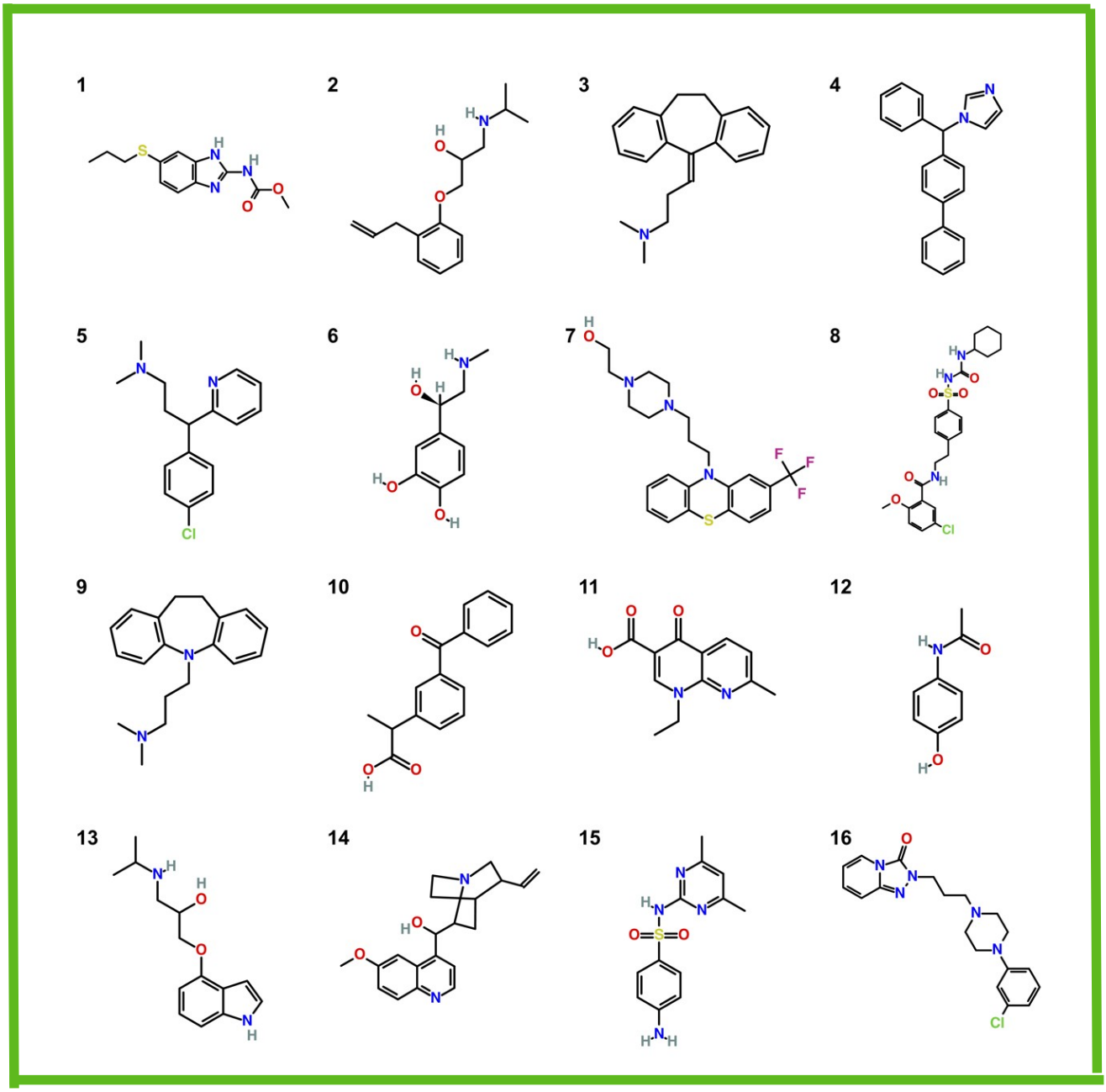
Joint UNIFI-ENEA participation in the international SAMPL9 blind challenge for predicting toluene-water LogP partition coefficients using massively parallel non equilibrium alchemical simulations

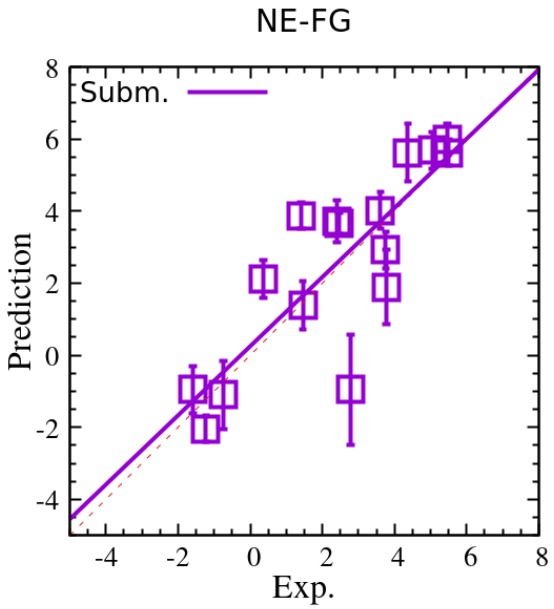
Piero Procacci(UNIFI) and Guido Guarnieri (ENEA)

The Challenge: *predicting the $LogP_{tw}$ for a series of 16 compounds with disparate flexibility or molecular weight and coarsely spanning a significant portion of the ``drug-like'' chemical space, including moieties such as carboxyl, carbonyl, sulfonic, oxydryl, amino, amide, halogen, phenyl, hetero-cyclic, alkyl*



Method: calculation of the solvation energies using **nonequilibrium** alchemical molecular dynamics simulations **(NE-FG)** where the solute-solvent interaction potential is rapidly switched on (fast growth) in *explicit* solvent (450 ps in water and toluene). GAFF2 force field for solute and solvent. Each LogP required two NE simulations (jobs) using 576 cores (12 nodes) on **CRESCO6** *requiring a total of two wall clock time hours*

Results: correlation plot between calculated and experimental LogP



Metrics

CCC (Lin) = 0.82
R (Pearson) = 0.83
*slope = **0.96***
intercept = 0.25
tau (Kendall) = 0.58
MUE = 1.12
MSE = -0.16

Comparison with other MD methods:

Subm.	method	FF	Charges	Ranked	t_{sim}/ns	CCC	MAE	*R*
VoltzLab	FEP/alchemy/EE	OpenFF-2.0	AM1-BCC	T	≃3E4	0.82	1.26	0.90
NE-FG	NE/alchemy	GAFF2	AM1-BCC	T	90	0.82	1.12	0.83
Beckstein-Iorga	FEP/alchemy	GAFF/TIP3P	AM1-BCC	F	≃ 250	0.75	1.50	0.79
Beckstein-Iorga	FEP/alchemy	OPLS/M24	mol2ff	T	≃ 250	0.73	1.72	0.88
Beckstein-Iorga	FEP/alchemy	OPLS/TIP4P	CM1A	F	≃ 250	0.72	1.88	0.91
Sprick	FEP/alchemy/HREM	GAFF/TIP3P	IpolQ-Mod	T	240	0.47	3.03	0.66
Oxford	E-S/MCC	GAFF2/TIP3P	AM1-BCC	T	1200	0.42	1.78	0.44
MD(Patel)	FEP/alchemy	CGenFF/TIP4P	cGenFF	F	216	0.21	2.63	0.25

*NE-FG resulted as **the best MD-based SAMPL9 prediction** for the Mean Unsigned Error (MUE), the Lin Concordance coefficient (CCC) and the benefit-cost ratio. The computational cost of NE-FG (t_{sim}/ns) was the **lowest** among all MD-based approaches*