JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

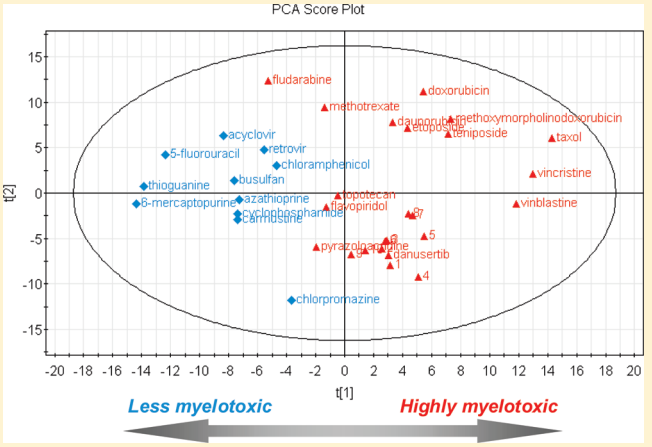
Predicting Myelosuppression of Drugs from in Silico Models

Patrizia Crivori,[*,†,§] Giulia Pennella,[‡] Miriam Magistrelli,[†] Pietro Grossi,[†] and Anna Maria Giusti[‡]

[†]Attrition Reducing Technologies and [‡]Pathology, Accelera, viale Pasteur 10, 20014 Nerviano, Italy

Supporting Information

ABSTRACT: Anticancer agents targeting proliferating cell populations in tumor as well as in normal tissues can lead to a number of side effects including hematotoxicity, a common dose-limiting toxicity associated with oncology drugs. Myelosuppression, regarded as unacceptable for other therapeutic indications, is considered a clinical risk also for new targeted anticancer drugs acting specifically on tumor cells. Thus, it becomes important not only to evaluate the potential toxicity of such new therapeutics to human hematopoietic tissue during preclinical development but also to anticipate this liability in early drug discovery. This could be achieved by using in silico models to guide the design of new lead compounds and the selection of analogs with reduced myelosuppressive potential. Hence, the purpose of this study was to develop computational models able to predict the potential myelotoxicity of drugs from their chemical structure. The data set analyzed included 38 drugs. The structural diversity and the drug-like space covered by these molecules were investigated using the ChemGPS methodology. Two sets of potentially relevant descriptors for modeling myelotoxicity (i.e., 3D Volsurf+ and 2D structural and electrotopological E-states descriptors) were selected and a Principal Component Analysis was carried out on the entire set of data. The first two PCs were able to discriminate the highest from the least myelotoxic compounds with a total accuracy of 95%. Then, a quantitative PLS model was developed by correlating a selected subset of in vitro hematotoxicity data with Volsurf+ descriptors. After variable selection, the PLS analysis resulted in a one-latent-variable model with r^2 of 0.79 and q^2 of 0.72. The inclusion of 2D descriptors in the PLS analysis improved only slightly the robustness and quality of the model that predicted the pIC_{50} values of 21 drugs not included in the model with a RMSEP of 0.67 and a squared correlation coefficient (r_0^2) of 0.70. Furthermore, in order to investigate whether the highly myelotoxic compounds are characterized by common structural features, which should be taken into consideration in the design of new candidate drugs, the entire data set was analyzed using GRIND toxicophore-based descriptors. One toxicophore emerged from the interpretation of the model. The toxicophore elements, at least determined by the molecules used in this study, are a pattern of H-bond acceptor groups, presence of a H-bond donor and H-bond acceptor regions at ~15 Å distance and a hydrophobic and H-bond acceptor interacting regions separated by a distance of ~12.4 Å. Moreover, the dimensions of the molecule play a role in its recognition as a myelotoxic compound.



INTRODUCTION

The field of hematotoxicology includes the study of adverse effects of xenobiotics on mature blood cells and also on progenitor and precursor cells (myelotoxicity) in the hematopoietic tissues. In humans, the hematopoietic tissues that produce new blood cells are primarily located in the bone marrow. Since this system is actively involved in cell division, it results one of the most common dose-limiting target tissues for conventional cytotoxic drugs. Consequently, myelosuppression in patients treated with anticancer drugs has been generally viewed as a "manageable" form of toxicity, even though it is one of the major limitations to the full therapeutic exploitation of these drugs. Moreover, the occurrence of myelosuppression, which is not tolerated for other drug therapies, should be minimized for new target-specific antitumor agents that are developed under the paradigm of high specificity and reduced toxicity. Thus, the early assessment of drug-induced hematotoxicity appears extremely important to evaluate the safety margin of new drugs.[1−5] The availability of in silico models would be a valuable decision-making tool to predict the myelotoxic potential of candidate drugs from their structural properties during early discovery phase. Combined with other relevant in silico filters for predicting biopharmaceutic and pharmacokinetic characteristics, they could be useful to design focus libraries and optimize lead compounds in terms of desired drug-like properties and safety

Received: September 29, 2010
Published: January 28, 2011

Table 1. Experimental pIC_{50} (M) Myelotoxicity Data, Actual Toxicity Class[a], Therapeutic Indications, and Main Mechanisms of Action of Drugs Selected As Training Set

code	drug	pIC_{50} (M) for CFC-GEMM	pIC_{50} (M) for GM-CFC[b]	actual toxicity class	therapeutic indication	main recognized mechanism of action
Tr-01	doxorubicin	7.68	8.06	HM[c]	anticancer	prevent DNA replication; inhibitor of topoisomerase II; free radical generation
Tr-02	vinblastine	7.66		HM	anticancer	inhibitor of microtubule formation
Tr-03	topotecan	7.64	7.37	HM	anticancer	inhibitor of topoisomerase I
Tr-04	gemcitabine	7.57	7.72	HM	anticancer	inhibitor of ribonucleotide reductase
Tr-05	taxol (paclitaxel)	7.39	8.33	HM	anticancer	inhibition of mitosis by stabilization of microtubules. Block the function of the apoptosis inhibitor protein Bcl-2
Tr-06	daunorubicin	7.38		HM	anticancer	topoisomerase-mediated interaction with DNA
Tr-07	vincristine	7.24		HM	anticancer	irreversibly binding to microtubules
Tr-08	danusertib	6.67		HM	anticancer	aurora kinase inhibitor
Tr-09	etoposide	6.32	5.78	HM	anticancer	inhibitor of topoisomerase II
Tr-10	methotrexate	5.38	3.99	HM	anticancer, immunosuppressant	inhibitor of the dihydrofolate reductase
Tr-11	chlorpromazine	4.93	4.71	LM[d]	antiemetic, antipsychotic	block of dopaminergic receptors
Tr-12	5-fluorouracil	4.54	5.26	LM	anticancer	antimetabolite. inhibitor of cell cycle progression. Inhibitor of DNA synthesis
Tr-13	chloramphenicol	3.89		LM	antibiotic	interfering with peptidyl transferase activity
Tr-14	cyclophosphamide	3.27	3.81	LM	anticancer, immunosuppressant	binding to DNA. inhibitor of DNA synthesis
Tr-15	acyclovir (zovirax)	3.06	3.57	LM	antiviral	inhibitor of herpes virus DNA polymerase

[a] Drug was considered as highly or less myelotoxic when it showed a pIC_{50} greater or less than 5.3, respectively. [b] Data taken from refs 18 and.19 [c] Highly myelotoxic. [d] Less myelotoxic.

profile. The in silico evaluation of myelotoxicity could be also helpful to identify and select the least toxic analogs and consequently to reduce the number of compounds to be evaluate using more time-consuming and resource intensive in vitro and in vivo screens.

This work represents the first attempt to build an in silico model to predict the myelosuppressive activity of drugs. The lack of computational models in this area is likely due to the paucity of the available in vitro experimental data. Although the hemopoietic colony-forming assay (CFA) was introduced in 1966,[6,7] and there are many reports of its application in drug development, it has never been routinely used in pharmaceutical research.[4,5] There are, indeed, a number of drawbacks to the CFA that have limited its use in drug development.[2] The assay is time-consuming and requires a high degree of technical skill to count and differentiate colonies manually. It is subjective and cannot be standardized. The introduction of a new assay, named HALO (hematopoietic/hemotoxicity assay via luminescence output),[1,2,8] which does not suffer from the disadvantages of the manual CFA, allows the use of the assay for screening discovery compounds during lead optimization. Currently, in our laboratories the potential hematotoxicity of candidate drugs is evaluated using human CFC-GEMM cells (colony-forming cell granulocyte, erythroid, megakaryocyte, macrophage) via luminescence output.[1,2,8] The use of mature stem cells has the advantage to provide information relevant to the whole hematopoietic system. A first set of known drugs, mainly anticancer agents, was initially tested to setup the assay and the data generated were used to develop predictive quantitative structure−myelotoxicity relationships (QSTR). The drug-like properties and structural diversity of the entire data set were investigated using the ChemGPS tool (Chemical Global Positioning System)[9,10] combined with Volsurf+ descriptors (GPSVS).[11,12] The electrotopological E-state indices,[13] as well as 3D molecular interaction field (MIF) Volsurf+[11] descriptors have been already successfully employed to model various toxicity end points.[14,15] Hence, these descriptors were selected to model the relevant in vitro myelotoxicity end point. Various statistical parameters were calculated and compared to evaluate the quality and robustness of the models generated. Then, the models were further tested by predicting the myelosuppressive activities of molecules not included in the training set. Furthermore, the structural requirements that most likely characterize myelotoxic compounds were also investigated using a 3D-QSTR approach based on GRIND-toxicophoric descriptors.[16,17]

MATERIALS AND METHODS

Drug Selection. A set of 17 commercial compounds, representing a variety of anticancer drugs and few additional compounds with different therapeutic indications, and 10 proprietary molecules synthesized by Nerviano Medical Sciences (NMS) Medicinal Chemistry (Table 1 and 2) were chosen and tested in vitro for bone marrow toxicity. All the commercial compounds, excluding topotecan that was provided by APAC Pharmaceutical, were obtained from Sigma Aldrich.

To dissolve the test items with low water solubility, DMSO (Sigma Aldrich) was used at the concentration of 0.05%.

All further dilutions analyzed in the assays were performed in culture medium and each dilution was added to the wells at a volume of 10% of the total culture volume of 100 μL.

In Vitro Bone Marrow Toxicity. The HALO platform (hematopoietic/hemotoxicity assay via luminescence output) based on the classical colony-forming assay procedure was used to perform the in vitro experiments.[1,2,8] HALO kits were purchased from HemoGenix Inc. (HALO-96 MeC Research Kit).

dx.doi.org/10.1021/ci1003834 |J. Chem. Inf. Model. 2011, 51, 434–445

Table 2. Experimental and Predicted pIC_{50} Myelotoxicity Data, Actual Toxicity Class[a], Therapeutic Indications, and Main Mechanisms of Action of Drugs Selected As the Test Set

code	drug	pIC_{50} (M) for CFC-GEMM	pIC_{50} (M) for GM-CFC[b]	predicted (model B) pIC_{50} (M)	predicted (model C) pIC_{50} (M)	predicted (model E) pIC_{50} (M)	actual toxicity class	therapeutic indication	main recognized mechanism of action
Ts-01	compound **1**	6.85		6.21	5.71	6.05	HM[c]	anticancer	kinase inhibitor
Ts-02	compound **2**	6.72		6.22	5.52	6.00	HM	anticancer	kinase inhibitor
Ts-03	compound **3**	6.65		6.21	5.45	5.95	HM	anticancer	kinase inhibitor
Ts-04	compound **4**	6.52		6.61	5.39	6.13	HM	anticancer	kinase inhibitor
Ts-05	compound **5**	6.27		6.69	5.83	6.35	HM	anticancer	kinase inhibitor
Ts-06	compound **6**	6.2		6.28	5.41	5.98	HM	anticancer	kinase inhibitor
Ts-07	compound **7**	6.04		6.61	5.43	6.22	HM	anticancer	kinase inhibitor
Ts-08	compound **8**	6.01		6.61	5.58	6.26	HM	anticancer	kinase inhibitor
Ts-09	compound **9**	5.9		5.77	5.32	5.62	HM	anticancer	kinase inhibitor
Ts-10	compound **10**	5.75		5.94	5.39	5.76	HM	anticancer	kinase inhibitor
Ts-11	6-mercaptopurine (6-MP)	5.1		3.70	4.21	3.89	LM[d]	anticancer immunosuppressant	antimetabolite; inhibitor of DNA synthesis
Ts-12	azathioprine	4.46		4.40	4.87	4.64	LM	anticancer immunosuppressant	prodrug of 6-MP
Ts-13	methoxy-morpholinyldoxorubicin		7.63	7.29	6.98	7.21	HM	anticancer	inhibitor of topoisomerase I and II
Ts-14	cytosar-U (cytarabine)		7.55	4.56	5.01	4.72	HM	anticancer	antimetabolite; inhibitor of DNA prolymerase
Ts-15	teniposide		7.1	7.10	6.68	6.97	HM	anticancer	DNA degradation; inhibitor of topoisomerase II; block
Ts-16	flavopiridol		6.85	5.55	5.45	5.50	HM	anticancer	inhibitor of cyclin-dependent kinase
Ts-17	fludarabine		5.98	5.09	4.83	5.05	HM	anticancer	antimetabolite; inhibitor of DNA prolymerase; inhibitor of ribonucleotide reductase and DNA primase
Ts-18	pyrazoloacridine		5.96	5.35	5.22	5.30	HM	anticancer	inhibitor of RNA synthesis, DNA synthesis, and inhibitor of topoisomerases I and II
Ts-19	thioguanine		5.15	3.79	4.22	3.97	LM	anticancer	antimetabolite; Inhibitor of DNA and RNA syntheses
Ts-20	busulfan		4.95	4.09	4.73	4.29	LM	anticancer	mechanism not fully understood; DNA alkylation and DNA breaks and inhibition of DNA replication and RNA transcription
Ts-21	carmustine		4.25	4.05	4.00	3.98	LM	anticancer	alkylation and cross-linking of DNA
Ts-22	retrovir (zidovudine)		4.04	4.86	4.78	4.81	LM	antiviral	inhibitor of HIV reverse transcriptase; Inhibitor of DNA polymerase

[a] Drug was considered as highly or less myelotoxic when it showed a pIC_{50} greater or less than 5.3, respectively. [b] Data taken from refs 18 and 19. [c] Highly myelotoxic. [d] Less myelotoxic.

dx.doi.org/10.1021/ci1003834 |J. Chem. Inf. Model. 2011, 51, 434–445

Human bone marrow (BM) mononuclear cells (MNCs) from different donors were supplied by Poietics Technologies Inc. and kept frozen in liquid nitrogen at −196 °C. Thawing of the BM MNCs was performed immediately prior to the experiment. The cells were then diluted with Iscove's modified Dulbecco's medium (IMDM, Gibco/Invitrogen S.r.l) containing 20% fetal bovine serum (FBS) (Euroclone) and 120 units/mL DNase (Sigma-Aldrich), added dropwise while gently swirling the cells. The cells were centrifuged at 200*g* for 15 min at room temperature and washed once with IMDM containing DNase. The pellets were resuspended in IMDM, and the cell suspension was filtered through a 100 μm nylon sterile filter to remove cell clumps; then the cell concentration was adjusted to 2×10^6/mL.

For the myelotoxicity determination, test items were dispensed in replicates into a 96-well plate then added with the different incubation components (serum species specific, methyl cellulose and growth factors cocktail) in specific ratios to cell suspensions according to the manufacturer procedures. At least six different concentrations were tested for each test item. The culture plates were transferred to a 37 °C, fully humidified incubator with an atmosphere of 5% CO_2 and incubated for 7 days. During the incubation period, the cells, stimulated by the growth factor combinations, proliferate and divide. At the end of the incubation period, intracellular adenosine triphosphate (ATP) level, indicative of the proliferation stage of the cells, was determined by a bioluminescence assay based on a reaction between ATP + luciferin + oxygen. Luminescence measurement was performed using a Packard Lumicount spectrophotometer.

Data Analysis. The inhibitory activity of each test item was calculated from the dose−response curves and expressed as IC_{50}, i.e., the concentration causing a 50% inhibition of the cell growth in treated cultures relative to untreated controls. IC_{50} values were obtained by linear regression analysis.

Data Sets. The training set included 15 known drugs. The experimental results reported in Table 1 are expressed as the negative logarithm of IC_{50} molar concentration (pIC_{50}). The myelotoxic activity spanned from 7.68 to 3.06 covering 4.62 log-unit. Two known anticancer drugs and ten new kinase inhibitors were used as test set to evaluate the in silico models (Table 2, compounds Ts-1 to Ts-12). In order to increase structural diversity, activity range and data distribution, additional in vitro hematotoxicity published results taken from an international validation study supported by the European Centre for the Validation of the Alternative Methods (ECVAM),[18,19] were included in the test set (Table 2, compounds Ts-13 to Ts-22). Since the international validation study[18,19] focused only on drug-induced neutropenia, the toxic effect of the studied compounds was assessed on a specific population of progenitor cells, the Granulocytes Monocytes (GM-CFC), using the in vitro colony forming assay. Despite the different experimental conditions (i.e., CFA on human GM-CFC cells versus HALO on human CFC-GEMM cells), the comparison of pIC_{50} values of 10 compounds belonging to the training set (Table 1) and tested in both experimental systems resulted in a correlation coefficient (r_0^2) of 0.86 (Figure 1) with a root-mean-squares error of prediction (RMSEP) of 0.67. On the basis of such results we were confident to use the remaining published data[18,19] to test the developed quantitative structure−toxicity relationships (QSTR). Bleomycin and cyclosporin reported in the original manuscript[18] were excluded from the analysis since the structure of the former is not completely known, whereas the latter molecule was an outlier in our models (i.e., structurally dissimilar to the training set). The activity range

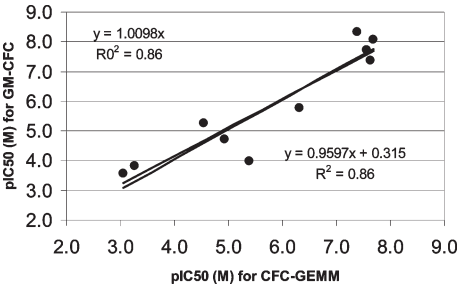


Figure 1. Experimental pIC_{50} for CFC-GEMM vs experimental pIC_{50} for GM-CFC of 10 drugs reported in Table 1.

of the entire test set spanned from 7.63 to 4.04 covering 3.5 log-unit (Table 2).

All the molecules were modeled in their neutral form and the three-dimensional (3D) structures were generated using the Corina program.[20]

VolSurf+ Descriptors. The VolSurf+ program[21] is a computational procedure producing and exploring the physicochemical property space of a molecule, starting from 3D maps of the interaction energies (MIF) with four different chemical probes (water, hydrophobic, carbonyl oxygen, and amide nitrogen probes).[11] The Volsurf+ descriptors[11] can be classified as size and shape molecular descriptors, descriptors related to hydrophilic, hydrophobic, hydrogen-bond regions, interaction energy moments, charge state descriptors and 3D pharmacophoric descriptors. A total of 120 Volsurf+ descriptors were calculated and analyzed to generate predictive QSTR. The most important and significant descriptors identified in this study are described more in details in Table 3.

GRIND-Toxicophoric Descriptors. The generation of alignment-independent GRIND descriptors involves different steps that are reported in detail by Pastor et al.[16,17] The selection of highly relevant regions in the MIFs was done using a new developed algorithm named AMANDA.[22] In this work, the GRIND descriptors were obtained as implemented in program Pentacle[23] using the default settings (DRY, O, N1 and TIP probes, with 0.5 Å grid step, dynamic parametrization, default AMANDA MIF discretization and default MACC-2 with 0.8 smoothing window). A total of 10 correlograms were calculated for each compound. Four of them are autocorrelograms obtained by analyzing node−node interactions belonging to the same MIF (i.e., DRY−DRY, N1−N1, O−O, and TIP−TIP), whereas the remaining 6 are cross-correlograms calculated using node−node interactions of different pairs of MIFs.

2D Structural and Atom-Type Electrotopological State Indices Descriptors. To model hematotoxicity data, 48 conformational-independent bidimensional descriptors, including 36 atom-type electrotopological state indices (E-state indices),[13] were calculated using Pipeline Pilot 7.5.[24] The E-state indices encode information about both the topological environment and the electronic interaction of an atom due to all other atoms in the molecule. To calculate the E-state indices, each atom in a molecule is categorized according to a valence state classification scheme, then all the atoms of the same type are grouped and their E-state values are summed-up to define the E-state index for the atom type.[13] The most important and significant electrotopological E-state descriptors identified in the QSTR are reported in Table 4. Other relevant 2D descriptors calculated and included in the analysis are related to size, polar and hydrophilic−lipophilic characteristic of the molecules, such as molecular weight, volume (Molecular_Volume), surface area

Table 3. Meaning of the Most Important and Significant 3D Volsurf+ Descriptors Identified in the Myelotoxicity Models

Volsurf descriptors	definition
	pharmacophoric descriptors
DRACAC, ACACAC, ACACDO, DRDRAC, DRACDO, DRDRDR, DRDRDO, ACDODO, DRDODO	The atoms of a structure are classified as dry (DR), H-bond donor (DO), and acceptor (AC); then all possible triplet of distances betweens these atoms are generated. The derived descriptors are the maximum area (over all possible conformers) of the triangles derived from every class of pharmacophoric triplets.
	descriptors obtained from the hydrophilic (H_2O) interaction fields
V	Volume of the water molecule interaction field at 0.2 kcal/mol energy level
S	Surface of the water interaction field at 0.2 kcal/mol energy level
R	Rugosity: ratio between the volume (V) and the surface (S)
G	Globularity: ratio between the surface (S) and the surface of a sphere with the same volume (V)
W1−W8	Volumes of the water molecule interaction fields at eight different energy levels: −0.2, −0.5, −1.0, −2.0, −3.0, −4.0, −5.0, and −6.0 kcal/mol
IW1-IW4	Integy moments: express the unbalance between the center of mass of a molecule and the center of mass of its hydrophilic regions. They are vectors pointing from the center of mass of the molecule to the center of mass of its hydrophilic regions.
	descriptors obtained from the amide nitrogen (N1) interaction fields
WN1-WN6	H-bond acceptor volumes may be defined as the molecular envelope generating attractive H-bond acceptor interactions at six different energy levels (from −1 to −6 kcal/mol)
	mixed descriptors
POL	Polarizability, is an estimation of the average molecular polarizability.
MW	Molecular weight
DIFF	Diffusivity. Computed using a modified Stokes−Einstein equation, control the dispersion of chemical in water fluid at 25 °C.
PSA, HSA	Polar and hydrophobic surface areas calculated as the sum of polar and hydrophobic regions, respectively.
SOLY	Intrinsic solubility calculated via a PLS model derived by fitting Volsurf+ descriptors to the logarithm of experimental intrinsic solubility data
	charge state descriptors
%FU4−%FU10	Percentage of un-ionized species calculated at pH 4, 5, 6, 7, 8, 9 and 10.

(Molecular_SurfaceArea), polar surface area (Molecular_PolarSurfaceArea), fractional polar surface area (Molecular_FractionalPSA; ratio of the polar surface area divided by the total surface area.), total solvent accessible surface area (Molecular_SASA), polar SASA (Molecular_PolarSASA), fractional polar SASA (Molecular_FractionalPolarSASA; ratio of the polar solvent accessible surface area divided by the total solvent accessible surface area), solvent accessible volume (Molecular_SAVol), number of hydrogen bond donors (Num-H-Donors) and acceptors (Num-H-Acceptors), and Ghose and Crippen octanol−water partition coefficient (AlogP).

Principal Component Analysis (PCA) and Partial Least Squares (Discriminant) Analysis (PLS, PLS-DA). PCA[25] was performed on the complete set of calculated VolSurf+ and 2D descriptors to evaluate the structural variance of the training and test sets. The descriptors were centered and scaled to unit variance. PCA was carried out using the SIMCA-P+[26] software.

PLS and PLS-DA[27,28] were applied to identify correlations between the calculated descriptors and the experimental toxicity data of the training set. In these analyses the 2D and VolSurf+ descriptors were centered and scaled to unit variance, whereas the GRIND descriptors were unscaled.

The predictive ability of the obtained PLS models were first evaluated using the default cross-validation method 7-fold leave-

Table 4. Meaning of the Most Important and Significant 2D Descriptors Identified in the Myelotoxicity Models

E-state indices[a]	definition (atom-type)
ES_Sum_sCH3	$-CH_3$
ES_Count_sCH3	number of group atom $-CH_3$
ES_Count__assC	number of group atom aC<
ES_Sum_sOH	−OH
ES_Count_sOH	number of group atom −OH
ES_Sum_ssssC	>C<
ES_Count_ssssC	number of group atom >C<
ES_Sum_sCl	−Cl
ES_Count_sCl	number of group atom −Cl
ES_Sum_dO	=O
ES_Count_dO	number of group atom =O
ES_Sum_ssNH	−NH−
ES_Count_ssNH	number of group atom −NH−
ES_Count__dssC	number of group atom ═C<
ES_Sum_ssO	−O−
ES_Count_ssO	number of group atom −O−

[a] The set of bonds to a skeletal atom is given by a string of lower case letters: s (single), d (double), t (triple), and a (aromatic).

dx.doi.org/10.1021/ci1003834 |J. Chem. Inf. Model. 2011, 51, 434–445

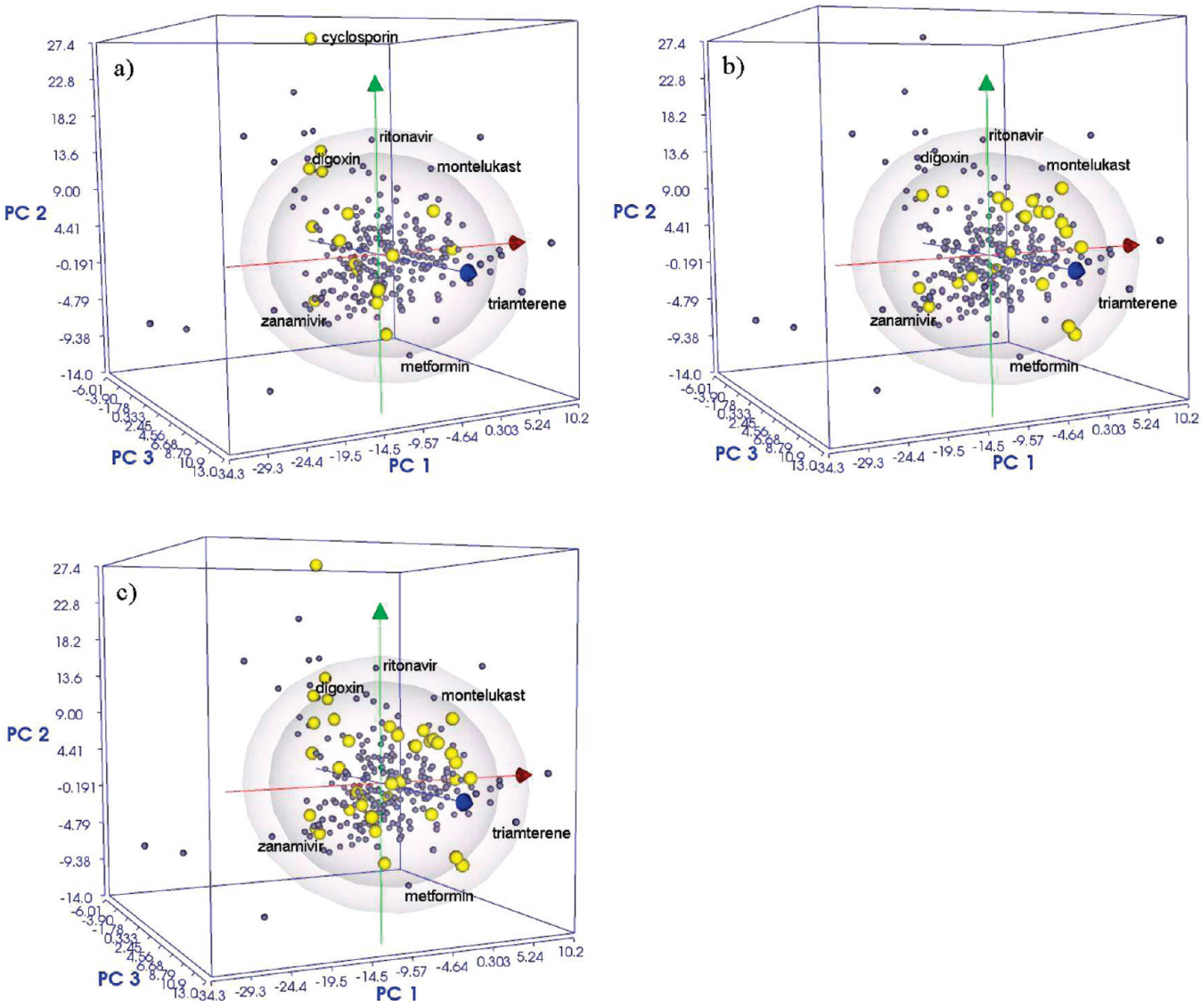


Figure 2. The (a) training set, (b) test set, and (c) entire data set shown as yellow dots projected on the three most important principal components (t1, t2, and t3) of the GPSVS map. The GPSVS satellites and core structures are shown as gray dots. Hotelling ellipsoids represent the drug-like model domain. Cyclosporin falls outside the model domain. As reference molecules, six core structures (i.e., digoxin, ritonavir, montelukast, triamterene, metformin, and zanamivir) are labeled in the GPSVS maps.

many-out implemented in SIMCA-P+[26] program. Then, the myelosuppressive activity of compounds not included in the models (i.e., test set reported in Table 2) was predicted and the obtained values compared with the experimental ones. The squared correlation coefficient values between the observed and predicted values of the test set compounds with intercept (r^2) and without intercept (r_0^2) were calculated to assess the quality of the predictions and consequently the predictive ability of the models. Two additional statistical parameters called root-mean-squares error (RMSEE) and root-mean-squares error of prediction (RMSEP), were also calculated.

PLS and PLS-DA were carried out using the SIMCA-P+[26] and Pentacle[23] software packages, respectively.

ChemGPS. It is a tool[9,10,12] that combines general properties such as, size, shape, lipophilicity, and hydrogen bond capacity and objects (i.e., chemical structures) to provide a consistent chemical space map. The ChemGPS drug space map coordinates were t-scores extracted via PCA from 120 Volsurf+ descriptors (GPSVS; i.e., ChemGPS using Volsurf+ descriptors)[12] that defined the physicochemical properties of a total set of around 300 core[29] (i.e., representative drugs selected from the literature) and satellite (non drug-like) structures. The projection of the compound descriptors on the ChemGPS map, allows to describe the diversity and the chemical space explored by the studied data set in comparison to a set of know drugs (i.e., ChemGPS drug space).

RESULTS AND DISCUSSION

Drug-like Properties and Structural Diversity Evaluation of the Data Set. The structural diversity of the data set used for modeling the in vitro myelosuppression data, was evaluated using the ChemGPS approach. The GPSVS analysis showed that the training set selected (Table 1) was characterized by a broad structural diversity, with molecules scattered in the overall drug-like chemical space defined by the ChemGPS core and satellite structures (Figure 2a). Only cyclosporin was positioned outside

dx.doi.org/10.1021/ci1003834 |J. Chem. Inf. Model. 2011, 51, 434–445

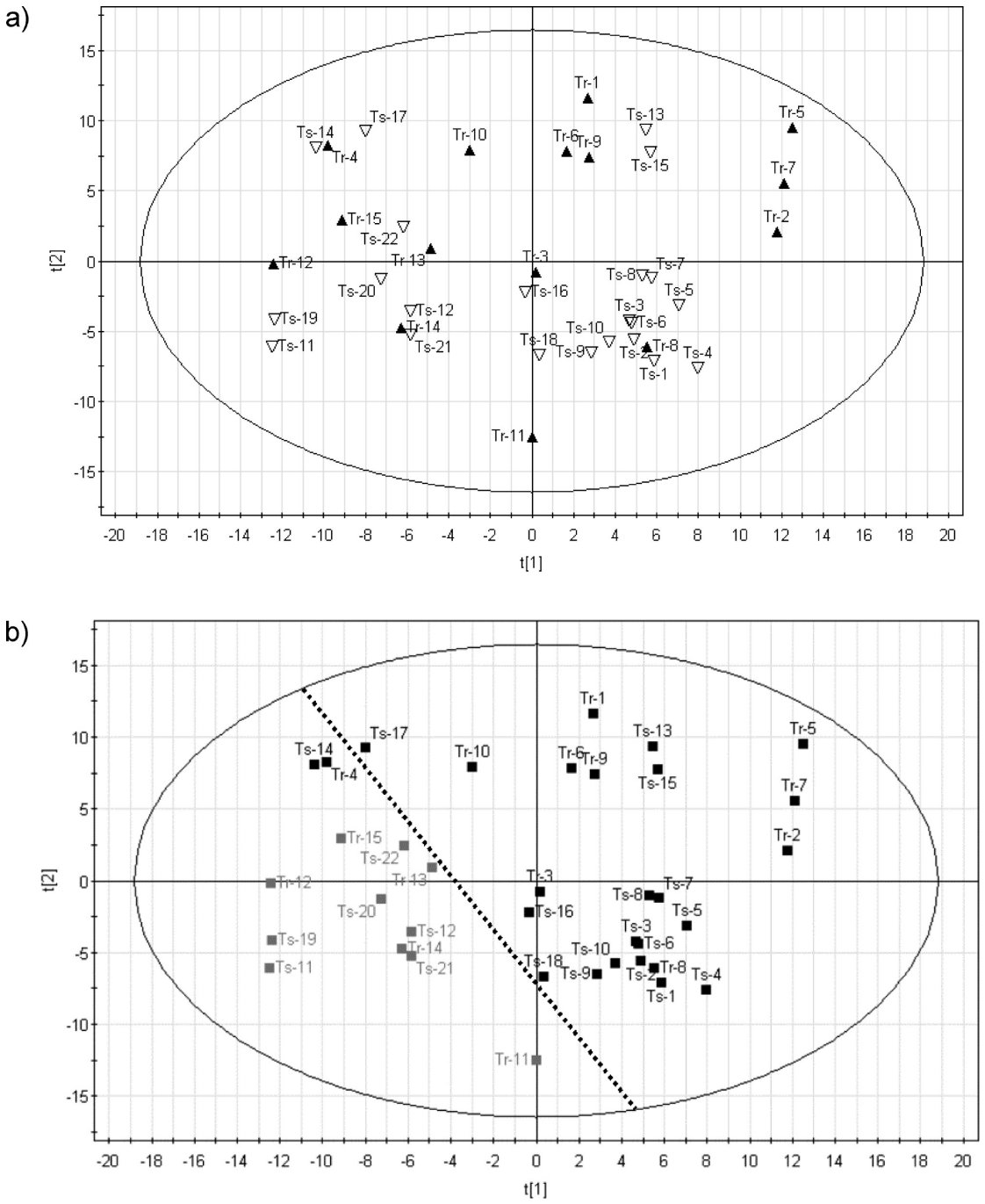


Figure 3. PCA t1-t2 score plot derived from the analysis of the 2D and 3D descriptors calculated for the entire training and test sets. The compound codes correspond to the compound names listed in Tables 1 and 2. (A) Black triangle and open inverted triangle denote training and test set compounds, respectively. (B) Black box and gray box denote compounds with pIC_{50} greater and less than 5.37, respectively.

of the model domain defined by the hotelling ellipsoids, as shown in Figure 2a. Since this molecule could negatively affect the model performance it was removed from the subsequent analyses. Furthermore, the ChemGPS analysis revealed a similar distribution of the test set (Table 2) in the GPSVS map (Figure 2b). Although a relatively small number of drugs was studied in this work, their structural properties covered a large volume of the drug-like space as shown in Figure 2c. Hence, the models developed using such structurally diverse data set could be considered of wide applicability. It is important to note that the similarity or dissimilarity of new compounds toward the molecules forming the training set can be estimate by projecting and analyzing their relative positioning on the GPSVS drug space map. Consequently it can be evaluated whether the models could properly be applied to predict the myelotoxicity of a new set of molecules.

Principal Component Analysis on the Data Set. The structural variance of the data set as well as the relevance of the descriptors for modeling myelosuppressive activity were analyzed with principal component analysis performed on the complete set of 2D and 3D-Volsurf+ descriptors calculated for the training and test sets. The two first PCs explained 53% of the

dx.doi.org/10.1021/ci1003834 |*J. Chem. Inf. Model.* 2011, 51, 434–445

Table 5. Statistical Data of the Models

		training set (14 compounds, gemcitabine removed)					test set (21 compounds, cytosar-U removed)		
model name	model type	number of variables	number of LV	r^2	q^2(7-fold leave-many-out)	RMSEE	r^2	r_0^2	RMSEP
A	3D-MIF Volsurf+	120	1	0.79	0.64	0.82	0.65	0.65	0.72
B	after variable selection	61	1	0.79	0.72	0.82	0.67	0.67	0.69
C	2D descriptors	48	1	0.72	0.63	0.94	0.70	0.58	0.82
D	3D Volsurf+ plus 2D descriptors	167	1	0.78	0.65	0.83	0.69	0.69	0.68
E	after variable selection	99	1	0.78	0.72	0.84	0.70	0.70	0.67

structural variance of the data set. The score plot for the first two PCs (Figure 3A) shows that the training and test sets share similar chemical space. The initially selected test set (Table 2, compounds Ts-1 to Ts-12) includes two known drugs and ten new molecules that share similar myelotoxicity activity and are clustered in a narrow chemical space as shown in Figure 3A. For this reason, well-documented in vitro hematotoxicity data (Table 2, compounds Ts-13 to Ts-22) were selected from the literature[18,19] and used to further validate the models. The final test set spans more than three-log unit activity and includes molecules with structural characteristics that cover all the chemical space explored by the training set (Figure 3A).

As shown in Figure 3B, the first two PCs are able to qualitatively discriminate between molecules with high and low myelosuppression potency (pIC_{50} cutoff value of 5.3). Such a result appears of high interest given that neither classification of compounds nor any information on myelosuppressive activity was introduced into the PCA model; thus the descriptors selected appeared to be relevant to model this specific toxicity. On the basis of the PCA score plot positions, 94.6% of the entire data set resulted correctly classified. A more detailed inspection of the score plot (Figure 3B) pointed out that gemcitabine (Tr-4), which belongs to the training set, and cytosar-U (Ts-14) are the only misclassified compounds. In other words, on the basis of their score plot positions these molecules appear less potent than expected. This behavior was also confirmed in the PLS models that predicted the compounds to be less toxic compared to what observed in vitro. Actually, it is likely that the in vitro values of these two are not reliable, as also reported by Pessina et al.,[18] who carried out an international blind trial to validate a model for estimate the maximum tolerated dose (MTD) of myelosuppressive xenobiotics using in vitro hematotoxicy data. In this trail, the calculated human MTDs for cytosar-U and gemcitabine using in vitro data were underestimated and the compounds predicted to be more toxic in comparison to what observed in the clinic.[18] Pessina et al.[18] suggested that differences in the levels of endogenous pyrimidines of GM-CFC cell cultures that can antagonize drug toxicity could not be accurately modeled in vitro. Since gemcitabine and cytosar-U are both pyrimidine analogues, the in vitro assay probably failed in this case to correctly evaluate the myelotoxicity of these compounds that resulted, erroneously, to be highly toxic.[18] Thus the predicted in silico hematotoxicity values in comparison to the in vitro data could better reflect the in vivo toxicity of these drugs.

Quantitative In Silico Myelotoxicity Model Developed Using 3D Volsurf+ Descriptors. The in vitro CFC-GEMM hematotoxicity data (Table 1) were correlated with the entire set of Volsurf+ descriptors (120) by PLS analysis. The PLS analysis resulted in a one-latent-variable (LV) model with an $r^2 = 0.62$; the cross-validation of the model yielded q^2 value of 0.46. Removing the outlier gemcitabine from the training set, the PLS analysis resulted in one LV model with an $r^2 = 0.79$ and a q^2 value of 0.64 (model A). After variable selection based on VIP[27,28,30] (variable importance in the projection) values the active variable decreased from 120 to 61. A new PLS multivariate data analysis was performed yielding a one-latent-variable model with an r^2 of 0.79 and a q^2 value of 0.72. Quality and robustness of the obtained model (model B) were tested by predicting the myelosuppressive activity of the test set previously defined as shown in Table 2. Confirming the PCA result, cytosar-U is predicted to be less potent than expected with more than two-log unit error. The model predicts the remaining 21 compounds with a RMSEP of 0.69 and a squared correlation coefficient value r_0^2 equal to 0.67. Three compounds, i.e., 6-mercatopurine, thioguanine and flavopiridol, are predicted less potent in comparison to their actual in vitro data and with more than one-log unit error. However, considering both the variability of the experimental data due to the use of different sources (i.e., umbilical cord blood[18,19] or bone marrow cells) and donors of human cells, and that the compounds Ts-13 to Ts-22 were tested using a different experimental protocol in comparison to the training set, the result in terms of predicted pIC_{50} can be considered almost accurate. The statistical parameters calculated for the training and test set are summarized in Table 5.

With the exception of gemcitabine and cytosar-U, the entire set is modeled without any other significant outlier behavior, although structurally different classes of compounds are present in the data set. Thus, this suggests a common mechanism of action and consequently common structural features required for induce myelotoxicity.

From a visual analysis of the PLS coefficients and VIP value profiles of the model (Figure 4 and 5), it was possible to detect the descriptors that have a greater importance in the model. Variables contributing positively toward myelosuppressive activity are depicted with positive bars and those contributing negatively with negative bars. In Figure 5, the descriptors are sorted in decreasing order of importance (i.e., VIP values) to explain myelosuppressive activity. The "pharmacophoric" descriptors (DRACAC, ACACAC, ACACDO, DRDRAC, DRACDO, DRDRDR, DRDRDO, and ACDODO; see Table 3) represented by the maximum area of the triangles (calculated over all possible conformers) formed by the combination of three different structural features (i.e., hydrophobic region, H-bond donor and acceptor groups) identified in each molecule, significantly contribute to the description of the variance of the toxicity data. Other important descriptors positively correlated with the myelosuppressive activity are mostly related to the size and shape of the molecules, such as molecular surface (S),

dx.doi.org/10.1021/ci1003834 |J. Chem. Inf. Model. 2011, 51, 434–445

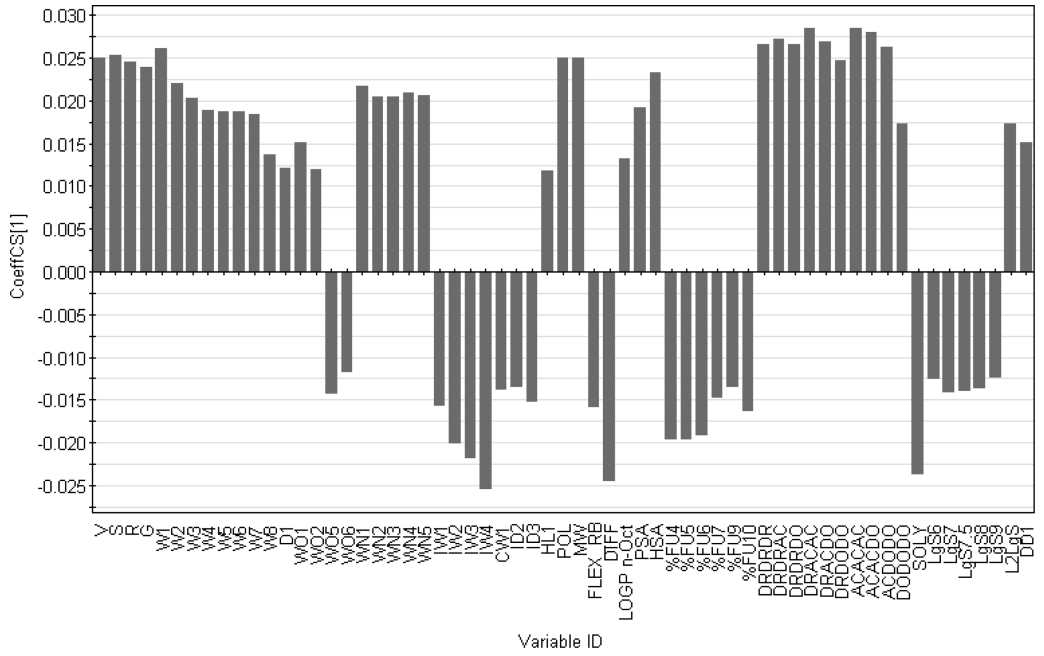


Figure 4. PLS coefficients profile of the model developed using 3D Volsurf+ descriptors (model B). The plot highlights the descriptors that are directly (positive values) or inversely (negative values) correlated to the myelotoxicity.

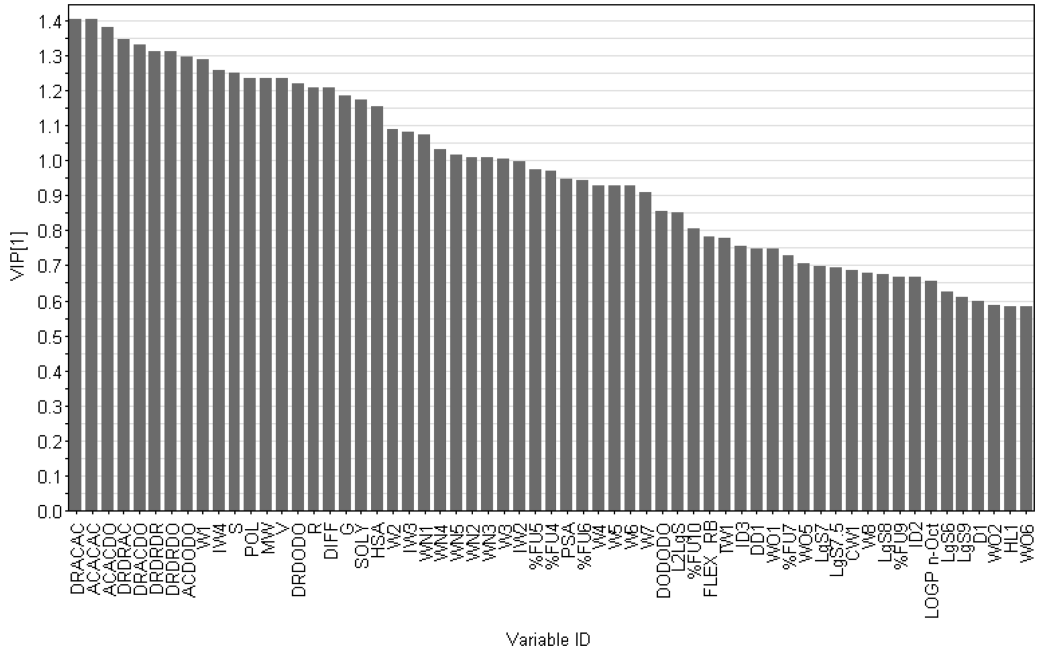


Figure 5. Variable importance on the projection (VIP) plot of the model developed using 3D Volsurf+ descriptors (model B). The descriptors are sorted in decreasing order of importance.

volume (V), rugosity (R), and globularity (G) (Table 3). The analysis of the PLS coefficients (Figure 4) also pointed out that descriptors related to presence of H-bond acceptor groups, wide polar and hydrophobic regions (Table 3, see WN1−WN6, W1−W7, PSA, and HSA descriptors) contribute positively to explain the myelosuppressive activity. In summary, myelotoxic compounds are in general bulky molecules characterized by wide polar interacting regions, as well as hydrophobic areas and presence of H-bond groups located at relatively high distance. In opposite, less toxic compounds seem characterized by localized polar interacting regions (see PLS coefficients and VIPs for IW1−IW4 in Figures 4 and 5, respectively), lower molecular weight, high diffusivity, high solubility, and high percent of uncharged species at physiological pH (see PLS coefficients and VIPs for DIFF, SOL and %FU descriptors in Figures 4 and 5, respectively).

Quantitative In Silico Myelotoxicity Model Developed Using 2D Descriptors. The selected set of 2D descriptors correlated by PLS analysis with the myelosuppressive data (Table 1)

dx.doi.org/10.1021/ci1003834 |J. Chem. Inf. Model. 2011, 51, 434–445

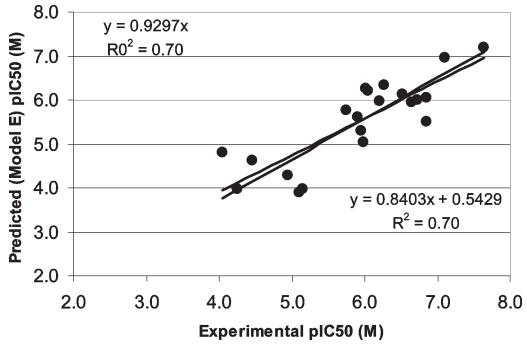


Figure 6. Experimental vs predicted (model E) pIC_{50} for 21 test compounds reported in Table 2.

yielded a model with one-latent-variable, a correlation coefficient r^2 of 0.62 and a q^2 value of 0.52. As for the model based on 3D descriptors, gemcitabine was an outlier and thus was removed from the final analysis. The resulting model (model C) consisted of one LV with an $r^2 = 0.72$ and a q^2 value of 0.63. The quality of the PLS model was then assessed predicting the myelosuppressive activity of the entire test set. Variable selection done on the basis of VIP values did not improve the predictive ability of the model. Also in this case the compound cytosar-U was predicted less toxic than expect with more than two-log unit error (Table 2). The statistical parameters calculated for the training and test set are summarized in Table 5. Although 2D descriptors have the advantage to be rapidly calculated and do not need conformation generation and selection, their use for modeling myelotoxicity penalized to some extent the goodness and robustness of the model that appears of lower quality in comparison to that developed using 3D descriptors (RMSEP = 0.82 and $r_0^2 = 0.58$, see Table 5). In particular, compounds Ts-1 to Ts-4, flavopiridol, and fludarabine, which are all kinase inhibitors except the latter molecule, were predicted with more than one-log unit error. However, the model interpretation was almost in good agreement and confirmed the findings previously described. The descriptors related to the size of the molecule, such as molecular weight, volume, surface area, SASA and SAVol, contribute to explain myelosuppressive activity. Other important descriptors correlated with myelotoxicity data are, number of H-bond acceptor descriptor, polar surface area and electrotopological E-states indices ES_Count_sCH3 and ES_Sum_sCH3, respectively. The electrotopological E-state indices encode information about both the topological environment and the electronic interaction of an atom resulting from all other atoms in the molecule. Consequently, their interpretation is not so straightforward. The coefficients and the high VIP values of E-states sp2 and sp3 oxygen bonded to non-hydrogen atoms (ES_Sum_dO, ES_Count_dO, ES_Sum_ssO, ES_Count_ssO) and hydroxyl group (ES_Sum_sOH, ES_Count_OH), confirm that the presence of hydrogen-bond groups may increase myelosuppressive activity.

Quantitative In Silico Myelotoxicity Model Developed Combining 2D and 3D-Volsurf+ Descriptors. The 2D descriptors calculated for the training set (after exclusion of gemcitabine) were merged with 119 descriptors (molecular weight was removed since already present in 2D descriptor set) produced by the Volsurf+ program, giving a total of 167 variables. The PLS analysis carried out on the resulting matrix yielded a one-latent-variable model with an $r^2 = 0.78$. The cross-validation of the model gave q^2 value of 0.65 (model D). On the basis of VIP values, a variable selection was applied to reduce the variable number. The resulting number of active variable decreased from 167 to 99. A new PLS multivariate data analysis was performed yielding a one-latent-variable model with an r^2 of 0.78 and a q^2 value of 0.72. The inclusion of both types of descriptors, 2D and 3D, improves only slightly the predictive power and the robustness of the model (model E) that was able to predict the test set (Table 2), after exclusion of cytosar-U, with a RMSEP of 0.67 and a r_0^2 of 0.70 (Figure 6). The statistical parameters calculated for the training and test set are reported in Table 5. The pharmacophoric 3D descriptors followed by size and shape descriptors as well as descriptors related to presence of H-bond acceptor groups result the most important variables to explain myelosuppressive activity.

Toxicophoric Hypothesis For Myelotoxic Drugs. To identify the toxicophore elements that characterize the highly myelotoxic drugs analyzed in this study, the entire compound data set was classified according to the in vitro pIC_{50} data as myelotoxic (pIC50 greater than 5.3) and less toxic compounds. The relationship between the two defined classes (Table 1) and the 620 GRIND not-scaled descriptors was determined using PLS-DA analysis. After filtering out the descriptors having no variability in the two compound classes, the number of variables decreased to 499. A variable selection using FFD factorial selection implemented in the Pentacle program did not contribute significantly to increase the performance of the final model; therefore, all the descriptors were retained in the analysis. Confirming the previous results, gemcitabine was an outlier and was removed from the final analysis. The model obtained, after cross-validation, had one significant LV and predicted the test set (Table 2) with an overall accuracy of 86%. Only three compounds were misclassified as low toxic molecules; cytosar-U, which resulted an outlier as in all previously described models, fludarabine and pyrazoloacridine.

From the analysis of the PLS-DA coefficients, the most important variables to explain myelosuppressive activity were identified (Figure 7, variables N1−N1−154, N1−TIP−596, N1−N1−159, O−N1−471, and DRY−N1−341) and an approximate toxicophore was outlined. Myelotoxic drugs are characterized by high values of the N1−N1 descriptors (N1−N1−154 and N1−N1−159 in Figure 7), which are related to the presence of favorable interacting regions placed 12.4 and 14.4 Å apart around H-bond acceptor groups. The descriptor of the N1−TIP cross-correlogram (N1−TIP−596 in Figure 7) represents the optimal distance (15.2−15.6 Å) that should separate one H-bond acceptor group from one edge of the molecule. This finding confirmed that the size of the molecules also play a role in the myelosuppression mechanism. Toxic compounds show a H-bond donor and H-acceptor groups at about 15 Å distance (O−N1−471 in Figure 7). Favorable interacting regions around a hydrophobic region and a H-bond acceptor group separated by a distance of 12.4−12.8 Å (DRY−N1−341 in Figure 7) also differentiate highly myelotoxic from less toxic compounds. A picture summarizing the relevant toxicophore features for the doxorubicin, as reference compound, is reported in Figure 8. With the exception of azathioprine, acyclovir and retrovir that, however, show only one of the identified toxicophore elements; i.e. a hydrophobic interacting region and a H-bond acceptor group at about 12.5 Å distance, it was interesting to note that the less toxic compounds lack the entire toxicophore moieties characteristic for highly myelosuppressive drugs.

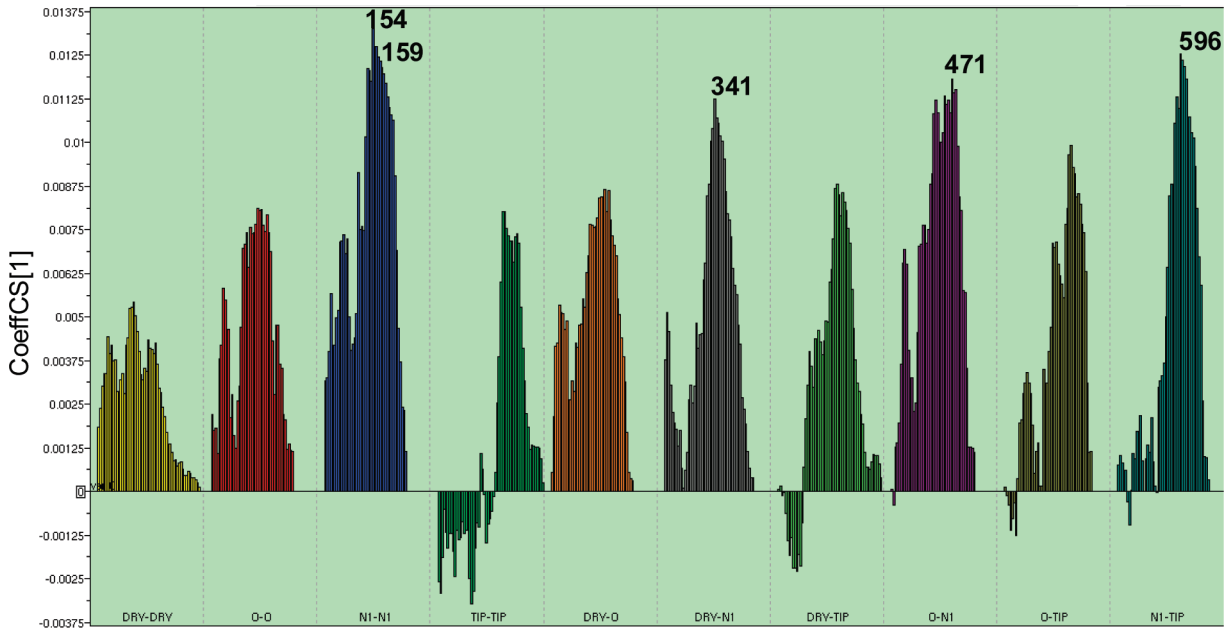


Figure 7. PLS-DA coefficients profile of the model developed using GRIND-toxicophore descriptors. Highly myelotoxic drugs are characterized by high values of the N1−N1−154, N1−TIP−596, N1−N1−159, O−N1−471, and DRY−N1−341 descriptors.

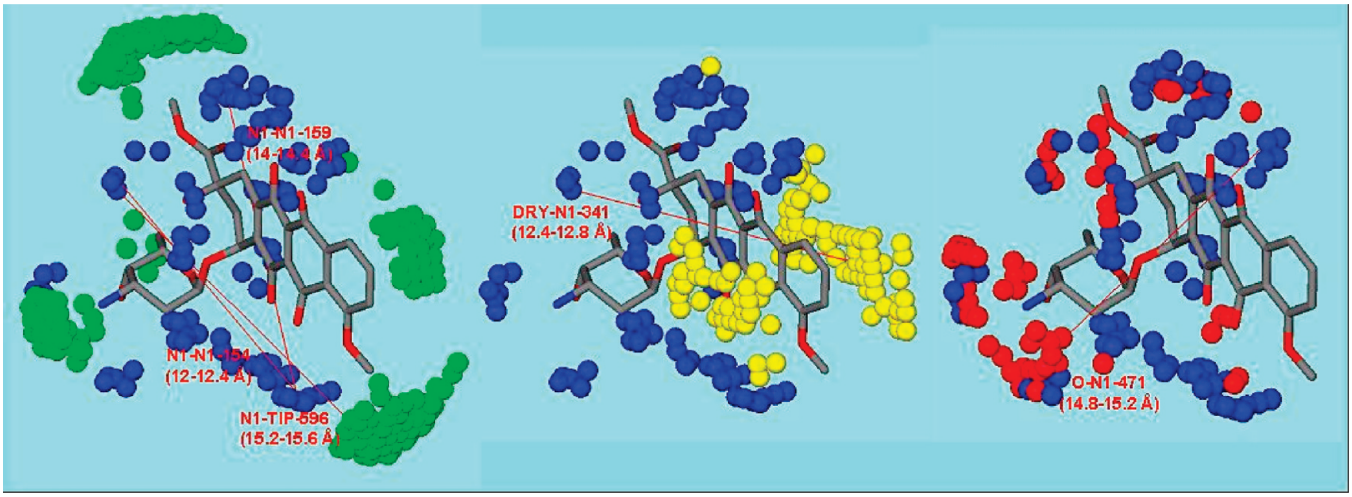


Figure 8. Toxicophore elements for highly myelotoxic molecules. The depicted molecule is the anticancer doxorubicin. Hydrogen atoms are not presented for clarity. The areas around the molecule are the Molecular Interaction Fields (MIF) computed with a donor hydrogen-bonding probe (N1; blue region), an acceptor hydrogen-bonding probe (O; red region), a hydrophobic probe (DRY; yellow region), and a steric probe (TIP; green region) probes, respectively. The N1, O, DRY, and TIP MIF generated around the molecule reflect the hydrogen-bonding acceptor, donor, hydrophobic and steric characteristics of the molecule, respectively. The red lines represent the distance in Å between couple of toxicophore elements.

CONCLUSION

The developed in silico models show an acceptable accuracy in predicting the potential in vitro myelotoxicity of compounds not included in the models. Furthermore, the broad drug-like properties of the training and test sets lend global validity to these models. Since the chemical structures, experimental data and modeling techniques are provided in the present paper, these virtual screens can be easily implemented and used for predicting the myelosuppression of drug-like compounds commonly investigated in drug discovery projects. Notably, the ChemGPS methodology allows the potential model's users to compare the structural characteristics of the tested molecules with those of the training set and therefore to estimate the reliability of the predicted values.

Moreover, the estimation of the myelosuppressive activity directly from the molecular structures does not require any experimental data and thus can be performed both for virtual molecules and when the amount of compound is not enough to be evaluate in vitro, as often happens in early drug discovery phase.

A qualitative GRIND-based model was also built to elucidate the toxicophore patterns that mainly differentiate the highly myelotoxic from less toxic molecules. As for the quantitative models, the robustness was confirmed by the evaluation of an external set

of 22 drugs. Although structurally diverse compounds were investigated and multiple mechanisms could induce myelotoxicity, a single "toxicophore" emerged from the analysis of the model. The identification of the most relevant functional groups and their spatial arrangement leading to myelosuppressive effects provides valuable information that would be helpful in the design of new candidate drugs.

Despite the relative small number of molecules investigated in this study, the structural diversity of the entire data set as well as the encouraging results obtained in prediction compensate for this initial limitation. However, the inclusion of additional experimental data, as well as the identification of specific mechanisms that cause myelotoxicity, which allows modeling each of them separately, could improve the predictive ability of the new generated models.

■ ASSOCIATED CONTENT

Supporting Information. PLS coefficients profile and VIP value plots for the QSTR models C and E, respectively. PLS-DA GRIND-toxicophore model: t1 vs t2 score plot of the training and projected (predicted) test sets.This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: patrizia.crivori@nervianoms.com. Telephone: +39 0331 581641.

Present Addresses

[§]Statistics and Programming, MIO Nerviano Medical Sciences Viale Pasteur 10, 20014 Nerviano, Italy.

■ REFERENCES

(1) Rich, I. N. High-throughput in vitro hemotoxicity testing and in vitro cross-platform comparative toxicity. *Expert. Drug Metab. Toxicol.* **2007**, *3*, 295–307.

(2) Rich, I. N. In vitro hematotoxicity testing in drug development: A review of past, present and future applications. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 100–109.

(3) Parchment, R. E; Huang, M.; Erickson-Miller, C. L. Roles for *in vitro* myelotoxicity tests in preclinical drug development and clinical trial planning. *Toxicol. Pathol.* **1993**, *21*, 241–250.

(4) Parchment, R. E.; Gordon, M.; Grieshaber, C. K.; Sessa, C.; Volpe, D.; Ghielmini, M. Predicting hematological toxicity (myelosuppression) of cytotoxic drug therapy from *in vitro* tests. *Ann. Oncol.* **1998**, *9*, 357–364.

(5) Parchment, R. E. Alternative testing systems for evaluating noncarcinogenic, hematologic toxicity. *Environ. Health Perspect.* **1998**, *106* (Suppl 2), 541–557.

(6) Bradley, T. R.; Metcalf, D. The growth of mouse bone marrow cells *in vitro*. *Aust. J. Exp. Biol. Med. Sci.* **1966**, *44*, 287–300.

(7) Pluznik, D. H.; Sachs, L. The induction of clones of normal mast cells by a substance from conditioned medium. *Exp. Cell Res.* **1966**, *43*, 553–563.

(8) Rich, I. N.; Hall, K. M. Validation and development of a predictive paradigm for hemotoxicology using a multifunctional bioluminescence colony-forming proliferation assay. *Toxicol. Sci.* **2005**, *87*, 427–441.

(9) Oprea, T. I.; Gottfries, J.; Sherbukhin, V.; Svensson, P.; Kuhler, T. C. Chemical information management in drug discovery: optimizing the computational and combinatorial chemistry interfaces. *J. Mol. Graphics Modell.* **2000**, *18*, 512–524.

(10) Oprea, T. I.; Gottfries, J. Chemography: The Art of Navigating in Chemical Space. *J. Comb. Chem.* **2001**, *3*, 157–166.

(11) Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. Molecular fields in quantitative structure-permeation relationships: The VolSurf approach. *J. Mol. Struct. (THEOCHEM)* **2000**, *503*, 17–30.

(12) Oprea, T. I.; Zamora, I.; Ungell, A.-L. Pharmacokinetically based mapping device for chemical space navigation. *J. Comb. Chem.* **2001**, *4*, 258–266.

(13) Hall, L. H.; Kier, L. B. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039–1045.

(14) Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A.; Xie, Q.; Tong, W. Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis* **2004**, *19*, 365–377.

(15) Arrigoni, C.; Crivori, P. Assessment of QT liabilities in drug development. *Cell Biol. Toxicol.* **2007**, *23*, 1–13.

(16) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): a novel class of alignment independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *17*, 3233–3243.

(17) Fontaine, F.; Pastor, M.; Sanz, F. Incorporating molecular shape into the alignment-free Grid Independent Descriptors. *J. Med. Chem.* **2004**, *11*, 2805–2815.

(18) Pessina, A.; Albella, B.; Bayo, M.; Bueren, J.; Brantom, P.; Casati, S.; Croera, C.; Gagliardi, G.; Foti, P.; Parchment, R.; Parent-Massin, D.; Schoeters, G.; Sibiril, Y.; Van Den Heuvel, R.; Gribaldo, L. Application of the CFU-GM Assay to Predict Acute Drug-Induced Neutropenia: An International Blind Trial to Validate a Prediction Model for the Maximum Tolerated Dose (MTD) of Myelosuppressive Xenobiotics. *Toxicol. Sci.* **2003**, *75*, 355–367.

(19) Pessina, A.; Albella, B.; Bueren, J.; Brantom, P.; Casati, S.; Gribaldo, L.; Croera, C.; Gagliardi, G.; Foti, P.; Parchment, R.; Parent-Massin, D.; Sibiril, Y.; Van Den Heuvel, R. Prevalidation of a model for predicting acute neutropenia by colony forming unit granulocyte/macrophage (CFU-GM) assay. *Toxicol. in Vitro* **2001**, *15*, 729–740.

(20) *CORINA*, version 3.46; Molecular Networks, GmbH, Computerchemie: Erlangen, Germany, 2008

(21) *Volsurf+*, version 1.0.4; Molecular Discovery Ltd.: London, U.K., 2010.

(22) Duran, A.; Martinez, G. C.; Pastor, M. Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in Molecular Interaction Fields. *J. Chem. Inf. Model.* **2008**, *9*, 1813–1823.

(23) *Pentacle*, version 1.0.4; Molecular Discovery Ltd.: London, U.K., 2010.

(24) *Pipeline Pilot*, version 7.5; Accelrys Inc.: San Diego, CA, U.S.A., 2009.

(25) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.

(26) *SIMCA-P+*, version 12.0.1; Umetrics AB.: Umeå, Sweden, 2009.

(27) Wold, S.; Johansson, E.; Cocchi, M. PLS-partial least-squares projections to latent structures. In *3D QSAR in Drug Design-Theory Methods and Applications*; Kubyni, H., Ed.; Escom Science Publishers: Leiden; 1993; pp 523−550.

(28) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. Multi- and megavariate data analysis. In *Principles and Applications*, Umetrics Academy, Umeå; 2001.

(29) Thummel, K. E.; Shen, D. D. Design and optimization of dosing regimens: pharmacokinetic data. In *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 10th ed.; Hardman, J. G., Limbird, L. E., Gilman, A. G., Eds.; Mc Graw-Hill: New York, 2001; pp 1917−2023.

(30) Chong, I.-G.; Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112.