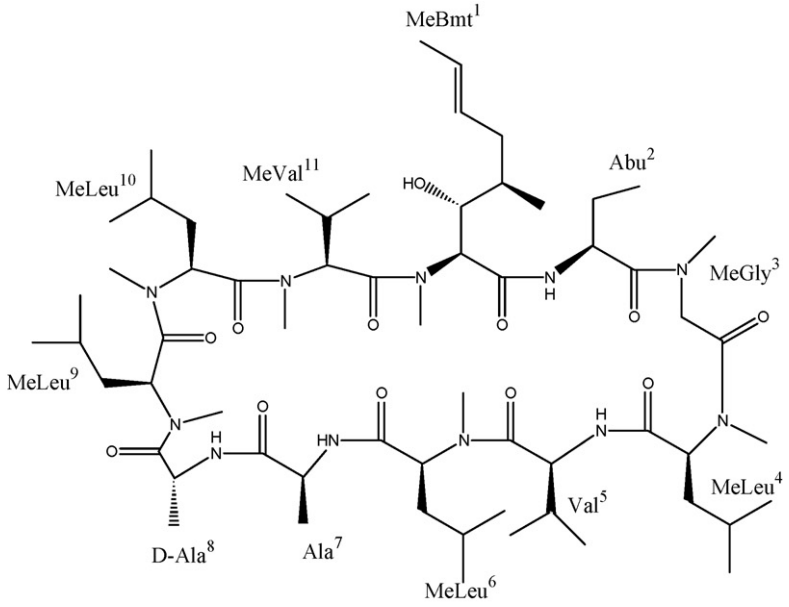
The 3D structures of the aligned inhibitors were utilized to generate 210 different molecular descriptors employing the DESCRIPTOR+ module of CERIUS2, and these include: molecular volume, dipole moment, polarizability, polar surface area, ovality, alignment similarity indices, electrotopological (E-state) indices, AlogP, semi-empirical quantum mechanical descriptors (HOMO and LUMO), connectivity indices, Kappa indices, Weiner indices, JURS descriptors, and Shadow descriptors [46].

2.5. Generation of receptor surface model

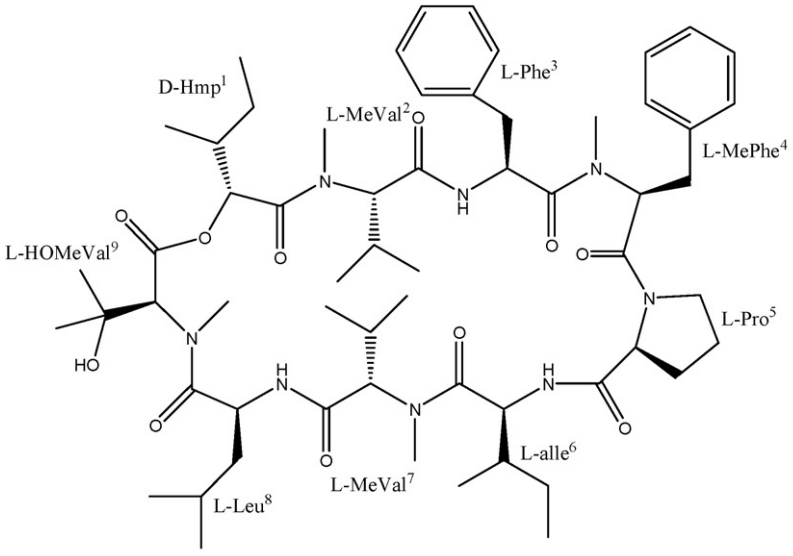
In RSA, a volumetric field characterizing the molecular shape is constructed for each of the aligned structures, termed the "shape field" [39–42,54]. The shape fields of the aligned molecules are combined to give "an isosurface" of the field, which is an explicit object with well-defined shape. In this project, the shape fields were generated using van der Waals' field function, which allows a receptor surface to be created inside or outside the van der Waals surface of a particular set of atoms (belonging to the aligned molecules) at a specified distance from the van der Waals sphere edge. The function "$\upsilon(r) = r - \mathrm{VDWr}$" is used to create shape fields at surface points around the molecules, at which the atomic coordinates of the contributing atoms of the aligned molecules are used to compute field values. In this function, r is the distance from each point to the corresponding atom and VDWr is the van der Waals radius of the atom. This function allows a surface to be constructed at any arbitrary isovalue of density on the grid. Thus at the van der Waals surface $\upsilon(r)$ is zero, inside the surface $\upsilon(r)$ is negative, and outside the volume $\upsilon(r)$ is positive. The isovalue chosen is called the surface fit. Surface points were created using the Marching Cubes algorithm [42], which generates a triangulated surface fit with an average surface density of about six points per square angstrom.

2.6. Mapping properties on the surface

In RSA each surface point has an associated charge. The assumption made is that the charge at any point on the surface is complementary to the partial atomic charge of any atom in contact with the surface. In the case of a set of aligned molecules each



Cyclosporine A (**CsA**, inhibitor **1**, as in table 1):
The basic scaffold of compounds **1-59**



Aureobasidine A (**AbA**, inhibitor **60**, as in table 1):
The basic scaffold of compounds **60-84**.

Fig. 1. The chemical scaffolds of different Pgp inhibitors shown in Table 1 (group abbreviations are shown in Table 1).

surface point is given a charge, which is equal but opposite in sign to the average partial charges of the closest atoms in the aligned molecules. In the current project the atoms were assigned Gasteiger partial charges implemented in CERIUS2 [51,52].

In addition to charge, surface points also store electrostatic potential values (**ELE**) calculated as the sum of electrostatic potentials exerted by the closest atoms of the aligned bundle against the corresponding surface point. This is computed by summing up E_{ele} for all aligned atoms (see Eq. (2) below) [39–42,54].

Furthermore, each surface point stores hydrogen bonding propensity value, which corresponds to the tendency of the point to be involved in a hydrogen bond. A value of +1.0 is assigned to volume vertex points where a hydrogen bond acceptor would be desirable in the receptor. A value of −1.0 is assigned to volume vertex points where hydrogen bond donor hydrogen might be desirable. These vertex points are found by projecting a cone away from each hydrogen bond donor or acceptor atom in the ensemble of molecules [39–42,54].

Additionally, each surface point is assigned a hydrophobic or hydrophilic attribute. This is distinguished by giving a value of 1.0 to a hydrophobic point and zero otherwise. A hydrophobic point is a point with low partial charge (absolute value less than 0.15), a low electrostatic potential (absolute value less than 0.01), and a low

hydrogen bond-donating or-accepting propensity (absolute value less than 0.1) [54].

The RSM can be constructed as open for the consideration of solvent accessible regions or can be generated as a closed surface enclosing some region of space. We employed the closed model with a solvation correction factor that penalizes polar groups placed in hydrophobic regions by 0.30 kcal/(mol Å^2) [54].

The total nonbonded interaction energies (**TOT**) are calculated between atoms of each molecule in the aligned set and the points on the receptor surface by means of van der Waals and electrostatic terms defined by Eqs. (1) and (2), respectively [39–42,54]:

$$E_{\mathrm{VDW}} = K\left[\left(\frac{\mathrm{RA}}{r}\right)^{12} - 2D\left(\frac{\mathrm{RA}}{r}\right)^{6}\right] \quad (1)$$

where RA = (VDWr) × C_h, with RA is the hybridization-corrected van der Waals radius of the atom, VDWr is the van der Waals radius, C_h is the hybridization factor, r is the distance between atom and surface point, K is the well-depth constant (set to 0.1 for all atom pair interaction), and D is empirically derived point density scaling factor. D scales VDW energy and forces so that ideal atom/surface interactions yield a typical value of 0.0125 kcal/Å^2 of surface contact [41,42,54].

The electrostatic term is based on charge–charge Coulombic function, as follows:

$$E_{\mathrm{ele}} = \frac{(322.1 Q_a Q_p) DS(r)}{r} \quad (2)$$

where r is the distance between the particular surface point and the corresponding atom, Q_a is the charge on the atom, and Q_p is the charge on the surface point, $S(r)$ is an atom-based switching function with a cutoff of 8 Å [39–42,54].

2.7. Development of receptor surface models

RSMs are normally generated from the most active structures in the data set. The rationale is that the most active molecules tend to explore the best spatial and electronic interactions within the receptor, while the least active don't and tend to have unfavorable steric and electrostatic interactions [39–42,54]. Therefore, we initiated our study by evaluating receptor models generated based on the most active two (**31** and **81**), four, five, and six inhibitors. However, we noticed that upon fitting [54] the collected inhibitors into the resulting RSMs, they suffered from excessively high strain energies. Therefore, we employed all the collected inhibitors to build corresponding RSMs using the "QSAR activity data" option. In this methodology, RSA uses activity data from training molecules to generate corresponding receptor models. The activity data are used to weight the relative importance of each template structure in generating the receptor model [54]. 16 RSMs were generated based on surface fit values ranging from −0.5 Å to +1.0 Å and a step of 0.10 Å (i.e., 16 models).

2.8. Evaluation of energies of molecules within RSM

Receptor models are normally used to evaluate how particular compounds fit into the hypothetical receptor and to calculate the energetics of binding. The ligand molecule is minimized by adjusting the geometry of its 3D structure into a "best-fit configuration" based on the constraints imposed by the receptor model [39–42,54]. Three global energy terms can be calculated by evaluating the interactions of each molecule within the RSM: the internal strain energy of the molecule as it sits in and constrained by the receptor (E_{total}), the interaction energy of the molecule with the receptor (E_{interact}), and the internal strain energy of a molecule as it sits in the receptor without being subject to receptor model constraints (E_{strain}) [41,42,54]. These energy readings were examined for each inhibitor against all 16 RSMs. It was noticed that all molecules had unfavorable (positive) E_{strain} and E_{interact} values when evaluated against RSMs of tight surface fits (<-0.20). However, upon fitting the molecules against RSMs of looser surface-fits the resulting energy values became less positive and sometimes negative, suggesting more favorable binding. On the other hand, too loose surface fits extended the distances (r) between the atoms of the aligned inhibitors and the nearest corresponding points on the receptor surface, which seemed to lead to less negative (i.e., less favorable) interactions with the RSM. Optimal energy readings were produced with RSMs of surface fits: −0.2 and −0.1 Å, and therefore, we employed these two RSMs to generate surface properties for QSAR analysis and excluded all others. The training molecules were energy-minimized within the receptor surface prior to generation of surface descriptors.

2.9. QSAR descriptors

RSM models can be used to generate surface descriptors towards QSAR analysis [41,46,54]. The descriptors are normally stored within the surface points of each RSM (see Section 2.6 above) such that each surface point stores three types of interactions: nonbonded van der Waals' interactions, electrostatic interactions, and the summation of interactions existing between a particular fitted molecule and the nearby surface point (**TOT**). However, in the current project, the numbers of surface points were prohibitively large (more than 20,000) prompting us to restrict their numbers to render the subsequent statistical modeling (Genetic partial least squares, G/PLS, see below) less computationally intensive. Therefore, it was decided to restrict the number of descriptors by selecting every 6th surface point followed by removing 50% of the remaining descriptors of lowest variance. Moreover, we added the additional global energy terms generated by fitting each inhibitor into each RSM, namely, E_{total}, E_{strain}, and E_{interact} (see Section 2.8).

The 2D descriptors, RSM surface descriptors and adjunct energy terms produced by fitting the inhibitors against the RSMs at −0.1 and −0.2 Å surface fits, were combined together for QSAR modeling. Subsequently, we employed the statistical technique of genetic partial least squares to select optimal combination of descriptors capable of explaining bioactivity variation in the training and testing sets.

2.10. QSAR modeling

We employed genetic partial least squares analysis to search for optimal regression equations capable of correlating the variations in biological activities of training compounds with variations in the corresponding RSM surface points [46]. G/PLS is derived from two methods: genetic function algorithm and partial least square regression. GFA originated from Holland's genetic algorithm [44] and Friedman's multivariate adaptive regression splines (MARS) algorithm [45,51]. GFA relies on the evolutionary operations of "crossover and mutation" to select optimal combinations of descriptors (i.e., chromosomes) capable of explaining bioactivity variation among training compounds from a large pool of possible descriptor combinations (i.e., chromosomes population). Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions [44]. G/PLS algorithm uses GFA to select appropriate basis functions to be used to model data, and PLS regression as the fitting technique to weigh the basis functions' relative contributions in the final model. Application of G/PLS allows the construction of larger QSAR equations while avoiding overfitting and eliminating most variables [46].

GFA develops an initial population of individuals, and a fitness function (lack of fit, LOF) [45,51] is used to estimate the quality of each individual. Individuals with the best fitness scores are allowed to mate and propagate their genetic material to offsprings through the crossover operation. After repeatedly performing these steps, the average fitness of the individuals in the population reaches certain maximal plateau indicating convergence at the optimal chromosome(s) [44,45,51].

Our preliminary diagnostic trials suggested the following optimal G/PLS parameters: scan linear equations of varying number of terms ranging from 6 to 15 and number of PLS latent variables (i.e., principal components) ranging from 2 to 6, at mating and mutation probabilities of 50%; population size of 500; number of generations (iterations) = 30,000 and LOF smoothness parameter = 1.0 (smoothness function controls the bias in the scoring factor between equations with different numbers of terms). Optimal QSAR equations were identified by assessing the corresponding predictive r^2 (r^2_{PRESS}) calculated from the test set of 16 inhibitors that were not included in the training set. These molecules were aligned according to the implemented alignment configuration, and their activities were predicted by corresponding G/PLS models generated from the training set (68 compounds). The optimum QSAR models were defined as those leading to the highest predictive r^2_{PRESS} and lowest sum of squared deviations between predicted and actual activity values for test molecules in the test set (PRESS). Predictive r^2_{PRESS} is defined as

$$r^2_{\text{PRESS}} = \frac{\text{SD} - \text{PRESS}}{\text{SD}} \quad (3)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules.

All QSAR models were cross-validated employing leave one-out (LOO) cross-validation and bootstrapping [46,55,56,58]. However, high-ranking QSAR models (i.e., corresponding to the best predictive r^2_{PRESS} values) were further validated by an additional validation to rule out the possibility of chance correlations: all biological activities were randomized 99 times (confidence level 99%) and were subjected to regression analysis, and the mean randomization r^2 was calculated [57].

3. Results

In the absence of 3D-structure for a particular receptor, a hypothetical model of the binding site can be built using receptor surface analysis. This approach differs from pharmacophore models in that it attempts to postulate and represent the essential features of a receptor site itself, rather than the common features or the essential 3D arrangement of the molecules that bind to it, as in the pharmacophore models [39–41]. Receptor models can easily and directly represent information, such as excluded areas and the shape of hydrophobic regions, which are difficult to represent using pharmacophore models [39–41]. Furthermore, RSMs' surface points encode for mutual interaction energies within the hypothetical receptor/ligand complex and can be searched by statistical methods (e.g., G/PLS) to identify key interactions involved in ligand-receptor binding and to achieve predictive QSAR models [39–41].

In the current project, the structural similarity of cyclosporines (**1**–**59**, Table 1 and Fig. 1) and aureobasidines (**60**–**84**, Table 1 and Fig. 1) combined with their comparable anti-Pgp potencies prompted us to hypothesized that both classes of compounds bind to certain common binding pocket within Pgp. Accordingly, it was decided to flexibly align the most potent representatives from each inhibitory category (i.e., compounds **31** and **81**) followed by

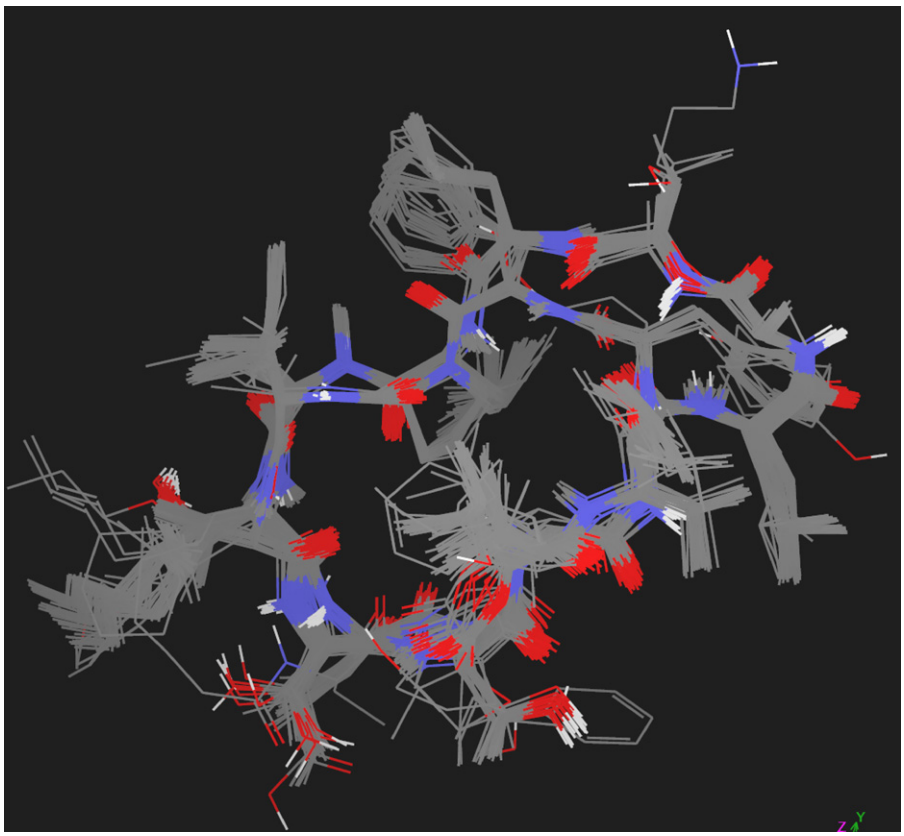


Fig. 2. Alignment of the whole set of collected Pgp inhibitors (**1**–**84**).

overlaying the remaining compounds onto these reference molecules, as shown in Fig. 2. Subsequently, the aligned ensemble was employed as a template to construct several RSMs. A number of possible combinations of adjustable parameters such as atomic partial charge, surface fits, charge complementarities, and evaluation with or without minimization of molecules within the receptor surface were evaluated. Subsequently, G/PLS was used to derive meaningful relationships between anti-Pgp bioactivities and RSA surface descriptors.

The best QSAR correlations were achieved based on linear variables from surface fit values of −0.1 and/or −0.2 Å, using Gasteiger atomic charges, electrostatic charge complementarity for surface point potential, and the nonbonding interaction energies encoded within surface points (**VDW**, **ELE**, **TOT**) in addition to the adjunct energy terms described earlier (i.e., E_{total}, E_{strain}, and $E_{interact}$, see Section 2.8 under Section 2). Still, it was necessary to incorporate additional two-dimensional (2D) physicochemical descriptors to achieve optimal self-consistent and predictive QSAR models.

The collected inhibitors were divided into two groups: a training subset of 68 compounds and a randomly selected test subset of 16 compounds. The test compounds were selected in such a way to represent a range of biological activities similar to that of the training set (see Section 2.2 under Section 2).

The quality of each QSAR model was assessed via four statistical criteria: (i) The conventional regression coefficients against the 68 training compounds (r^2_{68}), (ii) the leave-one-out regression coefficients (r^2_{LOO}), (iii) the bootstrapping regression coefficients (r^2_{BS}), and (iv) the predictive regression coefficients against the external set of 16 test compounds (r^2_{PRESS}). Table 2 shows the statistical criteria of the resulting QSAR models. However, despite that traditional crossvalidation tests such as r^2_{LOO} are very useful [57,59], they do not always pick up poor equations [55]. Accordingly, it was decided to further validate superior QSAR models by calculating their randomization correlation coefficients (r^2_{random}) at 99% confidence. This test was carried out by repeatedly permuting the activity values of the data set and evaluating the r^2 values. The r^2 values for 99 trials are shown in Table 2. This test ensures that the generated regression models were not produced by chance [55]. In fact, G/PLS is a powerful search technique that might yield apparently significant chance correlations necessitating such extensive validation [55].

Interestingly, optimal RSMs maintained high-quality statistical models regardless to the settings of the G/PLS analysis, as shown in

Table 2
Statistical criteria of the best 3D-QSAR models developed for training Pgp inhibitors employing several RSA surface fits

No. of RSMs[a]	Surface fits (Å)[b]	Success rate[c]	Properties of the best QSAR equations produced for each case						
			r^2_{68}[d]	r^2_{LOO}[e]	r^2_{BS}[f]	r^2_{PRESS}[g]	r^2_{Random}[h]	Number of terms[i]	2D terms[j]
1	−0.1	7/50	0.701	0.537	0.667	0.589	0.193	15	7
			0.654	0.572	0.655	0.664	0.198	14	6
			0.655	0.534	0.655	0.632	0.179	13	6
			0.671	0.556	0.668	0.622	0.148	11	5
			0.671	0.556	0.668	0.511	0.117	10[k]	4
			0.651	0.514	0.651	0.638	0.132	10[k]	5
			0.679	0.574	0.677	0.590	0.120	9	6
1	−0.2	9/50	0.785	0.584	0.734	0.527	0.188	14	6
			0.699	0.520	0.666	0.534	0.125	10	5
			0.761	0.511	0.675	0.552	0.104	9[k]	4
			0.728	0.534	0.594	0.507	0.115	9[k]	4
			0.665	0.549	0.665	0.573	0.110	9[k]	5
			0.698	0.560	0.682	0.548	0.099	8[k]	5
			0.673	0.545	0.658	0.576	0.097	8[k]	5
			0.657	0.544	0.657	0.572	0.103	8[k]	6
			0.620	0.515	0.620	0.596	0.072	6	4
2	−0.2, −0.1	14/50	**0.800**[l]	**0.673**[l]	**0.800**[l]	**0.600**[l]	**0.220**[l]	**16**[l]	**3**[l]
			0.737	0.622	0.715	0.535	0.171	15[k]	3
			0.678	0.534	0.678	0.609	0.195	15[k]	5
			0.721	0.554	0.696	0.597	0.190	14[k]	2
			0.689	0.560	0.689	0.624	0.190	14[k]	5
			0.800	0.662	0.789	0.548	0.169	13	4
			0.732	0.603	0.715	0.617	0.160	12	6
			0.768	0.696	0.760	0.511	0.140	11[k]	3
			0.655	0.527	0.653	0.551	0.141	11[k]	2
			0.769	0.598	0.739	0.561	0.126	10[k]	4
			0.687	0.505	0.665	0.506	0.123	10[k]	5
			0.652	0.525	0.652	0.548	0.111	9	4
			0.735	0.604	0.725	0.572	0.107	8	2
			0.671	0.568	0.669	0.512	0.160	6	4

The table also shows the effects of variable number of principle components on the statistical criteria of each model.
[a] Number of RSM models employed in building the combination QSARs.
[b] The isosurface fits of selected RSMs for QSAR analyses.
[c] Number of successful QSAR models (of $r^2_{LOO} \geq 0.5$ and $r^2_{PRESS} \geq 0.50$) per total number of scanned models.
[d] Non-cross-validated correlation coefficient determined against 68 training compounds.
[e] Cross-validation correlation coefficients determined by the leave-one-out technique.
[f] Bootstrapping correlation coefficient.
[g] Predictive r^2 determined for 16 test compounds.
[h] The average randomization correlation coefficients: the biological activities were randomized 99 times, and the mean randomization r^2 was calculated.
[i] Total number of terms in the optimal QSAR model.
[j] Number of the two-dimensional descriptors selected by G/PLS to achieve optimal QSAR model.
[k] QSAR models based on different number of latent variables.
[l] The statistical criteria of best model are underlined and highlighted in bold.

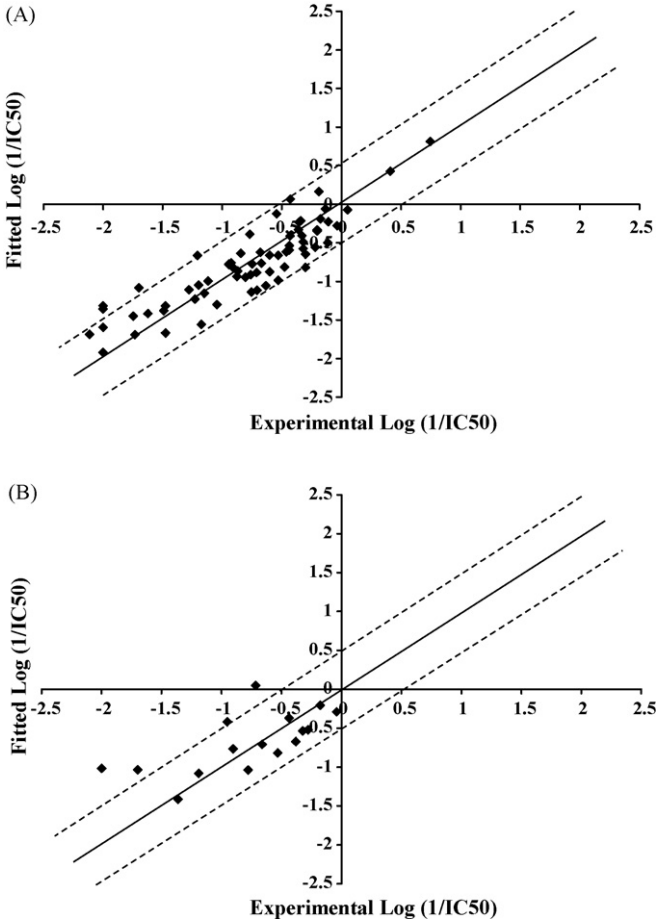


Fig. 3. Experimental versus fitted (**A**, 68 compounds, $r^2_{LOO} = 0.673$) and predicted (**B**, 16 compounds, $r^2_{PRESS} = 0.600$) bioactivities calculated from the best QSAR model obtained after 30,000 iterations of G/PLS performed on energy descriptors calculated from a combination of two RSMs of −0.1 and −0.2 Å surface fits (Table 2). The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the ±0.5 log point error margins.

Table 2. On the other hand, QSAR models resulting from other RSA configurations were quite sensitive to alterations in G/PLS settings, particularly to the number of latent variables and genetic iterations.

Intriguingly, most high-ranking QSAR models exhibited combinations of surface descriptors from 2 RSMs (as shown in Table 2). The best QSAR model, i.e., of best r^2_{LOO} and r^2_{PRESS}, included surface descriptors from RSMs corresponding to isosurfaces at −0.2 and −0.1 Å. Table 2 shows the statistical criteria of the highest-ranking QSAR models. Eq. (4) shows the optimal QSAR model, while Fig. 3 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors based on this model.

$$\begin{aligned}\log(1/IC_{50}) = {} & 0.24 + 0.29(\mathbf{TOT}/2887)_{-0.2} + 0.38(\mathbf{TOT}/4777)_{-0.2} \\ & + 1.32(\mathbf{ELE}/2845)_{-0.2} - 0.21(\mathbf{ELE}/4249)_{-0.2} \\ & - 1.48(\mathbf{TOT}/2845)_{-0.2} - 0.17(\mathbf{TOT}/5137)_{-0.2} \\ & - 0.17(\mathbf{TOT}/2815)_{-0.2} - 2.30(\mathbf{VDW}/3451)_{-0.1} \\ & + 1.04(\mathbf{VDW}/3031)_{-0.1} + 1.14(\mathbf{TOT}/5995)_{-0.1} \\ & + 0.10(\mathbf{TOT}/4831)_{-0.1} - 0.11(\mathbf{ELE}/2731)_{-0.1} \\ & + 0.11(k_3) - 0.37(\text{AtypeC6}) - 0.36\,\text{HBD}\end{aligned}$$

$$r^2_{68} = 0.800, \quad r^2_{LOO} = 0.673, \quad r^2_{BS} = 0.800, \quad r^2_{PRESS} = 0.600, \quad n = 68, \quad F = 13.834 \qquad (4)$$

where r^2_{68} is the correlation coefficient of the model against the training list, r^2_{LOO} is the leave-one-out correlation coefficient, r^2_{BS} is bootstrap r^2, r^2_{PRESS} is the predictive r^2 determined for the 16 test compounds (via Eq. (3), see Section 2.10 under Experimental) [46,60] and n is the number of training molecules. κ_3 is the third order Kier's shape index, AtypeC6 is one of the atom type logP descriptors, HBD is the number of hydrogen-bond doners [46]. The **ELE**, **VDW** and **TOT** terms describe the electrostatic, van der Waals' and total nonbonded interaction energies at the specified surface points, respectively. The terms $(\mathbf{VDW}/3451)_{-0.1}$, $(\mathbf{VDW}/3031)_{-0.1}$, $(\mathbf{TOT}/5995)_{-0.1}$, $(\mathbf{TOT}/4831)_{-0.1}$ and $(\mathbf{ELE}/2731)_{-0.1}$ correspond to RSM points generated at isosurface of −0.1 Å, while the terms $(\mathbf{TOT}/2887)_{-0.2}$, $(\mathbf{TOT}/4777)_{-0.2}$, $(\mathbf{ELE}/2845)_{-0.2}$, $(\mathbf{TOT}/2845)_{-0.2}$, $(\mathbf{ELE}/4249)_{-0.2}$, $(\mathbf{TOT}/2815)_{-0.2}$, and $(\mathbf{TOT}/5137)_{-0.2}$ correspond to a second RSM at isosurface fit of −0.2 Å. Fig. 4 shows the 3D distribution of all surface points in Eq. (4) relative to the aligned training molecules, while Figs. 5 and 6 show the spatial distribution of the surface points corresponding to these terms within the context of their respective RSMs. Emergence of surface points originating from two overlaying RSMs is quite interesting and has not been reported before. This behavior suggests certain complex interaction patterns within the ligand–receptor complex.

It remains to be mentioned that Eq. (4) represents other high-ranking models: The 2D and 3D descriptors of Eq. (4) frequented among other high ranking QSAR equations. In fact, if one of the surface descriptors in Eq. (4) failed to emerge in a particular high ranking QSAR model, another spatially close equivalent emerged instead.

4. Discussion

In RSA, the information about the receptor site is displayed graphically by mapping properties onto the surface. Regions of the surface are color-coded to indicate particular chemical properties. The intensity of the color on the surface corresponds to the magnitude of the property. The location and magnitude of a descriptor can be used as a guideline to improve the activity of molecules. Figs. 5 and 6 show graphical representations of the two successful receptor models constructed over the training set of molecules as have been aligned by consensus dynamics, i.e., at −0.1 and −0.2 Å isosurface fits (see Section 2.3 under Section 2). The electrostatic (Fig. 5) and steric (Fig. 6) fields are depicted on the receptors' surfaces as colored regions. The magenta color represents negative energy values as favorable interaction sites, while the green-colored regions represent positive energy values that are unfavorable sites for binding of a molecule on the receptor surface.

The descriptor that appears in a particular RSM equation can be associated with a structural feature of a molecule located closest to the position of that variable. The locations of some important descriptors that have emerged from the model of the receptor surface analysis described by Eq. (4) are displayed in Figs. 4–6. For the ease of discussing the effect of the nonbonded interactions at these surface points on the activity, we have demarcated these points around the basic scaffold from each class (compounds **1** and **60**), as shown in Figs. 7 and 8.

Emergence of a favorable van der Waals' interaction term at $(\mathbf{VDW}/3451)_{-0.1}$ in Eq. (4) indicates a direct relationship between the van der Waals' potential at surface point 3451 and anti-Pgp activity. The fact that this point is situated directly above the amidic methyl of MeLeu[9] in CsA scaffold (compound **1** in Table 1 and Figs. 1 and 7) agrees with the observed reduction in anti-Pgp activities in association with polar substituents at this position, e.g., the amidic hydrogen of Leu[9] in **40** (IC_{50} = 100 μM) and **50** (IC_{50} = 100 μM, Table 1 and Fig. 1). It seems that this point represents certain section in the van der Waals' surface within the

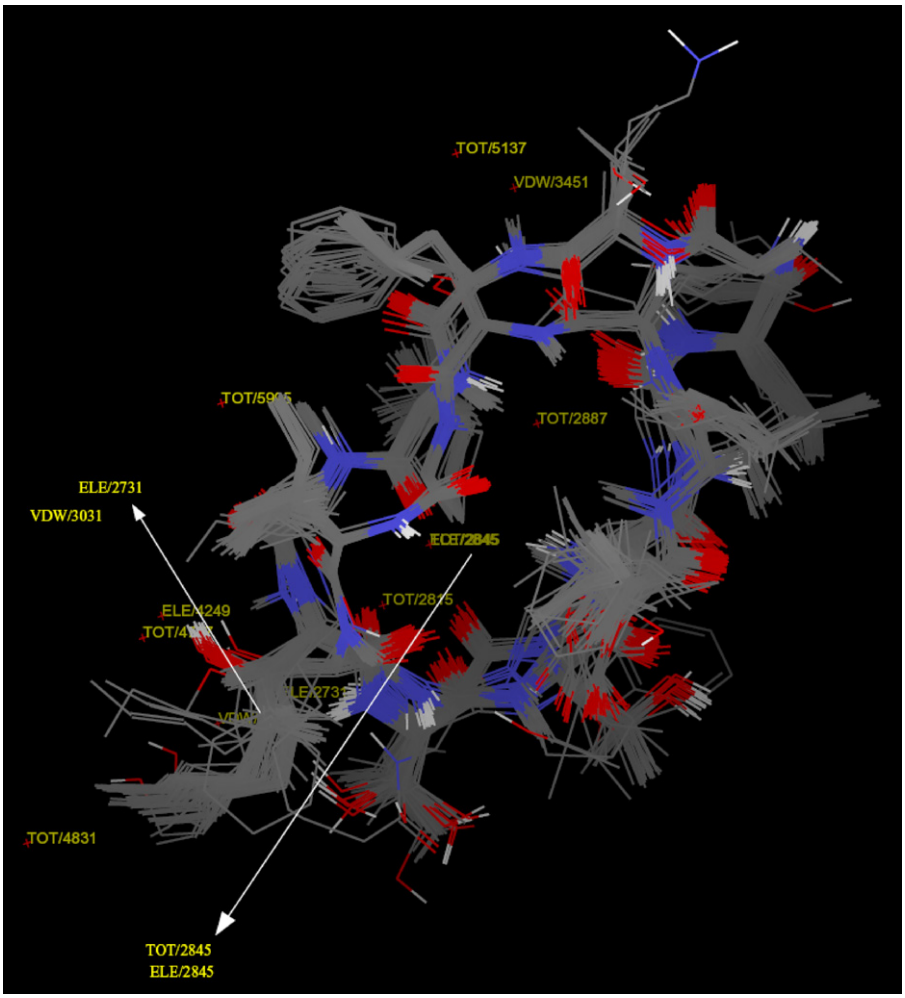


Fig. 4. The three-dimensional distribution of all surface points in Eq. (4) relative to the aligned training molecules.

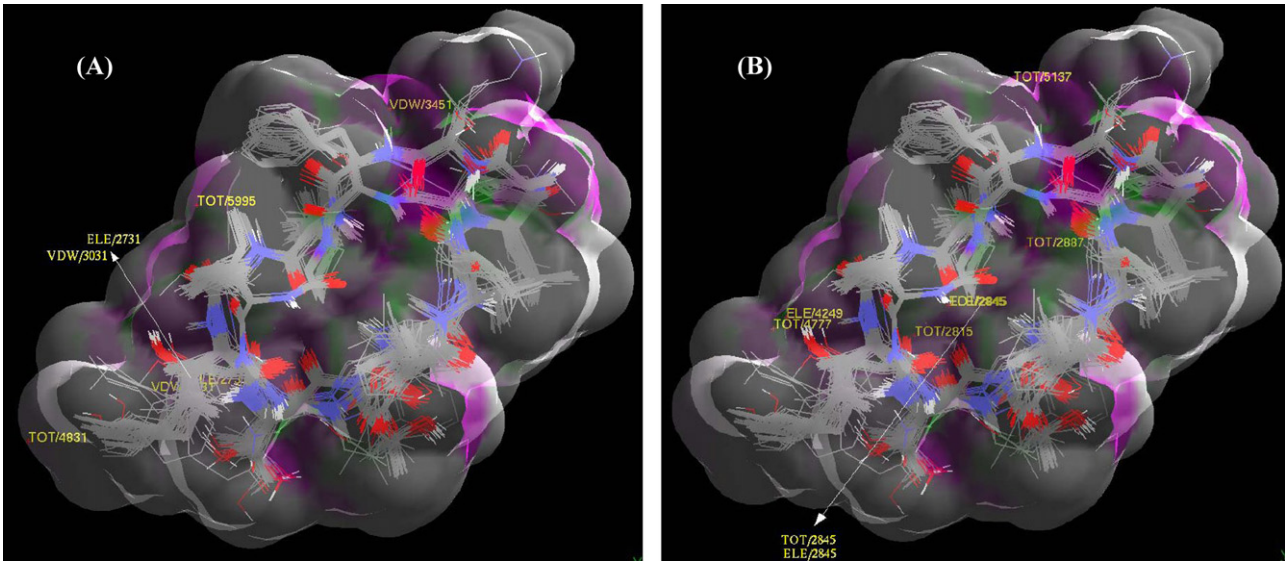


Fig. 5. Graphical representations of the electrostatic potential of the RSMs generated around the aligned inhibitors at surface fits of (A) −0.1 Å and (B) −0.2 Å. The electrostatic potentials of the proposed receptors are color-coded as follows: negative potential is shown in magenta, positive potential is show in green, and gray color represents neutral potentials. The most significant RSA descriptors (as selected by G/PLS) with their locations at both RSMs' surfaces are shown in yellow text.

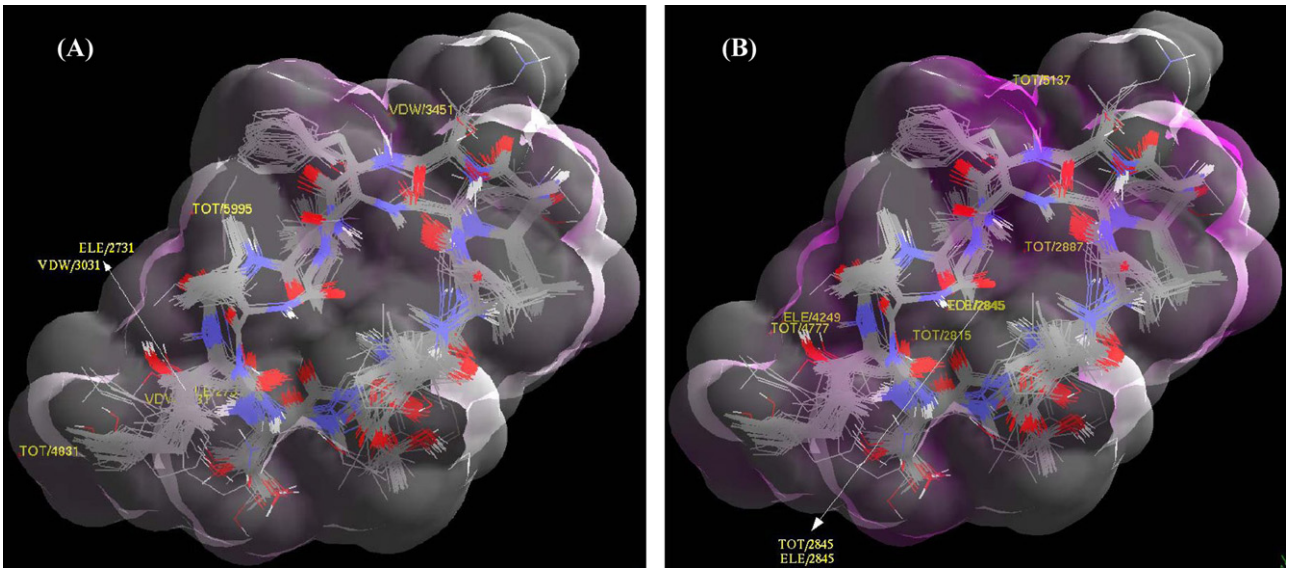


Fig. 6. Graphical representations of the steric fields of the RSMs generated around the aligned inhibitors at surface fits of (A) −0.1 Å and (B) −0.2 Å. The steric fields of the proposed receptors are color-coded as follows: negative potential is shown in magenta, positive potential is show in green, and gray color represents neutral potentials. The most significant RSA descriptors (as selected by G/PLS) with their locations at both RSMs' surfaces are shown in yellow text.

binding pocket capable of hydrophobic binding with corresponding features in the ligands. However, this surface point is too far from any moiety within the AbA scaffold suggesting other compensating interactions for this category.

Similarly, (**TOT**/5137)$_{-0.2}$ emerged in Eq. (4) combined with a favorable negative coefficient. The facts that this point is spatially situated adjacent to the β carbons of MeLeu[9] and L-MePhe[4] moieties of CsA and AbA scaffolds, respectively (Figs. 7 and 8), and that this term encodes for the total nonbonded interaction energy, i.e., **VDW** and **ELE** potentials, between template molecules and the nearest point on the receptor surface (i.e., 5137), suggest favorable mixed hydrophobic/polar substitution at the β carbons of the 9th and 4th moieties of CsA and AbA scaffolds, respectively. This trend explains the increase in anti-Pgp bioactivities associated with bulky and polar substituents at this point, e.g., **63** (β-HOMePhe[4] on AbA scaffold, IC_{50} = 2.15 μM), and the reduction in P-gp affinity in cases of compact corresponding side chains, e.g., **66**–**70** and **72** (Table 1 and Fig. 1).

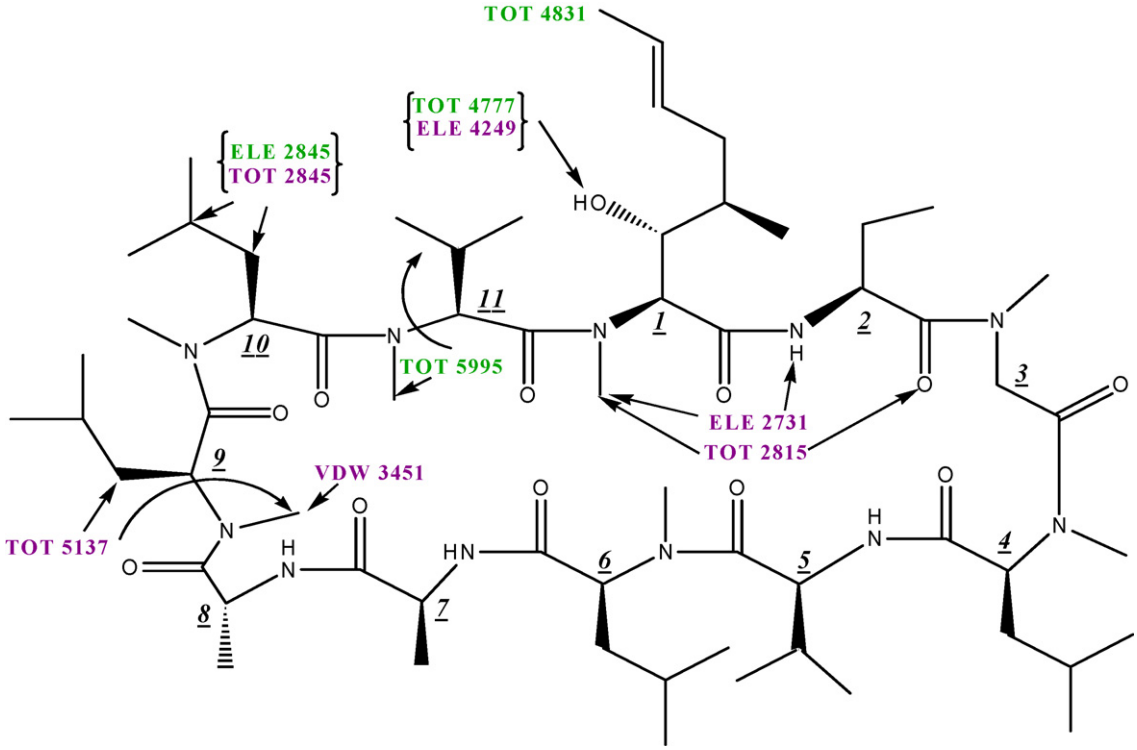


Fig. 7. Schematic representation showing the surface points of Eq. (4) and their closest atomic points on CsA (**1**, Table 1). Underlined bold numbers represent different moieties within the CsA scaffold. Green-colored fonts represent surface descriptors of positive regression coefficients that are unfavorable for binding of a molecule on the receptor surface, while magenta-colored fonts represent surface descriptors of negative regression coefficients favorable for ligand binding within the receptor surface.

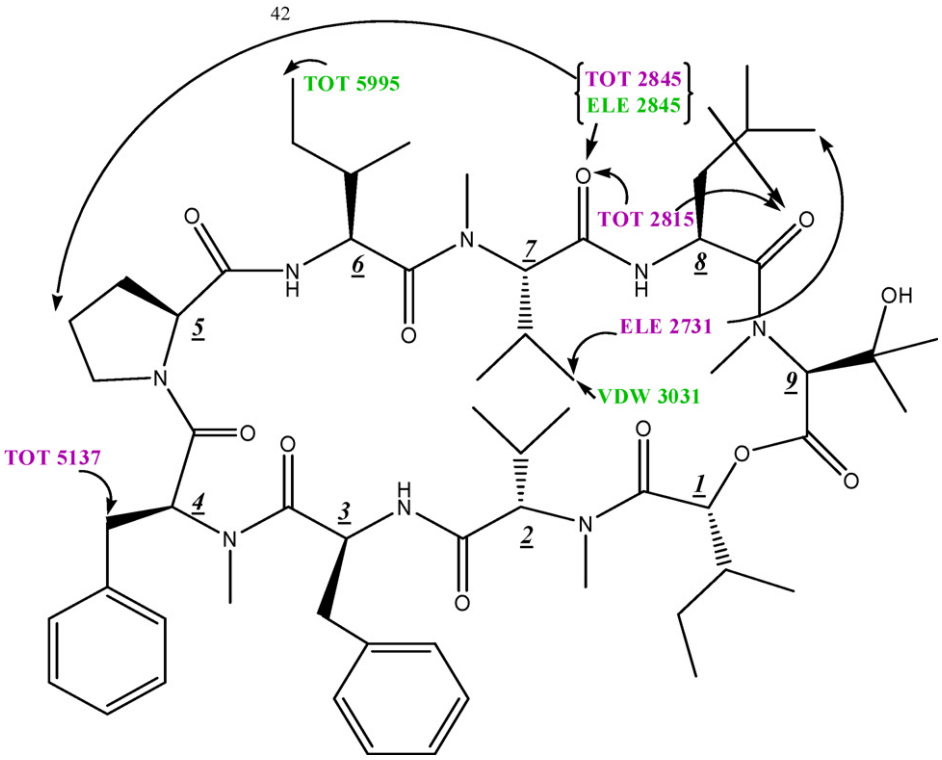


Fig. 8. Schematic representation showing the surface points of Eq. (4) and their closest atomic points on AbA (**60**, Table 1). Underlined bold numbers represent different moieties within the AbA scaffold. Green-colored fonts represent surface descriptors of positive regression coefficients that are unfavorable for binding of a molecule on the receptor surface, while magenta-colored fonts represent surface descriptors of negative regression coefficients favorable for ligand binding within the receptor surface.

Intriguingly, (**TOT**/2845)$_{-0.2}$ and (**ELE**/2845)$_{-0.2}$ emerged in Eq. (4) with contradicting coefficients. The equation proposes that (**TOT**/2845)$_{-0.2}$ favors binding to the receptor model, while (**ELE**/2845)$_{-0.2}$ disfavors binding. The fact that this point (i.e., 2845) is positioned adjacent to the side chain of the 10th moiety of the CsA backbone (Fig. 7), and midway between the side chain of the 5th residue and the amidic carbonyls of the 8th and 7th moieties of the AbA system (Fig. 8), suggest that Pgp-ligand binding favors bulky substituents at these positions, as proposed by Eq. (4). On the other hand, polar substituents at these positions seem to disfavor Pgp/ligand binding. This trend is clearly seen in the CsA category (Table 1 and Fig. 1) by comparing the enhanced affinity of **53** (Thr^{2} Leu^{5} *Leu*^{10}, IC_{50} = 6 μM) against **55** (Thr^{5} Leu^{5} *Ala*^{10}, IC_{50} = 56 μM). Similarly, this trend is clearly apparent in the AbA family (Table 1 and Fig. 1) and explains the enhanced Pgp affinity of **76** (bulky 7th residue: 2-hydroxy-isovaleric acid, IC_{50} = 1.66 μM) compared to **72** (compact glycine susbtituent at the 5th moiety, IC_{50} = 7.42 μM) or **77** (polar L-glutamic acid substituent at the 8th moiety, IC_{50} = 8.96 μM). Optimally oriented polar groups with bulky hydrophobic shields have the advantage of being protected from hydration additional to their ability to bind with complementary groups in the binding pocket.

On the other hand, the inverse relationship between (**TOT**/5995)$_{-0.1}$ and Pgp affinity proposed by Eq. (4) suggests that adjacent bulky and/or polar substituents, i.e., at the 11th moiety in CsA category and the terminal carbon of the 6th moiety in AbA category (Figs. 7 and 8) disfavor binding interactions with the binding pocket. This trend is clearly seen by comparing the superior performance of **1** (MeVal^{11}, IC_{50} = 3.40 μM) compared to **2** (D-MeVal^{11}, IC_{50} = 15.50 μM) in which the flip in the stereochemistry shifted the amidic carbonyl closer to 5995. Similarly, replacing methyl valine^{11} in **1** with methyl alanine^{11} in **6** (IC_{50} = 16.1 μM) shifted the amidic carbonyl closer to 5995 leading to apparent reduction in Pgp affinity. One of the obvious examples on the significance of (**TOT**/5995)$_{-0.1}$ in the AbA category is seen in the case of **73**: Replacing the bulky L-alloleucine^{6} of the basic AbA scaffold (**60**, IC_{50} = 2.27 μM) with smaller valine^{6} in **73** (IC_{50} = 1.29 μM) caused significant improvement in Pgp affinity (Table 1 and Fig. 1).

Emergence of positive (**TOT**/4777)$_{-0.2}$ and negative (**ELE**/4249)$_{-0.2}$ terms in Eq. (4) highlights the significance of the nearby hydroxyl of MeBmt^{1} (Fig. 1 and Table 1) in the CsA scaffold (Fig. 7). The contradicting coefficients of these spatially related terms suggest that Pgp binding favors compact, preferentially polar groups at the 1st moiety in CsA, i.e., instead of branched bulky groups. This trend can be clearly seen in the transition from **1** (MeBmt^{1}, IC_{50} = 3.4 μM), to **9** (MeLeu^{1}, IC_{50} = 8 μM), **10** (Me-cyclohexy-Ala^{1}, IC_{50} = 8.1 μM) and **11** (Me-cyclized-Bmt^{1}, IC_{50} = 13.1 μM, Table 1 and Fig. 1). However, these surface points are too far from any moiety within the AbA scaffold to have any influence on their binding, suggesting other compensating interactions for this category.

Emergence of disfavorable positive (**TOT**/4831)$_{-0.1}$ in Eq. (4) emphasizes the significance of the nearby terminal side chain atom of the 1st moiety in CsA scaffold (Fig. 7). Apparently, the corresponding region in Pgp binding pocket is sterically tight and therefore intolerant to bulky substituents regardless to their polarity status. This trend is clearly apparent by comparing Pgp affinity to **1** (MeBmt^{1}, IC_{50} = 3.4 μM) versus **12** and **56** (both have 8'-OH-MeBmt^{1}, IC_{50} = 50 μM and 15.8 μM, respectively, Fig. 1 and Table 1). However, surface point 4831 is too far from any moiety within the AbA scaffold to have any significant contribution regarding Pgp affinity to AbA compounds.

Emergence of favorable negative (**ELE**/2731)$_{-0.1}$ term in Eq. (4) signifies the nearby groups: the amidic nitrogens of the 1st and 2nd moieties within the CsA scaffold (Fig. 7), and the terminal side chain atoms of the 7th and 8th residues in the AbA scaffold (Fig. 8). Based

on this term, it seems that polar substituents at these positions enhance Pgp binding. This trend can be clearly seen via the enhancement of Pgp affinity associated with demethylating the amidic nitrogen of first moiety in **1** (MeBmt^1, IC_{50} = 3.4 μM) to give **35** (Bmt^1, IC_{50} = 2.3 μM) or **48** (Bmt^1 Val^2, IC_{50} = 2.1 μM, Table 1 and Fig. 1). However, although this point is situated close to the side chains of the 7th and 8th residues in the AbA scaffold, the significance of these side chains is not tested within AbA compounds. Nevertheless, we believe substituting these two residues with polar substituents will probably enhance their Pgp affinities.

Similarly, emergence of favorable surface interactions at (**TOT**/2815)$_{-0.2}$ in Eq. (4) emphasizes the significance of the adjacent amidic methyl and carbonyl groups of the 1st and 2nd residues of CsA scaffold, respectively (Fig. 7), and the amidic carbonyls of the 7th and 8th residues of the AbA scaffold (Fig. 8). The emergence of this surface point seems to encode for similar information to that of (**ELE**/2731)$_{-0.1}$ and (**TOT**/2815)$_{-0.2}$ in CsA- and AbA-based scaffolds, respectively, as shown in Figs. 7 and 8.

It remains to be mentioned that one of the surface terms in Eq. (4) is exclusively related to AbA scaffold, namely, (**VDW**/3031)$_{-0.1}$. This point is situated adjacent to the terminal carbon of L-MeVal^7 suggestive of disfavorable bulky substitution at this position. Although this particular point was not effectively evaluated in the training compounds, changing the substitution pattern at the 7th residue in **60** (L-MeVal^7, IC_{50} = 2.27 μM), **74** (Val^7, IC_{50} = 7.55 μM), **75** (D-MeVal^7, IC_{50} = 2.73 μM) and **76** (L-Hiv^7, IC_{50} = 1.66 μM) caused slight shifts in the positions of the isopropyl side chains, which allowed the regression analysis to detect this disfavored interaction.

5. Conclusions

Receptor surface models do not contain atoms but give a representation of the essential features of an active site by assuming complementarity between the shape and properties of the receptor site with respect to the set of binding molecules.

Out of the various receptor surface models generated in the present work, the best model Eq. (4) gave a predictive r^2 of 0.600 for a set of test molecules. The graphical representation of nonbonded interactions on the receptor surface provided a clear picture of regions on Pgp inhibitors that can be modified to improve its activity. The structure–activity relationships that have emerged from this study can be used advantageously in the development of new CsA and AbA-based Pgp antagonists. This study is the first attempt to construct 3D-QSAR model for this complex groups of Pgp inhibitors.

Acknowledgment

The authors wish to thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2008.07.005

References

[1] N. Aouali, L. Eddabra, J. Macadré, H. Morjani, Immunosuppressors and reversion of multidrug-resistance, Crit. Rev. Oncol. Hemat. 56 (2005) 61–70.
[2] K. Nooter, G. Stoter, Molecular mechanisms of multidrug resistance in cancer chemotherapy, Pathol. Res. Pract. 192 (1996) 768–780.
[3] E. Teodori, S. Dei, C. Martelli, S. Scapecchi, F. Gualtieri, The functions and structure of ABC transporters: Implications for the design of new inhibitors of Pgp and MRP1 to control multidrug resistance (MDR), Curr. Drug Targets 7 (2006) 893–909.
[4] A.R. Safa, Identification and characterization of the binding sites of P-glycoprotein for multidrug resistance-related drugs and modulators, Curr. Med. Chem. Anti-Canc. Agents 4 (2004) 1–17.
[5] S.V. Ambudkar, S. Dey, C.A. Hrycyna, M. Ramachandra, I. Pastan, M.M. Gottesman, Biochemical, cellular, and pharmacological aspects of the multidrug transporter, Annu. Rev. Pharmacol. Toxicol. 39 (1999) 361–398.
[6] P. Borst, R. Evers, M. Kool, J. Wijnholds, A family of drug transporters: the multidrug resistance-associated proteins, J. Natl. Cancer Inst. 92 (2000) 1295–1302.
[7] R.L. Juliano, V. Ling, A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants, Biochim. Biophys. Acta 455 (1976) 152–162.
[8] S.G. Dahl, I. Sylte, A.W. Ravna, Structures and models of transporter proteins, J. Pharmacol. Exp. Ther. 309 (2004) 853–860.
[9] K. Ueda, C. Cardarelli, M.M. Gottesman, I. Pastan, Expression of a full-length cDNA for the human 'MDR1' gene confers resistance to colchicine, doxorubicin, and vinblastine, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 3004–3008.
[10] A.H. Schinkel, P-Glycoprotein, a gatekeeper in the blood-brain barrier, Adv. Drug Deliv. Rev. 36 (1999) 179–194.
[11] M.M. Gottesman, I. Pastan, Biochemistry of multidrug resistance mediated by the multidrug transporter, Annu. Rev. Biochem. 62 (1993) 385–427.
[12] H. Lage, MDR1/P-glycoprotein (ABCB1) as target for RNA interference-mediated reversal of multidrug resistance, Curr. Drug Targets 7 (2006) 813–821.
[13] L.Z. Benet, T. Izumi, Y. Zhang, J.A. Silverman, V.J. Wacher, Intestinal MDR transport proteins and P-450 enzymes as barriers to oral drug delivery, J. Controlled Release 62 (1999) 25–31.
[14] M.V.S. Varma, K. Sateesh, R. Panchagnula, Functional role of P-glycoprotein in limiting intestinal absorption of drugs: contribution of passive permeability to P-glycoprotein meditated efflux transport, Mol. Pharm. 2 (2005) 12–21.
[15] A.H. Schinkel, J.W. Jonker, Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview, Adv. Drug Delivery Rev. 55 (2003) 3–29.
[16] K.L. Mealey, Therapeutic implications of the MDR-1 gene, J. Vet. Pharmacol. 27 (2004) 257–264.
[17] C. Avendano, J.C. Menéndez, Inhibitors of multidrug resistance to antitumor agents (MDR), Curr. Med. Chem. 9 (2002) 159–193.
[18] L. Schmitt, R. Tampe, Structure and mechanism of ABC transporters, Curr. Opin. Struct. Biol. 12 (2002) 754–760.
[19] M.F. Rosenberg, A.B. Kamis, R. Callaghan, C.F. Higgins, R.C. Ford, Three-dimensional structures of the mammalian multidrug resistance P-glycoprotein demonstrate major changes in the transmembrane domains conformational upon nucleotide binding, J. Biol. Chem. 287 (2003) 8294–8299.
[20] Y. Wang, Y. Li, S.L. Yang, L. Yang, Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach J, Chem. Inf. Model. 45 (2005) 750–757.
[21] P. Crivori, B. Reinach, D. Pezzetta, I. Poggesi, Computational models for identifying potential P-glycoprotein substrates and inhibitors, Mol. Pharm. 3 (2005) 33–44.
[22] L.J. Bain, J.B. McLachlan, G.A. LeBlanc, Structure-activity relationships for xenobiotic transport substrates and inhibitory ligands of P-glycoprotein, Environ. Health Perspect. 105 (1997) 812–818.
[23] J.E. Penzotti, M.L. Lamb, E. Evensen, P.D. Grootenhuis, A computational ensemble pharmacophore model for identifying substrates of P-glycoprotein, J. Med. Chem. 45 (2002) 1737–1740.
[24] Y. Li, Y. Wang, L. Yang, S.W. Zhang, C.H. Liu, S.L. Yang, Comparison of steroid substrates and inhibitors of P-glycoprotein by 3D-QSAR analysis, J. Mol. Struct. 733 (2005) 111–118.
[25] S. Ekins, R.B. Kim, B.F. Leake, A.H. Dantzig, E.G. Schuetz, L.B. Lan, K. Yasuda, R.L. Shepard, M.A. Winter, J.D. Schuetz, J.H. Wikel, S.A. Wrighton, Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein, Mol. Pharmacol. 61 (2002) 964–973.
[26] K. Ikai, K. Takesako, K. Shiomi, M. Moriguchi, Y. Umeda, J. Yamamoto, I. Kato, H. Naganawa, Structure of aureobasidin, A. J. Antibiot. 44 (1991) 925–933.
[27] R. Traber, Biosynthesis of cyclosporins, in: W.R. Strohl (Ed.), Biotechnology of Antibiotics, Marcel Dekker Inc., Basel, 1997, pp. 279–313.
[28] M.M. Dreyfuss, I.H. Chapela, Potential of fungi in the discovery of novel, low-molecular weight pharmaceuticals, in: V.P. Gullo (Ed.), The Discovery of Natural Products with Therapeutic Potential, Butterworth-Heinemann, Boston, 1994, pp. 49–80.
[29] M. Endo, K. Takesako, I. Kato, H. Yamaguchi, Fungicidal action of aureobasidin A, a cyclic depsipeptide antifungal antibiotic, against *Saccharomyces cerevisiae*, Antimicrob. Agents Chemother. 41 (1997) 672–676.
[30] F. Loor, F. Tiberghien, T. Wenandy, A. Didier, R. Traber, Cyclosporins: structure-activity relationships for the inhibition of the human MDR1 P-glycoprotein ABC transporter, J. Med. Chem. 45 (2002) 4598–4612.
[31] F. Tiberghien, T. Kurome, K. Takesako, A. Didier, T. Wenandy, F. Loor, Aureobasidins: structure-activity relationships for the inhibition of the human MDR1 P-glycoprotein ABC-transporter, J. Med. Chem. 43 (2000) 2547–2556.
[32] F. Tiberghien, F. Loor, Ranking of P-glycoprotein substrates and inhibitors by a calcein-AM fluorometry screening assay, Anti-Cancer Drug 7 (1996) 568–578.
[33] P. Crivori, B. Reinach, D. Pezzetta, I. Poggesi, Computational models for identifying potential P-glycoprotein substrates and inhibitors, Mol. Pharmaceutics 3 (2006) 33–44.
[34] Y. Xue, C.W. Yap, L.Z. Sun, Z.W. Cao, J.F. Wang, Y.Z. Chen, Prediction of P-glycoprotein substrates by a support vector machine approach, J. Chem. Inf. Comput. Sci. 44 (2004) 1497–1505.

[35] A. Garrigues, N. Loiseau, M. Delaforge, J. Ferte, M. Garrigos, F. Andre, S. Orlowski, Characterization of two pharmacophores on the multidrug transporter P-glycoprotein, Mol. Pharmacol. 62 (2002) 1288–1298.
[36] C.P. Reyes, F. Munoz-Martinez, I.R. Torrecillas, C.R. Mendoza, F. Gamarro, I.L. Bazzocchi, M.J. Nunez, L. Pardo, S. Castanys, M. Campillo, I.A. Jimenez, Biological evaluation, structure-activity relationships, and three-dimensional quantitative structure-activity relationship studies of dihydro-b-agarofuran sesquiterpenes as modulators of P-glycoprotein-dependent multidrug resistance, J. Med. Chem. 50 (2007) 4808–4817.
[37] Li. Wu-Xiong, Li. Leping, J. Eksterowicz, X.B. Ling, M. Cardozo, Significance analysis and multiple pharmacophore models for differentiating P-glycoprotein substrates, J. Chem. Inf. Model. 47 (2007) 2429–2438.
[38] R.B. Wang, C.L. Kuo, L.L. Lien, E.J. Lien, Structure-activity relationship: analyses of p-glycoprotein substrates and inhibitors, J. Clin. Pharm. Ther. 28 (2003) 203–228.
[39] M. Hahn, D. Rogers, Receptor surface models, Perspect. Drug Discov. Des. (1998) 117–133.
[40] M. Hahn, Receptor surface models. 1. Definition and construction, J. Med. Chem. 38 (1995) 2080–2090.
[41] M. Hahn, D. Rogers, Receptor surface models. 2. Application to quantitative structure-activity relationship studies, J. Med. Chem. 38 (1995) 2091–2102.
[42] W.E. Lorensen, H.E. Cline, Marching cubes: a high resolution 3-D surface construction algorithm, Comput. Graph. 4 (1987) 163–169.
[43] D. Rogers, A.J. Hopfinger, Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[44] J. Holland, Adaptation in Artificial and Natural Systems, University of Michigan Press, Ann Arbor, Michigan, 1975.
[45] J.H. Friedman, C.B. Roosen, An introduction to multivariate adaptive regression splines, Stat. Methods Med. Res. 4 (1995) 197–217.
[46] CERIUS2 4.10 QSAR, Accelrys Inc., San Diego, CA, 2005.
[47] S. Bhutoria, P.K. Mukherjee, S. Chhabra, N. Ghoshal, A 3D-QSAR of N-substituted 4-amino-3-3-dialkyl-2 (3H)-furanone GABAA receptor modulators based on receptor surface analysis, Lett. Drug Des. Discov. 3 (2006) 1–8.
[48] P.A. Datar, P.V. Desai, E.C. Coutinho, A 3D-QSAR of angiotensin II (AT1) receptor antagonists based on receptor surface analysis, J. Chem. Inf. Comput. Sci. 44 (2004) 210–220.
[49] A. Hirashima, T. Eiraku, E. Kuwano, M. Eto, Comparative receptor surface analysis of agonists for tyramine receptor which inhibit sex-pheromone production in Plodia interpunctella, Comb. Chem. High T. Scr. 7 (2004) 83–91.
[50] K. Kunal Roy, T. Leonard, QSAR analyses of 3-(4-benzylpiperidin-1-yl)-N-phenylpropylamine derivatives as potent CCR5 antagonists, J. Chem. Inf. Model. 45 (2005) 1352–1368.
[51] J. Gasteiger, M. Marsili, A new model for calculating atomic charges in molecules, Tetrahedron Lett. 34 (1978) 3181–3184.
[52] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativities: a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3228.
[53] CERIUS2 4.10 Conformational Search & Analysis, July 2005, Accelrys Inc., San Diego, CA, 2005.
[54] CERIUS2 4.00 Hypothesis & Receptor Models, April 1999, Accelrys Inc., San Diego, CA, 1999.
[55] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of qspr models, Quant. Struct.-Act. Relat. Comb. Sci. 22 (2003) 69–77.
[56] W. Sippl, Receptor-based 3D QSAR analysis of estrogen receptor ligands-merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods, J. Comput.-Aided Mol. Des. 14 (2000) 559–572.
[57] M. Clark, R.D. Cramer, The probability of chance correlation using partial least squares (PLS), Quant. Struct.-Act. Relat. 12 (1993) 137–145.
[58] C. Leger, D.N. Politis, J.P. Romano, Bootstrap technology and applications, Technometrics 34 (1992) 378–399.
[59] R.D. Cramer III, J.D. Bunce, D.E. Patterson, Crossvalidation, bootstrapping and PLS compared with multiple linear regression in conventional QSAR studies, Quant. Struct.-Act. Relat. 7 (1988) 18–28.
[60] L.F. Ramsey, W.D. Schafer, The Statistical Sleuth, 1st ed., Wadsworth Publishing Company, Belmont, CA, 1997.