UNIVERSITÀ DEGLI STUDI DI TORINO

Integrating Intramolecular Hydrogen Bonding (IMHB) Considerations in Drug Discovery Using ΔlogP As a Tool

*Marina Shalaeva *, Giulia Caron *, Yuriy A. Abramov , Thomas N. O'Connell, Mark S. Plummer, Geeta Yalamanchi, Kathleen A. Farley, Gilles H. Goetz, Laurence Philippe, and Michael J. Shapiro*

J. Med. Chem., 2013, 56 (12), pp 4870–4879

DOI: 10.1021/jm301850m

Publication Date (Web): May 27, 2013

Integrating intramolecular hydrogen bonding (IMHB) considerations in drug discovery using ΔlogP as a tool.

Marina Shalaeva[†], Giulia Caron[‡]*, Yuriy A. Abramov[†], Thomas N. O'Connell[†], Mark S. Plummer[†§], Geeta Yalamanchi[†!!], Kathleen A. Farley[†], Gilles H. Goetz[†], Laurence Philippe[†], Michael J. Shapiro[†]*

AFFILIATIONS:

[†]Worldwide Medicinal Chemistry, Pfizer Global Research & Development, Pfizer, Inc., Groton, CT, 06340, USA

[‡] Molecular Biotechnology and Health Sciences Department, University of Torino, via Quarello 15, 10135 Torino, Italy

KEYWORDS: Intramolecular Hydrogen Bonding, Hydrogen Bonding, Delta logP, logP toluene, calculated logP.

ABSTRACT

This study demonstrates that $\Delta logP_{oct\text{-}tol}$ (difference between logPoctanol and logPtoluene) describes compounds propensity to form intramolecular hydrogen bonds (IMHB) and may be considered a privileged molecular descriptor for use in drug discovery and for prediction of IMHB in drug candidates.

We identified experimental protocols for acquiring reliable $\Delta logP_{oct\text{-}tol}$ values on a set of compounds representing IMHB motifs most prevalent in Medicinal Chemistry, mainly molecules capable of forming 6-, 7-member IMHB rings.

Furthermore, computational $\Delta logP_{oct\text{-}tol}$ values obtained with COSMO-RS software provided a good estimate of experimental results and can be used prospectively to assess IMHB.

The proposed interpretation method based on $\Delta logP_{oct\text{-}tol}$ data allowed categorization of the compounds into 2 groups - with high propensity to form IMHB and poor propensity or poor relevance of IMHB.

The relative ^{1}H NMR chemical shift of an exchangeable proton was used to verify presence of IMHB and to validate the IMHB interpretation scheme.

1. INTRODUCTION

The incorporation of an intramolecular hydrogen bond (IMHB) into a molecule is gaining a great deal of interest in drug design as indicated by the number of papers recently published in key Medicinal Chemistry journals. [1-5] The presence of IMHB has been shown to significantly alter molecular properties due to formation of various conformers that in turn influence solubility, permeability, PK/ PD processes, and protein binding affinity. [6-9]

The IMHB as described by Desiraju [10] is an attractive interaction in which an electropositive hydrogen atom intercedes between two electronegative fragments of the same molecule and holds them together. A hydrogen bond is strong enough to restrict rotation of fragments by forming most commonly 5-8 membered rings. Importantly, IMHBs are weak enough to allow these fragments to come apart and lose their orientational specificity in high dielectric media such as water. The chameleon like nature of an IMHB becomes apparent when one realizes that in water an IMHB is unlikely to form and the polar groups may serve to increase solubility by readily forming *inter*molecular hydrogen bonds with water. Alternatively, molecules that can participate in IMHB shed water more readily when entering a low dielectric environment like a hydrophobic phospholipid bilayer. In this circumstance IMHB results in lipophilic, less polar molecular conformations which are expected to have higher passive membrane permeability. [11] In other words, a decrease in polarity is sometimes achieved through the formation of IMHBs, where the hydrogen bond donor (HBD) and acceptor (HBA) atoms are effectively shielded from water, thereby reducing the energetic penalty of desolvation required in moving from an aqueous environment through a phospholipid bilayer. [4]

The consequences of IMHBs to medicinal chemists are significant but often under-recognized and seldom predicted. For instance, lipophilicity may be underestimated when determined by calculated logP (clogP) in molecules with IMHBs, while hydrogen bond donor and acceptor counts are overestimated. Additionally, clogP, as well as hydrogen bond donor and acceptor counts, are part of the ubiquitous Ro5 parameters [12] , used to predict drug like properties and permeability. When

IMHB are present these Ro5 counts can be effectively stretched, broadening drug like property space allowing more diverse drug design. [9,13] Likewise preferred property space for Central Nervous System (CNS) drugs may be extended when IMHB are present, as hydrogen bond donor count and clogP are both parameters in the CNS Multi-Parameter Optimization (CNS MPO) score.[14] In support of this notion it was also found that $\Delta logP_{oct\text{-}alk}$ correlates with brain penetration and oral absorption. [15,16]

Recent systematic work incorporating IMHB considerations in drug design has been published by Kuhn and coworkers. [17] On the basis of pioneering work by Etter [18] and Bilton [19] and exhaustive searches of crystal structure databases, they derived propensities for IMHB formation of five- to eight-membered ring systems of relevance in drug discovery. The influence of IMHB on solubility, lipophilicity in octanol/water and permeability was also highlighted.

Unfortunately, one cannot simply examine a given 2D structure and immediately delineate the presence of one or more IMHBs and determine their strength because the thermodynamic equilibrium of closed *versus* open conformations depend on a number of complex factors (e.g. geometry, type of solvent and others) [20] acting simultaneously. The most common tools used to investigate IMHBs are spectroscopy (NMR, infrared and Raman, microwave), diffraction (X-ray and neutron diffraction), calorimetry and theoretical methods.[21] However, many of these techniques are not high throughput and data produced often require detailed interpretation by experts. These issues lead us to look for additional methods.

LogP is one of the most widely used parameters in drug design and it has been considered for evaluation of IMHB [8, 22, 23]. It has been demonstrated using solvatochromic equations that the difference between logP values obtained in different biphasic systems (ΔlogP), for example octanol/water and alkane/water ($\Delta logP_{oct\text{-}alk} = logP_{oct} - logP_{alk}$), is informative of IMHB when the solvents are very different from each other. [24] More on differences in logP systems is given in Supporting Information (Annex S1, S3).

The idea that $\Delta logP_{oct\text{-}alk}$ is informative of IMHB and the reports that $\Delta logP_{oct\text{-}alk}$ correlates with brain penetration and oral absorbtion [15, 16] lead us to explore $\Delta logP_{oct\text{-}tol}$ ($\Delta logP_{oct\text{-}tol} = logP_{oct} - logP_{tol}$. [25-29]

The investigation of IMHB by ΔlogP was proposed some time ago[30], however this approach was not widely implemented mainly because the practical tools, both experimental and theoretical, to obtain $logP_{alk}$ data for large series of compounds were limited.

The main goal of this study is to demonstrate that $\Delta logP_{oct\text{-}tol}$ ($\Delta logP_{oct\text{-}tol} = logP_{oct} - logP_{tol}$) distinguishes compounds with high propensity to form IMHB and to develop a protocol for its implementation in active Medicinal Chemistry projects where series of similar compounds are often available for relative comparisons.

In order to achieve this goal we needed to address three subgoals. The first subgoal was to identify experimental methods that provide reliable $\Delta logP_{oct\text{-}tol}$ for large series of compounds. To accomplish this we used an *ad hoc* dataset of commercially obtained compounds representing many prevalent IMHB motifs and used miniaturized shake-flask and HPLC methods to acquire the individual $logP_{oct}$ and $logP_{tol}$ values.

The second subgoal of the study was to validate $\Delta logP_{oct\text{-}tol}$ calculations where the molecular 3D structure is considered since it strongly influences the formation of IMHBs. We used the computational software COSMOtherm. The choice of COSMOtherm among a plethora of free and commercial computational tools available today for logP/logD calculations[31] was justified by two reasons: 1) unlike most logP calculators it uses the three dimensional structure of the molecules [32-34] and 2) COSMOtherm allows calculation of logP values in non octanol/water systems. Furthermore, COSMOtherm is the only commercial *a priori* (not restricted by the availability of experimental data) method [35] available today.

The third subgoal of the study was to validate the IMHB interpretation scheme based on measured and calculated $\Delta logP_{oct\text{-}tol}$ data with an independent technique such as relative ^{1}H NMR chemical shifts.

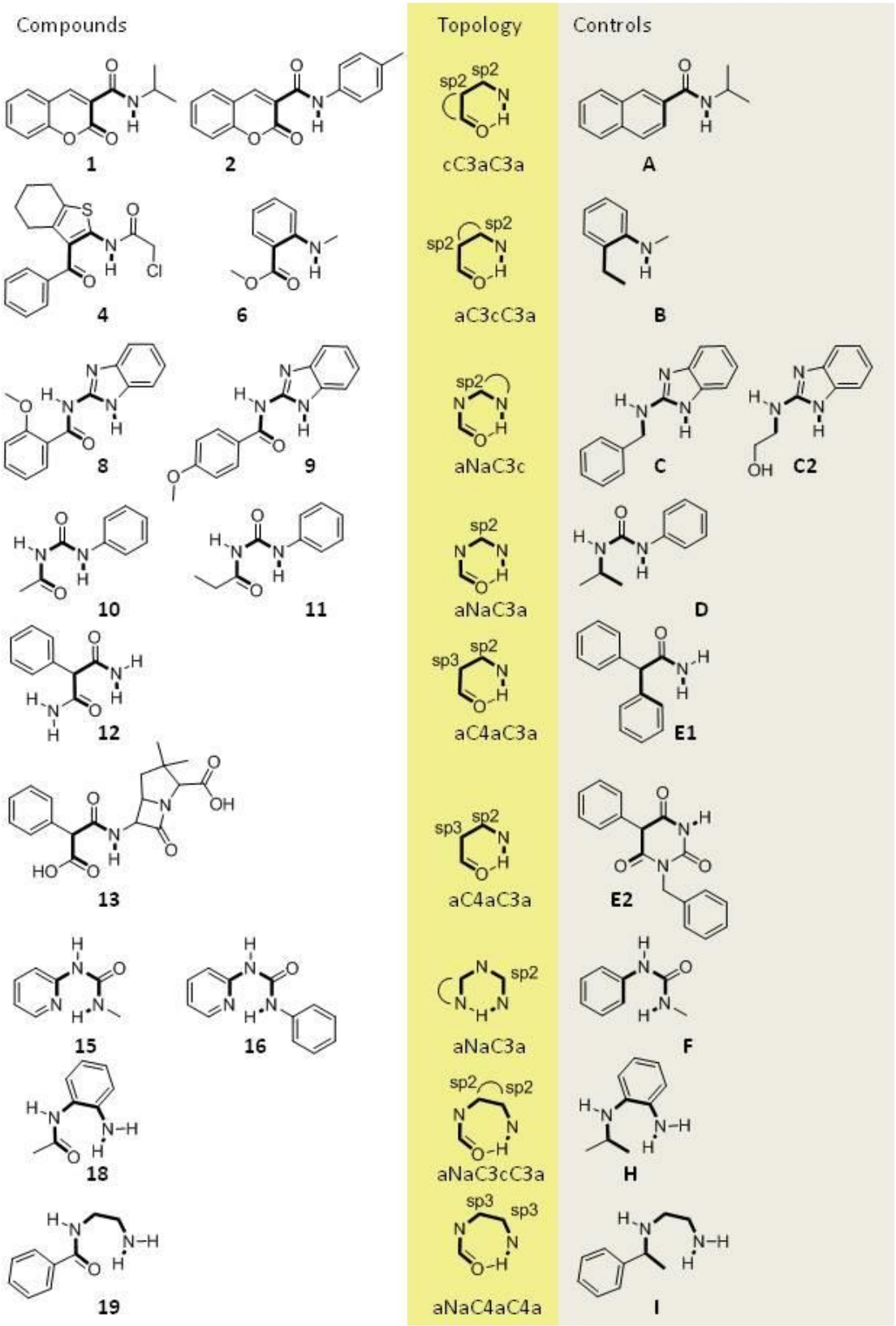
2. RESULTS AND DISCUSSION

2.1. DATASET SELECTION

The set of 24 compounds and controls, shown in Figure 1, was created following the topologies identified by Kuhn et al. [17] in their systematic study of IMHB based on the analysis of the Cambridge Structures Database (CSD). It was observed in their study that 6- and 7- member IMHB ring systems are, by far, the most prevalent motifs in Medicinal Chemistry. In our study we used the numbered commercial compounds that contain several of the topologies described by Kuhn et al. and compared them with similar lettered compounds (controls) that are unable to form IMHB. We attempted to have simple test structures with one possible IMHB.

In addition, the following aspects were taken into consideration while building the dataset to facilitate logP determination: solubility, ionization state, UV detection and commercial availability.

Figure 1. Chemical structures of investigated compounds; Kuhn's topologies [17] and controls



2.2. ΔLOGP DETERMINATION

Table 1. Lipophilicity data (controls are colored in grey)

	octanol/water			toluene/water		octanol/water – toluene/water	
	COSMO-RS logP	**SF_logP**	**ElogD**	COSMO-RS logP	**SF_logP**	COSMO-RS ΔlogP	**SF_ΔlogP**
1	2.53	1.8	2.13	2.78	1.9	-0.25	-0.10
2	4.0	****	4.25	5.05	****	-1.05	N/A
A	2.83	2.1*	2.68	2.44	1.6*	0.39	0.50
4	4.50	****	4.58	5.49	****	-0.99	N/A
6	3.05	2.1*	2.97	3.81	2.1*	-0.76	0.08
B	3.17	1.91	2.58	3.76	1.96	-0.59	-0.05
8	3.00	2.4*	3.20	3.69	1.9*	-0.69	0.51
9	2.71	***	2.89	2.73	***	-0.02	***
C	2.41	N/A	2.92	1.43	N/A	0.98	N/A
C2	0.59	0.46	1.1	-2.16	-2.46	2.75	2.92
10	1.49	1.43	1.76	1.57	0.71	-0.08	0.72
11	2.12	1.84	2.19	2.31	1.35	-0.19	0.49
D	2.01	1.76	1.84	1.09	0.13	0.92	1.63
12	-1.08	-0.56	-0.07	-2.73	-2.42	1.65	1.86
E1	1.50	1.56	2.11	0.99	0.61	0.51	0.95
13	2.40	0.15	**	-1.51	-1.63	3.91	1.78
E2	2.38	1.76	**	2.70	1.2	-0.32	0.56
15	0.62	0.83	0.8	-0.15	-0.31	0.77	1.14
16	2.47	2.17	3.1	2.73	1.64	-0.26	0.53
F	0.78	0.95	1.3	-0.43	-1.01	1.21	1.96
18	0.26	-0.17	0.13	-0.38	-1.75	0.64	1.58
H	2.37	1.08	1.65	2.69	0.91	-0.32	0.17
19	0.45	0.37	-0.69**	-0.08	0.01	0.53	0.36
I	2.41	0.97	-0.26**	1.43	1.06	0.98	-0.09

* uncertain value due to aggregation/quantification issues
** ionized at the pH of measurements
*** unstable compound
**** below quantification limits (too lipophilic)

Experimental logP values (SF_logP in Table 1) were obtained in the presence of DMSO (up to 10% total volume) in the solution, which assisted solubility while also mimicking a widely accepted practice of using DMSO stock solutions in high throughput assays in drug discovery programs, including logP/D measurements [36-39]. As a result, experimental $\Delta logP_{oct\text{-}tol}$ data (SF_ΔlogP in Table

1) were obtained for most of the molecules in the dataset. More information about the influence of DMSO on logP is given in the Supporting Information (Annex S2).

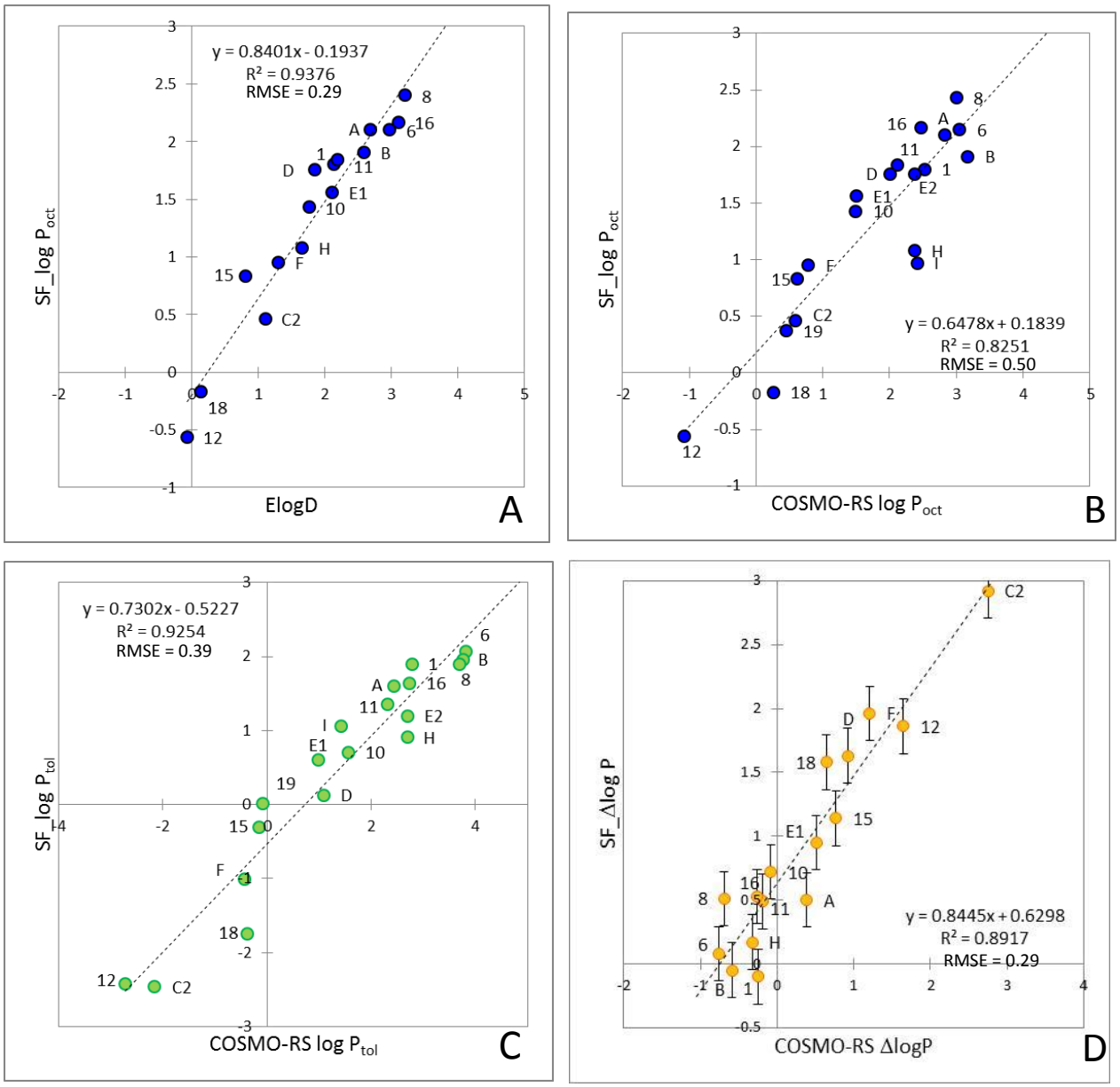
Theoretical lipophilicity values used in Table 1 were obtained with COSMOtherm and we refer to COSMO-RS $logP_{oct}$ and COSMO-RS $logP_{tol}$ as calculated values in octanol/water and toluene/water systems, respectively. Briefly, COSMOtherm software calculates logP from a chemical potential for any solvent system using the COSMO-RS method [40], which is based on a combination of the quantum chemical dielectric continuum solvation model (COSMO) with a statistical thermodynamic treatment of surface interactions. The method also allows prediction of the relative weight of each molecular conformation in the solvent using Boltzmann statistics. More details on the calculations are given in Section 4.4.2.

Finally, the accuracy of each logP value is especially critical in this study if the ΔlogP to be used as a descriptor of IMHB. Therefore, whenever possible, we tried to verify logP values obtained by one method with another experimental or computed value. The results of the cross validation of lipophilicity data are summarized in Figure 2 and show an excellent agreement between data produced by different methods.

The standard error of determination reported for SF_LogP [41] method is 0.1 (and thus 0.2 for ΔlogP). The Root Mean Square Errors (RMSE) derived from the regression analysis are listed on each plot. The RMSE for SF_logP vs. COSMO-RS logP for octanol (Fig. 2B) and toluene (Fig. 2C), are 0.5 and 0.4, respectively, demonstrating again, similarity in COSMO-RS modeling across the two systems. Apparently, the most significant bearing in this study is on the error in determination of ΔlogP values. The RMSE for SF_ΔlogP vs. COSMO-RS ΔlogP (Fig. 2D) is 0.29, which is slightly better than in each system separately, probably due to cancelling out of the DMSO effect. More details about the cross validation strategy [42-44] are given in the Supporting Information (Annex S2).

Figure 2. Cross validation of lipophilicity data. A) Validation of experimental SF_logP_{oct} with ElogD, B) Validation of computed COSMO-RS logP_{oct} with experimental SF_logP_{oct}, C) Validation

of experimental SF_logP_{tol} with computed COSMO-RS logP_{tol} and D) Validation of experimental SF_ΔlogP (error bars are shown) with computed COSMO-RS Δlog,.



2.3. ΔlogP ANALYSIS and an IMHB INTERPRETATION METHOD

No general guidelines are reported in the literature on the interpretation of ΔlogP in relation to the presence and the strength of IMHBs. Therefore the analysis of ΔlogP data was aimed at obtaining an IMHB interpretation method.

As discussed earlier, it is assumed that toluene, similar to apolar solvents, promotes folded conformations and formation of IMHB when possible, whereas the reverse is true for molecules in water and, to a lesser degree, in octanol. Therefore, the difference between $logP_{oct}$ and $logP_{tol}$ (i.e. ΔlogP) should reflect the propensity of a compound to form IMHB. The ΔlogP value by itself does not indicate the formation of IMHB. However, trends are observed if comparisons made in a pair-wise fashion for compounds in a series capable of forming IMHB (samples) and not capable of IMHB (controls). In particular, the comparative analysis of $\Delta logP_{oct\text{-}tol}$ in Table 1 reveals two possible situations:

- $\Delta logP_{oct\text{-}tol}$ of the control is > (larger) than $\Delta logP_{oct\text{-}tol}$ of the sample - Category I
- $\Delta logP_{oct\text{-}tol}$ of the control is < (smaller) than $\Delta logP_{oct\text{-}tol}$ of the sample - Category II

The first situation (Category I), when $\Delta logP_{oct\text{-}tol}$ of control is larger than that for the sample, is found in the following matched groups: **1** and Control **A**; **8** and Control **C2**; **10, 11** and Control **D; 15, 16** and Control **F**. In these groups the sample prefers toluene to octanol when compared with the control, presumably because of significant amounts of folded conformers with a high propensity to form IMHB.

The second situation (Category II), when the $\Delta logP_{oct\text{-}tol}$ of the control is lower than $\Delta logP_{oct\text{-}tol}$ of the sample, is found for the following groups: **6** and Control **B; 12** and Control **E1**, for **13** and Control **E2**, for **18** and Control **H** and **19** and Control **I**. In these groups the sample prefers octanol to toluene, probably, due to a significant presence of extended conformers. This suggests that either the sample has a poor propensity to form IMHBs or the IMHB has a poor relevance to $\Delta logP_{oct\text{-}tol}$. The influence of the experimental error associated with measured $\Delta logP_{oct\text{-}tol}$ value ($\pm$0.2 ; the error bars shown in Fig.2D) has to be considered in categorization of results. Apparently, in that "binning" categorization scheme, the consequence of the error is most significant for compounds with a small difference between ΔlogP of Sample and Control. Therefore, the classification threshold is defined by the error and the difference in ΔlogP has to be higher than 0.4, for clear

categorization. In other words, if ΔlogP values of Sample and Control are very close to each other the categorization becomes uncertain for that pair (example, **6** and Control **B).** However, substitution of the control with another molecule from the series could be recommended as a practical solution in a drug design project.

Importantly, the calculated COSMO-RS $\Delta logP_{oct\text{-}tol}$ values presented in Table 1 demonstrate that the same pairs of compounds fall into the same categories as described above for experimental $\Delta logP_{oct\text{-}tol}$ values. The only deviation was observed for compound **19** and Control **I**, which were placed into Category I by calculated $\Delta logP_{oct\text{-}tol}$ and into Category II by measured $\Delta logP_{oct\text{-}tol}$. This discrepancy stems mostly from the significant difference between predicted and experimental $logP_{oct}$ value for the control **I** (2.41 and 0.97, respectively). The COSMO-RS $logP_{oct}$ for control **I** is the most deviant predicted value in the set, as seen in Fig. 2B The deviant predicted value drives the error for COSMO-RS $logP_{oct}$ higher for the entire set. Recall, compound **6** and Control **B** categorization by experimental data are uncertain, as discussed above; however, COSMO-RS $\Delta logP_{oct\text{-}tol}$ calculated values assign this pair to Category I.

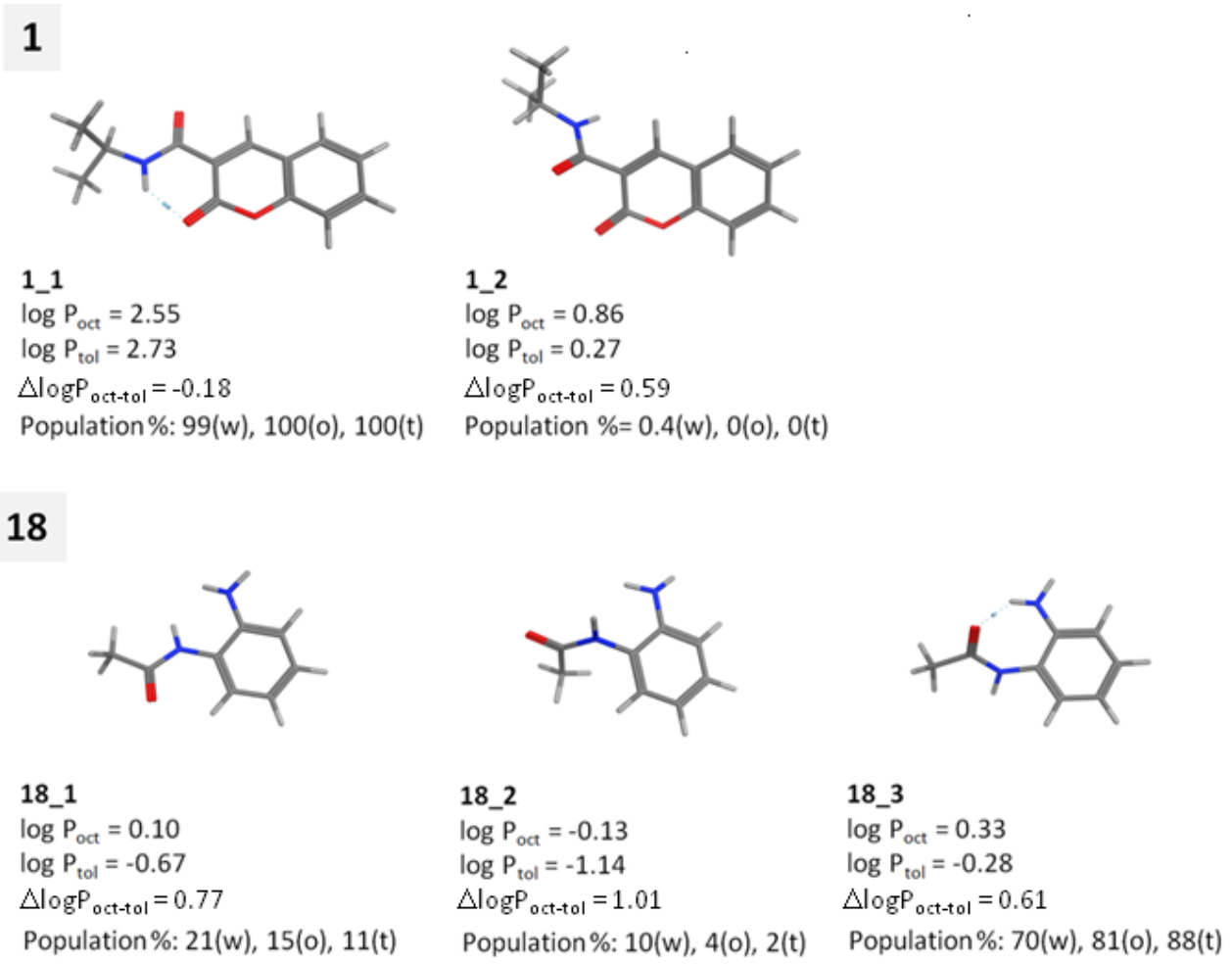
This illustrates one benefit of using computed values where compounds lacking measured logP values can be categorized based on the calculated logP values. Here in particular, **4** and Control **B; 9** and Control **C** are placed into the Category I using calculated $\Delta logP_{oct\text{-}tol}$ values, even though the low solubility and instability of compounds **4** and **9**, respectively, prevented us from obtaining measured $\Delta logP_{oct\text{-}tol}$ values. A second obvious benefit is the prospective use of calculated ΔlogP values in Medicinal Chemistry design on a series of virtual compounds.

The IMHB interpretation scheme was further supported by a closer inspection of the properties of the molecular conformers generated using COSMOS-RS. The example of conformers for compounds **1** (Category I, high propensity to form IMHB) and **18** (Category II, low propensity to form IMHB) are visualized in Figure 3. The conformer 1_1 (top left in Figure 3) forms IMHB in any solvent (populated about 100% as evaluated in water, octanol and toluene) and contributes most

to the logP value. The open conformers (as 1_2, top right in Figure 3) are poorly populated and largely more hydrophilic than the closed conformers.

The three most populated conformers of **18** (18_1, 18_2, 18_3 in Figure 4) may or may not form IMHB. They show similar $logP_{oct}$ values and their difference in $logP_{tol}$ is less important than that found for compound **1**.

*Figure 3. Examples of COSMO-RS results for conformations of compounds **1** (Category I, high propensity to form IMHB) and **18** (Category II, low propensity to form IMHB). The relative conformer populations in water (w), wet octanol (o) and toluene (t) are shown. For each conformer $logP_{oct}$ and $logP_{tol}$ were calculated ignoring all other conformations.*



These examples demonstrate that in contrast to 2D logP calculation methods, COSMO-RS gives a detailed view of conformational variability and supports that the presence of folded conformers

lowers ΔlogP value (1_1 and 18_3). As a consequence the comparison between ΔlogP of the sample and its control helps substantiate the propensity of the test molecules to form IMHB.

This approach was applied to ΔlogP data available from the literature in cyclohexane/water [25] and 1,2-dichloroethane/water systems [29, 45] and it revealed the same trends; data shown in Supporting Information (Annex S3).

2.4. IMHB verification by NMR and crystallographic data

To validate the IMHB interpretation scheme based on ΔlogP data, we set out to determine the relative ^{1}H NMR chemical shift of an exchangeable proton at a single temperature. Generally, an exchangeable proton that is hydrogen bonded will be more deshielded (higher chemical shift value) than a similar exchangeable proton that is not hydrogen bonded. As a consequence, and in analogy with ΔlogP analysis, it requires a comparative analysis between two compounds (sample and control) within the same chemical series in order to evaluate the propensity to form IMHB using the NMR ^{1}H chemical shift data of the exchangeable protons.

Table 2. Correlation of COSMO-RS ΔlogP and NMR ^{1}H Chemical Shift results and the percentage of crystallographic entries (%HB) with IMHB found by Kuhn and coworkers for the corresponding topologies [17]

	δCDCl$_3$ 300K	δCDCl$_3$ sample- δ CDCl$_3$ control	COSMO-RS ΔlogPcontrol- ΔlogPsample	%HB predicted by topology[17]	Category
1	8.69	2.63	0.64	93	I
2	10.78	4.72	1.44	93	I
A	6.06	-	-	-	
4	12.12	8.44	0.4	90	I
6	7.64	3.96	0.17	90	I
B	3.68	-	-	-	
10	10.37	4.36	1	85	I
11	10.44	4.43	1.11	85	I
D	6.01	-	-	-	
12	6.1	0.58	-1.14	18	II
E1	5.52	-	-	-	
15	9.24	4.66	0.44	81	I

16	11.76	7.18	1.47	81	I
F	4.58	-	-	-	
18	3.9	0.6	-0.96	37	II
H	3.3	-	-	-	

The presence of IMHB in the studied compounds is indicated by the chemical shift of the exchangeable proton involved in IMHB, which moves upfield in comparison with the control compound that cannot form IMHB. In particular, this upfield shift is clearly marked for all compounds that were amenable to measurements in $CDCl_3$ (Table 2) and DMSO (Supporting Information Table S2). Compounds **12** and **18** do not follow this trend and have only minor chemical shift differences. This further supports that **12** and **18** are correctly binned in Category II indicating that these compounds are unlikely to form IMHB to a significant extent. More details about NMR strategy [46] can be found in the Supporting Information (Annex S4).

In order to examine ΔlogP and NMR data for trends, the differentials between ΔlogP and δ $CDCl_3$ values for "samples" and "controls" are calculated (Table 2) and graphically presented in Fig.4, A and B, respectively.

Figure 4. COSMO-RS ΔlogP vs. NMR chemical shift trends demonstrated on differentials between values for sample and control. A) COSMO-RS (ΔlogPcontrol - ΔlogPsample); B) δ $CDCl_3$ sample – δ $CDCl_3$ control.

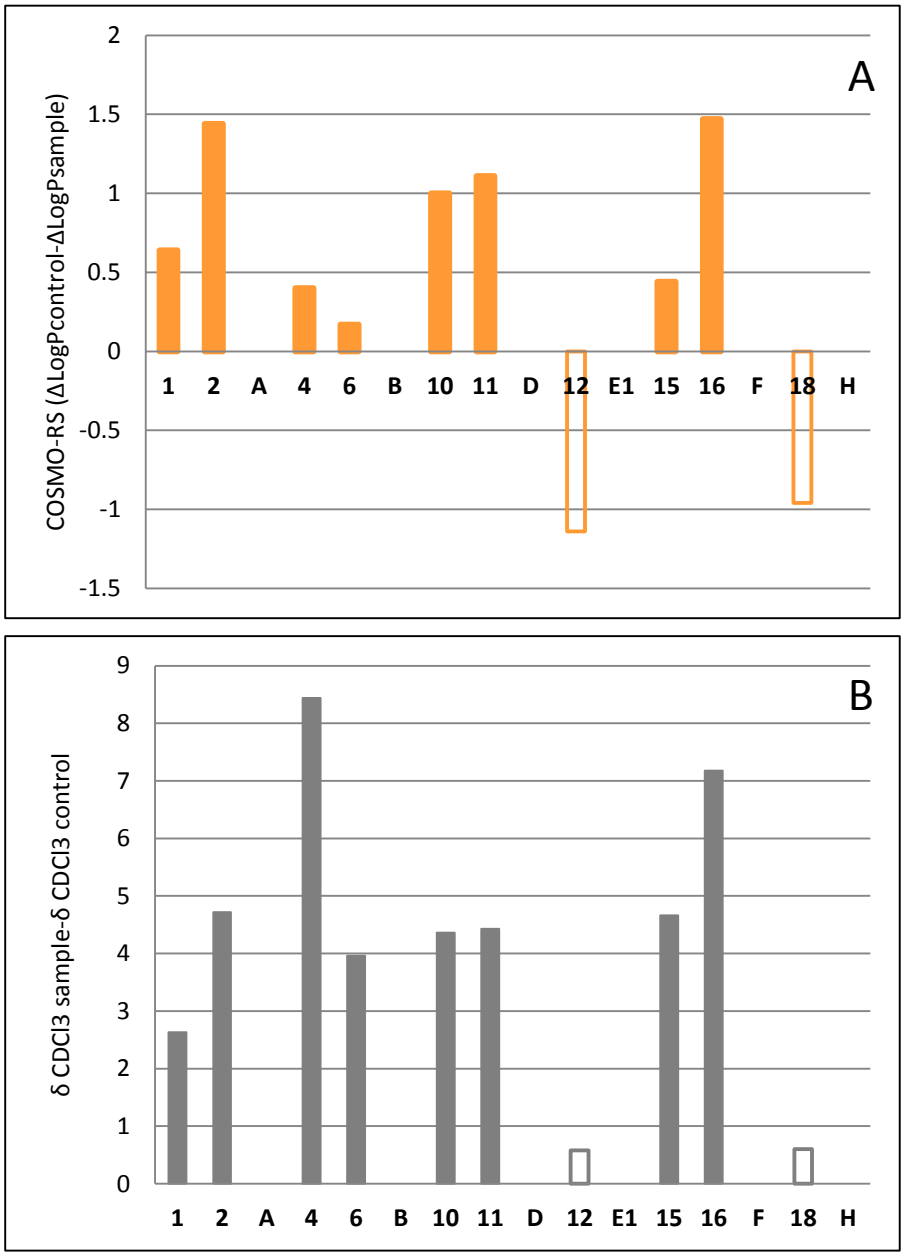


Figure 4 shows that the ΔlogP differentials (control-sample) for series of compounds **1, 2, A**; **4, 6, B**; **10, 11, D**; **15, 16, F** are positive. The differential for the δ $CDCl_3$ values for these samples is between 2 and 8. These compound pairs are from Category I, where samples have high propensity to form IMHB as described earlier.

The ΔlogP differentials for compounds **12, E1** and **18, H** are negative. And the differential of the δ $CDCl_3$ values for these samples is relatively low, below 0.6. These pairs are from Category II, characterized by poor propensity to form IMHB.

Therefore, both differentials - COSMO-RS (ΔlogPcontrol- ΔlogPsample) and (δ $CDCl_3$ sample- δ $CDCl_3$ control) - separated compounds into the same two groups, Category I and Category II, respectively.

Furthermore, the frequency of hydrogen bond occurrences associated with each topology, %HB, as defined by Kunh and coworkers [17,] is much higher for molecules we identified as having higher propensity to IMHB (Category I), than for structures with poor propensity to IMHB (Category II).

The crystallographic structures of compounds **9** and **12** were retrieved from the CSD (accessed on May 6, 2013) with codes XENKAD and FULSUA, respectively. Both compounds showed presence of IMHB in the solid state according to Kuhn's criteria. The %HB for topologies represented by **9** and **12** are 93.5% and 18%, respectively. This suggests that **9** (Figure 1) has high propensity to form IMHB not only in the solid state, but also in solution. The reverse is true for **12** which is expected to have poor propensity to form IMHB in solution. In our studies, no experimental ΔlogP data could be determined for **9**, while calculated ΔlogP values are in line with category I compounds (high propensity to form IMHB). The experimental ΔlogP and NMR data for **12** enabled its classification in category II (low propensity to form IMHB). These findings demonstrate differences in the formation of IMHB in solid state and liquid environments.

The nature of the division into Category I and II should be studied further for the purpose of identifying structures with high propensity to IMHB.

3. CONCLUSIONS

In this work $\Delta logP_{oct\text{-}tol}$ was used to evaluate the propensity of compounds to form IMHB. In particular we propose an IMHB interpretation scheme that enables categorization of compounds in two categories

- $\Delta logP_{oct\text{-}tol}$ of the control is > (larger) than $\Delta logP_{oct\text{-}tol}$ of the sample - Category I
- $\Delta logP_{oct\text{-}tol}$ of the control is < (smaller) than $\Delta logP_{oct\text{-}tol}$ of the sample - Category II

where compounds with high propensity to IMHB fall into Category I and compounds with poor propensity to IMHB fall into Category II.

This approach could be applied in early discovery projects using fast shake flask $\Delta logP_{oct\text{-}tol}$ measurements from DMSO solutions using small quantities of compound. Furthermore, we suggest that calculated values can be used prospectively on virtual compounds within a series to evaluate IMHB and potentially stretch the druggability mnemonics Ro5 and CNS MPO score to include more diverse structures.

Furthermore, it was observed that:

1. Determination of the IMHB presence requires analysis of data obtained in the same matrix environment on structures capable and not capable of IMHB.
2. $\Delta logP_{oct\text{-}tol}$ calculations by COSMOtherm provided good estimation of ΔlogP values and could be applied for the IMHB interpretation on virtual compounds, including "ideal" virtual controls, as they need not be made or tested experimentally. That approach is especially attractive in case of prospective design of IMHB and on compounds not lending themselves to experimental studies by NMR or logP due to solubility or other issues.
3. The frequency of hydrogen bond occurrences associated with each topology, %hb, as defined by Kuhn and coworkers [17] is much higher for molecules we identified as having higher propensity to IMHB (Category I), than for structures with poor propensity to IMHB (Category II).

According to the results reported here, $\Delta logP_{oct\text{-}tol}$ should be included in Medicinal Chemistry design for optimization of physical chemical properties, as a privileged molecular descriptor for delineating the propensity of compounds to form IMHB.

4. EXPERIMENTAL SECTION

4.1. MATERIALS

All compounds were obtained commercially from Sigma-Aldrich (#,**1,2,8,9,10,11,15,16, 18,C,D,F**), Alfa Aesar (#**12, 13, E1, E2**), Asinex (#**4**), SAFC/Japan (#**6**), ChemBridge (**A**), ChemDiv (**B**), ACE Synthesis (**H**), Life Chemicals (#**19, I**) and used as received. The purity of compounds was ≥95%, as specified in accompanying documentation, The degradation observed on compound #**9** lead to exclusion of all experimental measurements on that compound.

4.2. LOGP MEASUREMENTS

4.2.1. Miniaturized shake-flask. An automated, miniaturized shake flask method in a 96-well format using the automated liquid handling capabilities of the Analiza Automated Discover Workstation (ADW) and ANTEK chemilimunescent Detector was used. [41]

Sample Preparation: The compounds were prepared at 3 concentrations spanning one order of magnitude in Universal Buffer (composed of 0.15 M NaCl and 0.1 M each of phosphoric, boric, and acetic acids, adjusted to the appropriate pH with NaOH). The pH was chosen based on the pK_a of the compounds to have them as neutral species for the determination of logP values.

- *$\Delta logP_{oct\text{-}tol}$ obtained from dry powder dissolved in Universal Buffer*. The resulting samples were added as 250µL aliquots into 250µL of organic phase (saturated with the corresponding buffer), in triplicate.
- *$\Delta logP_{oct\text{-}tol}$ obtained from samples dissolved in DMSO*. DMSO stock solutions (30, 15 and 3mM) were prepared. Samples (50µL) were added to systems containing 200µL buffer phase and 250µL organic phases, in triplicate.

LogP measurements were done at pH=7.4 on most compounds. A few compounds were measured at other pHs where they were in neutral form. For example, **1**, **2**, **4** and **A** were considered to be fully neutral over pH range 1 to 9. At the same time, **6** and **B** are fully neutral in pH range above 6-7 and so are **15**, **16** and **F**. Neutrality was not fully verified for **8**, **13** (two acidic moieties) and **E2**

(acidic pKa about 2.8) and **19** and **I** (basic pKa about 9-10). pKa and pH values listed in the Supplemental Information.

Partitioning and Quantitation

The plates were sealed, vortexed, and centrifuged to aid in phase settling.

The equimolar nitrogen response of the chemiluminescent nitrogen detector was calibrated using standards which span the dynamic range of the instrument, from 0.08 to 4500µg/mL nitrogen. The ADW is used to withdraw aliquots from both the top and bottom phases in each system. These aliquots were quantified using the calibration curve and the logarithm of the ratio of the concentration in the top phase to the concentration in the bottom phase is calculated as logP.

Quantification limits in compound detection in one of the phases, which is general limitation of shake-flask method for logP values above 3.0 or below -3.0, were found. In our set the concentration of compounds **6, 8** and control **A** in the aqueous phase was near or below the quantification limit in both octanol/water and toluene/water systems, and therefore, an accurate $logP_{oct}$ or $logP_{tol}$ could not be determined for these compounds.

4.2.2. ElogD An automated RP-HPLC method developed earlier in our laboratories was used for logD_pH7.4 measurements. [43, 44] The retention of each compound is obtained at 3 different concentrations of methanol in the mobile phase and extrapolated to 0% methanol (100%water). The column is Supelcosil LC-ABZ, 5 um, 4.6 x50 mm. The Mobile Phase is comprised of 15-75% of A and 85-25% of B, consequently, where A is methanol with addition of 0.25% 1-octanol. The Mobile Phase B consists of 20 mM of MOPS (morpholine-propane-sulfonic acid) buffer prepared in octanol saturated water. Samples were prepared in 1/1 Methanol/water at about 100 ug/ml. UV detection at 5 wavelength – 210, 225, 245, 275 and 310 was used.

4.3. NMR

Samples were prepared at 250uM, 150uM or 100uM directly in NMR solvent and ^1H spectra were collected every 5 K over a 35K range. Temperature range of 280 K to 315K was used for $CDCl_3$ and a temperature range of 300K to 335K was used for DMSO-d_6. Spectra were collected on a

Bruker 600MHz equipped with a 5mm inverse TCI cryoprobe or a Bruker 500MHz spectrometer with a 5mm SmartProbe; both with a BVT3200 temperature control unit and BCU-05 cooling unit. Temperature was equilibrated within 0.1K for 10-15 minutes prior to each experiment. The solvent peak was used as reference. $CDCl_3$ was treated with base immediately prior to analysis to remove residual acid.

4.4. IN-SILICO TOOLS

4.4.1. COSMO-RS calculations Diverse sets of conformations were generated using mixed MCMM/Low-Mode algorithm and OPLS_2005 forcefield as implemented in MacroModel program (MacroModel; Schrodinger, LLC, New York, NY. 2011. http://www.schrodinger.com). The generated conformations were further optimized in aqueous media by the Turbomole package (TURBOMOLE, TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com) using the BP86 density functional [47-49] with a TZVP [50] basis set (BP-TZVP-COSMO level of theory). The generated screening charge densities of all conformations were used for COSMOtherm (Eckert F, Klamt A. 2012. COSMOtherm, Version C3.0 Release 12.01) calculations of $logP_{oct}$, $logP_{tol}$ and $\Delta logP_{oct\text{-}tol}$ properties.

4.4.2. pKa calculations. pKa calculations were performed using MoKa (Molecular Discoveries, Ltd, v.1.1.0) and ACDlabs v. 12.1.

AUTHOR INFORMATION

Corresponding authors:

For M. S.[†] *phone 1-860-441-0569;* E-mail: marina.shalaeva@pfizer.com

For G C.[‡] phone +390116708337; E-mail: giulia.caron@unito.it

Present Addresses:

For M.P. [§] Molecular Cellular and Developmental Biology Department, Yale University, New Haven, CT, 06520, USA

For G.Y. [//] Mass-Spectrometry Division, Waters Inc., 34 Maple St. Milford, MA, 01757, USA

ACKNOWLEDGEMENT

Authors would like to thank Dr. Ye Che for helpful discussions at early stages of the study and Mr. James Bradow for the analysis and purification of compound#9. We also grateful to Mrs. Aimee Kestranek and Analiza Inc. staff for the quality of logP measurements.

ABBREVIATIONS USED: IMHB- Intra Molecular Hydrogen Bonding, HB – Hydrogen bonding, 2D – two Dimensional, 3D – three Dimensional, Ro5 – Rule of Five, CNS MPO – Central Nervous System Multi-Parameters Optimization; TC – Temperature Coefficient.

ASSOCIATED CONTENT

Supporting information available: Calculated logP, logD and pKa values for the investigated dataset, pH of SF_logP measurements, as well as more details on logP data cross-validation, NMR and correlations with other partitioning systems, are available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Rezai, T., Yu, B., Millhauser, G. L., Jacobson, M. P. & Lokey, R. S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J. Am.Chem.Soc.* **2006**, 128, 2510–2511.
2. Madrid, P. B., Liou, A. P., DeRisi, J. L. & Guy, R. K. Incorporation of an intramolecular hydrogen-bonding motif in the side chain of 4-aminoquinolines enhances activity against drug-resistant P. falciparum. *J.Med.Chem.* **2006**, 49, 4535–4543.
3. McDonagh, A. F. & Lightner, D.A.Influence of conformation and intramolecular hydrogen bonding on the acyl glucuronidation and biliary excretion of acetylenic bis-dipyrrinones related to bilirubin. *J.Med.Chem.* **2007**, 50, 480–488.
4. Rafi, S. B., Hearn, B. R., Vedantham, P., Jacobson, M. P. & Renslo, A. R. Predicting and improving the membrane permeability of peptidic small molecules. *J.Med.Chem.* **2012,** 55, 3163–3169.
5. Labby,K.J.; Xue, J.M.; Ji, H.; Mataka, J.; Li, H.; Martásek, P.; Roman, L.J.; Poulos, T.L.; Silverman, R.B. Intramolecular hydrogen bonding: A potential strategy for more bioavailable inhibitors of neuronal nitric oxide synthase. *Bioorg.&Med.Chem*, **2012**, 20, 2435-2443.
6. Desai, P. V., Raub, T. J.; Blanco, M.-J.; How hydrogen bonds impact P-glycoprotein transport and permeability, *Bioorg.&Med.Chem. Let.* **2012**, 22, 6540-6548.
7. D'Andrea, P., Mauro, S., Porcelloni, M., Rossi, C., Altamura, M., Catalioto, R.M., Giuliani, S., Maggi, C.A.,Fattori, D. hNK2 receptor antagonists. The use of intramolecular hydrogen bonding to increase solubility and membrane permeability. *Bioorg.&Med.Chem. Let.* **2011,** 21, 1807–1809.
8. von Geldern, T.W.; Hoffman, D.J.; Kester, J.A.; Nellans, H.N.; Dayton, B.D.; Calzadilla, S.V.; Marsh, K.C.; Hernandez, L.; Chiou, W.; Dixon, D.B.; Wu-Wong, J.R.; Opgenorth, T.J.

Azole endothelin antagonists. 3. Using delta logP as a tool to improve absorption. *J.Med.Chem.* **1996,** 39, 982–991.

9. Alex, A., Millan, D. S., Perez, M., Wakenhut, F. & Whitlock, G.A. Intramolecular hydrogen bonding to improve membrane permeability and absorption in beyond rule of five chemical space. *Med.Chem.Comm.* **2011,** 2, 669-674.
10. Desiraju, G.R. A Bond by Any Other Name. *Angew. Chem. Int. Ed.* **2011,** 50, 52–59.
11. Carrupt, P.-A.; Testa, B.; Bechalany, A.; El Tayar, N.; Descas, P.; Perrissoud, D.; Morphine 6-glucuronide and morphine 3-glucuronide as molecular chameleons with unexpected lipophilicity. *J.Med. Chem.* **1991**, *34*, 1272-1275.
12. Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv.Drug Deliv.Rev.* **2001**, 46(1-3), 3-26.
13. Guimarães, C. R. W., Mathiowetz, A. M., Shalaeva, M., Goetz, G. & Liras, S. Use of 3D Properties to Characterize Beyond Rule-of-5 Property Space for Passive Permeation. *J.Chem. Inf.Model.* **2012**, 52, 882-890.
14. Wager, T.T,, Hou X,, Verhoest P.R,, Villalobos A. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. *ACS Chem. Neurosci.* **2010,** 1(6), 435-449.
15. Abraham, M.H.; Acree, W.E. Jr; Leo, A.J.; Hoekman, D.; Cavanaugh, J.E. Water – Solvent Partition Coefficients and DLog P Values as Predictors for Blood – Brain Distribution; Application of the Akaike Information Criterion. *J.Pharm. Sci*. **2010**, 99, 2492–2501.
16. Liu, X., Testa, B. & Fahr, A. Lipophilicity and its relationship with passive drug permeation. *Pharm.Res.* **2011**, 28, 962–977.
17. Kuhn, B., Mohr, P. & Stahl, M. Intramolecular hydrogen bonding in medicinal chemistry. *J.Med. Chem.* **2010,** 53, 2601–2611.

18. Etter, M. C. Encoding and decoding hydrogen-bond patterns of organic compounds. *Acc. Chem. Res.* **1990**, 23 (4), 120–126.

19. Bilton, C., Allen, F. H., Shields, G. P. & Howard, J. A. K. Intramolecular hydrogen bonds: common motifs, probabilities of formation and implications for supramolecular organization research papers. *Acta Cryst.* **2000**, B56, 849-856.

20. Grabowski, S. J. Hydrogen bonding strength—measures based on geometric and topological parameters. *J.Phys.Org. Chem.* **2004**, 17, 18–31.

21. Jeffrey, G. *An Introduction to Hydrogen Bonding*. Oxford University Press: New York, **1997**, pp. 218-219.

22. Young,R.C.: Mitchell,R.C.: Brown,T.H.; Ganellin,C.R.; Griffiths,R.; Jones,M.; Rana,K.K.; Saunders, D.; Smith, I.R,; Sore, N.E.; Wilks, T. J. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H 2 receptor histamine antagonists. *J.Med.Chem.* **1988**, 31, 656–671.

23. Goodwin, J.; Conradi, R.; Ho, N.; Burton, P. Physicochemical determinants of passive membrane permeability: role of solute hydrogen-bonding potential and volume. *J.Med.Chem.* **2001**, 44(22), 3721-3729.

24. El Tayar, N.; Tsai, R. S.; Testa, B.; Carrupt, P. A. & Leo, A. Partitioning of solutes in different solvent systems: the contribution of hydrogen-bonding capacity and polarity. *J.Pharm.Sci.* **1991**, 80, 590–598.

25. Zissimos, A.M.; Abraham, M.H.; Barker, M.C.; Box, K.J. & Tam, K.Y. Calculation of Abraham descriptors from solvent-water partition coefficients in four different systems; evaluation of different methods of calculation. *J. Chem. Soc., Perkin Trans.* **2002**, 2, 470-477.

26. Toulmin, A.; Wood, M.; Kenny, P. Toward prediction of alkane/water partition coefficients. *J. Med. Chem.* **2008**, 51, 3720-3730.

27. Avdeef, A. Physicochemical Profiling (Solubility, Permeability and Charge State). *Curr. Top. Med. Chem.* **2001**, 1, 277–351.

28. Steyaert, G.; Lisa, G.; Gaillard, P.; Boss, G.; Reymond, F.; Girault, H.; Carrupt, P.-A. and Testa, B. Intermolecular Forces Expressed in 1,2-Dichloroethane/Water Partition Coefficients: A Solvatochromic Analysis. *J. Chem. Soc., Faraday Trans.,* **1997**, 93, 401-406.

29. Novaroli, L.; Bouchard, G.; Caron, G.; Fruttero, R.; Carrupt, P.A.; Testa, B. The lipophilicity behaviour of COMT inhibitors, *CHIMIA*. **2002**, 56, 344- 344,

30. Caron, G. and Ermondi, G. A comparison of calculated and experimental parameters as sources of structural information: the case of lipophilicity-related descriptors. *Mini reviews in Med. Chem.* **2003**, 3, 821–830.

31. Mannhold, R.; Poda, G. I.; Ostermann, C. & Tetko, I. V. Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of Log P Methods on More Than 96,000 Compounds. *J.Pharm.Sci.* **2009**, 98, 861–893.

32. Carrupt, P. A.; Testa, B.; Gaillard, P. Computational approaches to lipophilicity: methods and applications. In: *Reviews in Computational Chemistry*, Ed.: Lipkowitz, K. B.; Boyd, D. B. Vol.11; Wiley: **1997**, 241–315. http://onlinelibrary.wiley.com/doi/10.1002/9780470125885.ch5/summary

33. Caron, G. & Ermondi, G. Calculating virtual log P in the alkane/water system (log P(N)(alk)) and its derived parameters deltalog P(N)(oct-alk) and log D(pH)(alk). *J.Med.Chem.* **2005**, 48, 3269–79.

34. Sakuratani, S., Kasai, K., Noguchi, Y. & Yamada, Y. Comparison of Predictivities of Log P Calculation Models Based on experimental Data of 134 Simple Organic Compounds. *QSAR & Combinatorial Science.* **2007**, 26, 109–116.

35. Wittekindt, C. & Klamt, A. COSMO-RS as a Predictive Tool for Lipophilicity. *QSAR & Combinatorial Science.* **2009**, 28, 874–877.

36. Alelyunas, Y. W., Pelosi-Kilby, L., Turcotte, P., Kary, M.-B. & Spreen, R. C. A high throughput dried DMSO LogD lipophilicity measurement based on 96-well shake-flask and

atmospheric pressure photoionization mass spectrometry detection. *J.Chrom.A.* **2010**, 1217, 1950–1955.

37. Ford, H., Merski, C. & Kelley, J. A rapid microscale method for the determination of partition coefficients by HPLC. *J.Liq.Chrom.* **1991**, 14, 3365–3386.
38. Mason, B. Chapter 6. High-Throughput Measurement of Physicochemical Properties. In: *Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability*. Ed.: Waterbeemd H.; Testa, B. Vol.40, 2nd ed,; Wiley: **2009**, 118. http://onlinelibrary.wiley.com/doi/10.1002/9783527623860.ch6/summary
39. Gulyaeva N, Zaslavsky A, Lechner P, Chlenov M, Chait A, Zaslavsky B. Relative hydrophobicity and lipophilicity of beta-blockers and related compounds as measured by aqueous two-phase partitioning, octanol-buffer partitioning, and HPLC. *Eur. J. Pharm. Sci.* **2002**, 17, 81-93.
40. Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews: Comp. Mol. Sci.* **2011**, 1, 699–709.
41. http://analiza.com/index.html
42. Shalaeva, M.; Philippe, L.; Plummer, M.; Che, Y.; Abramov, Y.; Kestranek, A.; Ford, M. ΔLogP as a Molecular Descriptor - Measurement and Prediction. *Keystone Symposia "Computer-Aided Drug Design",* **2010,** Whistler, Canada.
43. Lombardo, F., Shalaeva, M. Y., Tupper, K. A., Gao, F. & Abraham, M. H. ElogP oct : a tool for lipophilicity determination in Drug Discovery. *J.Med.Chem*. **2000**, 43, 2922–2928.
44. Lombardo, F., Shalaeva, M. Y., Tupper, K. A. & Gao, F. ElogD oct: A Tool for Lipophilicity Determination in Drug Discovery. 2. Basic and Neutral Compounds. *J.Med.Chem.* **2001**, 44, 2490–2497.
45. Caron, G.; Steyaert, G.; Pagliara, A.; Reymond, F.; Crivori, P.; Gaillard, P.; Carrupt, P.A.; Avdeef, A.; Comer, J.E.; Box, K.J.; Girault, H.H.; Testa, B. Structure-lipophilicity

relationships of neutral and protonated β−blockers. Part I. Intra- and Intermolecular Effects in Isotropic Solvent Systems. *Helv. Chim. Acta.* **1999**, 82, 1211-1222.

46. Baxter, N. J. and Williamson, M. P. Temperature dependence of 1H chemical shifts in proteins. *J. Biomol. NMR.* **1997**, 9, 359–369.

47. *Becke, A. Density*-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev A.* **1988**, **38,** 3098–3100.

48. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev B.* **1986**, 34, 8822–8824.

49. Perdew, J. P. Erratum: density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev B.* **1986**, 34, 7406–7406.

50. Schäfer, A., Huber, C. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.,* **1994**, 100, 5829–5835.