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

This is the author's manuscript

Original Citation:

Availability:
This version is available  http://hdl.handle.net/2318/143547 since

Published version:
DOI:10.1021/jm301850m

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)
This is an author version of the contribution published on:

Questa è la versione dell’autore dell’opera:

Integrating Intramolecular Hydrogen Bonding (IMHB) Considerations in Drug Discovery Using $\Delta \log P$ 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

DOI: 10.1021/jm301850m
Publication Date (Web): May 27, 2013

The definitive version is available at:
La versione definitiva è disponibile alla URL:

http://pubs.acs.org/doi/abs/10.1021/jm301850m
Integrating intramolecular hydrogen bonding (IMHB) considerations in drug discovery using ΔlogP as a tool.

Marina Shalaeva†*, Giulia Caron‡*, Yurii 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 $\Delta \log P_{\text{oct-tol}}$ (difference between $\log P_{\text{octanol}}$ and $\log P_{\text{toluene}}$) 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 $\Delta \log P_{\text{oct-tol}}$ 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 $\Delta \log P_{\text{oct-tol}}$ 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 $\Delta \log P_{\text{oct-tol}}$ 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 intermolecular 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. 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. In support of this notion it was also found that $\Delta \log P_{\text{oct-alk}}$ correlates with brain penetration and oral absorption.

Recent systematic work incorporating IMHB considerations in drug design has been published by Kuhn and coworkers. On the basis of pioneering work by Etter and Bilton 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) 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. 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. It has been demonstrated using solvatochromic equations that the difference between logP values obtained in different biphasic systems ($\Delta \log P$), for example octanol/water and alkane/water ($\Delta \log P_{\text{oct-alk}} = \log P_{\text{oct}} - \log P_{\text{alk}}$), is informative of IMHB when the solvents are very different from each other. More on differences in logP systems is given in Supporting Information (Annex S1, S3).
The idea that $\Delta \log P_{\text{oct-alk}}$ is informative of IMHB and the reports that $\Delta \log P_{\text{oct-alk}}$ correlates with brain penetration and oral absorption \cite{15, 16} lead us to explore $\Delta \log P_{\text{oct-tol}}$ ($\Delta \log P_{\text{oct-tol}} = \log P_{\text{oct}} - \log P_{\text{tol}}$).\cite{25-29}

The investigation of IMHB by $\Delta \log P$ was proposed some time ago \cite{30}, however this approach was not widely implemented mainly because the practical tools, both experimental and theoretical, to obtain $\log P_{\text{alk}}$ data for large series of compounds were limited.

The main goal of this study is to demonstrate that $\Delta \log P_{\text{oct-tol}}$ ($\Delta \log P_{\text{oct-tol}} = \log P_{\text{oct}} - \log P_{\text{tol}}$ ) 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 $\Delta \log P_{\text{oct-tol}}$ 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 $\log P_{\text{oct}}$ and $\log P_{\text{tol}}$ values.

The second subgoal of the study was to validate $\Delta \log P_{\text{oct-tol}}$ 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 \cite{31} was justified by two reasons: 1) unlike most logP calculators it uses the three dimensional structure of the molecules \cite{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 \cite{35} available today.
The third subgoal of the study was to validate the IMHB interpretation scheme based on measured and calculated $\Delta \log P_{\text{oct-tol}}$ 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 and controls.
2.2. ΔLOGP DETERMINATION

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

<table>
<thead>
<tr>
<th></th>
<th>octanol/water</th>
<th>toluene/water</th>
<th>octanol/water – toluene/water</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COSMO-RS logP</td>
<td>SF_logP</td>
<td>ElogD</td>
</tr>
<tr>
<td>1</td>
<td>2.53</td>
<td>1.8</td>
<td>2.13</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
<td>****</td>
<td>4.25</td>
</tr>
<tr>
<td>A</td>
<td>2.83</td>
<td>2.1*</td>
<td>2.68</td>
</tr>
<tr>
<td>4</td>
<td>4.50</td>
<td>****</td>
<td>4.58</td>
</tr>
<tr>
<td>6</td>
<td>3.05</td>
<td>2.1*</td>
<td>2.97</td>
</tr>
<tr>
<td>B</td>
<td>3.17</td>
<td>1.91</td>
<td>2.58</td>
</tr>
<tr>
<td>8</td>
<td>3.00</td>
<td>2.4*</td>
<td>3.20</td>
</tr>
<tr>
<td>9</td>
<td>2.71</td>
<td>***</td>
<td>2.89</td>
</tr>
<tr>
<td>C</td>
<td>2.41</td>
<td>N/A</td>
<td>2.92</td>
</tr>
<tr>
<td>C2</td>
<td>0.59</td>
<td>0.46</td>
<td>1.1</td>
</tr>
<tr>
<td>10</td>
<td>1.49</td>
<td>1.43</td>
<td>1.76</td>
</tr>
<tr>
<td>11</td>
<td>2.12</td>
<td>1.84</td>
<td>2.19</td>
</tr>
<tr>
<td>D</td>
<td>2.01</td>
<td>1.76</td>
<td>1.84</td>
</tr>
<tr>
<td>12</td>
<td>-1.08</td>
<td>-0.56</td>
<td>-0.07</td>
</tr>
<tr>
<td>E1</td>
<td>1.50</td>
<td>1.56</td>
<td>2.11</td>
</tr>
<tr>
<td>13</td>
<td>2.40</td>
<td>0.15</td>
<td>**</td>
</tr>
<tr>
<td>E2</td>
<td>2.38</td>
<td>1.76</td>
<td>**</td>
</tr>
<tr>
<td>15</td>
<td>0.62</td>
<td>0.83</td>
<td>0.8</td>
</tr>
<tr>
<td>16</td>
<td>2.47</td>
<td>2.17</td>
<td>3.1</td>
</tr>
<tr>
<td>F</td>
<td>0.78</td>
<td>0.95</td>
<td>1.3</td>
</tr>
<tr>
<td>18</td>
<td>0.26</td>
<td>-0.17</td>
<td>0.13</td>
</tr>
<tr>
<td>H</td>
<td>2.37</td>
<td>1.08</td>
<td>1.65</td>
</tr>
<tr>
<td>19</td>
<td>0.45</td>
<td>0.37</td>
<td>-0.69**</td>
</tr>
<tr>
<td>I</td>
<td>2.41</td>
<td>0.97</td>
<td>-0.26**</td>
</tr>
</tbody>
</table>

* 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. As a result, experimental ΔlogP_{oct-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<sub>oct</sub> and COSMO-RS logP<sub>tol</sub> 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<sup>40</sup>, 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<sup>41</sup> 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<sup>42-44</sup> are given in the Supporting Information (Annex S2).

*Figure 2. Cross validation of lipophilicity data. A) Validation of experimental SF_logP<sub>oct</sub> with ElogD. B) Validation of computed COSMO-RS logP<sub>oct</sub> with experimental SF_logP<sub>oct</sub> 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 ΔlogP.

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 $\log P_{\text{oct}}$ and $\log P_{\text{tol}}$ (i.e. $\Delta \log P$) should reflect the propensity of a compound to form IMHB. The $\Delta \log P$ value by itself does not indicate the formation of IMHB. However, trends are observed if comparisons made in a pairwise fashion for compounds in a series capable of forming IMHB (samples) and not capable of IMHB (controls). In particular, the comparative analysis of $\Delta \log P_{\text{oct-tol}}$ in Table 1 reveals two possible situations:

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

The first situation (Category I), when $\Delta \log P_{\text{oct-tol}}$ 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 $\Delta \log P_{\text{oct-tol}}$ of the control is lower than $\Delta \log P_{\text{oct-tol}}$ 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 $\Delta \log P_{\text{oct-tol}}$.

The influence of the experimental error associated with measured $\Delta \log P_{\text{oct-tol}}$ 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 $\Delta \log P$ of Sample and Control. Therefore, the classification threshold is defined by the error and the difference in $\Delta \log P$ has to be higher than 0.4, for clear
categorization. In other words, if $\Delta \log P$ 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 $\Delta \log P_{\text{oct-tol}}$ values presented in Table 1 demonstrate that the same pairs of compounds fall into the same categories as described above for experimental $\Delta \log P_{\text{oct-tol}}$ values. The only deviation was observed for compound 19 and Control I, which were placed into Category I by calculated $\Delta \log P_{\text{oct-tol}}$ and into Category II by measured $\Delta \log P_{\text{oct-tol}}$. This discrepancy stems mostly from the significant difference between predicted and experimental logP value for the control I (2.41 and 0.97, respectively). The COSMO-RS logP 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 higher for the entire set. Recall, compound 6 and Control B categorization by experimental data are uncertain, as discussed above; however, COSMO-RS $\Delta \log P_{\text{oct-tol}}$ 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 $\Delta \log P_{\text{oct-tol}}$ values, even though the low solubility and instability of compounds 4 and 9, respectively, prevented us from obtaining measured $\Delta \log P_{\text{oct-tol}}$ values. A second obvious benefit is the prospective use of calculated $\Delta \log P$ 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<sub>oct</sub> values and their difference in logP<sub>tol</sub> 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<sub>oct</sub> and logP<sub>tol</sub> 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\textsuperscript{25} and 1,2-dichloroethane/water systems\textsuperscript{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\textsuperscript{17}

<table>
<thead>
<tr>
<th></th>
<th>δ\textsubscript{CDCl\textsubscript{3}}\textsuperscript{300K}</th>
<th>δ\textsubscript{CDCl\textsubscript{3} sample} - δ\textsubscript{CDCl\textsubscript{3} control}</th>
<th>ΔlogPcontrol - ΔlogPsample</th>
<th>%HB predicted by topology\textsuperscript{17}</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.69</td>
<td>2.63</td>
<td>0.64</td>
<td>93</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>10.78</td>
<td>4.72</td>
<td>1.44</td>
<td>93</td>
<td>I</td>
</tr>
<tr>
<td>A</td>
<td>6.06</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12.12</td>
<td>8.44</td>
<td>0.4</td>
<td>90</td>
<td>I</td>
</tr>
<tr>
<td>6</td>
<td>7.64</td>
<td>3.96</td>
<td>0.17</td>
<td>90</td>
<td>I</td>
</tr>
<tr>
<td>B</td>
<td>3.68</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10.37</td>
<td>4.36</td>
<td>1</td>
<td>85</td>
<td>I</td>
</tr>
<tr>
<td>11</td>
<td>10.44</td>
<td>4.43</td>
<td>1.11</td>
<td>85</td>
<td>I</td>
</tr>
<tr>
<td>D</td>
<td>6.01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>6.1</td>
<td>0.58</td>
<td>-1.14</td>
<td>18</td>
<td>II</td>
</tr>
<tr>
<td>E1</td>
<td>5.52</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>9.24</td>
<td>4.66</td>
<td>0.44</td>
<td>81</td>
<td>I</td>
</tr>
</tbody>
</table>
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 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 $\Delta$logP vs. NMR chemical shift trends demonstrated on differentials between values for sample and control. A) COSMO-RS ($\Delta$logP<sub>control</sub> - $\Delta$logP<sub>sample</sub>$); B) $\delta$ CDCl<sub>3</sub> sample – $\delta$ CDCl<sub>3</sub> control.

Figure 4 shows that the $\Delta$logP 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 $\delta$ CDCl<sub>3</sub> 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 CDCl₃ 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 CDCl₃ sample- δ CDCl₃ 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 $\Delta \log P_{\text{oct-tol}}$ 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:

- $\Delta \log P_{\text{oct-tol}}$ of the control is $>$ (larger) than $\Delta \log P_{\text{oct-tol}}$ of the sample - Category I
- $\Delta \log P_{\text{oct-tol}}$ of the control is $<$ (smaller) than $\Delta \log P_{\text{oct-tol}}$ 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 $\Delta \log P_{\text{oct-tol}}$ 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. $\Delta \log P_{\text{oct-tol}}$ calculations by COSMOtherm provided good estimation of $\Delta \log P$ 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, $\Delta \text{log}P_{\text{oct-tol}}$ 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 chemiluminescent 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\textsubscript{a} of the compounds to have them as neutral species for the determination of logP values.

- \( \Delta \log P_{\text{oct-tol}} \) 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.

- \( \Delta \log P_{\text{oct-tol}} \) 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.
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<sub>oct</sub> or logP<sub>tol</sub> 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<sub>pH7.4</sub> measurements. 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 µm, 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 µg/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 <sup>1</sup>H spectra were collected every 5 K over a 35K range. Temperature range of 280 K to 315K was used for CDCl<sub>3</sub> and a temperature range of 300K to 335K was used for DMSO-<i>d</i><sub>6</sub>. 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. CDCl₃ was treated with base immediately prior to analysis to remove residual acid.

4.4. IN-SILICO TOOLS


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.


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


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


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


