

Review

Block relevance (BR) analysis and polarity descriptors in property-based drug design

Giuseppe Ermondi and Giulia Caron*

Molecular Biotechnology and Health Sciences Dept., Università degli Studi di Torino, via Quarello 15, 10135 Torino, Italy

*Corresponding Author: E-mail: giulia.caron@unito.it; Tel.: +39-011-6708337;

Received: April 06, 2018; Revised: May 23, 2018; Published: May 29, 2018

Abstract

Block Relevance (BR) analysis is a tool to interpret QSPR/PLS models which can provide the information content of any physicochemical determinant used in property-based drug discovery; its application for the characterization of experimental polarity descriptors is discussed.

Keywords

ADME; hydrogen bonding; IAM; QSPR; PLS; PSA; VIP

Property-based drug design: the need for the information content of physicochemical descriptors

The pool of physicochemical properties including size, lipophilicity, solubility, permeability, polarity, and hydrogen bond (HB) capacity defines the drug-like profile of potential candidates. Thus, during the lead optimization process, medicinal chemists control the absorption, distribution, metabolism and excretion (ADME) fate of a candidate by monitoring the variation in the physicochemical profile following chemical modifications [1].

Any molecular property can be quantified by a few physicochemical descriptors. For instance, lipophilicity can be described by the log of the distribution coefficient in the octanol/water system (i.e. $\log D_{\text{oct}}$) but also by the log of the distribution coefficient in the toluene/water system (i.e. $\log D_{\text{tol}}$). $\log D_{\text{oct}}$ and $\log D_{\text{tol}}$ encode diverse information [2] and thus are not interchangeable. Therefore, it could be relevant for a medicinal chemist to know how a structural change impacts both $\log D_{\text{oct}}$ and $\log D_{\text{tol}}$. More generally, to select the most relevant set of descriptors for any drug discovery program, it is crucial to know the information content of any descriptor. Abraham's solvation equations (their review is beyond the aim of this paper) are the most known Quantitative Structure-Property Relationships (QSPRs) tools that provide the information content of any molecular property [3,4]. Briefly, by measuring a linear free-energy related solute property for a set of at least 25 compounds with known 2D molecular descriptors (E, S, A, B and V, optimised for small mono- or bifunctional organic molecules in unionized form) the solvation equation can be built using a multiple linear regression (MLR).

To provide an alternative tool to Abraham's equations, we have designed and implemented a

chemoinformatic strategy named Block Relevance (BR) analysis [5] which affords an interpretation “at a glance” of QSPRs models based on a selected pool of descriptors and a Partial Least Square (PLS) algorithm. To fully understand BR analysis and thus its relevance in property-based drug design, we have briefly reviewed [6,7] some basic concepts of QSPR.

QSPR for modelling physicochemical properties

Quantitative Structure-Property Relationships can be formulated using the eq. 1 [8]:

$$Property = f(Descriptors) \quad (1)$$

Where a *Property* is in principle any physicochemical property, the *Descriptors* are numerical values that are related to the structure of the chemical compounds and, *f* is a tailored function able to quantitatively correlate the *Property* and the *Descriptors*. In a context of properties-based drug design, the endpoint is modelling the pool of physicochemical properties which define drug-like profile of the candidate compounds, e.g. lipophilicity, solubility, permeability, polarity and hydrogen bond (HB) capacity. In this paper we will use *Property* to indicate what we want to model i.e. experimental descriptors like $\log P_{oct}$ which quantifies the molecular property of lipophilicity.

Although a plethora of *Descriptors* has been reported in the literature [9] there is still no agreement on what constitutes the “best” pool for molecular design [10] since it depends on the property that is under study. The subset of VolSurf+ (VS+) descriptors (v. 1.1.2, www.moldiscovery.com) based on 3D molecular fields (MIFs, see below) are a reasonable choice for modelling physicochemical properties. Shortly, a 3D MIF is calculated using the GRID force field [11,12] which may be viewed as a 3D matrix, with attractive and repulsive energy values between a chemical probe and a target molecule (e.g. a drug candidate). By using four probes and various cut-off values of energy, the capacity of the drug to form intermolecular interactions can be quantified and converted into molecular descriptors. To model physicochemical properties five blocks of VS+ descriptors corresponding to five types of intermolecular interactions are essential: 1) the DRY block: the DRY probe is used to calculate descriptors related to the hydrophobic interaction taking into account both the entropic and enthalpic effects [13]; 2) the OH2 block: the OH2 probe mimics water interaction with the target surface and thus is used to calculate descriptors related to hydrophilic properties of the target; 3) the NH block: the hydrogen bond acceptor (HBA) properties of the target are evaluated using a hydrogen bond donor (NH) probe 4) the O block: the hydrogen bond donor (HBD) properties of the target are evaluated using an hydrogen bond acceptor (O) probe, 5) the Others block: additional molecular descriptors are obtained using a combination of information derived from different probes to represent the unbalance of hydrophilic and hydrophobic regions on the surface target. Finally, in line with the evidence that molecular dimensions are crucial to characterize drug-like candidates [14], a sixth block of size and shape descriptors (the Size block) cannot be neglected. Overall the six blocks of VS+ descriptors listed above define the capacity of the drug to form intermolecular interactions and thus can be reasonably used in QSPR for modelling physicochemical properties.

In some cases, the properties are related to chemical descriptors through physicochemical theories, (e.g. absorbance vs concentration, Lambert-Beer equation); more often, properties and descriptors relations can be found through a statistical treatment. When a statistical approach is required, a mechanistic interpretation of the validated model is not always straightforward since the presence of a correlation between two variables does not imply causation. Therefore, to obtain a mechanistic interpretation of the model (this is an essential step for property-based drug discovery purposes), it is necessary to weight the contribution of each descriptor to the final model. Many algorithms could be used to correlate properties

with descriptors, showing different capacities to accurately describe the contribution of the descriptors to the model [15]. Multilinear regression (MLR) is the simplest tool to obtain and to interpret models but its application is appropriate only with a few models and not intercorrelated descriptors are needed. The Partial Least Squares (PLS) method data analysis is also widely used in medicinal chemistry. PLS manages a large number of potential intercorrelated descriptors-by projecting the data into a low dimensional and not intercorrelated “latent variable” space. Components in PLS are constructed to maximize the covariance between the dependent variable y and the original independent variables x [8]. To extract the contribution of any descriptor from PLS models, the analysis of VIPs and coefficients plots is needed (see below). Recently, nonlinear approaches, such as artificial neural networks (ANN) or support vector machines (SVM), have attracted the attention of many practitioners. These methods can “catch” hidden nonlinearities between properties and descriptors providing better predictors than the linear models. However, these nonlinear methods are not as easy to interpret as the MLR or PLS models [8], since they are by definition “black boxes” preventing any interpretation of underlying mechanisms.

Block Relevance (BR) analysis as a tool to interpret QSPR/PLS models

As mentioned above PLS is a widely used algorithm to generate QSPR models which, however, are often not easy to interpret. It is described below how BR analysis can overcome this limit.

Theory

Variable Importance in Projection (VIP) plots are often used to find the relative contribution of each descriptor to the final PLS model. The VIP for the descriptor j , VIP_j is defined according to Wold et al.[16]:

$$VIP_j = \left\{ \frac{\sum_{f=1}^F w_{jf}^2 \cdot SSY_{f \cdot J}}{SSY_{total} \cdot F} \right\}^{\frac{1}{2}} \quad (2)$$

where w_{jf} is the weight value for variable j component f , SSY_f is the sum of squares of explained variance for the f^{th} component and J the number of variables. SSY_{total} is the total sum of squares explained by the dependent variable, and F is the total number of components. The weights in a PLS model reflect the covariance between the independent and dependent variables and the inclusion of the weights is what allows VIP to reflect not only how well the dependent variable is described but also how important that information is for the model of the independent variables. Note that the sum of squared VIP values is equal to the number of descriptors (J):

$$\sum_{j=1}^F VIP_{i=1}^2 = J \quad (3)$$

To make easier PLS interpretations we assume that descriptors, if conveniently chosen, can be grouped in blocks. The Block Relevance (BR) of each block of descriptors was defined as the ratio of the sum of the squared VIP values of a given block of descriptors to the number of those descriptors.

$$BR_i = \frac{\sum_{j=1}^{N_i} VIP_j^2}{N_i} \quad (4)$$

where i is the number of blocks (6), N is the number of descriptors for any block, VIP_j is the value of VIP for each predictor fitting the PLS model. BR shows the relevance of a certain block of descriptors in the model; the higher the value of BR, the more important is that block.

Depending on the sign of the PLS coefficient, BR was broken down into BR (+) and BR (–), Eq. (5).

$$BR_i = BR_i(+)+ + BR_i(-) \quad (5)$$

The current BR analysis version

According to its definition, BR analysis requires a supervised grouping of ad hoc selected descriptors. In the current version we implement 82 of the original 128 VS+ descriptors since, as discussed above, they define the capacity of the drug to participate in intermolecular interactions. The 82 descriptors are grouped in six blocks (Fig. 1a) and each block contains information about the propensity of the compound to be involved in a specific intermolecular interaction, e.g. the OH2 block reveals the propensity of the molecule to interact with water.

BR analysis workflow is shown in Fig. 1b. In the first step, the dataset is set-up. This consists in collecting physicochemical data (e.g. $\log D_{\text{Oct}}$) and obtaining the corresponding compounds structures with standard procedures. Secondly, the dataset is submitted to VS+ to calculate the 82 descriptors and to obtain and validate the PLS models (internal and external validations are always performed as reported in the literature [17]). Finally, VIPs and coefficients values are exported to Excel, elaborated and submitted to BR analysis which produces an easy-to-interpret graphical output.

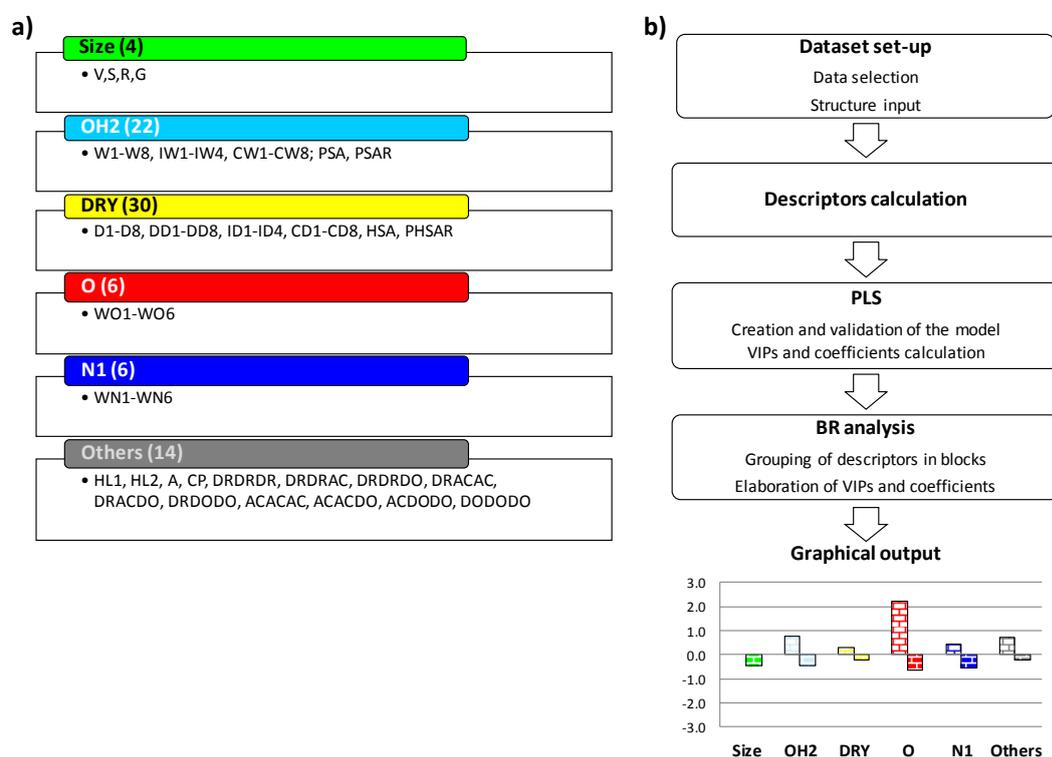


Figure 1. (a) Selected VolSurf+ descriptors organized in the six blocks as implemented in the current version of BR analysis (see text for more details) **(b)** BR analysis workflow.

Our recent modelling study of Immobilized Artificial Membrane (IAM) chromatography [2] will be used to illustrate details of the BR approach. In this case, BR analysis was applied to extract the relative contribution of the intermolecular forces governing the IAM.PC.DD2 $\log K_w^{\text{IAM}}$ parameter from PLS models.

The model was built using the whole dataset of compounds used in [2]: ($N^{\circ}(\text{training}) = 189$, $N^{\circ}(\text{test}) = 64$, $R^2 = 0.82$, $Q^2 = 0.71$). In Fig. 2, the VIPs are shown (Fig. 2a) and the coefficients plots (Fig. 2b) which should be combined to interpret the PLS model. In fact, the VIPs plot reveals the relevance of a descriptor in the model, whereas the coefficients plot establishes if the increase of variables corresponds to an increase

(positive weights) or a decrease (negative weights) of the property. Using Fig. 2a and Fig. 2b to interpret the PLS model is far from being user-friendly.

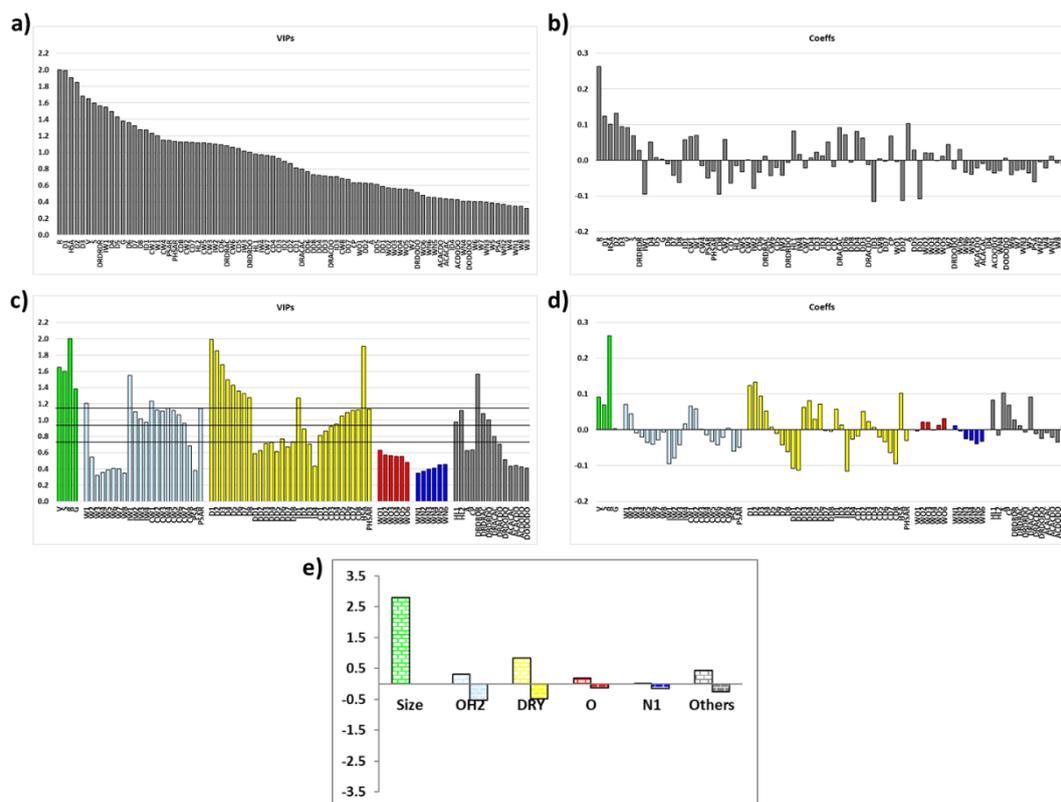


Figure 2. Log K_w^{IAM} model [2]: **(a)** the PLS VIP plot (standard output), **(b)** the PLS coefficients plot (standard output), **(c)** the PLS VIP plot (output coloured by blocks), the lines correspond to three reasonable threshold values (1.21, 1.00, 0.83), **(d)** the PLS coefficients plot (output coloured by blocks) **(e)** BR graphical output: the Size block (green) is related to the influence of size and shape; the Water block (cyan) reflects the role of the hydrophilicity whereas the DRY block (yellow) the role of hydrophobicity; red and blue blocks are related to the solutes' HBD and HBA properties, respectively; and, finally, the Others block (grey) is related to the unbalance of hydrophilic and hydrophobic regions.

The two plots, Fig. 2c (VIPs) and 2d (coefficients) have been represented by using the colour codes of BR blocks. Fig. 2c and 2d are clearer than Fig. 2a and 2b but the identification of the most relevant VIPs remains a crucial step for PLS interpretation since several threshold criteria have been reported in the literature [18]. A low threshold could result in the selection of too many descriptors and in a problematic interpretation, whereas a high threshold could result in the loss of important descriptors and in an inaccurate interpretation. For instance, in Fig. 2c three reasonable threshold values (1.21, 1.00 and 0.83) are reported which allows the identification of 18, 36 and 43 descriptors, respectively. If the 18 descriptors obtained with the largest threshold value are considered, it can be verified that most (but not all) of them belong to the Size and DRY blocks. Size descriptors coefficients are all positive (Fig. 2d), thus the larger the compound, the larger $\log K_w^{IAM}$. The behavior of hydrophobic descriptors is less clear since strong hydrophobic interactions (D5-D8) increase $\log K_w^{IAM}$ values, whereas weak interactions (D1-D4) are not favourable (Fig. 2d). Overall, the VIPs and coefficients plot together do not allow a clear and unique PLS model interpretation.

BR analysis graphical output for the same BR model is shown in Fig. 2e. For a correct interpretation of the BR analysis graphical results the following rules are applied: a) blocks with positive weighting (e.g., the green block) show how much the property increases $\log K_w^{IAM}$, whereas those with negative weighting indicate how much the property decreases $\log K_w^{IAM}$; b) according to our experience, block significance was

set at 0.2 and thus the blue and red blocks in Fig. 2e are not significant and c) a block with comparable positive and negative contributions (e.g. OH2 and Others in Fig. 2e), indicates that it is poorly relevant in the description of the investigated property. Overall, BR analysis shows that $\log K_w^{IAM}$ is essentially related to the dimensions of the molecules.

Table 1. Molecular properties analysed with the BR Analysis

Descriptor type	Descriptor	Reference
Biomimetic chromatographic index	$\log K_w^{IAM}$	[2]
Biomimetic chromatographic index	$\log K$ mucin	[19]
Biomimetic combined chromatographic index	$\Delta \log K_w^{IAM}$	[2]
Biophysical method	$\text{Log } K_D$ (SPR)	[20]
Chromatographic index	$\log k'$ C18	[5]
Chromatographic index	$\log k'$ Ph	[5]
Chromatographic index	$\log k'$ CN-RP	[5]
Chromatographic index	$\log k'$ OH-RP	[5]
Chromatographic index	$\log k'$ CN-RP	[5]
Chromatographic index	$\log k'$ OH-RP	[5]
Chromatographic index	$\log k'$ NH2-NP	[5]
Chromatographic index	$\log k'$ HILIC	[5]
Chromatographic index	$\log k'$ EPSA	[21]
Chromatographic index	$\log k'$ 80 PLRP-S	[22]
Chromatographic index	ElogD	[22]
Combined partition coefficient	$\Delta \log P_{\text{oct-tol}}$	[6]
Computed descriptor	TPSA	[21]
Computed distribution coefficient	Calc $\log D_{\text{oct}}$	[23]
Computed partition coefficient	Calc $\log P_{\text{tol}}$	[22]
Distribution coefficient	$\log D_{\text{lip}}$	[20]
Distribution coefficient	$\log D_{\text{oct}}^{7.4}$	[2]
Partition coefficient	$\log P_{\text{oct}}$	[5]
Partition coefficient	$\log P_{\text{oct}}$	[7]
Partition coefficient	ElogP	[7]
Partition coefficient	$\log P_{\text{oct}}$	[6]
Partition coefficient	$\log P_{\text{tol}}$	[6]
Partition coefficient	$\log P_{\text{oct}}$	[2]
Permeability rate	$\log P_{\text{app}}$ (RRCK)	[21]
Permeability rate	$\log P_{\text{app}}$ (Caco-2)	[21]
Permeability rate	$\log P_{\text{app}}$ (6.5/7.4)	[23]
Permeability rate	$\log P_{\text{app}}$ (7.4/7.4)	[23]

As expected, the VIPs/coefficients plots and the BR analysis provide a similar picture, but the BR analysis immediately focuses on the main features that drive the model. It is noteworthy to underline that the two approaches can be combined; using the BR analysis to obtain a general idea of the main interaction governing the property under study and then the VIPs/coefficients plots for a more detailed analysis of the results.

Up to now BR analysis has been applied to various properties related to the drug discovery process (Table 1) to deconvolute the different balance of intermolecular forces governing the properties under investigation. Overall, we verified the interchangeability of descriptors obtained with different methods and claimed, or not to provide the same information. This was done for lipophilicity, permeability and polarity descriptors. The next section describes application of BR analysis to the identification of experimental polarity descriptors.

BR analysis and polarity descriptors

Polarity is a molecular property of remarkable interest in the characterization of the propensity of compounds to form an intramolecular hydrogen bond (IMHB) [24] and thus in the prediction of permeability [25]. It is often estimated using the polar surface area (PSA) which can be calculated either using the 2D structure, Topological Polar Surface Area (TPSA), or the 3D structure, the Molecular Polar Surface Area (MPSA). Both descriptors have severe limitations since a) HBD and HBA strength and HB directionality are not considered b) HBA and HBD contributions are not distinguished and c) some atoms with partial charge different from zero are not considered to be polar [26]. Moreover, as discussed in a previous paper, increasing the size of the molecules, the PSA takes into account steric effects that TPSA does not reproduce [19]. These limitations suggest that measuring rather than predicting polarity could improve property-based drug discovery.

To verify whether an experimental physicochemical descriptor is a clean polarity determinant, we can apply BR analysis and check if the considered descriptor is mainly driven by the three blocks related to polarity (light blue, red and blue blocks, see Fig. 3a for a concise colour code scheme). Below we describe how BR analysis provides insight into three descriptors that in principle can be considered polarity descriptors.

$\Delta \log P_{\text{oct-tol}}$ is the difference between $\log P_{\text{oct}}$ (the logarithm of the partition coefficient P in the octanol/water system) and $\log P_{\text{tol}}$ (the logarithm of the partition coefficient P in the toluene/water system). BR analysis (Fig. 3b) indicates that $\Delta \log P_{\text{oct-tol}}$ is a clean descriptor of exposed HBD properties [6]. Notably, the solutes' HBA properties do not significantly contribute to $\Delta \log P_{\text{oct-tol}}$ which therefore cannot be strictly considered a polarity descriptor.

EPSA is an exposed polarity measurement which assesses polarity by retention time using controlled Supercritical Fluid Chromatography (SFC) conditions [24,27,28]. BR analysis was applied to understand the balance of intermolecular forces governing EPSA (Fig. 3c). Results showed that EPSA is governed by the solutes' HBD (red, positive) and HBA (blue, negative). Because HBD and HBA blocks have the opposite sign, BR analysis gives evidence that the presence of HBA groups could be considered as an interference and therefore the HBD properties of the molecule can be underestimated by EPSA when the structure includes more HBA groups. [21] These findings suggest that also EPSA cannot be strictly considered a polarity descriptor.

According to Grumetto et al. [29] $\Delta \log K_w^{\text{IAM}}$ is the difference between the logarithm of the experimental chromatographic retention factor ($\log K_w^{\text{IAM}}$) and the value here named $\text{clog } K_w^{\text{IAM}}$ calculated using eq. 7.

$$\text{clog } K_w^{\text{IAM}} = 0.92 * \log P_{\text{oct}}^N - 1.03 \quad (7)$$

BR analysis showed that $\Delta \log K_w^{\text{IAM}}$ is a clean polarity descriptor since it is due to both HBD and HBA properties (Fig. 3d). [2]

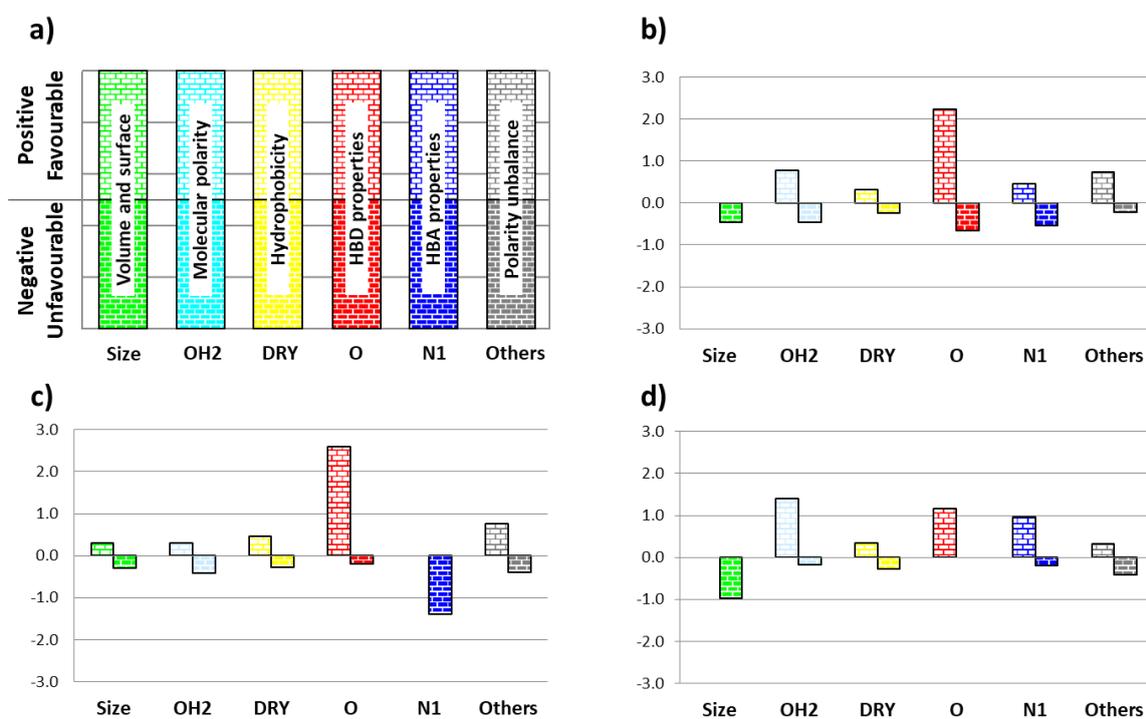


Figure 3. (a) Blocks color codes and significance to help BR graphical output interpretation. BR analysis graphical output for (b) $\Delta \log P_{\text{oct-tol}}$, (c) EPSA [19], (d) $\Delta \log K_w^{\text{IAM}}$ [2]

Since the determination of experimental polarity is crucial for large and flexible structures, $\Delta \log P_{\text{oct-tol}}$, EPSA and $\Delta \log K_w^{\text{IAM}}$ data was collected for indinavir and nelfinavir (structures and data in Fig. 4). In a previous study it was shown that nelfinavir has a larger propensity to form IMHBs than indinavir [30] and thus it was expected that $\Delta \log P_{\text{oct-tol}}$, EPSA and $\Delta \log K_w^{\text{IAM}}$ of nelfinavir are smaller than the correspondent values determined for indinavir. This was verified for $\Delta \log P_{\text{oct-tol}}$ and $\Delta \log K_w^{\text{IAM}}$ but not for EPSA, which fails in ranking the polarity of the two protease inhibitors; probably because of the high and different number of HBA groups present in the structures.

	Indinavir	Nelfinavir
MW	614	568
HBA	7	5
HBD	4	4
TPSA	118	101
Rotatable bonds	12	10
pK _a	3.87 (B); 5.68 (B)	5.96 (B); 12.97 (A)
log P _{oct}	2.85	4.02
log P _{tol}	0.86	3.49
$\Delta \log P_{\text{oct-tol}}$	1.99	0.53
EPSA	81	101
$\Delta \log K_w^{\text{IAM}}$	0.99	0.15

Figure 4. Indinavir and Nelfinavir: comparison of their physicochemical profile (in bold the polarity descriptors discussed in the text)

Conclusions

The design of drug candidates with the right balance of physicochemical properties (property-based drug design) calls for a tool like the Block Relevance (BR) analysis which can interpret QSPR/PLS models and provide an easy-to-read map of the information content of any experimental molecular descriptor. Up to now BR analysis has been applied to about 30 molecular descriptors mostly related to lipophilicity, polarity and permeability and some of these applications were the outcome of a collaboration with scientists in Pfizer (Groton, CT, USA). In this publication, after reviewing some basic concepts of QSPR and the BR analysis methodology, we have shown how BR analysis can identify the different features of three polarity descriptors ($\Delta\log P_{\text{oct-tol}}$, EPSA and $\Delta\log K_{\text{W}}^{\text{IAM}}$) and thus help to define their applicability domain. In the near future, we plan to apply BR analysis to molecular determinants obtained through biomimetic chromatography and biophysical methods for ADME-related protein-ligand interactions. To further improve the potential of BR analysis, the implementation of new sets of descriptors are in course of study in our laboratories.

Acknowledgements: We thank Gilles Goetz (Pfizer, Groton, CT, USA) for providing EPSA values for indinavir and nelfinavir.

References

- [1] K. Valko, Physicochemical and biomimetic properties in drug discovery, John Wiley & Sons, Hoboken, N.J., USA, 2014.
- [2] G. Ermondi, M. Vallaro, G. Caron, Learning how to use IAM chromatography for predicting permeability. *Eur. J. Pharm. Sci.* **114** (2018) 385–390.
- [3] M.H. Abraham, Hydrogen bonding. 31. Construction of a scale of solute effective or summation hydrogen-bond basicity. *J. Phys. Org. Chem.* **6** (1993) 660–684.
- [4] M.H. Abraham, H.S. Chadha, R. Leitao, R.C. Mitchell, W.J. Lambert, R. Kaliszan, A. Nasal, P. Haber, Determination of solute lipophilicity, as $\log P(\text{octanol})$ and $\log P(\text{alkane})$ using poly(styrene–divinylbenzene) and immobilised artificial membrane stationary phases in reversed-phase high-performance liquid chromatography. *J. Chromatogr. A* **766** (1997) 35–47.
- [5] G. Ermondi, G. Caron, Molecular interaction fields based descriptors to interpret and compare chromatographic indexes. *J. Chromatogr. A* **1252** (2012) 84–89.
- [6] G. Ermondi, A. Visconti, R. Esposito, G. Caron, The Block Relevance (BR) analysis supports the dominating effect of solutes hydrogen bond acidity on $\Delta\log P(\text{oct-tol})$. *Eur. J. Pharm. Sci.* **53** (2014) 50–54.
- [7] G. Caron, M. Vallaro, G. Ermondi, The Block Relevance (BR) analysis to aid medicinal chemists to determine and interpret lipophilicity. *Med. Chem. Commun.* **4** (2013) 1376–1381.
- [8] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative correlation of physical and chemical properties with chemical structure: Utility for prediction, *Chem. Rev.* **110** (2010) 5714–5789.
- [9] R. Todeschini, V. Consonni, Handbook of molecular descriptors, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2008.
- [10] D.J. Livingstone, The Characterization of Chemical Structures Using Molecular Properties. A Survey. *J. Chem. Inf. Comput. Sci.* **40** (2000) 195–209.
- [11] G. Cruciani, P. Crivori, P.A. Carrupt, B. Testa, Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. *J. Mol. Struct. (Theochem)* **503** (2000) 17–30.
- [12] P. Crivori, G. Cruciani, P.-A. Carrupt, B. Testa, Predicting Blood–Brain Barrier Permeation from Three-Dimensional Molecular Structure, *J. Med. Chem.* **110** (2000) 2204–2216.

- [13] R.C. Wade, Calculation and application of Molecular Interaction Fields, in: G. Cruciani (Ed.), *Mol. Interact. Fields*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2006: pp. 27–42.
- [14] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23** (1997) 3–26.
- [15] L. Michielan, S. Moro, Pharmaceutical Perspectives of Nonlinear QSAR Strategies. *J. Chem. Inf. Model.* **50** (2010) 961–978.
- [16] S. Wold, E. Johansson, M. Cocchi, PLS: Partial Least Squares Projections to Latent Structures, in: H. Kubinyi (Ed.), *3D QSAR Drug Des.*, Kluwer Academic Publisher, Dordrecht (The Netherlands), 2000: pp. 523–550.
- [17] A. Tropsha, P. Gramatica, V. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **22** (2003) 69–77.
- [18] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **118** (2012) 62–69.
- [19] G. Caron, S. Visentin, C. Pontremoli, G. Ermondi, Profile of the intermolecular forces governing the interaction of drugs with mucin. *Int. J. Pharm.* **488** (2015) 67–69.
- [20] G. Ermondi, G. Caron, Do surface-based match solution-based techniques? the case of drug-liposome interaction. *Int. J. Pharm.* **508** (2016) 123–127.
- [21] G.H. Goetz, M. Shalaeva, G. Caron, G. Ermondi, L. Philippe, Relationship between passive permeability and molecular polarity using block relevance analysis. *Mol. Pharm.* **14** (2017) 386–395.
- [22] G. Caron, M. Vallaro, G. Ermondi, G.H. Goetz, Y.A. Abramov, L. Philippe, M. Shalaeva, A Fast Chromatographic Method for Estimating Lipophilicity and Ionization in Nonpolar Membrane-Like Environment. *Mol. Pharm.* **13** (2016) 1100–1110.
- [23] T. Potter, G. Ermondi, G. Newbury, G. Caron, Relating Caco-2 permeability to molecular properties using block relevance analysis. *Med. Chem. Commun.* **6** (2015) 626–629.
- [24] G.H. Goetz, W. Farrell, M. Shalaeva, S. Sciabola, D. Anderson, J. Yan, L. Philippe, M.J. Shapiro, High throughput method for the indirect detection of intramolecular hydrogen bonding. *J. Med. Chem.* **57** (2014) 2920–2929.
- [25] C.R.W. Guimarães, A.M. Mathiowetz, M. Shalaeva, G. Goetz, S. Liras, Use of 3D Properties to Characterize Beyond Rule-of-5 Property Space for Passive Permeation. *J. Chem. Inf. Model.* **52** (2012) 882–890.
- [26] G. Caron, G. Ermondi, Molecular descriptors for polarity: the need of going beyond PSA. *Fut. Med. Chem.* **8** (2016), 2013–2016.
- [27] G.H. Goetz, L. Philippe, M.J. Shapiro, EPSA: A Novel Supercritical Fluid Chromatography Technique Enabling the Design of Permeable Cyclic Peptides. *ACS Med. Chem. Lett.* **5** (2014) 1167–1172.
- [28] S. Sciabola, G.H. Goetz, G. Bai, B.N. Rogers, D.L. Gray, A. Duplantier, K.R. Fonseca, M.A. Vanase-Frawley, N.M. Kablaoui, Systematic N-methylation of oxytocin: Impact on pharmacology and intramolecular hydrogen bonding network. *Bioorganic Med. Chem.* **24** (2016) 3513–3520.
- [29] L. Grumetto, G. Russo, F. Barbato, Polar interactions drug/phospholipids estimated by IAM-HPLC vs cultured cell line passage data: Their relationships and comparison of their effectiveness in predicting drug human intestinal absorption. *Int. J. Pharm.* **500** (2016) 275–290.
- [30] G. Caron, M. Vallaro, G. Ermondi, Log P as a tool in intramolecular hydrogen bond considerations. *Drug Discov. Today: Technol.* (2018) doi.org/10.1016/j.ddtec.2018.03.001.