

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Molecular Interaction Fields vs. Quantum-Mechanical-based descriptors in the modelling of lipophilicity of platinum(IV) complexes**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/128300> since

*Published version:*

DOI:10.1039/C2DT32360E

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



# Molecular Interaction Fields vs. Quantum-Mechanical-based descriptors in the modelling of lipophilicity of platinum(IV) complexes

Giuseppe Ermondi,<sup>a</sup> Giulia Caron,<sup>a</sup> Mauro Ravera,<sup>b</sup> Elisabetta Gabano,<sup>b</sup> Sabrina Bianco,<sup>b</sup> James A. Platts,<sup>c</sup> Domenico Osella<sup>\*b</sup>

5 We report QSAR calculations using VolSurf descriptors to model the lipophilicity of 53 Pt(IV) complexes with a diverse range of axial and equatorial ligands. Lipophilicity is measured using an efficient HPLC method. Previous models based on a subset of this data are shown to be inadequate, due to incompatibility of whole molecule descriptors between axial carboxylate and hydroxo ligands. Instead, the interaction surfaces of complexes with various probes are used as independent descriptors. Partial  
10 least squares modelling using three latent variables results in an accurate ( $R^2 = 0.92$ ) and robust model ( $Q^2 = 0.87$ ) of lipophilicity, that moreover highlights the importance of size and hydrophobicity terms and the modest relevance of hydrogen bonding.

## Introduction

It is well known that similar molecules, with just minor structural variations, can exhibit dramatically different biological behaviour. Consequently, the development of the quantitative structure-activity relationships (QSAR), *i.e.* mathematical relationships linking chemical structure and pharmacological activity, help the chemists to design out negative properties and incorporate positive attributes to the molecules under investigation. The final goal is to design potentially active compounds prior to their synthesis, and to discharge drug candidates with unfavourable pharmacokinetic/pharmacodynamic profiles, thus limiting the amount of *in vitro* and *in vivo* experiments required in drug discovery and optimisation.

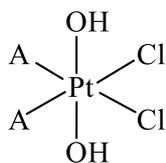
Considering the complexity of the ADME (Absorption, Distribution, Metabolism and Excretion) prediction of drug candidates, today the original dependent variable “A” in QSAR, representing the desired therapeutic effect, may be replaced by other, generic properties “P” (QSPR, *i.e.* quantitative structure-property relationship) in order to limit the extension of calculations.<sup>1</sup> In this context lipophilicity is a key feature because it is directly related to the ability of a molecule to cross passively cell membranes.<sup>2,3</sup> The partition coefficient  $\log P_{o/w}$  is widely used to represent molecular lipophilicity because it measures the differential solubility of a compound between n-octanol (a model of the lipid bilayer of a cell membrane) and water (the solvent in and out of cells). The importance of this parameter is confirmed by the fact that  $\log P_{o/w}$  is one of the properties identified by Lipinski in the “Rule of 5” for drug-like molecules.<sup>4</sup> Lipophilicity of a molecule is generally thought to arise from its size and polarity. Thus, a large, apolar molecule will have a tendency to partition into the organic phase, where formation of cavities is easier, while a more polar molecule will tend to partition into water, the more polar and hydrogen bonding solvent.<sup>5</sup>

Although a lipophilicity experiment, by using the traditional

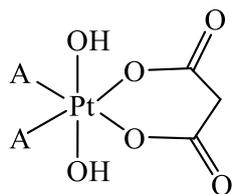
shake-flask method seems at a first glance simple and easy, it is not a trivial matter because data can vary significantly with experimental conditions. The main difficulty in the shake-flask method comes from the significant errors affecting the final  $\log P_{o/w}$  in the case of extreme values (*i.e.* very polar or very lipophilic compounds), that comes from an unbalanced partition in favour of the aqueous or the organic phase, respectively. This is the case of the well-known Pt(II)-based antitumor drug cisplatin and its analogues, characterized by quite negative  $\log P_{o/w}$  values. Moreover the shake-flask method is generally slow, expensive and, sometimes, poorly reproducible.

For these reasons, alternative experimental methods have been developed to measure the lipophilicity. Among them, the retention parameters in RP-HPLC (where n-octanol is ideally replaced by the C18 chains functionalizing silica as stationary phase, while the mobile phase consists of various mixtures of water and organic co-solvents) are often used, even if criticized by some authors as not truly replacing the shake-flask method.<sup>6,7,8</sup> Since RP-HPLC retention is due to partitioning between (polar) mobile and (apolar) stationary phases, there is a straightforward correlation between the partition coefficient and the HPLC capacity factor  $k'$  ( $k' = (t_R - t_0) / t_0$ , where  $t_0$  is the retention time for an unretained compound and  $t_R$  is the retention time of the analyte). The  $\log k'_0$  ( $k'_0$  is the HPLC capacity factor extrapolated to 100% water) values of compounds with known  $\log P_{o/w}$ , can be used to create a calibration curve ( $\log P_{o/w} = a \log k'_0 + b$ ) to evaluate partition coefficients from chromatographic data.

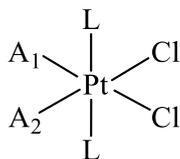
Historically,  $\log P_{o/w}$  values are preferred because one would like to have a comparison with standard experimental lipophilicity values or with calculated  $\log P$  values. However,  $\log k'_0$  is a good lipophilicity index *per se* and is generally used because of its direct correlation with  $\log P_{o/w}$  values. Moreover, being obtained only in an experimental way  $\log k'_0$  does not require any conversion, in such a way that no further experimental error is introduced in the final lipophilicity value.



- 1 A = NH<sub>3</sub>  
 2 CH<sub>3</sub>NH<sub>2</sub>  
 3 (CH<sub>3</sub>)<sub>2</sub>CHNH<sub>2</sub>  
 4 H<sub>2</sub>N(CH<sub>2</sub>)<sub>2</sub>NH<sub>2</sub>  
 5 CH<sub>3</sub>HN(CH<sub>2</sub>)<sub>2</sub>NH<sub>2</sub>  
 6 CH<sub>3</sub>HN(CH<sub>2</sub>)<sub>2</sub>NHCH<sub>3</sub>  
 7 (CH<sub>3</sub>)<sub>2</sub>N(CH<sub>2</sub>)<sub>2</sub>NHCH<sub>3</sub>  
 8 (CH<sub>3</sub>)<sub>2</sub>N(CH<sub>2</sub>)<sub>2</sub>N(CH<sub>3</sub>)<sub>2</sub>  
 9 C<sub>4</sub>H<sub>7</sub>NH<sub>2</sub>  
 10 C<sub>5</sub>H<sub>9</sub>NH<sub>2</sub>  
 11 C<sub>6</sub>H<sub>11</sub>NH<sub>2</sub>



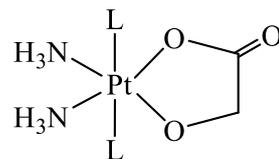
- 12 A = NH<sub>3</sub>  
 13 A = CH<sub>3</sub>NH<sub>2</sub>  
 14 A = H<sub>2</sub>N(CH<sub>2</sub>)<sub>2</sub>NH<sub>2</sub>



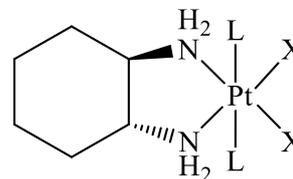
- 23 A<sub>1,2</sub> = NH<sub>3</sub>  
 L = Cl  
 24 OCOCH<sub>3</sub>  
 25 OCOCF<sub>3</sub>  
 26 OCOCH<sub>2</sub>CH<sub>3</sub>  
 27 OCO(CH<sub>2</sub>)<sub>2</sub>CH<sub>3</sub>  
 28 OCO(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>  
 29 OCO(CH<sub>2</sub>)<sub>4</sub>CH<sub>3</sub>  
 30 OCO(CH<sub>2</sub>)<sub>5</sub>CH<sub>3</sub>  
 31 OCO(CH<sub>2</sub>)<sub>2</sub>COOH

- A<sub>1,2</sub> = H<sub>2</sub>N(CH<sub>2</sub>)<sub>2</sub>NH<sub>2</sub>  
 L = Cl  
 45 OCOCH<sub>3</sub>

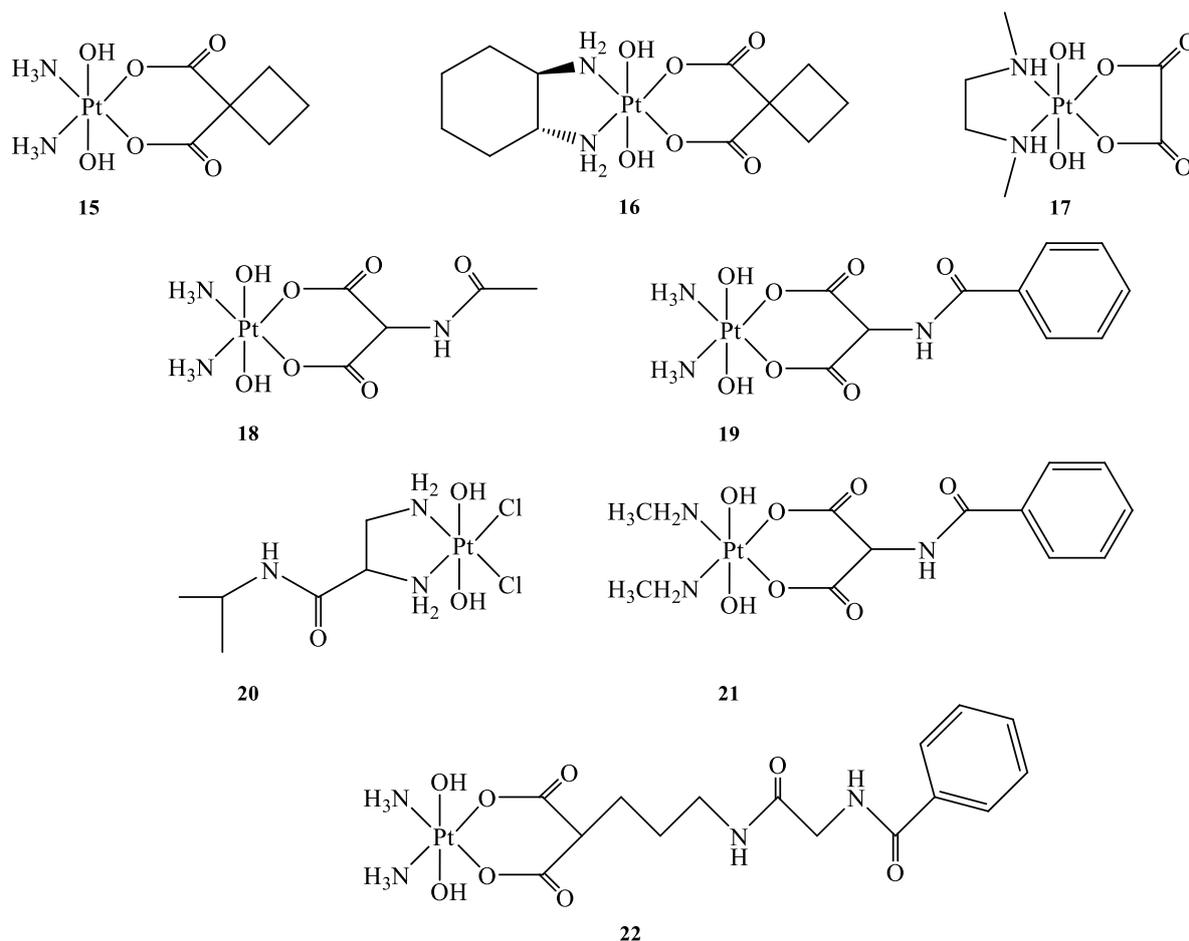
- 47 A<sub>1</sub> = C<sub>6</sub>H<sub>11</sub>NH<sub>2</sub> A<sub>2</sub> = NH<sub>3</sub>  
 L = Cl  
 48 OCOCH<sub>3</sub>  
 49 OCO(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>  
 50 OCO(CH<sub>2</sub>)<sub>5</sub>CH<sub>3</sub>



- 32 L = OCOCH<sub>3</sub>  
 33 OCOCF<sub>3</sub>  
 34 OCOCH<sub>2</sub>CH<sub>3</sub>  
 35 OCO(CH<sub>2</sub>)<sub>2</sub>CH<sub>3</sub>  
 36 OCO(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>  
 37 OCO(CH<sub>2</sub>)<sub>4</sub>CH<sub>3</sub>  
 38 OCO(CH<sub>2</sub>)<sub>5</sub>CH<sub>3</sub>  
 39 OCO(CH<sub>2</sub>)<sub>2</sub>COOH



- X = Cl  
 L = Cl  
 40 OCOCH<sub>3</sub>  
 41 OCOCH<sub>2</sub>CH<sub>3</sub>  
 42 OCO(CH<sub>2</sub>)<sub>2</sub>CH<sub>3</sub>  
 43 OCO(CH<sub>2</sub>)<sub>3</sub>CH<sub>3</sub>  
 44 X = L = OCO(CH<sub>2</sub>)<sub>2</sub>CH<sub>3</sub>



**Fig. 1** Sketch of the complexes under study: the dihydroxido complexes **1-22** are reported here for the first time, while complexes **23-53** were previously studied.<sup>16,25</sup>

In recent years octahedral Pt(IV) complexes have emerged as an alternative to traditional cisplatin-like compounds. They are generally considered antitumor pro-drugs that can be reduced *in vivo* to the active Pt(II) metabolite losing the axial ligands (L) in the hypoxic tumor milieu.<sup>9</sup> These compounds are more inert to ligand substitution reactions than Pt(II) counterparts, leading to lower systemic toxicity from unwanted side-reactions and increasing the likelihood of the drug reaching its cellular target.

Interestingly, Satraplatin (*trans,cis,cis*-bis(acetato)ammine(cyclohexylamine) dichloridoplatinum(IV), JM216) is actually on clinical trials as orally active antitumor drug.<sup>10</sup>

Dihydroxido Pt(IV) complexes represent an interesting sub-class of Pt(IV) compounds.<sup>11</sup> In clinical experience iproplatin (*cis,trans,cis*-dichloridodihydroxidobis(isopropylamine) platinum(IV), also nicknamed CHIP or JM9) demonstrated excellent activity in several phase II trials.<sup>12,13,14</sup> However, this compound was abandoned due to the lack of any superior performance with respect to cisplatin.<sup>15</sup>

In a previous study, a relationship between cytotoxicity and both reduction potential  $E_p$  and partition coefficient  $\log P_{o/w}$  of Pt(IV) complexes has been established: the easier the reduction and the higher the lipophilicity, the higher the cytotoxicity.<sup>16</sup> These two chemico-physical properties, in turn related to cellular uptake and activation by reduction in cytosol, can be tuned through choice of the six ligands around the Pt(IV) centre, especially the axial ones.<sup>11,17</sup> These ligands can also act as

carriers or adjuvant agents, offering interesting applications in the controlled release and in the drug targeting and delivery.<sup>18</sup>

The ability to predict relevant physicochemical properties of platinum complexes directly from their structures would represent an important step in rational design of new platinum drugs, allowing potential candidates to be proposed before lengthy synthesis and testing. In the literature a limited number of publications report on models to predict  $\log P_{o/w}$  values of Pt(II)<sup>19,20,21,22,23,24</sup> and Pt(IV)<sup>20,22,24</sup> complexes by using different descriptors and, more importantly, different mathematical approaches. In this framework, a statistically accurate model for the prediction of  $\log P_{o/w}$  of a series of Pt(IV) complexes containing carboxylates or chlorides as axial ligands has been recently published by us. A combination of surface area and atomic charges was necessary to build the model.<sup>25</sup> This study also demonstrated the ability of the PM6 semi-empirical method<sup>26</sup> to accurately reproduce the X-ray geometry of Pt(IV) complexes such as those discussed here.

In order to evaluate the applicability of the above-reported model to other Pt(IV) complexes, a new set of compounds has been studied. In the present work 22 dihydroxido Pt(IV) complexes (**1-22** in Figure 1) have been newly synthesised or, in some cases, re-synthesised according to standard methods.<sup>27,28,29,30</sup> Their extrapolated capacity factors  $\log k'_0$  were measured and *in silico* properties evaluated by QSRR (Quantitative Structure-Retention Relationship) methods together with the previously studied Pt(IV) complexes (**23-53** in Figure 1).

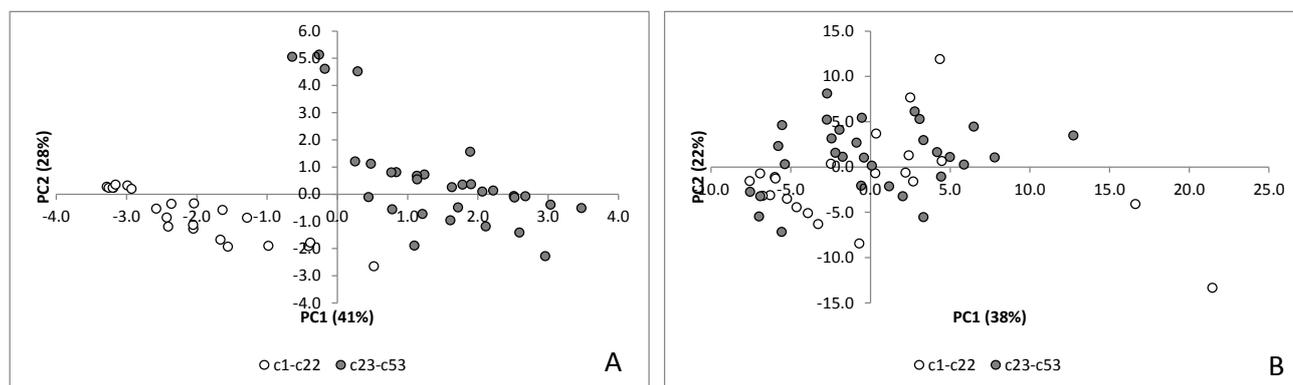


Fig. 2 The score plot of the first two main PCs (in parentheses the variance %): A) QM descriptors, B) VolSurf descriptors

## Results and Discussion

### 5 Synthesis and characterization of Pt(IV) complexes

The dihydroxido Pt(IV) complexes **1-22** were obtained by oxidation<sup>30</sup> of the parent Pt(II) compound<sup>27,28,29</sup> with hydrogen peroxide (see supporting information for the details of the syntheses). The samples contained residual H<sub>2</sub>O<sub>2</sub>, removed by washing the compounds several times with cold water. All the dihydroxido Pt(IV) complexes, except compounds **5-8**, are poorly soluble in water. Figure 1 shows the entire set of the studied complexes.

### Determination of extrapolated capacity factors, $\log k'_0$

15 All the 22 dihydroxido complexes under study were injected into RP-HPLC and their  $\log k'_0$  values, *i.e.* the capacity factor extrapolated to 0% of organic co-solvent, namely MeOH, were measured (Table 1). To obtain a more general QSRR relationship, the  $\log k'_0$  values of other 31 Pt(IV) complexes, determined in the same experimental conditions were added in this study.<sup>25</sup> The  $\log k'_0$  of the entire series of Pt(IV) complexes cover a wide range of values (ca. 8 log units). The increase of the lipophilicity is expected to enhance cellular uptake by passive diffusion.<sup>20,31,32</sup>

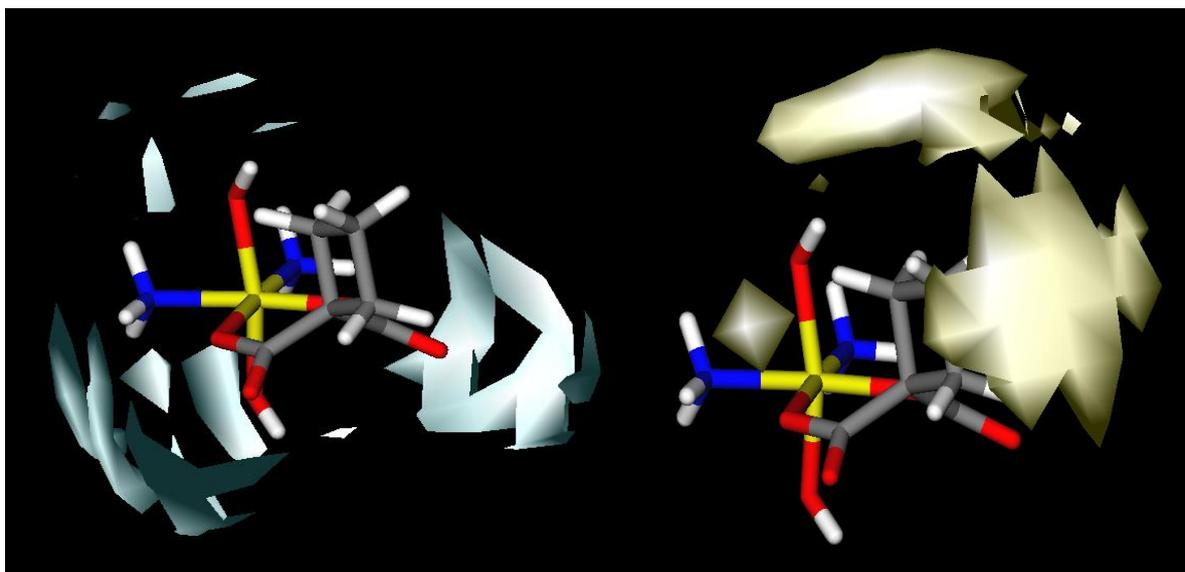
Table 1  $\log k'_0$  values for the complete **1-53** series of Pt(IV) complexes

| Cmpd | $\log k'_0$ <sup>a</sup> | Cmpd | $\log k'_0$ <sup>a</sup> | Cmpd | $\log k'_0$ <sup>a</sup> |
|------|--------------------------|------|--------------------------|------|--------------------------|
| 1    | -0.97                    | 19   | 0.23                     | 37   | 2.29                     |
| 2    | -0.50                    | 20   | -0.56                    | 38   | 3.53                     |
| 3    | 0.64                     | 21   | 0.98                     | 39   | -0.46                    |
| 4    | -0.90                    | 22   | 0.58                     | 40   | -0.16                    |
| 5    | -0.75                    | 23   | -0.78                    | 41   | 0.31                     |
| 6    | -0.55                    | 24   | -0.66                    | 42   | 1.98                     |
| 7    | -0.24                    | 25   | 0.37                     | 43   | 5.11                     |
| 8    | -0.08                    | 26   | -0.25                    | 44   | 3.83                     |
| 9    | 1.20                     | 27   | 0.68                     | 45   | -0.83                    |
| 10   | 2.16                     | 28   | 1.72                     | 46   | -0.51                    |
| 11   | 2.84                     | 29   | 2.94                     | 47   | 1.34                     |
| 12   | -0.96                    | 30   | 4.22                     | 48   | 1.60                     |
| 13   | -0.75                    | 31   | -0.60                    | 49   | 4.56                     |
| 14   | -0.89                    | 32   | -0.70                    | 50   | 6.98                     |
| 15   | -0.50                    | 33   | -0.06                    | 51   | 1.60                     |
| 16   | 0.19                     | 34   | -0.32                    | 52   | 3.70                     |
| 17   | -0.58                    | 35   | 0.40                     | 53   | 4.26                     |
| 18   | -0.63                    | 36   | 1.29                     |      |                          |

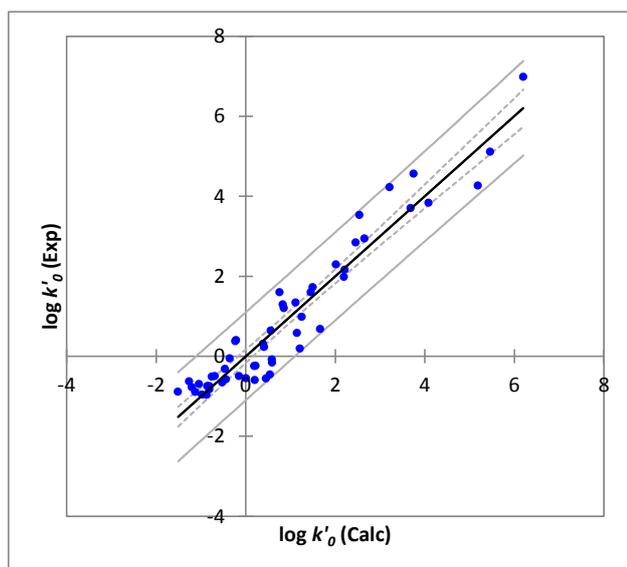
<sup>25</sup> <sup>a</sup>  $\log k'_0$  for complexes **23-53** are from ref. <sup>25</sup>

### QSRR analysis

The  $\log k'_0$  of some compounds of the series (**23-53**) was successfully modelled with a limited set of QM descriptors.<sup>25</sup> The same approach failed when applied to the whole series of **1-53** complexes. This is probably due to the increased heterogeneity of the training set, in turn due to the different axial ligands present in **1-22** with respect to **23-53** compounds (OH instead of carboxylates) that enormously enlarge the chemical domain of the molecules. This is shown by the analysis of the score plot of the Principal Component Analysis (PCA) based on QM descriptors (Figure 2). PCA is a general tool for the interpretation of large data tables, in which the number of the original variables is reduced by a projection of the objects (*i.e.* the molecules) onto a smaller number of new variables termed principal components (PC).



**Fig. 3** MIF generated by the OH2 probe at  $-5.0 \text{ kcal mol}^{-1}$  (left) and by the DRY probe at  $-0.2 \text{ kcal mol}^{-1}$  (right) for compound **15**.



**Fig. 4** Correlation between experimental (Exp) and calculated (Calc)  $\log k'_0$ .

The PCs are orientated so that the first PC describes as much as possible of the original variation between the objects. The second PC is orientated in an orthogonal manner to the first PC and is directed to describe as much as possible of the remaining variation and so on. The projection of objects onto a PC is called score. By plotting the scores for two PCs it is possible to graphically find similarities and differences between objects. Figure 2A shows that the two main PCs (PC1 and PC2) based on the QM descriptors split the whole series of compounds in the two groups (**1-22** and **23-53**) characterised by different kind of axial ligands. This result confirms that the previously selected series of QM descriptors (Table S1)<sup>25</sup> are not suited to exhaustively model the chromatographic behaviour of the whole series of platinum complexes.

A totally different approach has been attempted, replacing QM descriptors with descriptors based on 3D molecular fields, *i.e.* VolSurf descriptors.<sup>33,34</sup> Briefly, VolSurf is a computational

procedure to find out molecular descriptors from 3D molecular interaction fields (MIFs)<sup>35</sup> obtained employing the GRID force field.<sup>36,37,38</sup> Interaction fields are obtained with different probes and the surface of the regions that encompass interaction energy values under certain cutoff limits are calculated. In particular, water (OH2), hydrophobic (DRY), hydrogen bond donor (HBD, amide N1) and hydrogen bond acceptor (HBA, carbonyl O) probes have been considered in the present work.

Since VolSurf descriptors represent polarity and hydrophobicity (as well as size and shape) of molecules, they are generally well suited for the modelling of lipophilicity indexes.<sup>39</sup> This was shown in previous studies where robust QSRR models based on VolSurf descriptors were used to model the variation of different chromatographic lipophilicity indexes.<sup>40,41</sup> Figure 2B confirms the suitability of the application of VolSurf descriptors to the whole series of investigated complexes, with no discrimination of carboxylate, chloride or hydroxo species. As an example, Figure 3 shows the visual representation of two of the 92 Volsurf descriptors obtained with OH2 and DRY probes. They represent the volume of the molecular envelope which is accessible to, and interacts attractively with OH2 probe at  $-5.0 \text{ kcal mol}^{-1}$  and with the DRY probe at  $-0.2 \text{ kcal mol}^{-1}$  for compound **15** (see Experimental for details and Supporting Information for more envelopes).

A relationship between  $\log k'_0$  and the VolSurf descriptors was obtained by Partial Least Squares (PLS) regression technique. A three latent variables (3LVs) model was found: the three main components explained about 90% of the total variance ( $R^2 = 0.92$ ) while the root mean square of the errors (RMSE) was 0.3. To validate these models internal validation was firstly used. Whereas some researchers in the QSAR field support internal validation, others consider that internal validation is not a sufficient test to check the robustness of models, and external validation is necessary.<sup>42</sup> In this case, however, since the sample size is relatively small, and thus holding a portion of it back for testing would be wasteful, it was preferred to use cross-validation, with multiple rounds using different partitions.

In particular the 53 compounds were assigned in a random way

to N groups, each one containing an equal (or nearly equal) number of compounds. Then models were built keeping one of these groups out of the analysis until all of the compounds were kept out once. The formation of the groups and the validation was repeated 50 times. Results were similar with different partitions:  $Q^2$  (LOO) = 0.87;  $Q^2$  (N = 4) = 0.85;  $Q^2$  (N = 3) = 0.85.

Finally, to further validate PLS model the order of Y values which produced unacceptable  $R^2$  and  $Q^2$  values was randomized (data not shown).

The relationship between experimental and calculated values is shown in Figure 4. A slope of approximately 1 and an intercept of about 0 indicate a very good correspondence between experimental and calculated values.

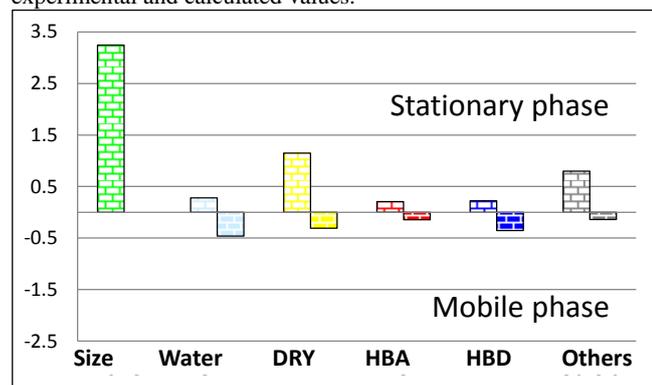


Fig. 5  $BR(+)$  and  $BR(-)$  for PLS model of  $\log k'_{0}$ .

The main drawback of PLS models is the difficulty of their interpretation from the chemical standpoint. The Variable Importance in Projection (VIP) plots are often used to overcome this limit.<sup>43</sup> Briefly, the VIPs show which descriptors are the most important for the model, whereas the sign of the PLS coefficients indicates the positive or negative contribution of the descriptor to the investigated variable. However the use of VIPs for model interpretation is somewhat subjective and therefore it remains of difficult understanding.

Recently, some of us described a new method to obtain a mechanistic interpretation of PLS models by introducing the concept of block of descriptors.<sup>41</sup> This approach can be applied to any set of molecular descriptors but is particularly suited for VolSurf+ descriptors. Since technical reasons prevent to use VolSurf+ software with platinum complexes, here we slightly modified the method above described to adapt it to VolSurf descriptors. The relevance of any block of descriptors to the PLS model is calculated by the Block Relevance ( $BR$ ) parameter (see Experimental part). Table 2 lists the  $BR$  values for the model discussed above, and demonstrates that  $\log k'_{0}$  is mainly characterized by the block of size descriptors ( $BR = 3.24$ ), which is at least twice as important as any other block.<sup>41</sup>

Table 2  $BR$  values for VolSurf PLS model of  $\log k'_{0}$ .

| Block  | N° Descriptors | $BR$ |
|--------|----------------|------|
| Size   | 7              | 3.24 |
| Water  | 23             | 0.74 |
| DRY    | 16             | 1.46 |
| HBA    | 15             | 0.35 |
| HBD    | 15             | 0.58 |
| Others | 16             | 0.94 |

To take into account the sign of the PLS coefficient,  $BR$  is split in  $BR(+)$  relating to retention in the stationary phase, and in  $BR(-)$ , relating to retention in the mobile phase (see Experimental part). Figure 5 shows the trend of  $BR(+)$  and  $BR(-)$  for the investigated system. A positive coefficient means that an increase of the block of the considered descriptors causes an increase in  $\log k'_{0}$  and

thus the retention in stationary phase); the reverse is true for negative coefficients. The relevance of the block of size and hydrophobic descriptors (Figure 5) indicates that the lipophilicity of these complexes is mainly driven by the ligands rather than by platinum(IV) core properties. The importance of this block of descriptors, and their positive sign, is consistent with established ideas of the molecular properties that affect lipophilicity.<sup>5</sup> It is noteworthy that this conclusion can be drawn only using MIFs based descriptors because of the limits of QM descriptors in the modelling of chromatography behavior of platinum complexes.

## Experimental

### General

$K_2[PtCl_4]$  (Johnson Matthey and Co.) and all other chemicals (Aldrich) were used without further purification.

The multinuclear NMR spectra were measured on a JEOL Eclipse Plus operating at 400 MHz ( $^1H$ ), 100.5 MHz ( $^{13}C$ ) and 85.9 MHz, respectively.  $^1H$  and  $^{13}C$  NMR chemical shifts were reported in parts per million (ppm) referenced to solvent resonances; for measurements in  $D_2O$  1% methanol was added as internal reference.  $^{195}Pt$  NMR spectra were recorded using a solution of  $K_2[PtCl_4]$  in saturated aqueous KCl as the external reference. The shift for  $K_2[PtCl_4]$  was adjusted to -1628 ppm from  $Na_2[PtCl_6]$  ( $\delta = 0$  ppm).

RP-HPLC and mass analysis were performed using a Waters HPLC-MS instrument equipped with Alliance 2695 separations module, 2487 dual lambda absorbance detector, and 3100 mass detector. Electrospray ionization mass spectra (ESI-MS) were obtained delivering a diluted solution of the compound in methanol directly into the spectrometer source at 0.01 mL  $min^{-1}$ .<sup>44</sup> The source and desolvation temperatures were set to 150 and 250 °C, respectively, with nitrogen used both as a drying and a nebulizing gas. The cone and the capillary voltages were usually 30 V and 2.70 kV, respectively. Quasi-molecular ion peaks  $[M+H]^+$  or sodiated  $[M+Na]^+$  peaks were assigned on the basis of the  $m/z$  values and of the simulated isotope distribution patterns.

Purity of compounds was assessed by analytical RP-HPLC (see below), elemental analysis and determination of Pt content by inductively coupled plasma-optical emission spectroscopy (ICP-OES). Elemental analyses were carried out with a EA3000 CHN Elemental Analyzer (EuroVector, Milano, Italy). Platinum was quantified by means of a Spectro Genesis ICP-OES spectrometer (Spectro Analytical Instruments, Kleve, Germany) equipped with a crossflow nebulizer. In order to quantify the platinum concentration the Pt 299.797 nm line was selected. A platinum standard stock solution of 1000 mg  $L^{-1}$  was diluted in 1.0% v/v nitric acid to prepare calibration standards.

### Synthesis of platinum complexes

The dihydroxido Pt(IV) complexes **1-22** (Figure 1) were obtained by oxidation with hydrogen peroxide<sup>30</sup> of the parent Pt(II) compound obtained from the Dhara's method and its adaptations.<sup>27,28,29</sup> Briefly,  $K_2[PtCl_4]$  was produced in solution by reaction of  $K_2[PtCl_4]$  (1 mmol) and KI (6 mmol), and reacted with an amine ligand (3.3 mmol for monodentate amines, A, and 1.7 mmol for bidentate ones, A<sub>2</sub>) to give the *cis*-[PtA<sub>2</sub>I<sub>2</sub>] precipitate. This was isolated by centrifugation and washed with water, ethanol and diethyl ether. Upon reaction with  $AgNO_3$  (1.96 mmol) or  $Ag_2SO_4$  (0.98 mmol), the corresponding diaqua intermediate was formed. After removal of AgI, the diaqua species reacted with chlorides or carboxylates to yield the final Pt(II) complex. The yields were from 70 to 85%. This complex

(0.5 mmol) reacted with a tenfold excess of 35% hydrogen peroxide (440  $\mu\text{L}$ ) in water at RT for 24 h. The solvent was then removed under reduced pressure and the dihydroxido product was collected as a pale yellow solid, which was washed with cold water, methanol and diethyl ether and dried in vacuum. The yields were from 50 to 70% (see Electronic Supplementary Information for details).

### RP-HPLC method

Chromatographic analysis were used to evaluate the purity and the capacity factors of the compounds. The chromatographic conditions were:<sup>19</sup> silica-based C18 stationary phase (5  $\mu\text{m}$  Gemini<sup>®</sup> C18 column 25 $\times$ 3 mm ID); mobile phase containing 15 mM HCOOH aqueous solution with different percentage of MeOH (flow rate = 0.75 ml min<sup>-1</sup>; isocratic elution, UV-visible detector set at 210 nm). KCl was the internal reference to determine the column dead-time ( $t_0$ ).<sup>45</sup> Platinum complex solutions were 0.25 mM. A chromatogram for each complex with every different eluant composition has been performed (the methanol fraction,  $\varphi$ , ranging from 20 to 70%) and the corresponding retention time  $t_R$  was used to calculate  $\log k'$  ( $k' = (t_R - t_0) / t_0$ ).

From these data, the extrapolation of the  $\log k'$  to 0% MeOH ( $\log k'_0$ ), corresponding to the capacity factor in pure water, for all compounds has been performed (Eq. 1).<sup>16,25,46</sup>

$$\log k' = \log k'_0 - S\varphi \quad (1)$$

where  $S$  is a solute-dependent solvent strength specific to the organic modifier on the stationary phase under consideration.

### Quantum-Mechanical (QM) descriptors

Surface area, volume, PSA (surface sum over oxygen and nitrogen atoms), frontier orbital energies, dipole moment and atomic partial charges were extracted from the PM6<sup>26</sup> (MOPAC2009, <http://openmopac.net/> and Spartan 08, Wavefunction, Irvine, CA, USA) fully optimised geometries of complexes **1-22** (see Table S1, Electronic Supplementary Information). The same set of data for **23-53** was taken from the literature.<sup>25</sup>

### Principal Component Analysis (PCA)

PCA was performed using Simca v.13 (Umetrics Sweden)

### VolSurf model

Pt(IV) parameterization in the GRID force field implemented in VolSurf v.4.1.2 software (<http://www.moldiscovery.com/>) was performed according to a procedure similar to that described elsewhere for Pt(II).<sup>39</sup> The geometries of fully optimized complexes were saved in mol2 format and submitted to GRID to obtain the molecular interaction fields (MIFs). The binary .kont files of all structures were then submitted to VolSurf for the calculation of the 92 VolSurf descriptors. These were extracted from MIFs obtained with the water (OH2), hydrophobic (DRY), hydrogen bond donor (HBD, amide N1) and hydrogen bond acceptor (HBA, carbonyl O) probes. Partial Least Squares (PLS) analysis was used as implemented in the VolSurf software.

Chemical interpretation of the PLS model was performed as described in the details elsewhere using VolSurf+ descriptors.<sup>39,40</sup> Since for technical reasons VolSurf+ descriptors for platinum complexes cannot be calculated, here we used VolSurf descriptors. These latter were grouped in six blocks: *a*) descriptors that characterize the size and shape of the solute (7 descriptors in the text, briefly called size, color-code green), *b*) 23 descriptors that express the solute's interaction with water

molecules (in the text indicated as water, colour-code light blue), *c*) 15 descriptors that describe the solute's ability to form hydrogen bond interactions with the donor group of the probe (that mimics the chromatographic system, colour-code blue, see below), *d*) 15 descriptors expressing the solute's ability to form hydrogen bond interactions with the acceptor group of the probe (that mimics the chromatographic system, colour-code red, see below), *e*) 16 descriptors describing the solute's propensity of the solute to participate in hydrophobic interactions (in the text called DRY for short, colour-code yellow), *f*) 16 descriptors mainly describing the imbalance between hydrophilic and hydrophobic regions (in the text called others, colour-code grey). More details about the significance of the descriptors can be found in the original paper.<sup>40</sup>

Block Relevance ( $BR$ ) was defined as the ratio of the sum of the squared Variable Importance Projection ( $VIP$ ) values of a given block of descriptors and the number of those descriptors (Eq. 2).

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

where  $i$  is the number of blocks (6 here, Table 2),  $N$  is the number of descriptors for any block,  $VIP$  is the value of each predictor fitting the PLS 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  $BR(+)$  relating to retention in the stationary phase, and  $BR(-)$ , relating to retention in the mobile phase (Eq. 3).

$$BR_i = BR_{i(+)} + BR_{i(-)} \quad (3)$$

## Conclusions

We report the HPLC measurement and QSRR modelling of 53 Pt(IV) complexes as models of potential anti-cancer pro-drugs. New values of  $\log k'_0$  for 22 dihydroxo complexes have been measured, and combined with 31 previously reported values to result in a large and diverse set of complexes whose lipophilicity spans 8 orders of magnitude, and hence act as a stringent test of possible statistical models. We show that conventional models based on whole molecule properties, extracted from semi-empirical PM6 calculations, are incapable of modelling this larger dataset, in contrast to an earlier study. In contrast, sampling to the interaction of each complex with a series of probes and using these interactions as independent variables results in an accurate and robust model. This VolSurf model, using three latent variables, not only correlates and predicts lipophilicity with good accuracy, but also highlights the molecular properties that determine lipophilicity, showing size and hydrophobic interactions to play major roles. This study confirms the superiority of VolSurf over QM descriptors to model lipophilicity of platinum (IV) complexes, and suggests that they have excellent potential in modelling biologically active transition metal complexes in general.

## Acknowledgments.

This research was carried out within the framework of the European Union CMST COST Action CM1105 "Functional metal complexes that bind to biomolecules" and Inter-University Consortium for Research on the Chemistry of Metal Ions in Biological Systems (CIRCMSB, Bari, Italy).

## Notes and references

- <sup>a</sup> CASMedChem Research Group at the Centre for Innovation, Dipartimento di Biotecnologie Molecolari e Scienza della Salute, Università di Torino, Via Quareello 11, 10135 Torino, Italy.
- <sup>b</sup> Dipartimento di Scienze e Innovazione Tecnologica (DiSIT), Università del Piemonte Orientale "Amedeo Avogadro", Viale Michel 11, 15121 Alessandria, Italy. E-mail: [domenico.osella@unipmn.it](mailto:domenico.osella@unipmn.it)
- <sup>c</sup> School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, Wales, UK.
- <sup>†</sup> Electronic Supplementary Information (ESI) available: details of the syntheses and characterization of complexes **1-22**; list of MOPAC2009 descriptors; MIFs of complex **15**. See DOI: 10.1039/b000000x/
- <sup>1</sup> H. Kubinyi, in *QSAR: Hansch Analysis and Related Approaches* (Methods and Principles in Medicinal Chemistry, vol. 1), ed. R. Mannhold, P. Krosggaard-Larse and H. Timmermann, VCH, Weinheim, 1993, ch. 1.
  - <sup>2</sup> G. Caron and G. Ermondi, in *Molecular Drug Properties Measurement and Prediction* (Methods and Principles in Medicinal Chemistry, vol. 37), ed. R. Mannhold, Wiley-VCH, Weinheim, 2008, ch 12.
  - <sup>3</sup> X. Liu, B. Testa and A. Fahr, *Pharm. Res.*, 2011, **28**, 962.
  - <sup>4</sup> C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 2001, **46**, 3.
  - <sup>5</sup> M. H. Abraham, *Chem. Soc. Rev.*, 1993, **22**, 73.
  - <sup>6</sup> W. J. Lambert, *J. Chromatogr. A*, 1993, **656**, 469.
  - <sup>7</sup> T. Braumann, *J. Chromatogr.*, 1986, **373**, 191.
  - <sup>8</sup> S. Martel, D. Guillarme, Y. Henchoz, A. Galland, J. L. Veuthey, S. Rudaz and P. A. Carrupt, in *Molecular Drug Properties Measurement and Prediction* (Methods and Principles in Medicinal Chemistry, vol. 37), ed. R. Mannhold, Wiley-VCH, Weinheim, 2008, ch 13.
  - <sup>9</sup> M. D. Hall, H. R. Mellor, R. Callaghan and T. W. Hambley, *J. Med. Chem.*, 2007, **50**, 3403.
  - <sup>10</sup> G. Doshi, G. Sonpavde and C. N. Sternberg, *Expert Opin. Drug Metab. Toxicol.* 2012, **8**, 103.
  - <sup>11</sup> M. D. Hall, R. D. Dolman and T. W. Hambley, in *Metal complexes in tumor diagnosis and as anticancer agents* (Metal Ions in Biological Systems, vol. 42), ed. A. Sigel and H. Sigel, Dekker, New York, 2004, pp. 298-322.
  - <sup>12</sup> V. H. C. Bramwell, D. Crowther, S. O'Malley, R. Swindell, R. Johnson, E. H. Cooper, N. Thatcher and A. Howell, *Cancer Treat. Rep.*, 1985, **69**, 409.
  - <sup>13</sup> N. J. Wheate, S. Walker, G. E. Craig and R. Oun, *Dalton Trans.*, 2010, **39**, 8113.
  - <sup>14</sup> L. Pendyala, J. Wayne Cowens, G. B. Chheda, S. P. Dutta, J. Patrick and J. Creaven, *Cancer Res.*, 1988, **48**, 3533.
  - <sup>15</sup> H. Anderson, J. Wagstaff, D. Crowther, R. Swindell, M. J. Lind, J. McGregor, M. S. Timms, D. Brown and P. Palmer, *Eur. J. Cancer Clin. Oncol.*, 1988, **24**, 1471.
  - <sup>16</sup> P. Gramatica, E. Papa, M. Luini, E. Monti, M. B. Gariboldi, M. Ravera, E. Gabano, L. Gaviglio and D. Osella, *J. Biol. Inorg. Chem.*, 2010, **15**, 1157.
  - <sup>17</sup> A. M. Montaña and C. Batalla, *Curr. Med. Chem.*, 2009, **16**, 2235.
  - <sup>18</sup> For general reviews see: (a) N. Graf and S. J. Lippard, *Adv. Drug Deliv. Rev.*, 2012, **64**, 993; (b) K. S. Lovejoy and S. J. Lippard, *Dalton Trans.*, 2009, 10651; (c) E. Gabano, M. Ravera and D. Osella, *Curr. Med. Chem.*, 2009, **16**, 4544. For recent, relevant publications on this topic see: (d) S. Mukhopadhyay, C. M. Barnés, A. Haskel, S. M. Short, K. R. Barnes, and S. J. Lippard, *Bioconjugate Chem.*, 2008, **19**, 39; (e) L. J. Parker, L. C. Italiano, C. J. Morton, N. C. Hancock, D. B. Ascher, J. B. Aitken, H. H. Harris, P. Campomanes, U. Rothlisberger, A. De Luca, M. Lo Bello, W. H. Ang, P. J. Dyson, and M. W. Parker, *Chem. Eur. J.*, 2011, **17**, 7806; (f) L. Gaviglio, A. Gross, N. Metzler-Nolte, and M. Ravera, *Metallomics*, 2012, **4**, 260; (g) J. Yang, X. Sun, W. Mao, M. Sui, J. Tang and Y. Shen, *Mol. Pharmaceutics*, 2012, **9**, 2793.
  - <sup>19</sup> J. A. Platts, S. P. Oldfield, M. M. Reif, A. Palmucci, E. Gabano and D. Osella, *J. Inorg. Biochem.*, 2006, **100**, 1199.
  - <sup>20</sup> J. A. Platts, D. E. Hibbs, T. W. Hambley and M. D. Hall, *J. Med. Chem.*, 2001, **44**, 472.
  - <sup>21</sup> I. V. Tetko, I. Jaroszewicz, J. A. Platts and J. Kuduk-Jaworska, *J. Inorg. Biochem.*, 2008, **102**, 1424.
  - <sup>22</sup> S. P. Oldfield, M. D. Hall and J. A. Platts, *J. Med. Chem.*, 2007, **50**, 5227.
  - <sup>23</sup> A. A. Toropov, A. P. Toropova and E. Benfenati, *J. Math. Chem.*, 2009, **46**, 1060.
  - <sup>24</sup> P. Sarmah and R. C. Deka, *J. Comput.-Aided. Mol. Des.*, 2009, **23**, 343.
  - <sup>25</sup> J. A. Platts, G. Ermondi, G. Caron, M. Ravera, E. Gabano, L. Gaviglio, G. Pelosi and D. Osella, *J. Biol. Inorg. Chem.*, 2011, **16**, 361.
  - <sup>26</sup> J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173.
  - <sup>27</sup> S. C. Dhara, *Indian J. Chem.* 1970, **8**, 193.
  - <sup>28</sup> R. C. Harrison and C. A. McAuliffe, *Inorg. Chim. Acta*, 1980, **46**, L15.
  - <sup>29</sup> F. D. Rochon and L. M. Gruia, *Inorg. Chim. Acta* 2000, **306**, 193.
  - <sup>30</sup> C. M. Giandomenico, M. J. Abrams, B. A. Murrer, J. F. Vollano, M. I. Rheinheimer, S. B. Wyer, G. E. Bossard and J. D. Higgins III, *Inorg. Chem.*, 1995, **34**, 1015.
  - <sup>31</sup> W. H. Ang, S. Pilet, R. Scopelliti, F. Bussy, L. Juillerat-Jeanneret and P. J. Dyson, *J. Med. Chem.*, 2005, **48**, 8060.
  - <sup>32</sup> A. R. Ghezzi, M. Aceto, C. Cassino, E. Gabano and D. Osella, *J. Inorg. Biochem.*, 2004, **98**, 73.
  - <sup>33</sup> G. Cruciani, P. Crivori, P. A. Carrupt and B. Testa, *J. Mol. Struct. (Theochem)*, 2000, **503**, 17.
  - <sup>34</sup> P. Crivori, G. Cruciani, P. A. Carrupt and B. Testa, *J. Med. Chem.*, 2000, **43**, 2204.
  - <sup>35</sup> G. Cruciani, R. Mannhold, H. Kubinyi and G. Folkers, *Molecular interaction fields. Applications in drug discovery and ADME prediction*. Wiley-VCH, Zurich, 2006.
  - <sup>36</sup> D. N. Boobbyer, P. J. Goodford, P. M. Mcwhinnie and R. C. Wade, *J. Med. Chem.*, 1989, **32**, 1083.
  - <sup>37</sup> P. J. Goodford, *J. Med. Chem.*, 1985, **28**, 849.
  - <sup>38</sup> R. C. Wade and P. J. Goodford, *J. Med. Chem.*, 1993, **36**, 148.
  - <sup>39</sup> G. Caron and G. Ermondi, in *Comprehensive Medicinal Chemistry*, ed. B. Testa and H. van de Waterbeemd, Elsevier, Oxford, 2nd edn, 2006, vol. 5, pp. 425-452.
  - <sup>40</sup> G. Caron, M. Ravera and G. Ermondi, *Pharm. Res.*, 2011, **28**, 640.
  - <sup>41</sup> G. Ermondi and G. Caron, *J. Chromatogr., A*, 2012, **1252**, 84.
  - <sup>42</sup> a) S. Wold, M. Sjostrom, and L. Eriksson, *Chemometr. Intell. Lab.*, 2001, **58**, 109; b) T. Scior, J. L. Medina-Franco, Q.-T. Do, K. Martínez-Mayorga, J. A. Yunes Rojas, and P. Bernard, *Curr. Med. Chem.*, 2009, **16**, 4297
  - <sup>43</sup> G. Caron and G. Ermondi, *J. Med. Chem.*, 2005, **48**, 3269.
  - <sup>44</sup> R. Bertani, R. Seraglia, D. Favretto, R. A. Michelin, M. Mozzon, S. Mazzega Sbovata and S. Sassi, *Inorg. Chim. Acta*, 2003, **356**, 357.
  - <sup>45</sup> O. Heudi, S. Mercier-Jobard, A. Cailleux and P. Allain, *Biopharm. Drug Dispos.*, 1999, **20**, 107.
  - <sup>46</sup> K. Valkó, L. R. Snyder and J. L. Glajch, *J. Chromatogr., A*, 1993, **656**, 501.