XXXII Symposium on Bioinformatics and Computer-Aided Drug Discovery (BCADD-2026)

Emerging Challenges and Opportunities for In Silico Drug Discovery

LIPOPHILICITY IN DRUG DESIGN
EVOLUTION FROM THE RULE OF FIVE TO AI-DRIVEN DISCOVERY

Dr. William J. Zamora R

October, 2026

UCR

Bio^3
COMPUTACIONAL BIOPHYSICS, BIOLOGY AND BIOINFORMATICS GROUP

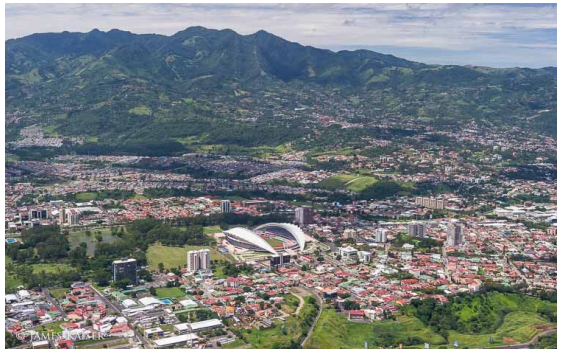
aToxCIA
Laboratorio de Toxicología Computacional e Inteligencia Artificial
LEBI Laboratorio de Ensayos Biológicos

UNA
UNIVERSIDAD NACIONAL
COSTA RICA

About Costa Rica









21.04.25

14:14

Outline



INTRODUCTION



OBJECTIVES



METHODS



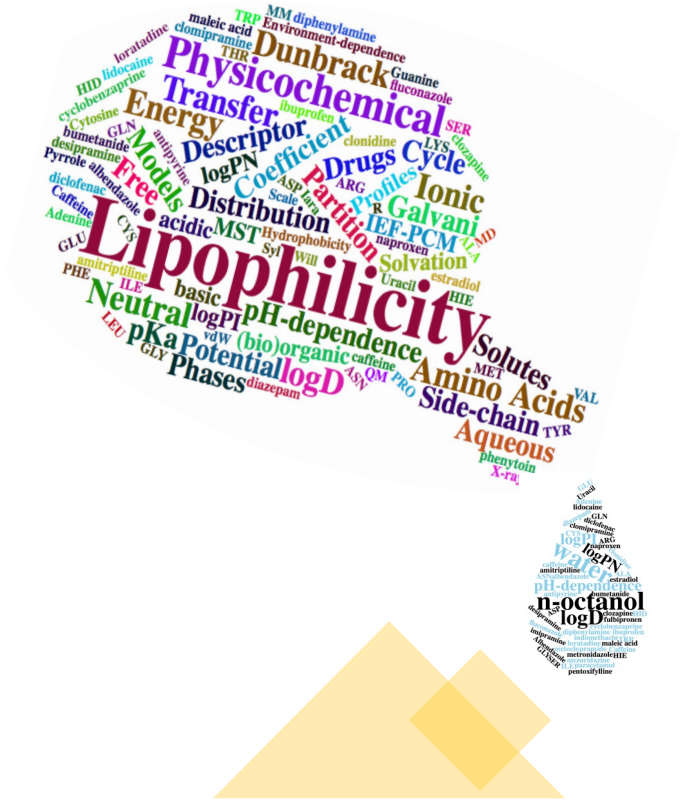
RESULTS

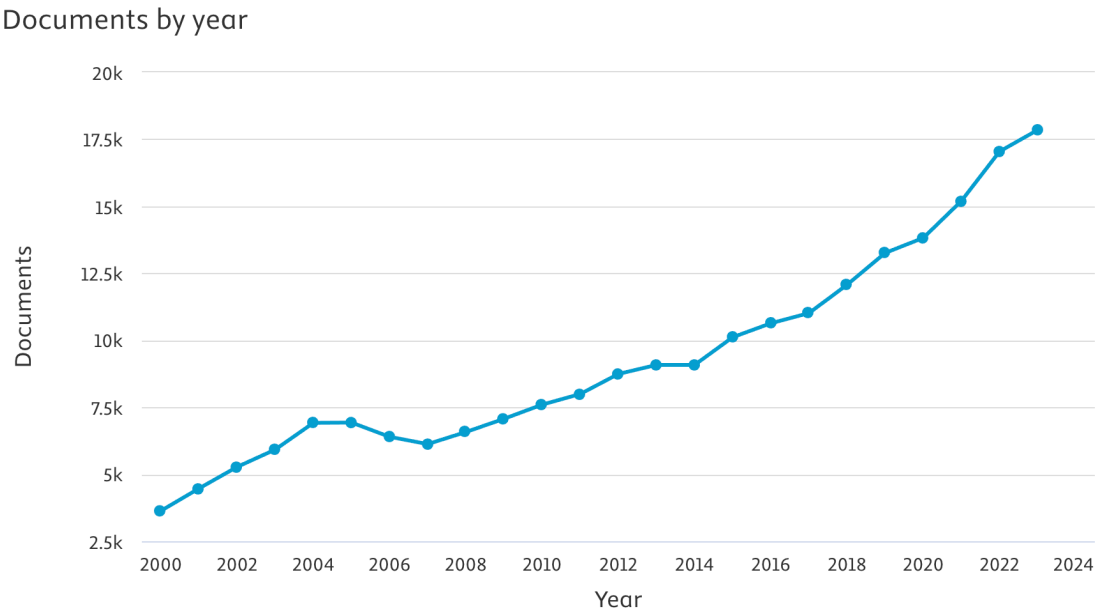


CONCLUSIONS AND CHALLENGES

Introduction

- *Timeline of Lipophilicity or Hydrophobicity Terms*

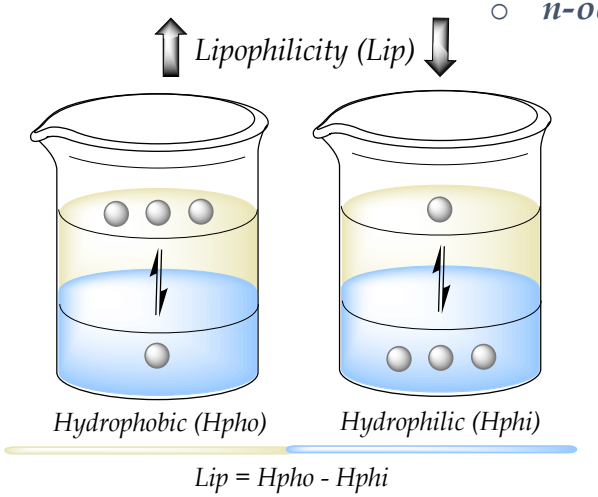




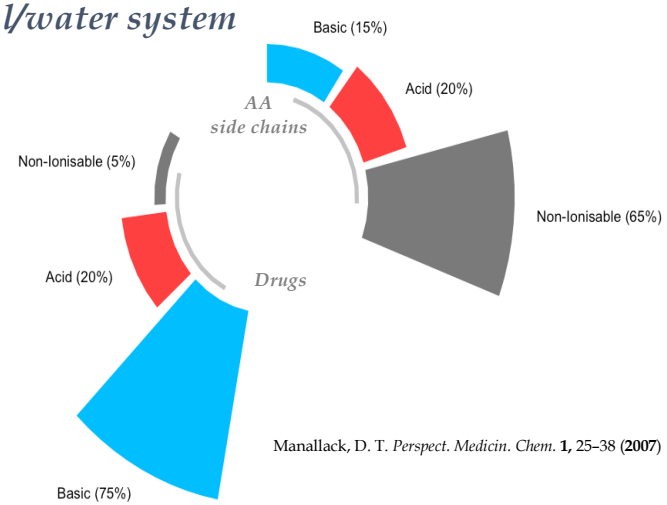
Source: www.scopus.com

Introduction

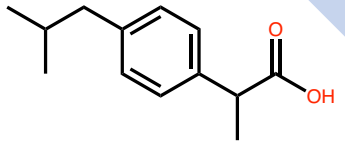
- *n-octanol/water system*



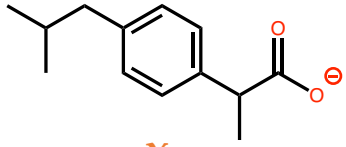
Lip = Hpho - Hphi



Manallack, D. T. *Perspect. Medicin. Chem.* **1,** 25–38 (**2007**)



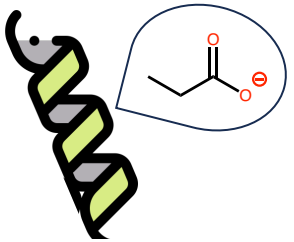
HX



X-

$$\log P_N = \log\left(\frac{\left[HX^N\right]_o}{\left[HX^N\right]_w}\right)$$

$$\log D_{pH} = \log\left(\frac{\left[\mathrm{HX{+}X^-}\right]_o}{\left[\mathrm{HX{+}X^-}\right]_w}\right)$$

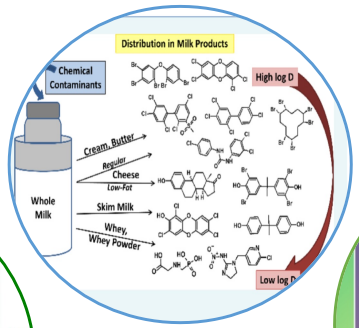


Introduction

$\log P_N$

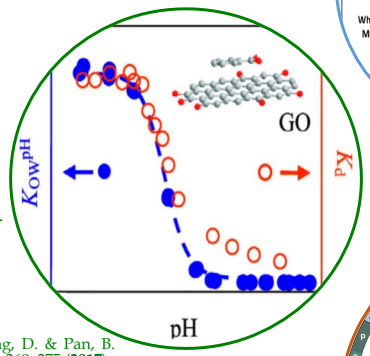
$\log D_{pH}$

Food Science



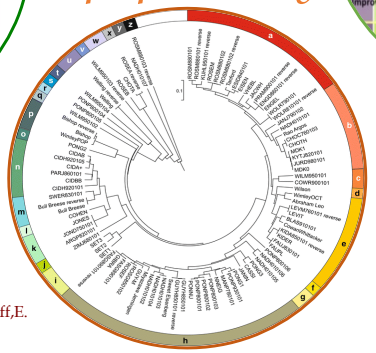
Lupton, S. J. et al.
J. Agric. Food Chem. 66, 306–314 (2018).

Materials & Environment

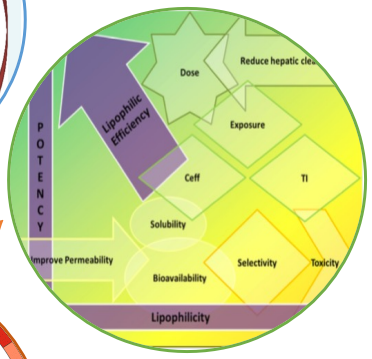


Li, H., Cao, Y., Zhang, D. & Pan, B.
Sci. Total Environ. **618,** 269–275 (**2017**).

Lipophilicity



Simm, S., Einloft, J., Mirus, O. & Schleiff,E.
Biol. Res. **49,** 31 (**2016**).



Drug Development

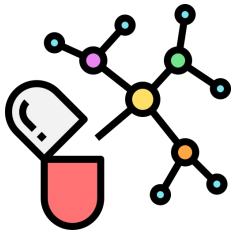
Johnson, T. W., Gallego, R. A. & Edwards, M.P.
J. Med. Chem. **61,** 6401–6420 (**2018**).

Introduction

$\log P_N$

$\log D_{pH}$

- *Drug Development*



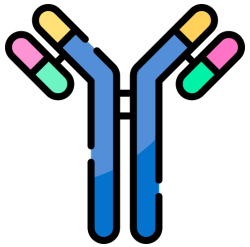
Small molecules

MW < 500-700 Da



Peptides

MW ~ 500–5000 Da



Therapeutic proteins

MW > 5000 Da

Introduction

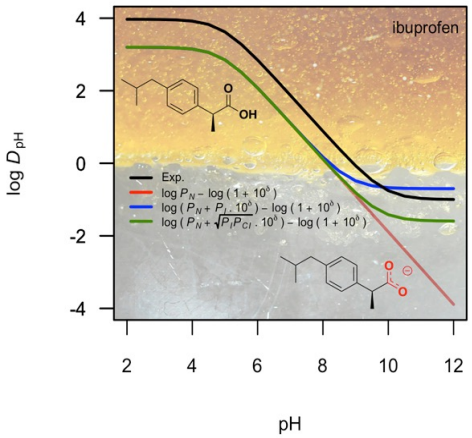
$\log P_N$ ○ *n-octanol/water system* $\log D_{pH}$

THE JOURNAL OF PHYSICAL CHEMISTRY B — Article — pubs.acs.org/JPCB

Prediction of pH-Dependent Hydrophobic Profiles of Small Molecules from Miertus–Scrocco–Tomasi Continuum Solvation Calculations

Published as part of The Journal of Physical Chemistry virtual special issue "Manuel Yáñez and Otilia Mó Festschrift".

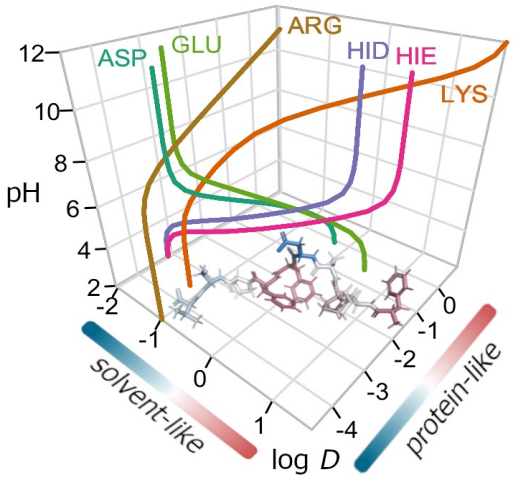
William J. Zamora,[†,‡] Carles Curutchet,[‡] Josep M. Campanera,[*,‡] and F. Javier Luque[*,†]



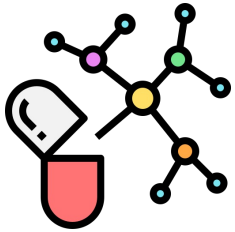
THE JOURNAL OF PHYSICAL CHEMISTRY Letters — Letter — Cite This: J. Phys. Chem. Lett. 2019, 10, 883–889 — pubs.acs.org/JPCL

Development of a Structure-Based, pH-Dependent Lipophilicity Scale of Amino Acids from Continuum Solvation Calculations

William J. Zamora, Josep M. Campanera,* and F. Javier Luque*



Introduction



Small molecules

MW < 500 Da

The Rule of 5

1. 5 H-bond donors
2. 10 H-bond acceptors
3. Molecular weight < 500 Da
4. **Log *P* < 5**.

Drug Development

Acceptable level of in vivo **clearance** for a drug is **log*D* of 0–3**

log *P* < 3.25 provided the best odds for obtaining >100 μM **solubility** for a typical drug molecule (melting point of 150 °C)

$$\log(\mathrm{Sol}_{aq}) = 0.5 - \log P - 0.01(\mathrm{MP}\text{-}25)$$

Acceptable **permeability log *D* of 1–3.5**

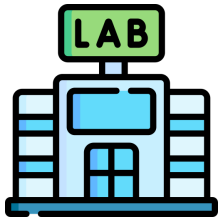
LipE should be prioritized based on the importance of **enthalpic optimization**

$$\mathrm{LipE} = \mathrm{LLE} = -\log(\mathrm{potency}) - \log D$$

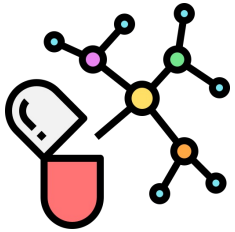
to achieve **low dose** and **dosing frequency**

Introduction

Drug Development







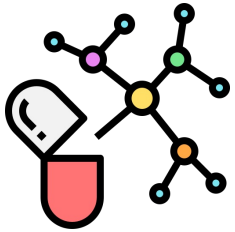
Small molecules

MW < 500 Da

	Shake-flask	pH-metric	Centrifugal partition chromatography [7][8]	Cyclic voltammetry [9][10]	RP-HPLC [11–17]
Measurement	Direct single points	Direct lipophilicity profiles	Direct single points	Indirect single points	Indirect single points
Log *P* range	–3 to +3	–1 to +8	–3 to +3	–8 to +1	0 to +5
Partition solvents	various	various	various	1,2-dichloroethane, nitrobenzene	methanol, acetonitrile
pH range	0 to 14	1.8 to 12.2	0 to 14	0 to 14	2 to 7.4
pH established by…:	…buffer solution	…full pH range measured during experiment	…buffer solution	…buffer solution	…buffer solution
Substance amount	10–50 mg	2–50 mg	<1 mg	1–10 mg	1–10 mg
Notes	Classical method, tedious, limited precision, phase-separation problems	Only ionizable solutes, fast	Detection problems, difficult stabilization of the system	Only ions	Interaction with the column

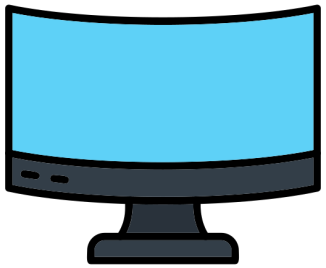
Introduction

- Drug Development

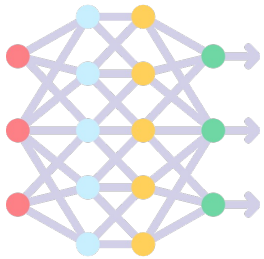


Small molecules

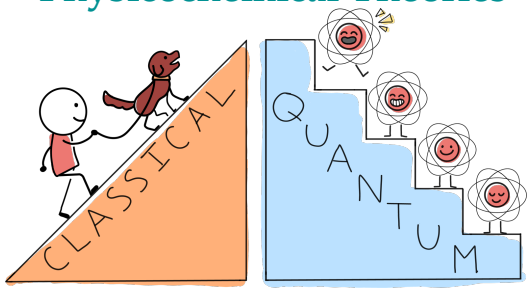
MW < 500 Da



Data exploration



Physicochemical Theories

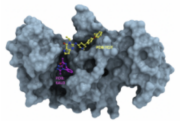


Introduction

SAMPL challenges

Statistical Assessment of Modeling of Proteins and Ligands (SAMPL)

CHALLENGE #6

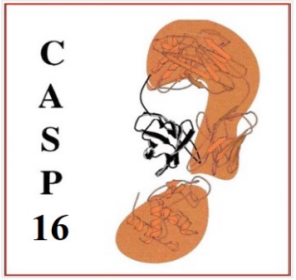


Finding ligands targeting the triple Tudor domain of SETDB1

The sixth CACHE challenge is focused on SETDB1, a multi domain protein involved in epigenetic mechanisms and an immuno-oncology target. Participants are asked to find molecules occupying one or multiple subcavities of the histone binding groove of the SETDB1 triple Tudor domain (TTD). A few ligands (PDB codes 7CJT, 8UWP, 6AU3) occupy one or two of three subcavities that accommodate methylated or acetylated lysine sidechains, as shown in this structure overview.

Critical Assessment of Computation Hit-finding Experiments (CACHE)

The Critical Assessment of Protein Structure Prediction (CASP)



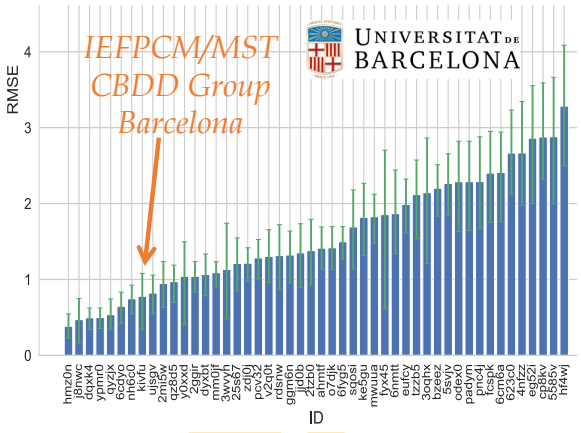
Introduction

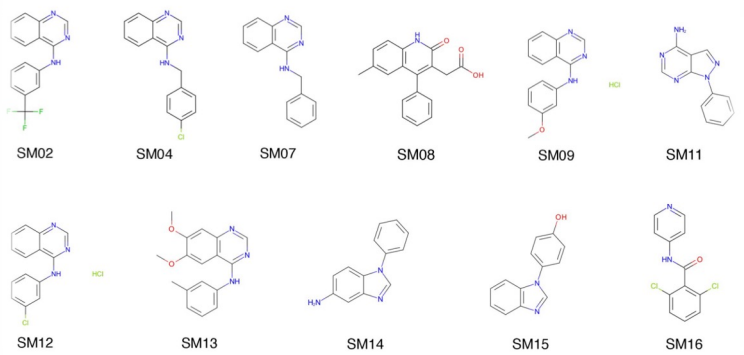
Journal of Computer-Aided Molecular Design (2020) 34:443–451
https://doi.org/10.1007/s10822-019-00262-4

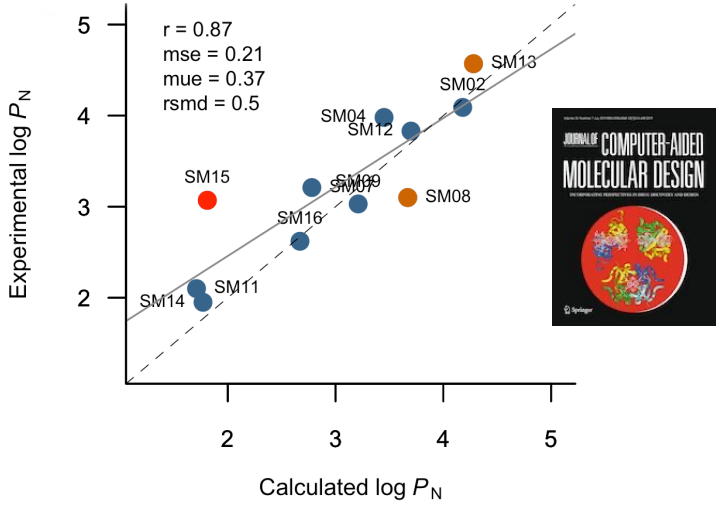
Prediction of the *n*-octanol/water partition coefficients in the SAMPL6 blind challenge from MST continuum solvation calculations

William J. Zamora[1,2] · Silvana Pinheiro[2] · Kilian German[3] · Clara Ràfols[3] · Carles Curutchet[2] · F. Javier Luque[1]

Received: 15 October 2019 / Accepted: 21 November 2019 / Published online: 27 November 2019
© Springer Nature Switzerland AG 2019





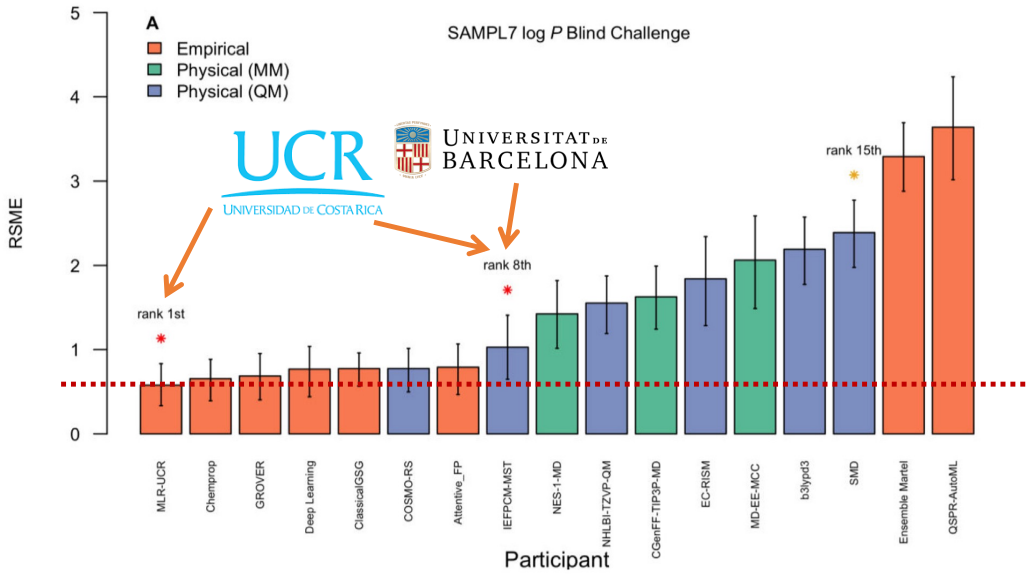


Introduction

Prediction of *n*-octanol/water partition coefficients and acidity constants (pK_a) in the SAMPL7 blind challenge with the IEFPCM-MST model

Antonio Viayna[1] · Silvana Pinheiro[2] · Carles Curutchet[3] · F. Javier Luque[1] · William J. Zamora[4,5]

Received: 23 February 2021 / Accepted: 4 June 2021 / Published online: 10 July 2021
© The Author(s) 2021, corrected publication 2021

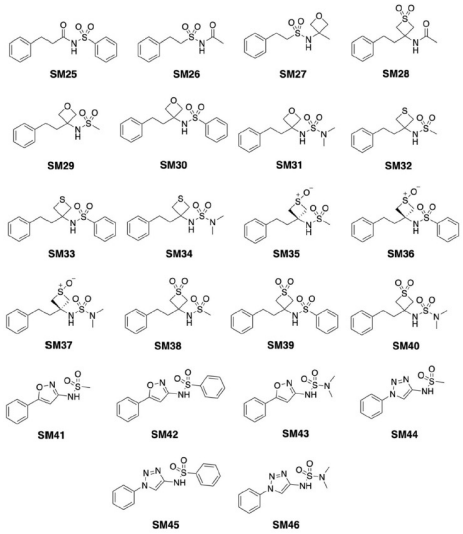


ELSEVIER
Artificial Intelligence in the Life Sciences
Volume 3, December 2023, 100077

Trends and challenges in chemoinformatics research in Latin America

Jazmín Miranda-Salas [a], Carlos Peña-Varas [b c], Ignacio Valenzuela Martínez [b], Dionisio A. Olmedo [d], William J. Zamora [e f i], Miguel Angel Chávez-Fumagalli [g], Daniela Q. Azevedo [h], Rachel Oliveira Castilho [h], Vinicius G. Maltarollo [h], David Ramírez [b], José L. Medina-Franco [a]

Fig. 1 Dataset of 22 small molecules proposed in the SAMPL7 log *P* challenge



Introduction

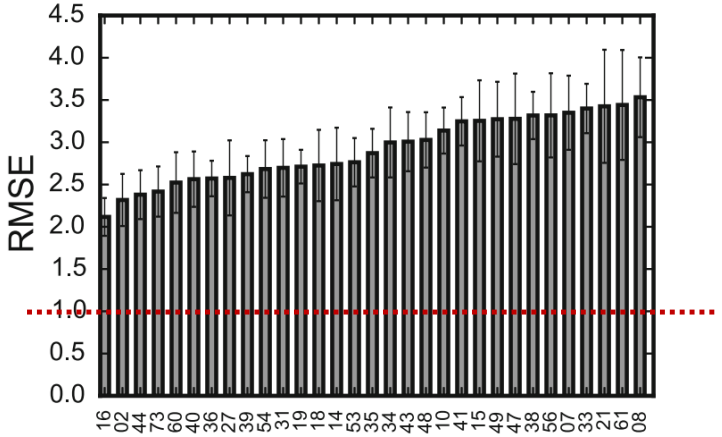
$\log D_{pH}$

J Comput Aided Mol Des
DOI 10.1007/s10822-016-9954-8

Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge

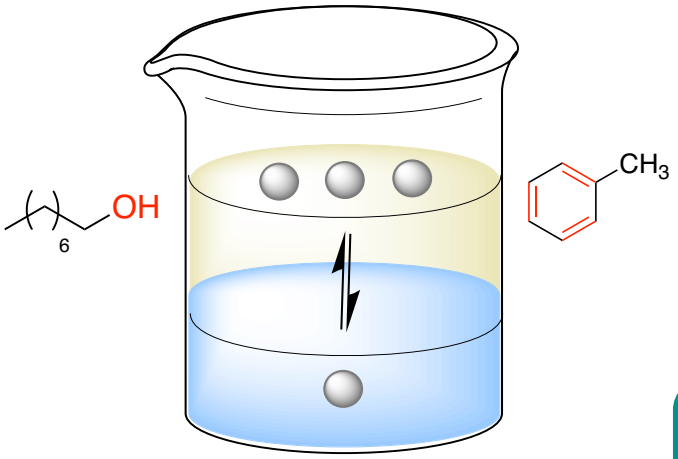
Caitlin C. Bannan[1] · Kalistyn H. Burley[2] · Michael Chiu[3] · Michael R. Shirts[4] · Michael K. Gilson[5] · David L. Mobley[1,2]

$$\log D = \log P_N - \log(1 + 10^{\delta})$$



Introduction

Beyond the n-octanol/water Partition Coefficient in Drug Discovery

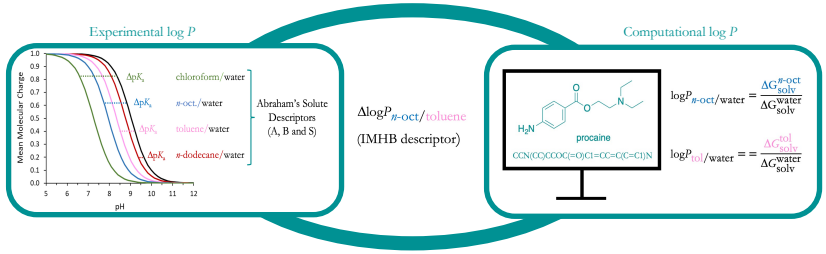


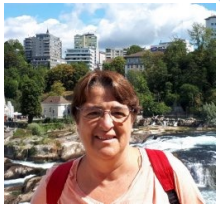
ELSEVIER

European Journal of Pharmaceutical Sciences
Volume 168, 1 January 2022, 106066

Molecular characteristics of several drugs evaluated from solvent/water partition measurements: Solvation parameters and intramolecular hydrogen bond indicator

Rebeca Ruiz [a], William J. Zamora [b c], Clara Ràfols [d], Elisabeth Bosch [d]





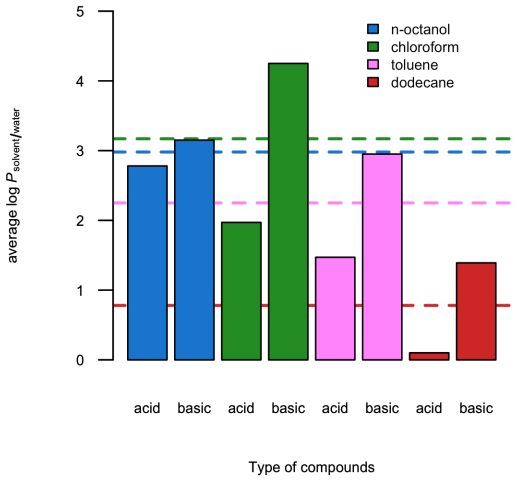
UNIVERSITAT DE BARCELONA

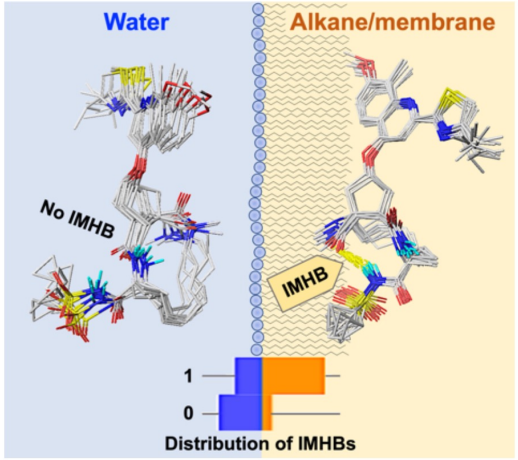


Pion

Introduction

Beyond the n-octanol/water Partition Coefficient in Drug Discovery





David L, Wenlock M, Barton P, Ritzén A. Prediction of Chameleonic Efficiency. ChemMedChem. **2021** Sep 6;16(17):2669-2685. doi: 10.1002/cmdc.202100306.

Objetives

Themed collection

The SAMPL Challenges

Physical Chemistry Chemical Physics

ROYAL SOCIETY OF CHEMISTRY

SAMPL challenges

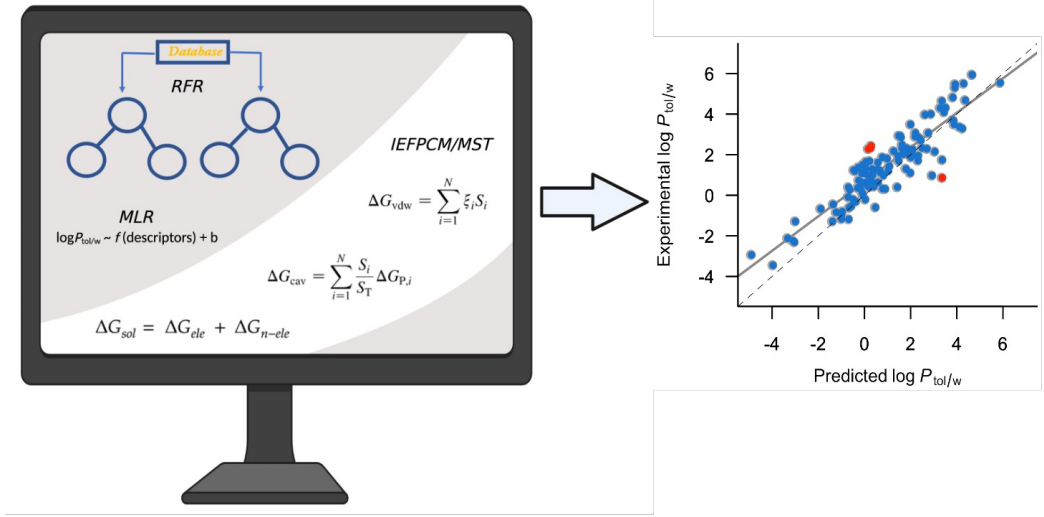
zenodo Search Upload Communities

May 1, 2023 Dataset Open Access

Toluene/Water Partition Coefficient Database

William J. Zamora; Silvana Pinheiro; Antonio Viayna; Carles Curutchet; Laia Bisbal; Rebeca Ruiz; Clara Ràfols; F. Javier Luque

In recent years the use of partition systems other than the widely used biphasic *n*-octanol/water has received increased attention to gain insight into the molecular features that dictate the lipophilicity of compounds. Thus, the difference between *n*-octanol/water and toluene/water partition coefficients has proven to be a valuable descriptor to study the propensity of molecules to form intramolecular hydrogen bonds and exhibit chameleon-like properties that modulate solubility and permeability. In this context, a dataset of **252 unique molecules** with experimentally determined **toluene/water partition coefficient** is deposited here.



PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

Check for updates

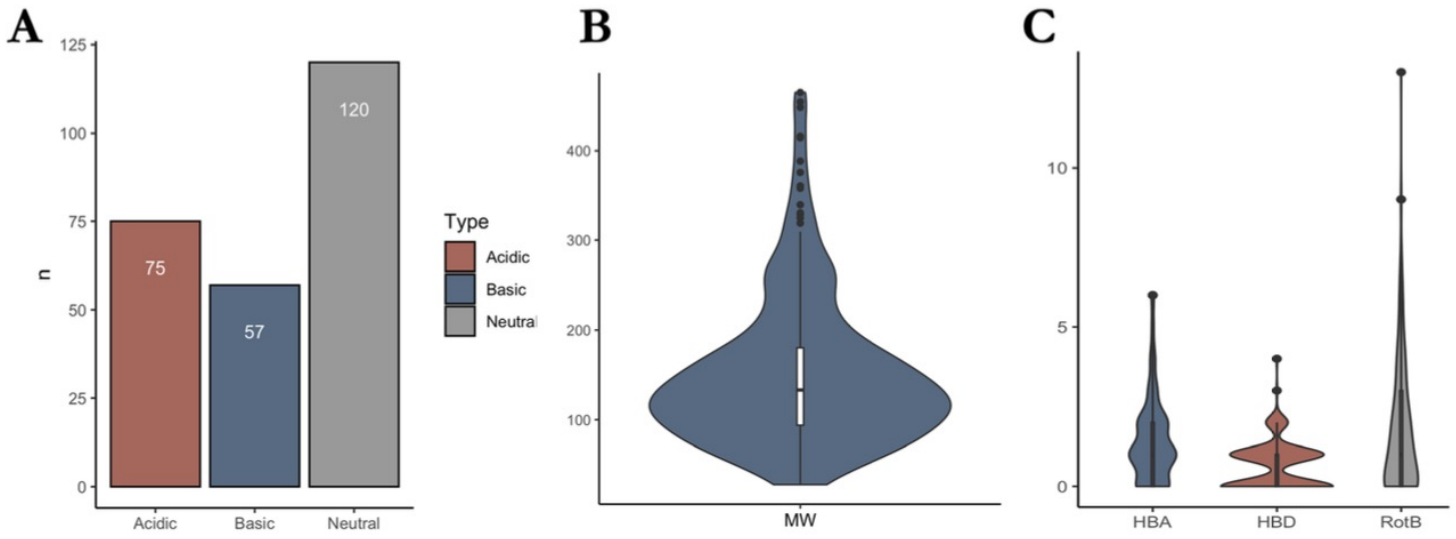
Cite this: *Phys. Chem. Chem. Phys.*, 2023, **25**, 17952

Prediction of toluene/water partition coefficients in the SAMPL9 blind challenge: assessment of machine learning and IEF-PCM/MST continuum solvation models†

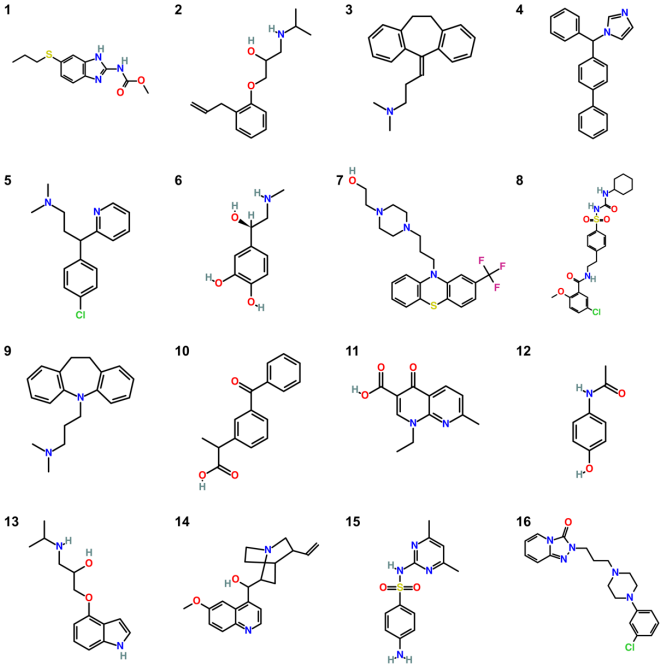
William J. Zamora, *[abc] Antonio Viayna, [def] Silvana Pinheiro, [ab] Carles Curutchet, [fg] Laia Bisbal,[eh] Rebeca Ruiz, [i] Clara Ràfols *[eh] and F. Javier Luque *[def]

Methods

- Distribution of the database of 252 small molecules



Methods



The External Set

Table 3 Experimental log $P_{tol/w}$ values determined for the SAMPL9 dataset

ID	Name	log $P_{tol/w}$	Standard deviation	Exp. equipment
1	Albendazole	3.76	0.03	Pion SiriusT3
2	Alprenolol	2.40	0.01	Pion SiriusT3
3	Amitriptyline	5.51	0.01	Sirius D-PAS & GLpKa
4	Bifonazole	5.47	0.06	Pion SiriusT3
5	Chlorpheniramine maleate salt	3.61	0.01	Sirius D-PAS & GLpKa
6	Epinephrine	-1.23	0.14	Pion SiriusT3
7	Fluphenazine dihydrochloride	4.37	0.04	Sirius D-PAS & GLpKa
8	Glyburide	2.79	0.02	Pion SiriusT3
9	Imipramine hydrochloride	5.05	0.02	Sirius D-PAS & GLpKa
10	Ketoprofen	2.47	0.02	Pion SiriusT3
11	Nalidixic acid	1.46	0.01	Sirius D-PAS & GLpKa
12	Paracetamol	-1.59	0.06	Pion SiriusT3
13	Pindolol	0.36	0.01	Pion SiriusT3
14	Quinine	1.41	0.01	Sirius D-PAS & GLpKa
15	Sulfamethazine	-0.74	0.03	Sirius D-PAS & GLpKa
16	Trazodone hydrochloride	3.77	0.07	Pion SiriusT3



log $P_{tol/w}$

UCR UNIVERSIDAD DE COSTA RICA

UNIVERSITAT DE BARCELONA

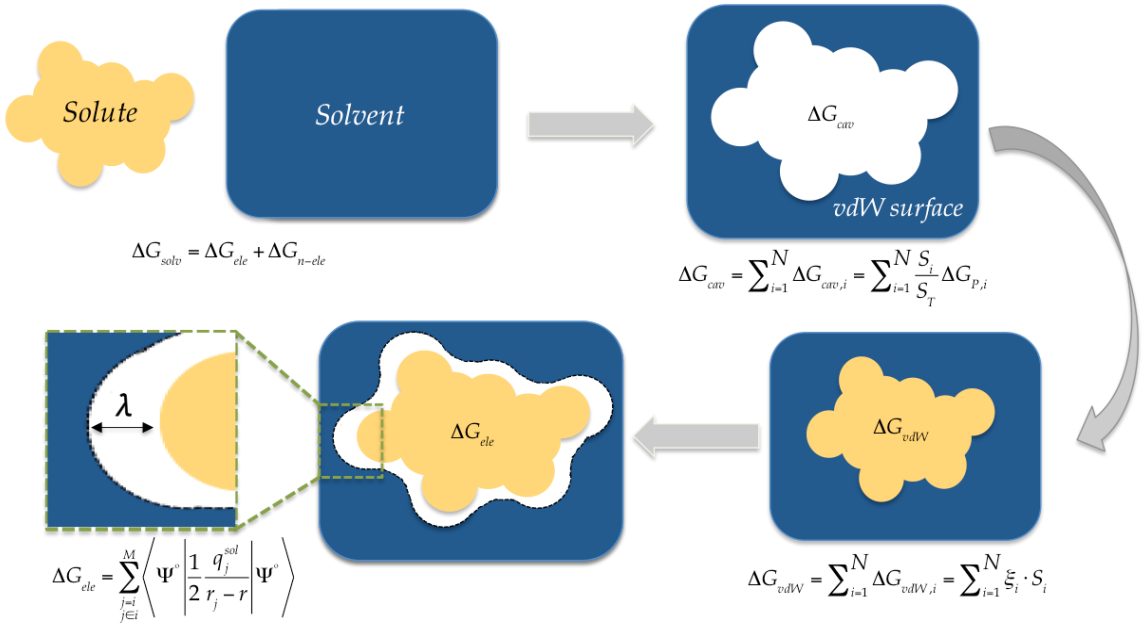
Pion

Methods

UNIVERSITAT DE BARCELONA

IEF-PCM/MST solvation model for toluene

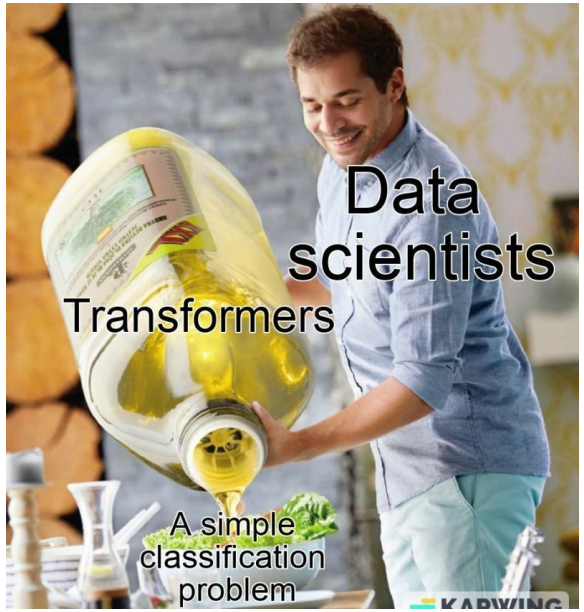
UCR UNIVERSIDAD DE COSTA RICA

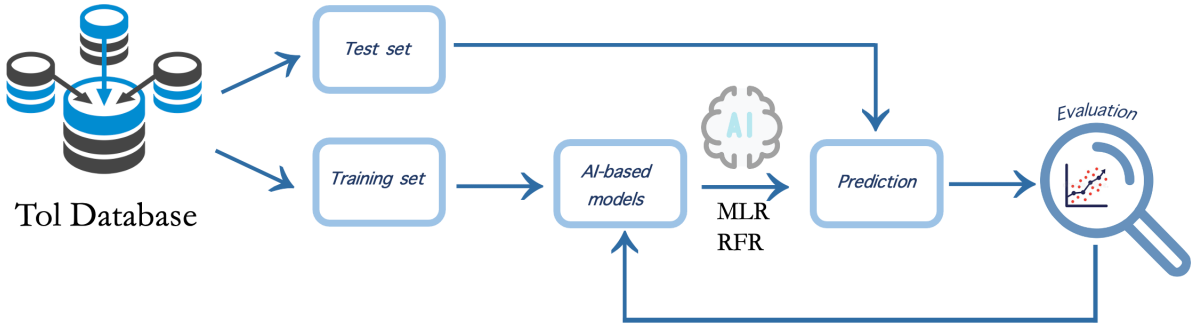


$$\Delta G_{sol}^{exp} = \Delta G_{sol}^{w} - 2.303 \cdot R \cdot T \cdot \log P_{org/w}$$

$$\Delta G_{sol}^{res} = \Delta G_{sol}^{exp} - \Delta G_{ele} - \Delta G_{cav}$$

Methods

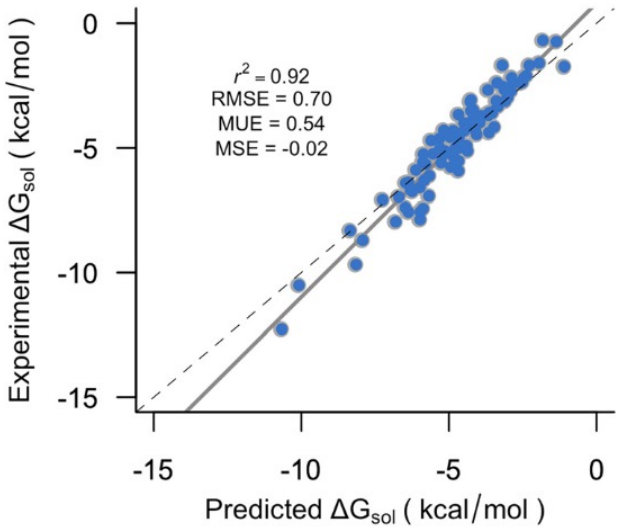




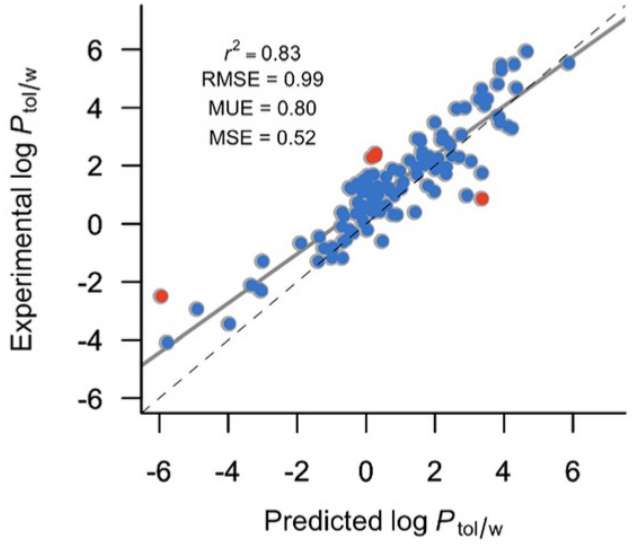
$$\log P_{\mathrm{tol/w}} = \sum_{i=1}^{n} c_i d_i + b \qquad \log P_{\mathrm{tol/w}} = \frac{1}{T}\sum_{i=1}^{n} t_i(X_{\mathrm{E}})$$

Results

- Solvation in Toluene



- Test Set Toluene/water partitions

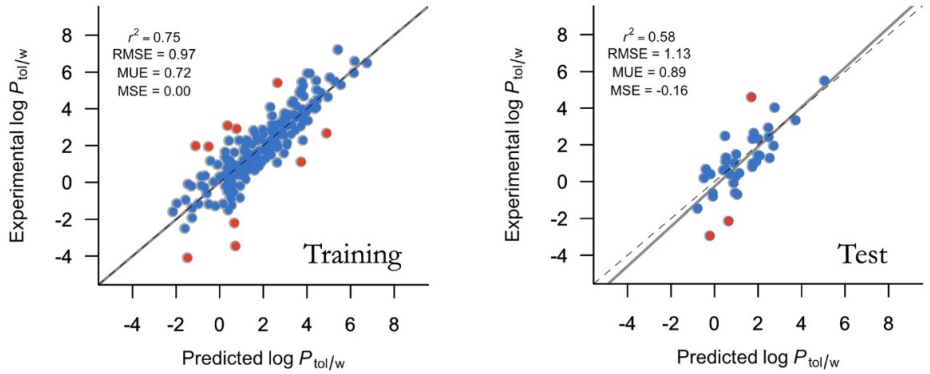


IEF-PCM/MST solvation model for toluene

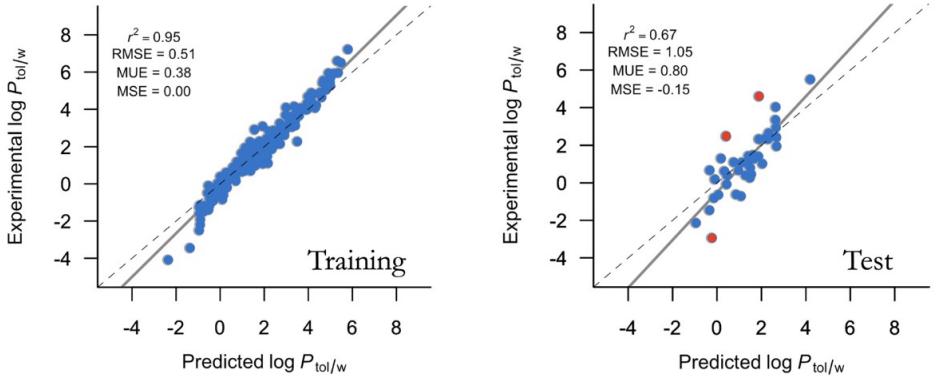
Results

Training and Test Sets

MLR



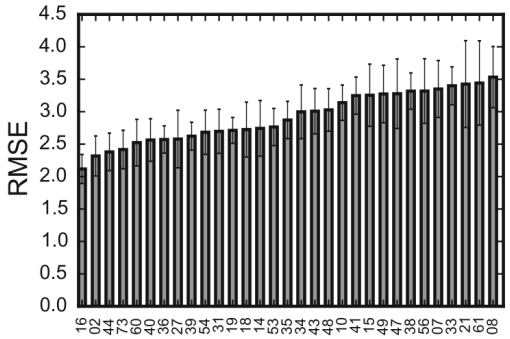
RFR



J Comput Aided Mol Des
DOI 10.1007/s10822-016-9954-8

Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge

Caitlin C. Bannan[1] · Kalistyn H. Burley[2] · Michael Chiu[3] · Michael R. Shirts[4] · Michael K. Gilson[5] · David L. Mobley[1,2]



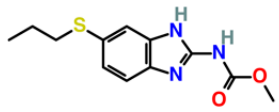
Results

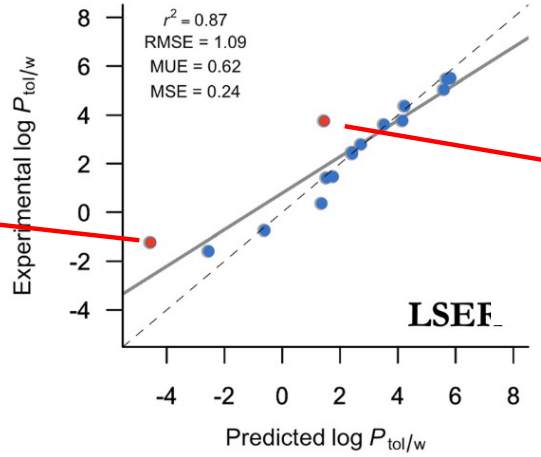
External Set

UFZ-LSER database

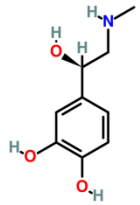
UFZ HELMHOLTZ Centre for Environmental Research

1





6

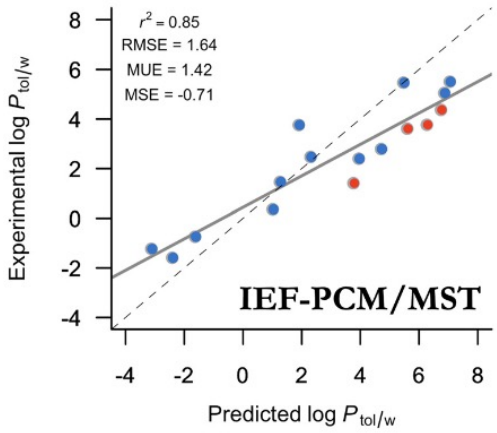


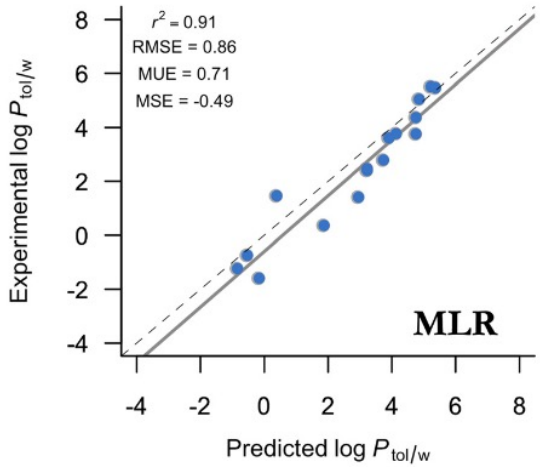
$$\log P_{tol/w} = 0.143 + 0.527 \cdot E - 0.720 \cdot S - 3.010 \cdot A - 4.824 \cdot B + 4.545 \cdot V$$

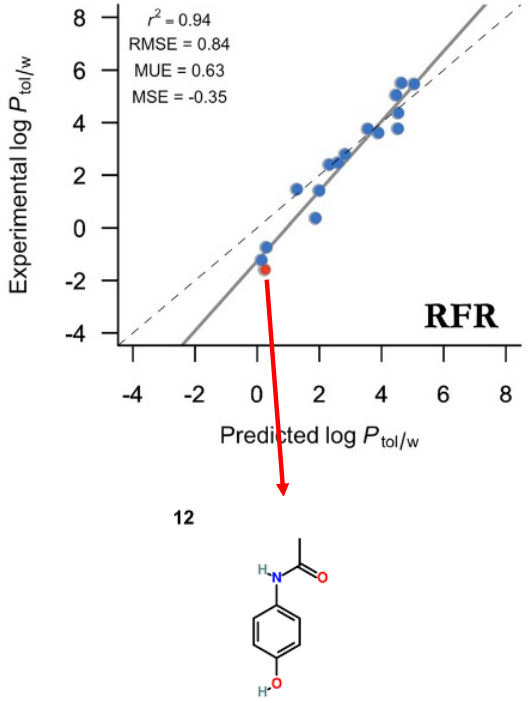
Table 2 Solute descriptor parameters and estimated $\log P_{tol/w}$ values of the 16 compounds used in the SAMPL9 blind challenge by using eqn (9)

Substance	E	S	A	B	V	$\log P_{tol/w}$ (eqn (9))
Albendazole	1.974	1.96	0.65	1.08	1.948	1.46
Alprenolol	1.250	1.09	0.15	1.44	2.159	2.43
Amitriptyline	1.920	1.42	0.00	1.08	2.400	5.83
Bifonazole	2.410	2.13	0.00	1.15	2.501	5.70
Chlorpheniramine	1.465	1.41	0.00	1.33	2.210	3.53
Epinephrine	1.400	1.17	1.21	1.53	1.415	−4.55
Fluphenazine	2.160	2.25	0.18	1.85	3.091	4.24
Glyburide	2.740	3.85	0.85	2.01	3.558	2.73
Imipramine	1.150	1.45	0.00	1.04	2.402	5.60
Ketoprofen	1.650	2.26	0.55	0.89	1.978	2.43
Nalidixic acid	1.630	1.86	0.01	1.16	1.700	1.76
Paracetamol	1.120	1.63	1.01	0.91	1.172	−2.54
Pindolol	1.700	1.59	0.30	1.40	2.009	1.37
Quinine	2.469	1.23	0.37	1.97	2.551	1.54
Sulfamethazine	2.130	2.53	0.59	1.53	2.004	−0.60
Trazodone	2.800	2.44	0.00	1.68	2.730	4.17

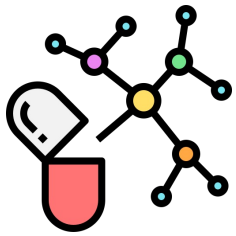
Results


r^2 = 0.85
RMSE = 1.64
MUE = 1.42
MSE = -0.71
IEF-PCM/MST
Experimental log $P_{tol/w}$
Predicted log $P_{tol/w}$


r^2 = 0.91
RMSE = 0.86
MUE = 0.71
MSE = -0.49
MLR
Experimental log $P_{tol/w}$
Predicted log $P_{tol/w}$

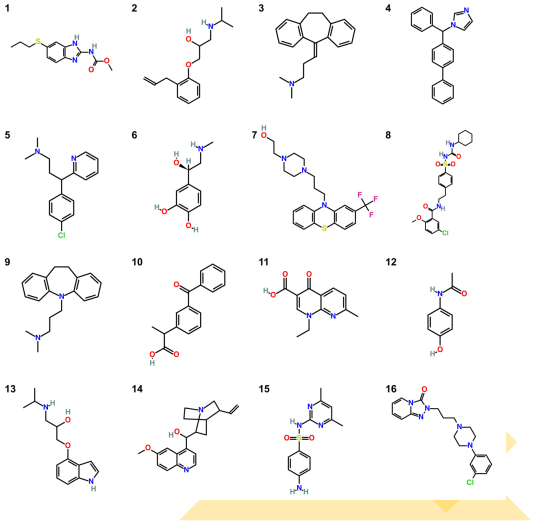

r^2 = 0.94
RMSE = 0.84
MUE = 0.63
MSE = -0.35
RFR
Experimental log $P_{tol/w}$
Predicted log $P_{tol/w}$
12

Results



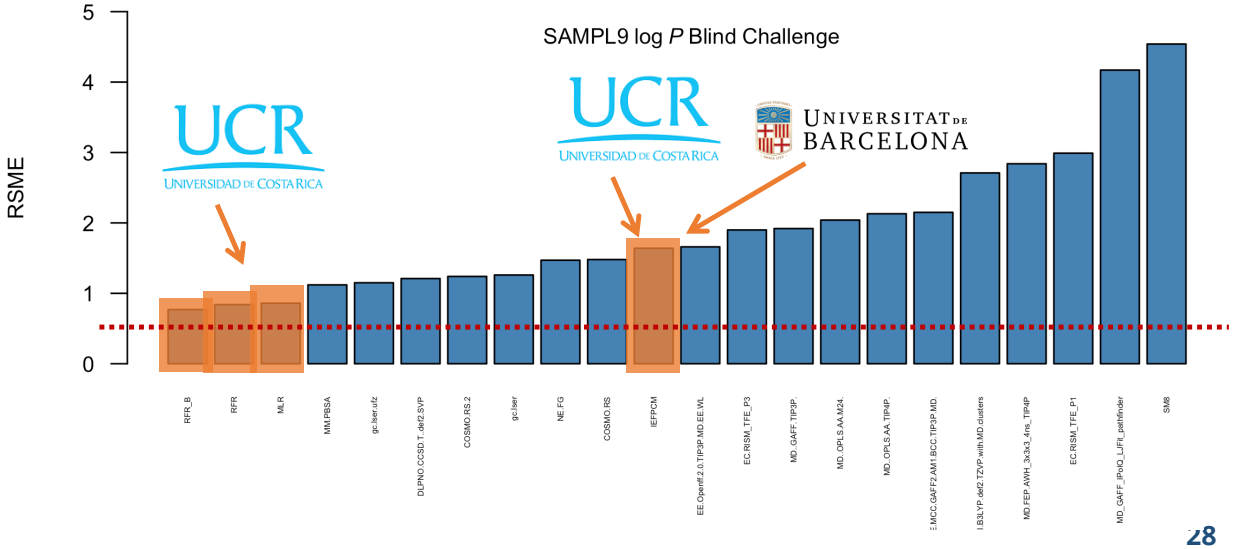
Small molecules

MW < 500 Da

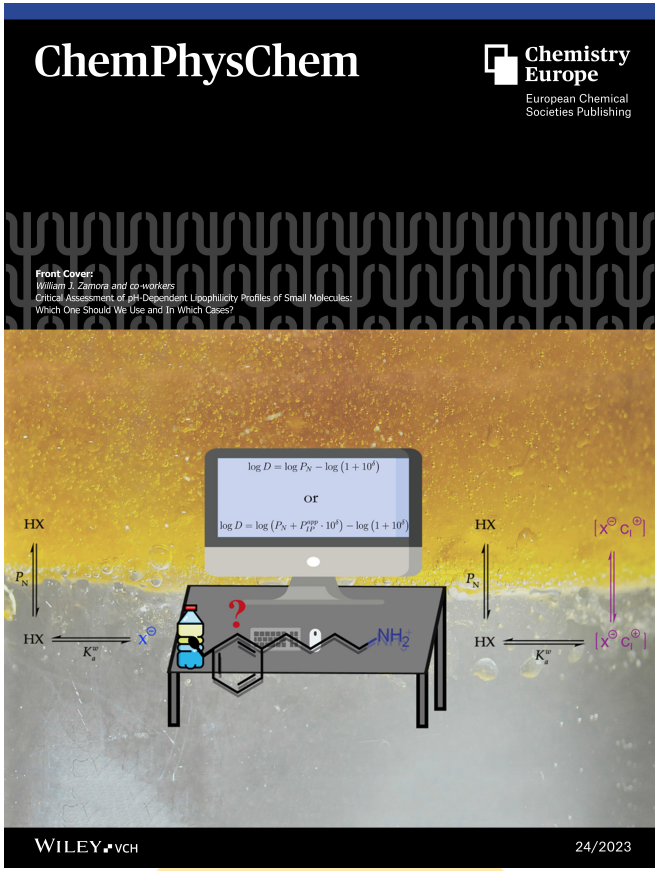


SAMPL challenges

highly developed area for small molecules



Results



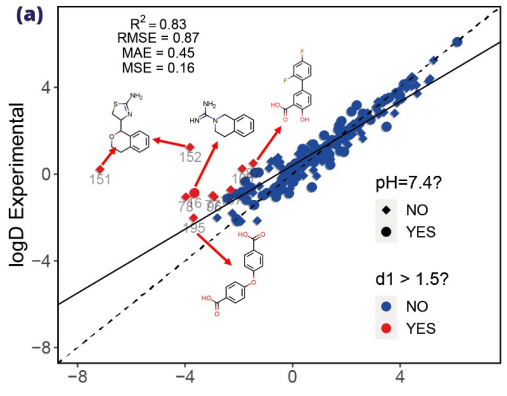
ChemPhysChem
Chemistry Europe
European Chemical Societies Publishing

Research Article

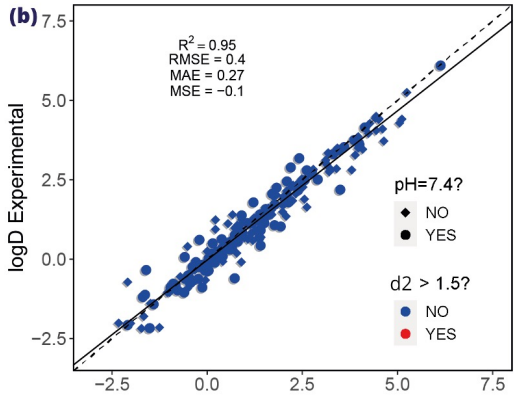
Critical Assessment of pH-Dependent Lipophilicity Profiles of Small Molecules: Which One Should We Use and In Which Cases?

Esteban Bertsch, Sebastián Suñer, Dr. Silvana Pinheiro, Prof. Dr. William J. Zamora

First published: 03 October 2023 | https://doi.org/10.1002/cphc.202300548



$$\log D = \log P_N - \log(1+10^{\delta})$$



$$\log D = \log\left(P_N + P_I \cdot 10^{\delta}\right) - \log(1+10^{\delta})$$

Results

Lipophilicity

Molecular Informatics

LiProS
Lipophilicity Profile Simulations

LiProS: Findable, Accessible, Interoperable, and Reusable Data Simulation Workflow to Predict Accurate Lipophilicity Profiles for Small Molecules

Esteban Bertsch-Aguilar[1,2] | Antonio Piedra[2] | Daniel Acuña[1] | Sebastián Suñer[1] | Sylvana Pinheiro[1,3] | William J. Zamora[1,2,3]

[1]CBio[3] Laboratory, School of Chemistry, University of Costa Rica, San Pedro, Costa Rica | [2]National Advanced Computing Collaboratory (CNCA), National High Technology Center (CeNAT), Costa Rica | [3]Laboratory of Computational Toxicology and Artificial Intelligence (LaToxCIA), Biological Testing Laboratory (LEBi), University of Costa Rica, San Pedro, Costa Rica

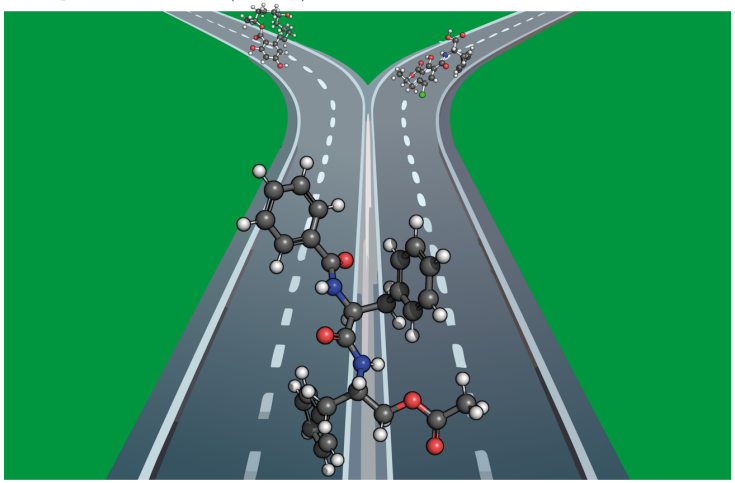
Correspondence: William J. Zamora (william.zamoraramirez@ucr.ac.cr)

Received: 5 May 2025 | **Revised:** 5 July 2025 | **Accepted:** 28 July 2025

Funding: University of Costa Rica, Grant/Award Numbers: 115-C2–126 908-C3–610

Keywords: chemoinformatics | FAIR | hydrophobicity | lipophilicity | machine learning

$$\log D_{\mathrm{pH}} = \log P_{\mathrm{N}} - \log\left(1 + 10^{\delta}\right) \qquad \log D_{\mathrm{pH}} = \log\left(P_{\mathrm{N}} + P_{\mathrm{I}}^{\mathrm{app}} \cdot 10^{\delta}\right) - \log\left(1 + 10^{\delta}\right)$$



Results

Solubility

$$\log S_0 = 0.5 - 0.01(\text{mp} - 25) - \log P_N$$

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Open Access

This article is licensed under CC-BY 4.0

pubs.acs.org/jcim Article

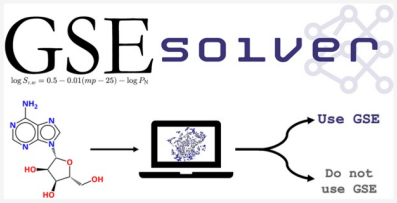
Data-Driven Critical Evaluation of the General Solubility Equation

Esteban Bertsch-Aguilar,* Nahomy Quirós, Frederick Schosinsky, Silvana Pinheiro, José L. Medina-Franco, and William J. Zamora*

Cite This: J. Chem. Inf. Model. 2026, 66, 11330−11347 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Solubility is a widely measured physicochemical property and a highly sought-after descriptor in the evaluation of molecules for drug design and development, medicinal chemistry, and agrochemical sciences. Consequently, a wide range of *in silico* models based on deductive and inductive reasoning have been developed. However, the General Solubility Equation (GSE) stands out as a simple, thermodynamically derived equation to predict intrinsic solubility ($\log S_0$, in mol/L) of molecules from their melting point and *n*-octanol/water partition coefficient ($\log P_N$). Despite its widespread use, the applicability of this equation has not been rigorously evaluated to establish when it can or cannot produce accurate predictions. Therefore, a large-scale evaluation of the GSE was conducted using an extensive database of fully experimental melting point and $\log P_N$ data for 2740, mostly drug-like compounds. A systematic analysis based on cheminformatics descriptors was conducted to identify the primary factors that allow the GSE to achieve predictions within the experimental uncertainty range (±1 log *S* unit). This information has been used to create GSESolver: a supervised machine learning classification model that, based on the selected descriptors, can predict whether the GSE will perform accurately solely from its SMILES string. GSESolver tool can be used by anyone in the scientific community to quickly assess molecule solubility and it was developed under the FAIR principles (Findable, Accessible, Interoperable, and Reusable) in a Google Colab notebook to ensure ease of use. Finally, the utility of the GSESolver tool was exploited in medicinal chemistry applications for the *in silico* prediction of the Maximum Absorbable Dose (MAD) of potential drugs, and in a rational agrochemical design application where solubility is a key factor.

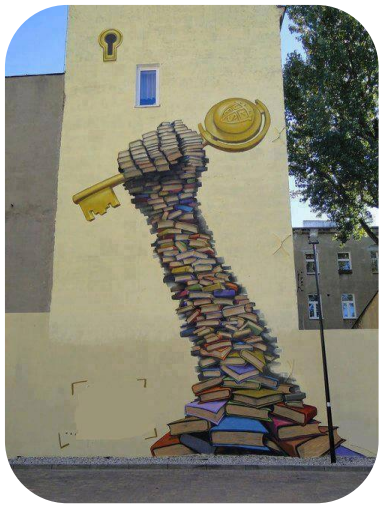




ACS Publications
Most Trusted. Most Cited. Most Read.
www.acs.org

Conclusions and Challenges





Acknowledgments

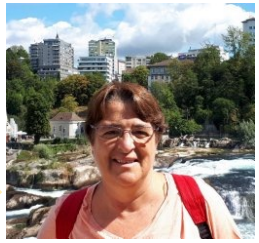
Bio3
COMPUTACIONAL BIOPHYSICS, BIOLOGY AND BIOINFORMATICS GROUP

UNIVERSITAT DE BARCELONA













THE UNIVERSITY OF CALIFORNIA IRVINE



aToxCIA
Laboratorio de Toxicología Computacional e Inteligencia Artificial
LEBI Laboratorio de Ensayos Biológicos

Pion



Gracias!

Obrigado! *Danke!*



william.zamoraramirez@ucr.ac.cr

Thanks! *Graciès!*