

□ SAMPL challenges

Statistical Assessment of the Modeling of Proteins and Ligands

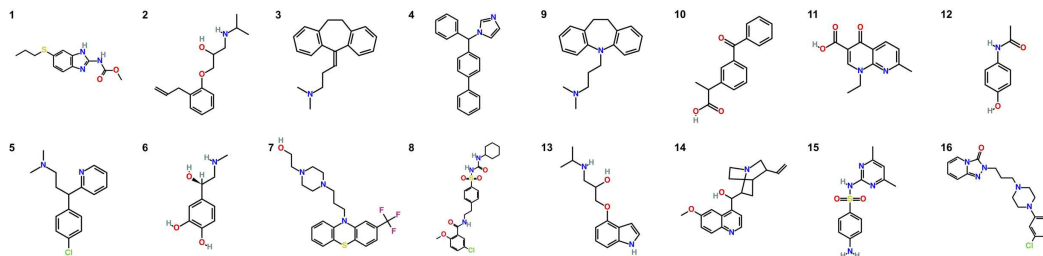
○ SAMPL9 challenges (2021년)

경진대회 요약 (3개 대학의 논문 출간을 미루고 공급받은 16개 화합물의 구조로부터 toluene-water partition coefficient 값 예측)

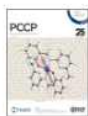
Dr. Clara Rafols (University of Barcelona), Dr. Rebeca Ruiz (Pion, UK) and William Zamora (University of Costa Rica), who measured the values used here and who are delaying publication

Experimental measurements were done for **sixteen compounds** with acid-base properties, most of them drugs showing a variety of therapeutical capabilities. These were commonly selected from Sigma-Aldrich $\geq 98\%$ (www.sigmaaldrich.com), with some selected because of potential internal hydrogen bonding. The partition solvent was from Sigma-Aldrich: toluene (ACS reagent, $\geq 99.5\%$, 179418).

ID	name	SMILES
SAMPL9-1	Albendazole	<chem>CCCS1ccc2c(c1)[nH]c(n2)NC(=O)OC</chem>
SAMPL9-2	Alprenolol	<chem>CC(C)NCC(O)COc1ccccc1CC=C</chem>
SAMPL9-3	Amitriptyline	<chem>CN(C)CCC=C2c1ccccc1CCc3ccccc23</chem>
SAMPL9-4	Bifonazole	<chem>c1ccc(cc1)C(c2ccc(cc2)c3ccccc3)n4ccnc4</chem>
SAMPL9-5	Chlorpheniramine maleate salt	<chem>CN(C)CCC(c1ccc(Cl)cc1)c2ccccc2</chem>
SAMPL9-6	Epinephrine	<chem>CNC[C@H](O)c1ccc(O)c(O)c1</chem>
SAMPL9-7	Fluphenazine dihydrochloride	<chem>OCCN4CCN(CCCN2c1ccccc1Sc3ccc(cc23)C(F)(F)F)CC4</chem>
SAMPL9-8	Glyburide	<chem>COc1ccc(Cl)cc1C(=O)NCCc2ccc(cc2)S(=O)(=O)NC(=O)NC3CC(CCC3)CCC3</chem>
SAMPL9-9	Imipramine hydrochloride	<chem>CN(C)CCCN2c1ccccc1CCc3ccccc23</chem>
SAMPL9-10	Ketoprofen	<chem>CC(C(=O)O)c1cccc(c1)C(=O)c2ccccc2</chem>
SAMPL9-11	Nalidixic acid	<chem>CCn1cc(C(=O)O)c(=O)c2ccc(C)nc12</chem>
SAMPL9-12	Paracetamol	<chem>CC(=O)Nc1ccc(O)cc1</chem>
SAMPL9-13	Pindolol	<chem>CC(C)NCC(O)COc1ccccc2[nH]ccc12</chem>
SAMPL9-14	Quinine	<chem>COc4ccc3nccc(C(O)C1CC2CCN1CC2C=C)c3c4</chem>
SAMPL9-15	Sulfamethazine	<chem>Cc2cc(C)nc(NS(=O)(=O)c1ccc(N)cc1)n2</chem>
SAMPL9-16	Trazodone hydrochloride	<chem>Clc1cccc(c1)N4CCN(CCCn3nc2ccccc2c3=O)CC4</chem>



19개팀 참가 결과 분석하여 논문화



From the journal:
Physical Chemistry Chemical Physics

Integrating multiscale and machine learning approaches towards the SAMPL9 log P challenge†

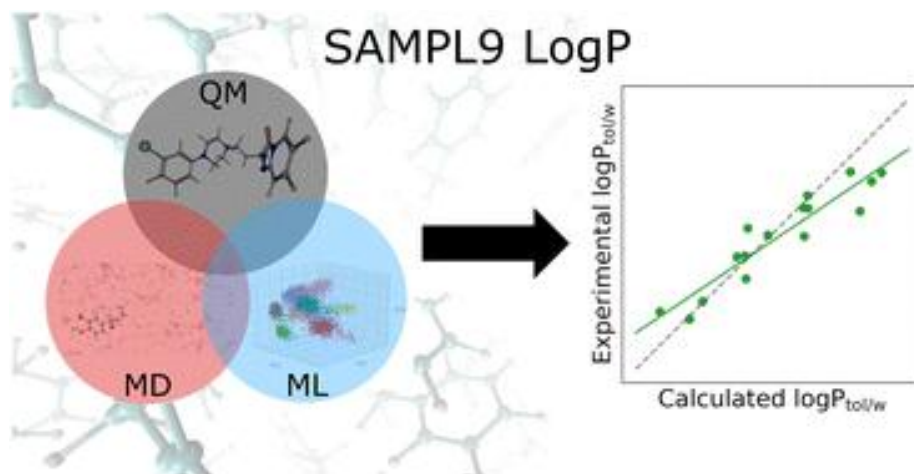


Michael R. Draper,^a Asa Waterman, IV,^{ib} ^a Jonathan E. Dannatt,^{ib} ^a and Prajay Patel,^{ib} ^a

Author affiliations

Abstract

The partition coefficient ($\log P$) is an important physicochemical property that provides information regarding a molecule's pharmacokinetics, toxicity, and bioavailability. Methods to accurately predict the partition coefficient have the potential to accelerate drug design. In an effort to test current methods and explore new computational techniques, the statistical assessment of the modeling of proteins and ligands (SAMPL) has established a blind prediction challenge. The ninth iteration challenge was to predict the toluene–water partition coefficient ($\log P_{\text{tol/w}}$) of sixteen drug molecules. Herein, three approaches are reported broadly under the categories of quantum mechanics



Category	Name	RSMD	R2
Physical (MM)	MM.PBSA	1.52	0.79
Empirical	gc.lser.ufz	1.57	0.85
Physical (QM)	DLPNO.CCSD.T..def2.SVP	1.64	0.75
Physical (QM)	COSMO-RS	1.68	0.86
Empirical	gc.lser	1.72	0.83
Physical (MM)	(NE-FG) NonEquilibrium Fast Growth	2.00	0.69
Physical (QM)	COSMO-RS	2.01	0.83
Physical (MM)	EE.Openff.2.0.TIP3P.MD.EE.WL	2.26	0.8
Physical (QM)	EC.RISM_TFE_P3	2.59	0.79
Physical (MM)	MD (GAFF/TIP3P)	2.61	0.62
Physical (MM)	MD (OPLS-AA/M24)	2.78	0.78
Physical (MM)	MD (OPLS-AA/TIP4P)	2.90	0.83
Physical (MM)	EE.MCC.GAFF2.AM1.BCC.TIP3P.MD.	2.92	0.19
Mixed (QM/MM)	RI.B3LYP.def2.TZVP.with.MD.clusters	3.69	0.17
Physical (MM)	FEP-AWH_3x3x3_4ns_TIP4P	3.86	0.06
Physical (QM)	EC.RISM_TFE_P1	4.07	0.75
Physical (MM)	MD_GAFF_IPOlQ_LJFit_pathfinder	5.67	0.43
Physical (QM)	sm8-tfe-multiple-conformations-basis	6.18	0.39

○ SAMPL8 challenges (2020년)

- ▷ Host-guest binding for drugs of abuse binding to CB8
- ▷ Host-guest binding for a series of compounds binding to Gibb deep cavity cavitand (GDCC) hosts
- ▷ **SAMPL8 Physical Properties**

경진대회 요약 (GSK로부터 공급받은 23개 화합물의 구조로부터 pKa값 예측)

A GSK pKa/logD challenge

- GSK on data collection for a physical properties challenge.
- pKa data for **24 compounds**
- pH-dependent solubility for these compounds

logD for 11 of these compounds for distribution between different phases: water-octanol, water-cyclohexane, water-ethyl acetate, water-heptane, water-MEK, water-TBME, and cyclohexane-DMF. Not all combinations of distribution coefficient are available because of compound solubility in the different phases. The total number of data points/combinations of (compound)x(phase identities) is between 40 and 50.

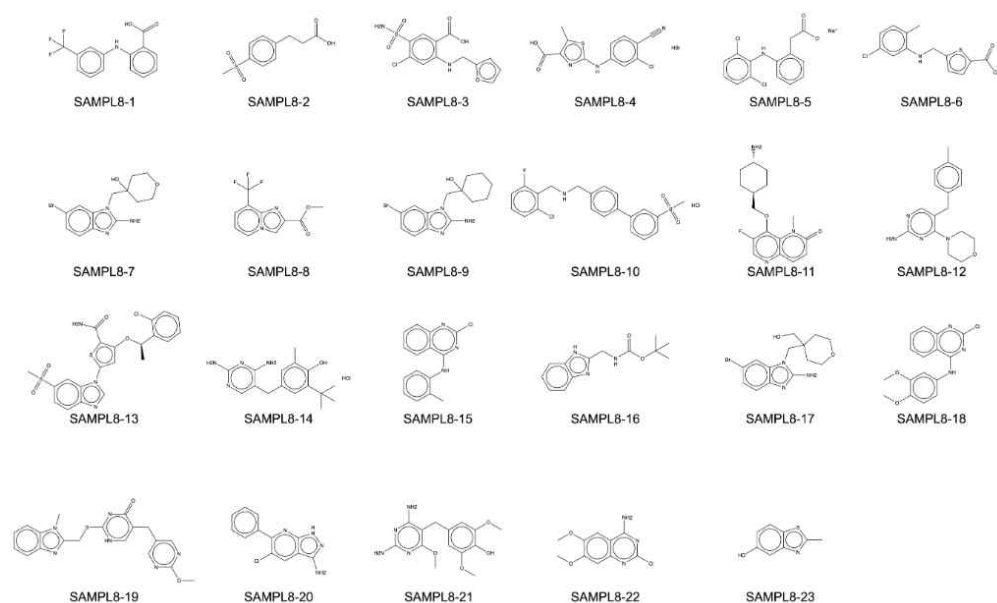


Fig 1. SAMPL8 Challenge molecules.

○ SAMPL7 challenges (2019)

- ▷ Protein-ligand binding to the Pleckstrin homology domain interacting protein (PHIP) second bromodomain (PHIP2), in three phases.
- ▷ Host-guest binding on octa acid derivatives (Gibb lab)
- ▷ Host-guest binding on modified cyclodextrin derivatives (Gilson lab)
- ▷ Host-guest binding on a glycouril clip-like host (Isaacs lab)
- ▷ **Physical property prediction (logP, pKa, permeability) for a congeneric series**

경진대회 요약 (UCSD Ballatore group으로부터 공급받은 22개 화합물의 구조로부터 logP, pKa, permeability값 예측)

Physical property prediction (logP, pKa, permeability) for a congeneric series

We are excited to announce a new set of SAMPL7 challenges focusing on pKa, partitioning, and permeability. The **Ballatore group at UCSD** is contributing a set of measured water-octanol logP, logD, and pKa values for 22 compounds. They also provided Parallel Artificial Membrane Permeability Assay (PAMPA) permeability and melting point values they measured.

☐ CACHE challenges

CRITICAL ASSESSMENT OF COMPUTATIONAL HIT-FINDING EXPERIMENTS



○ Grand Challenge 4 (August 27, 2018 Mike Chiu)

경진대회 요약 (얀센과 노바티스로부터 154개, 460개 결합친화도 데이터를 활용하여 구조로부터 랭킹 예측)

Overview

Grand Challenge 4 (GC4) is a blinded prediction challenge for the computational chemistry community, with components addressing pose-prediction, **affinity ranking**, and free energy calculations. GC4 is based on two different protein targets, Cathepsin S (CatS) and beta secretase 1 (BACE). The datasets were generously contributed by **Janssen Pharmaceuticals and Novartis**, respectively.

Subchallenge 1: BACE

This is a pose-prediction, **affinity ranking**, and free energy challenge, occurring in two stages. It is based on a dataset comprising **20 ligand-protein co-crystal structures**, and **binding data (IC50s)** spanning three orders of magnitude for **154 compounds**.

Inputs:

Stage 1a: SMILES strings of the 20 ligands to be docked and the FASTA sequence of the target, BACE. SMILES strings of the 154 compounds for affinity prediction or ranking. SMILES strings of the molecules in the free energy set (34 molecules) for the calculation of relative or absolute binding affinities.

Outputs:

In Stage 1a, your predicted poses for the 20 ligands, in a coordinate system of the participant's choosing (we will internally do the alignment). Your predicted affinities, or affinity rankings, for all 154 compounds and/or your predicted absolute or relative binding affinities (in kcal/mol) for the free energy set of 34 compounds.

Subchallenge 2: Cathepsin S

This is an affinity ranking, and free energy challenge. It is based on binding data (IC50s) spanning three orders of magnitude for **460 compounds**.

Inputs:

FASTA sequence for the CatS target and the corresponding SMILES strings of all 459 molecules for affinity prediction/ranking, SMILES strings of the 39 molecules in the free energy challenge

Outputs:

Predicted affinity/ranking for all 459 compounds, and the predicted affinities for all 39 molecules in the free energy set.

