University of Strathclyde Science

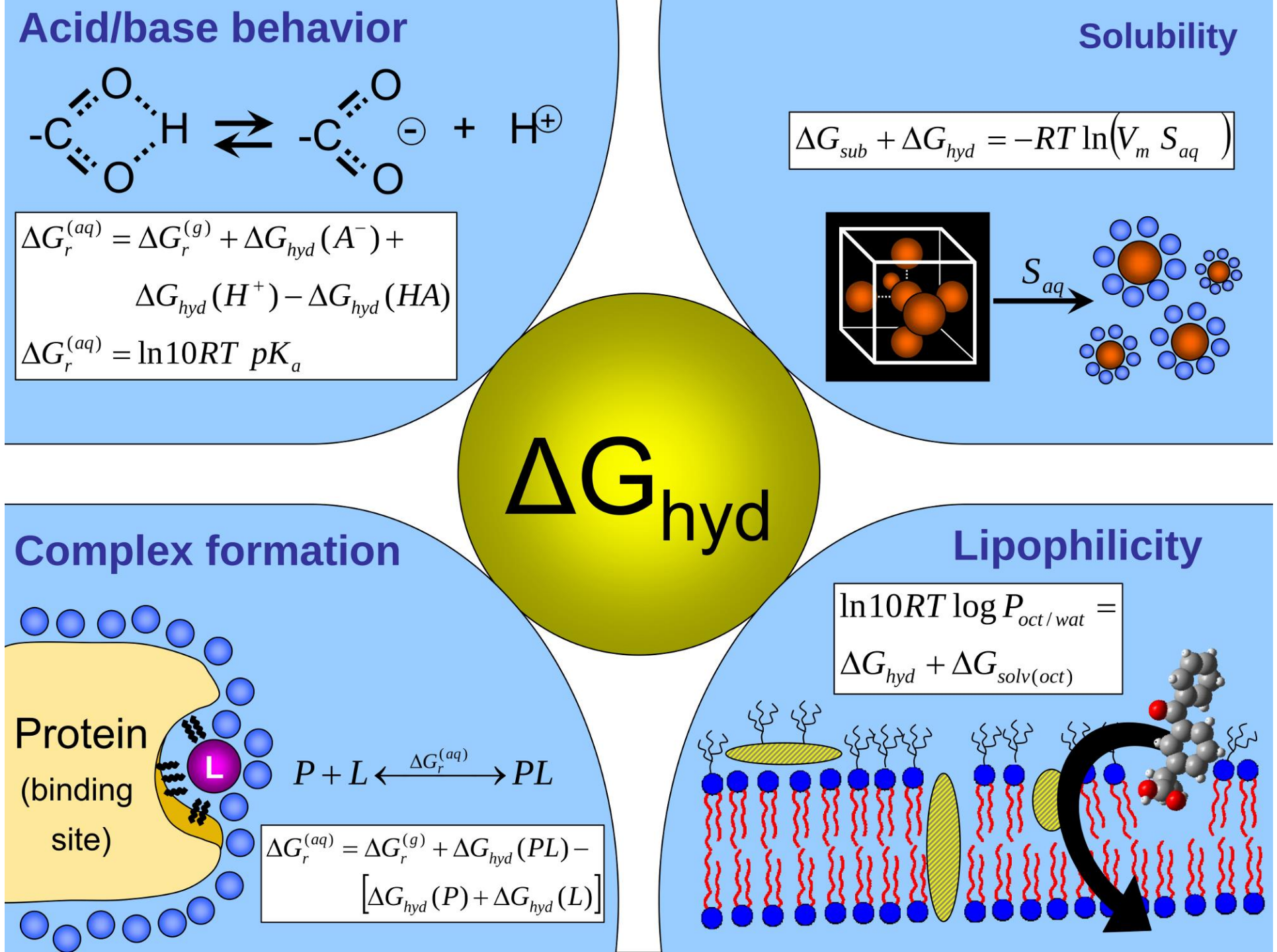
Development of Solvation-Based Molecular Descriptors for Predicting Physicochemical Properties and Chromatographic Retention Times

Dr David Palmer

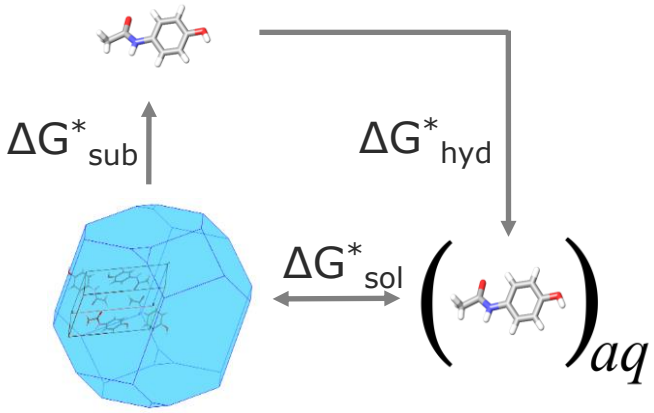
PhysChemForum Meeting
1st October 2024

“The solvent plays a key role in solution chemical systems. It influences practically all of the physical and chemical properties of the molecular species present: solubility, partitioning, reaction rate, chemical equilibrium, spectroscopic properties, …”

R. Cabot and C. A. Hunter, *Chem. Soc. Rev.*, **2012**, *41*, 3485–3492


Acid/base behavior
$\Delta G_r^{(aq)} = \Delta G_r^{(g)} + \Delta G_{hyd}(A^-) + \Delta G_{hyd}(H^+) - \Delta G_{hyd}(HA)$
$\Delta G_r^{(aq)} = \ln 10 RT\ pK_a$
Solubility
$\Delta G_{sub} + \Delta G_{hyd} = -RT \ln(V_m\ S_{aq})$
S_{aq}
ΔG_{hyd}
Complex formation
Protein
(binding
site)
L
$P + L \xleftrightarrow{\Delta G_r^{(aq)}} PL$
$\Delta G_r^{(aq)} = \Delta G_r^{(g)} + \Delta G_{hyd}(PL) - [\Delta G_{hyd}(P) + \Delta G_{hyd}(L)]$
Lipophilicity
$\ln 10 RT \log P_{oct/wat} = \Delta G_{hyd} + \Delta G_{solv(oct)}$

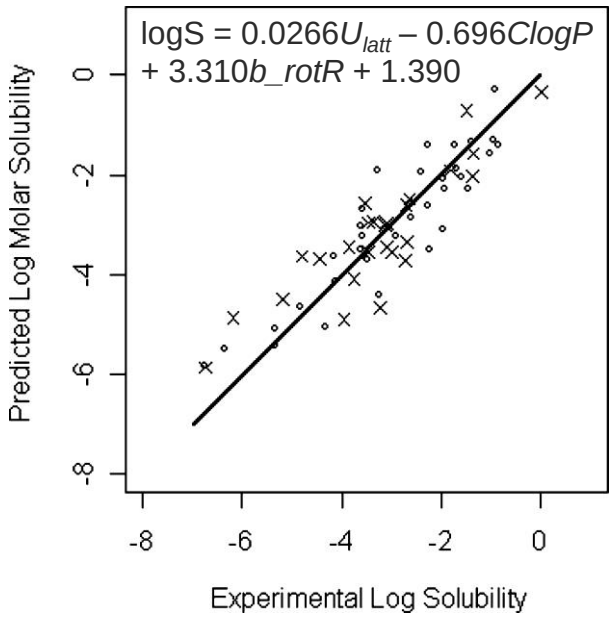
Physics-Based Solubility Prediction



$$\Delta G^*_{sol} = \Delta G^*_{sub} + \Delta G^*_{hyd} = -RT \ln S_o V_m$$

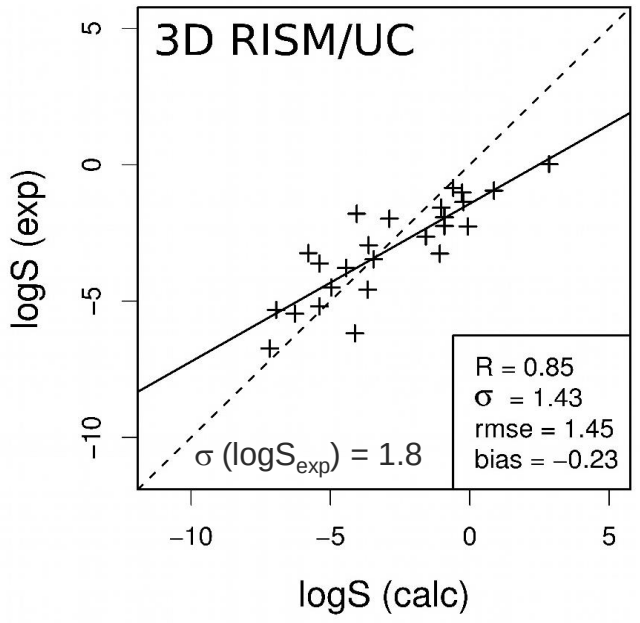
Accuracy

Computational Expense



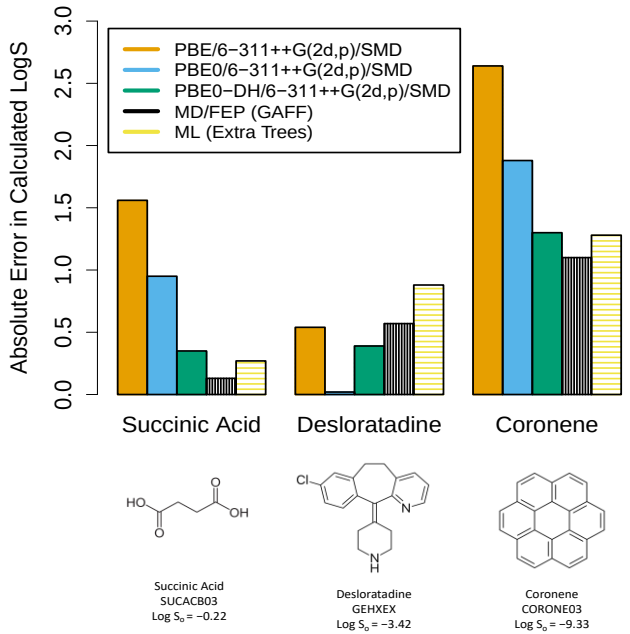
Empirical Parameterisation

Mol. Pharmaceutics **2008**, *5*, 2, 266



Classical Simulation

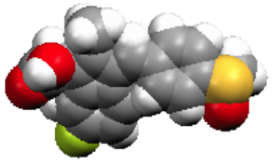
J. Chem. Theory Comput., ***2012***, *8*, 3322



Quantum Mechanics

J. Chem. Theory Comput. **2021**, *17*, 6, 3700

Data-Driven Methods



Molecule	LogS	X1	X2	…	XN
Molecule 1					
Molecule 2					
Molecule 3					
Molecule 4					
Molecule 5					
Molecule 6					
Molecule 7					
Molecule ..					

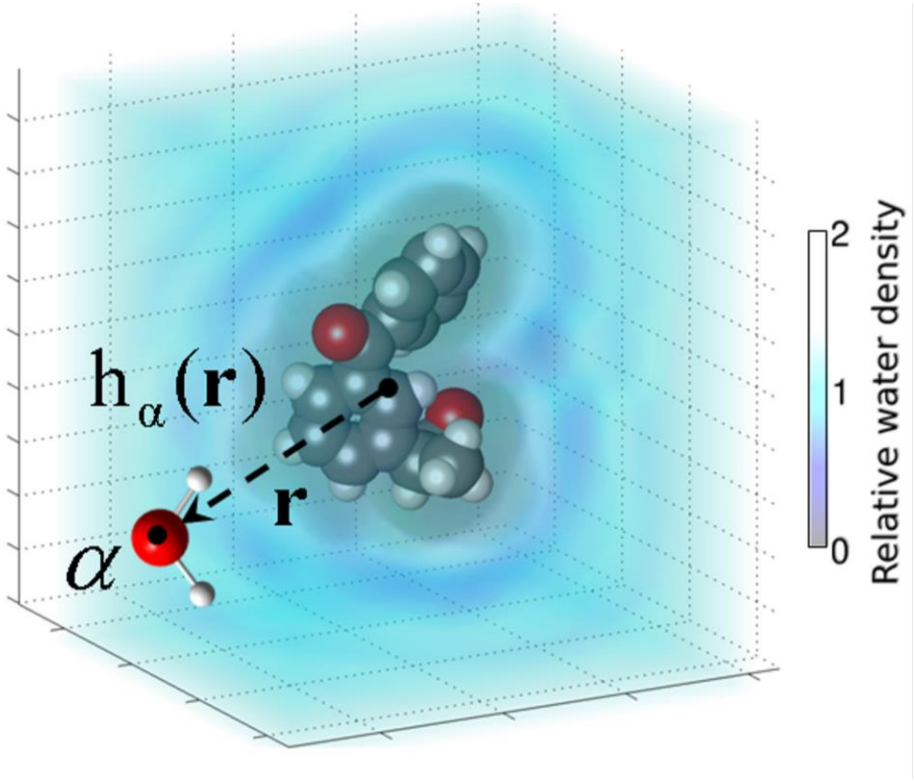
Molecule	LogS	X1	X2	…	XN
Molecule ..					
Molecule ..					
Molecule ..					
Molecule N					

$$\log S = f(x)$$

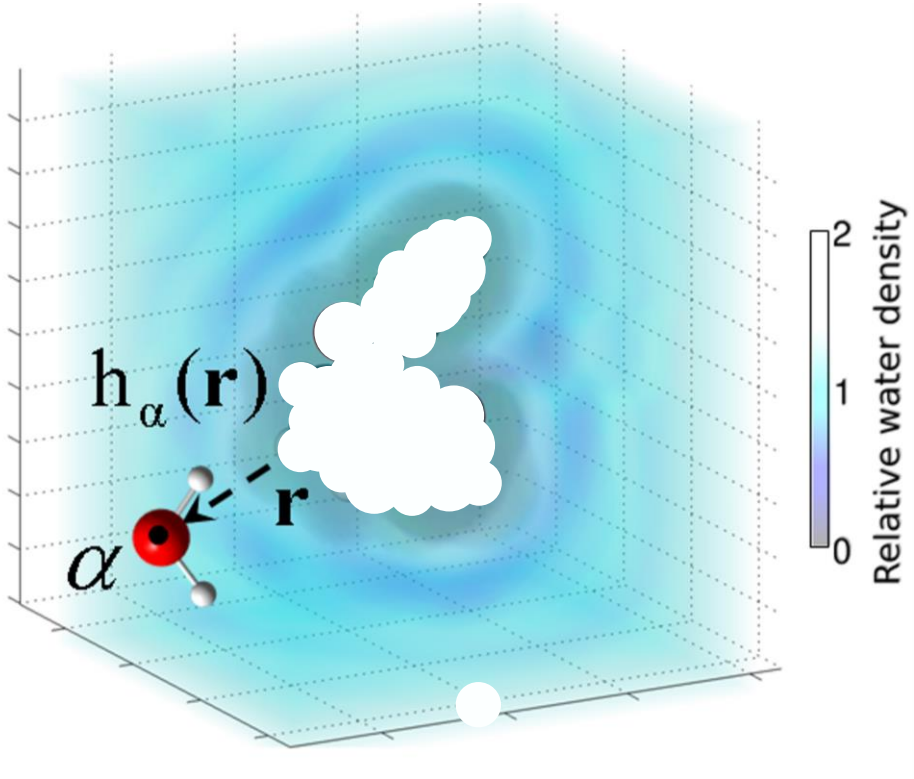
Advantages: quick, accurate (in favourable circumstances)

Disadvantages: unreliable for molecules/conditions not in the training set, predictions normally limited to a single set of environmental conditions, not grounded in physics = difficult to systematically improve

Molecular Descriptors



Solvation Descriptors



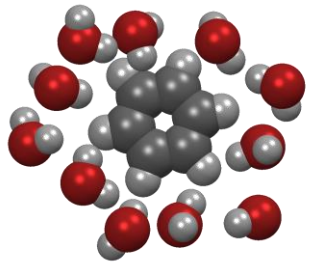
Thermodynamics: Statistical Mechanics of Molecular Solvation

Methods for calculating thermodynamic parameters differ by description of the solvent structure

Solvent Models

Explicit Solvent (MD,MC)

Each solvent molecule is a separate atomistic object



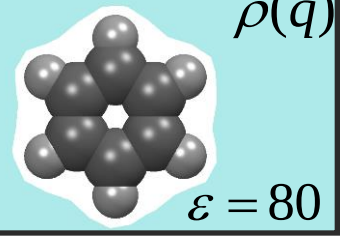
accurate (?) but computationally expensive

Implicit Solvent

- Solvent distribution functions

Continuum electrostatics

fast but not always accurate enough



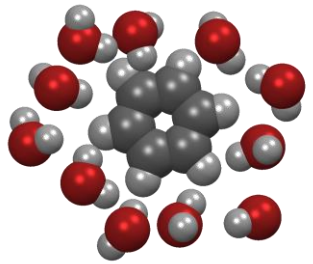
Thermodynamics: Statistical Mechanics of Molecular Solvation

Methods for calculating thermodynamic parameters differ by description of the solvent structure

Solvent Models

Explicit Solvent (MD,MC)

Each solvent molecule is a separate atomistic object



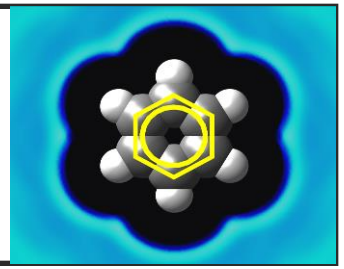
accurate (?) but computationally expensive

Implicit Solvent

- Solvent distribution functions

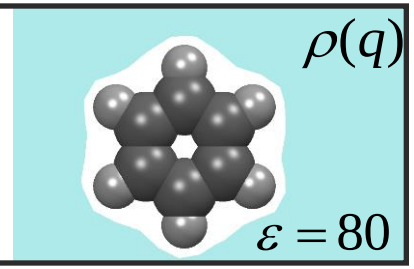
Integral Equations Theory of Liquids

we are working to make it fast and accurate



Continuum electrostatics

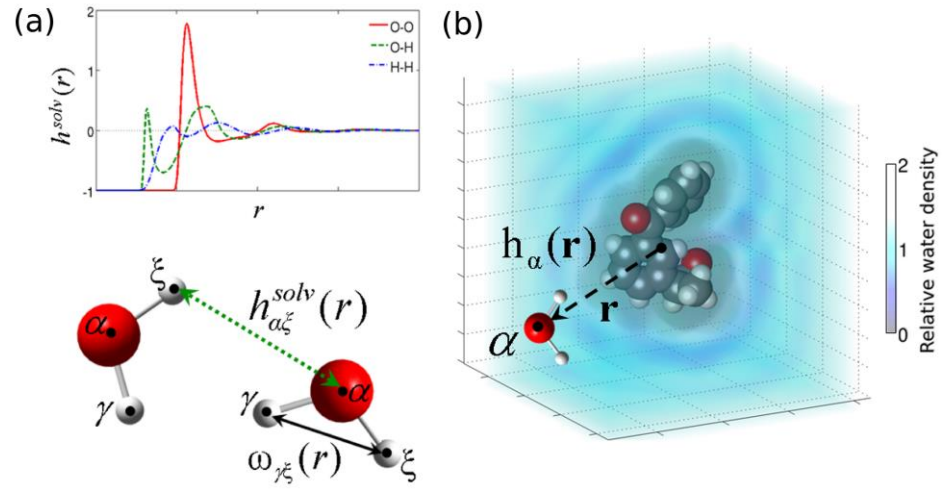
fast but not always accurate enough



3D RISM

3D Reference Interaction Site Model (3D RISM)

Solvent **Solution**



$$c_{xa} = w_{xa}(r) + r\,h_{xa}(r) \qquad h_a(\mathbf{r}) = g_a(\mathbf{r}) - 1$$

3D-RISM equation:

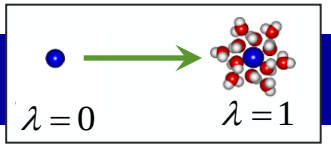
$$h_a(\mathbf{r}) = \sum_{x=1}^{N_{solv}} \int_{R^3} c_x(\mathbf{r} - \mathbf{r}')\, c_{xa}(|\mathbf{r}'|)\, d\mathbf{r}'$$

3D-RISM closure relationship:

$$h_a(\mathbf{r}) = \exp\left(-\beta u_a(\mathbf{r}) + h_a(\mathbf{r}) - c_a(\mathbf{r}) + B_a(\mathbf{r})\right) - 1$$

bridge function

End-Point Calculations



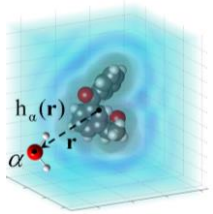
Hydration Free Energy:

$$\Delta G^{GF}_{hyd} = k_B T \sum_{a=1}^{N_{solv}} \rho_a \int_{R^3} \left[-c_a(\mathbf{r}) - \tfrac{1}{2} c_a(\mathbf{r}) h_a(\mathbf{r})\right] d\mathbf{r}$$

Partial Molar Volume:

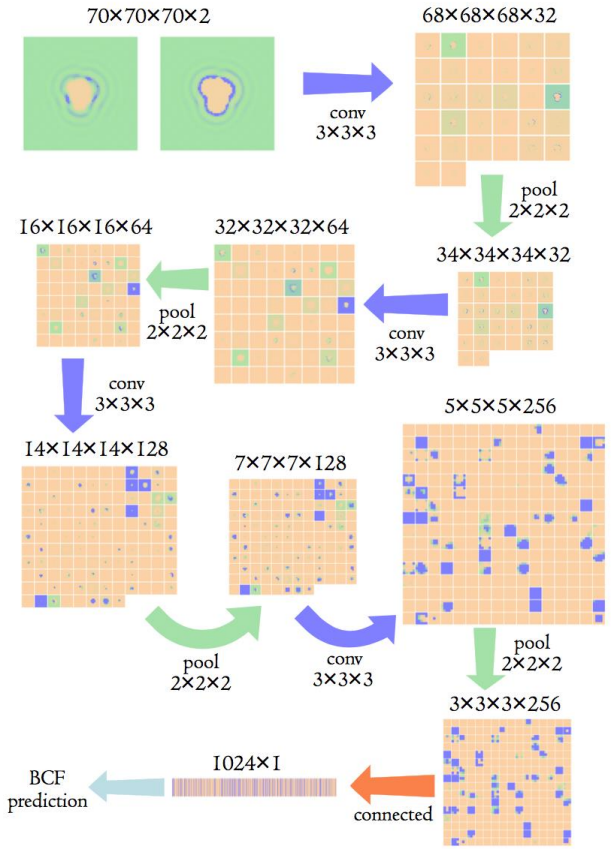
$$\bar{V} = k_B T \kappa_T \left(1 - \rho \sum_{a=1}^{N_{solv}} \int_{R^3} c_a(\mathbf{r})\, d\mathbf{r}\right)$$

Deep Learning + Statistical Mechanics: Environmental Impact of New Modalities



3D RISM input

3D Convolutional Neural Network



Bioaccumulation Factor

$$BCF = \frac{C_{biota}}{C_{water}}$$

C is concentration of compound in "*biota*" or "*water*" (measured under controlled laboratory conditions)

Model		RMSE	MAE	R^2
US EPA (baseline)	consensus model	0.66	0.51	0.76
	single model	0.68	0.64	0.74
ActivNet4 (3D data)	training/test	0.66	**0.48**	**0.77**
	5-fold CV	0.65	0.48	0.77
XGBoost (3D data)	training/test	0.85	0.70	0.61
	5-fold CV	0.91	0.72	0.54
Linear Regression (ΔG and $\bar{V}$)	training/test	1.11	0.92	0.32

J. Phys. Cond. Matt. **2018**, *30*, 32LT03

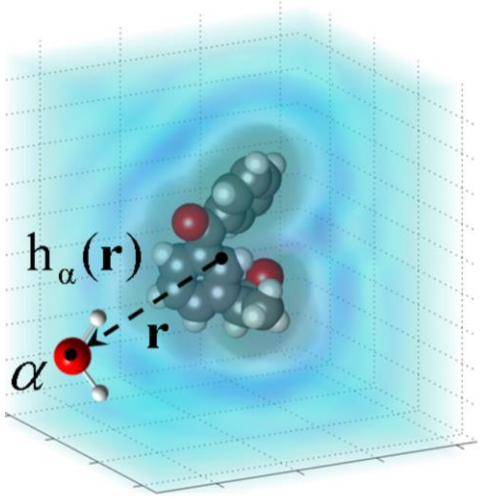
Drawbacks of using Solvation Density Distributions from 3D RISM for Machine Learning

Practical Issues

- calculations too slow for high-throughput screening (hours for some solutes)
- significant memory and disk space requirements to store full distribution functions for large datasets
- advanced machine learning methods required

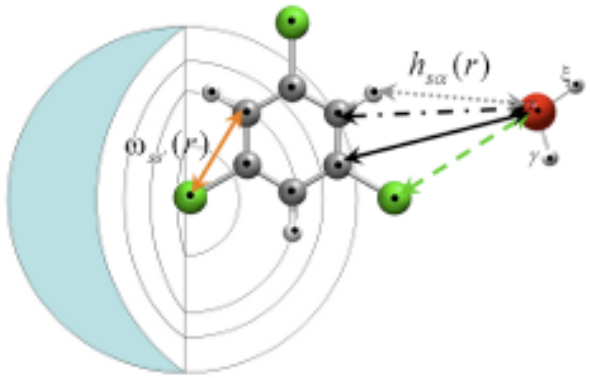
Scientific issues

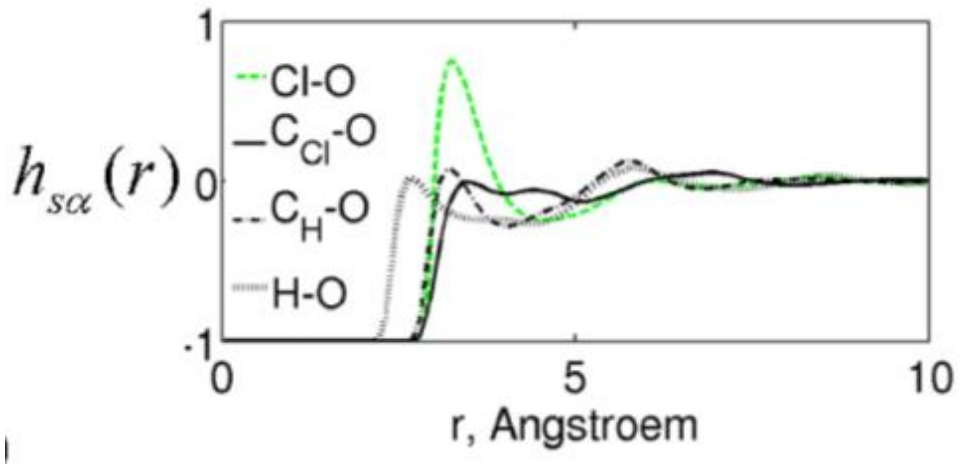
- distribution functions are not invariant to rotation of solute
- distributions are dependent on solute and solvent conformations (and to a lesser extent forcefield parameters too).



1D RISM

1D Reference Interaction Site Model





$$h(r) = g(r) - 1$$

$$\chi_{\alpha\xi} = \omega_{\alpha\xi}(r) + \rho h_{\alpha\xi}(r)$$

RISM equations:

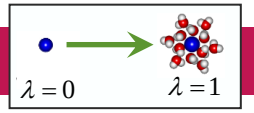
$$h_{s\alpha}(r) = \sum_{s'=1}^{N_{solute}} \sum_{\xi=1}^{N_{solvent}} \int_{R^3} \omega_{ss'}(r) * c_{s'\xi}(r) * \chi_{\xi\alpha}(r)$$

Closure relations:

$$h_{s\alpha}(r) = \exp\left[-\beta u_{s\alpha}(r) + h_{s\alpha}(r) - c_{s\alpha}(r) + B_{s\alpha}(r)\right] - 1$$

site-site bridge function

End-point calculations



$$\Delta G_{hyd}^{HNC} = \frac{2\pi\rho}{\beta} \sum_{s=1}^{N_{solute}} \sum_{\alpha=1}^{N_{solvent}} \int_0^\infty \left[-2c_{s\alpha}(r) - c_{s\alpha}(r)h_{s\alpha}(r) + h_{s\alpha}^2(r)\right] r^2 dr.$$

Chandler, Singh, J. Chem. Phys. (1984), 81, 1975

Advantages of using 1D RISM Distribution Functions for Machine Learning

- fast calculation time on single CPU (1-2 mins)
  - suitable for moderate throughput screening
- grounded in statistical physics
- adapt to changes in, e.g.:
  - environmental conditions (e.g. temperature)
  - solvents
  - co-solutes (e.g. brines)
- physically intepretable
  - descriptors have physical meaning
- no sampling noise
  - SFED descriptors have almost no statistical noise unlike e.g. descriptors derived from molecular simulation (e.g. molecular dynamics, Monte Carlo)
  - possible to converge 1D RISM equations to very small tolerance

Drawbacks of using 1D RISM Distribution Functions for Machine Learning

Practical Issues

- no open-source and actively maintained 1D RISM *solute-solvent* solvers available

- proof-of-concept work had used the RISM-MOL program[1], but this is:
  - written in Matlab
  - appears to no longer be maintained
  - only applicable to water at 298 K
  - difficult to extend

Scientific Issues

- $N_{solute} * N_{solvent}$ h(r) and c(r) functions per system (where N is number of atoms)
  - different chemical systems have different number of descriptors
  - solvation thermodynamics depend on the values of all h(r) and c(r) functions
  - *How do we train ML models when number of descriptors change for each system?*

- 1D RISM theory alone gives inaccurate solvation thermodynamics

[1]http://www.sergiievskyi.com/_rismmol.html

pyRISM: An open-source solver for 1D RISM equations

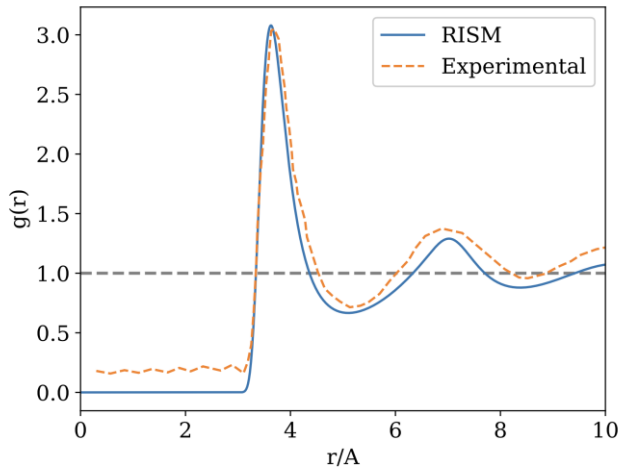


FIG. 1: The calculated[22] and experimental RDFs of Argon at 85K.

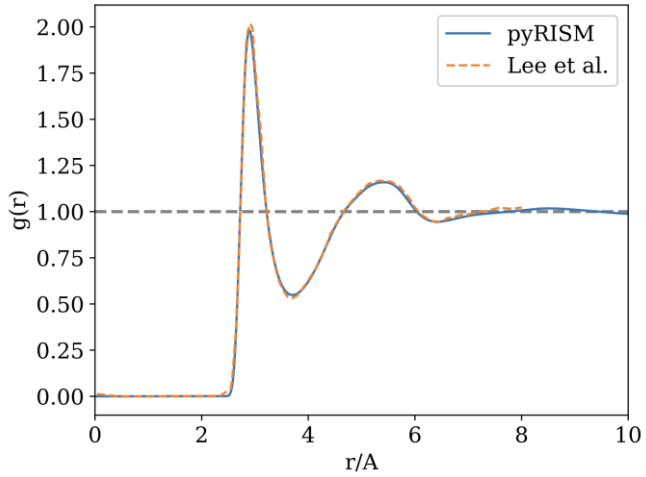


FIG. 3: The partial RDF of the oxygen sites between water and 2-propanol calculated via pyRISM compared to the XRISM result obtained by Lee et al.[25]

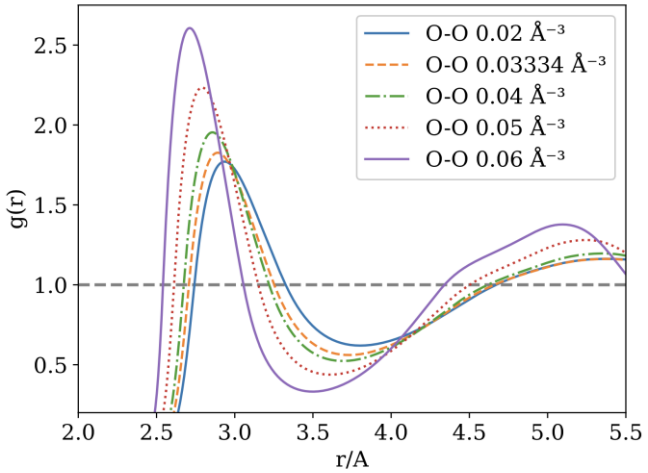
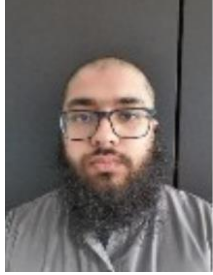


FIG. 5: The partial RDF of the oxygen sites between water and 2-propanol calculated via DRISM at varying densities.



- Developed by Abdullah Ahmad
- Open-source, free.
  - *github.com/2AUK/pyRISM*
- XRISM and DRISM
- More adaptable than existing codes
  - solute-solvent calculations
  - different pure and mixed solvents
  - different temperatures
- Fast
  - Ng acceleration
  - MDIIS algorithm
- Written in Rust and Python

1D RISM Solvation Free Energy Densities are Useful Machine Learning Descriptors

- The standard 1D RISM free energy formulae have the same general form:

Hydration Free Energy

Hyper-netted-chain:

$$\Delta G_{\mathrm{HNC}} = 2\pi\rho kT \sum_{s\alpha} \int_0^\infty [-2c_{s\alpha}(r) - h_{s\alpha}(r) \times (c_{s\alpha}(r) - h_{s\alpha}(r))] r^2 dr$$

Gaussian-Fluctuations:

$$\Delta G_{\mathrm{GF}} = 2\pi\rho kT \sum_{s\alpha} \int_0^\infty (-2c_{s\alpha}(r) - c_{s\alpha}(r) h_{s\alpha}(r)) r^2 dr$$

Kovalenko-Hirata:

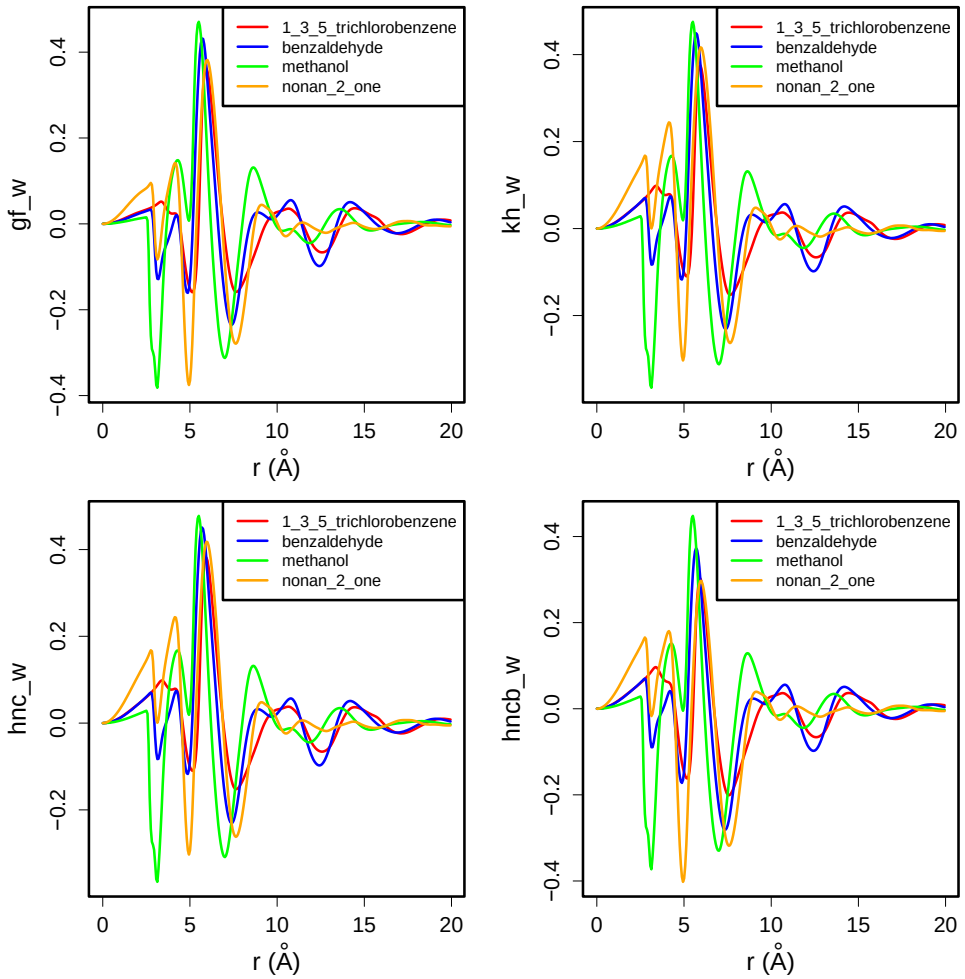
$$\Delta G_{KH} = 2\pi\rho kT \sum_{s\alpha} \int_0^\infty [-2c_{s\alpha}(r) - h_{s\alpha}(r)(c_{s\alpha}(r) - \Theta(-h_{s\alpha}(r)))] r^2 dr$$

General Form:

$$DG = c \sum_{Sa} \int_0^\infty f(r) r^2 \, dr$$

- Omitting the integral gives the solvation free energy density (SFED)

1D RISM Solvation Free Energy Densities are Useful Machine Learning Descriptors



- solvation free energy density (SFED) profiles are unique molecular fingerprints
- only weak dependence on choice of SFE functional
- continuous functions that depend on distance between solute and solvent sites only
  - rotationally invariant
- one SFED for each solute, but
  - some dependence on solute state (ionization, tautomerism), and conformer
  - varies with physical and environmental properties, e.g. solvent, temperature, etc.

Deep Learning Free Energy Functionals for 1D RISM

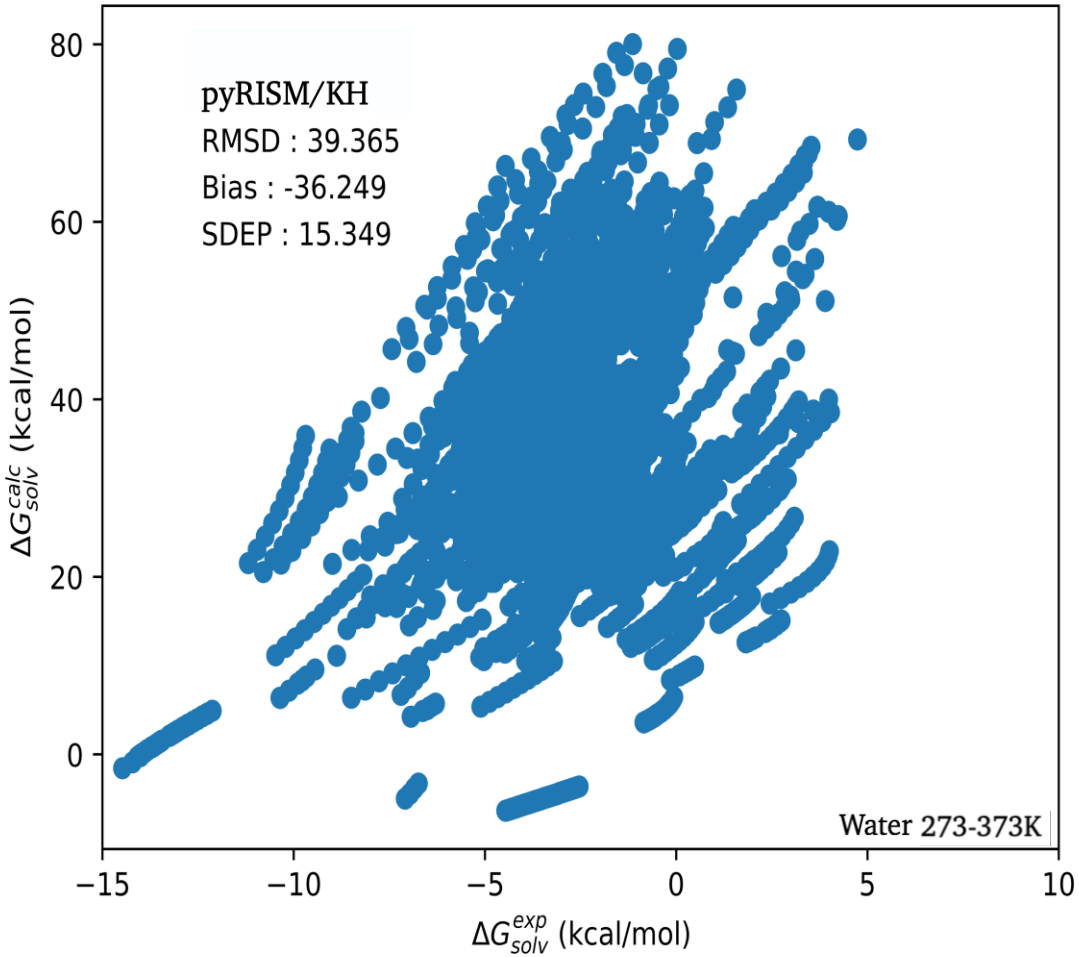
- solvation free energy data compiled from the published literature

- 3653 data points for neutral solutes*
  - 3440 in water (273 – 373 K)
  - 109 in chloroform
  - 79 in carbon tetrachloride
  - 25 in methanol

- 180 data points for ionized solutes
  - 103 in water
  - 77 in methanol

- 238 data points in water, chloroform, carbon tetrachloride and methanol with enthalpy, entropy and free energy of solvation

Standard 1D RISM theory (neutral solutes)



*temperatures are 298 K unless indicated

Deep Learning Free Energy Functionals for 1D RISM

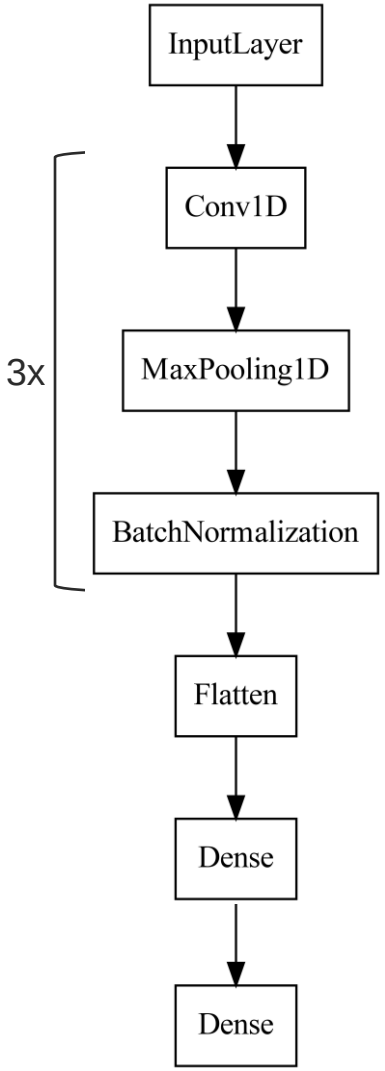
- RISM calculations
  - AMBER-GAFF forcefield for solute and solvent
  - except water (mSPC/E)

- Solvation Free Energy Density
  - every 40th grid point from r = 0 Å to r = 8 Å was used
  - 160 SFED descriptors.

- Nested cross-validation:
  - Outer loop: 50-fold Monte Carlo CV
  - Inner loop: 5-fold CV
  - Stratified by solute (or solvent)

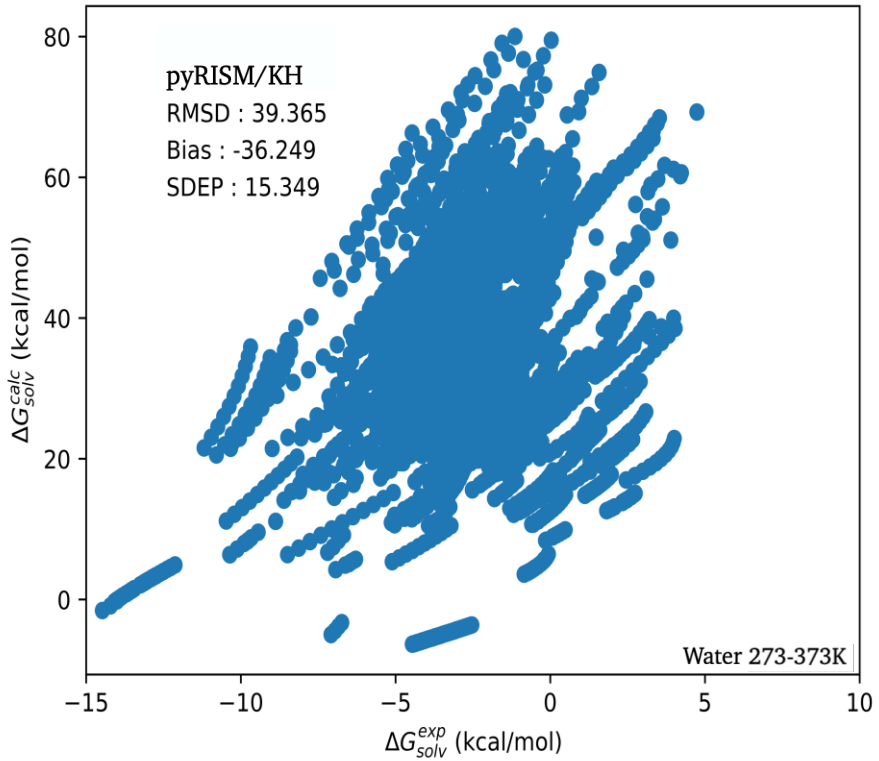
- 1D Convolutional Neural Network
  - (similar but slightly less accurate results from traditional ML algorithms, e.g. RF, PLS, SVM, etc).

1D Convolutional Neural Network (CNN)

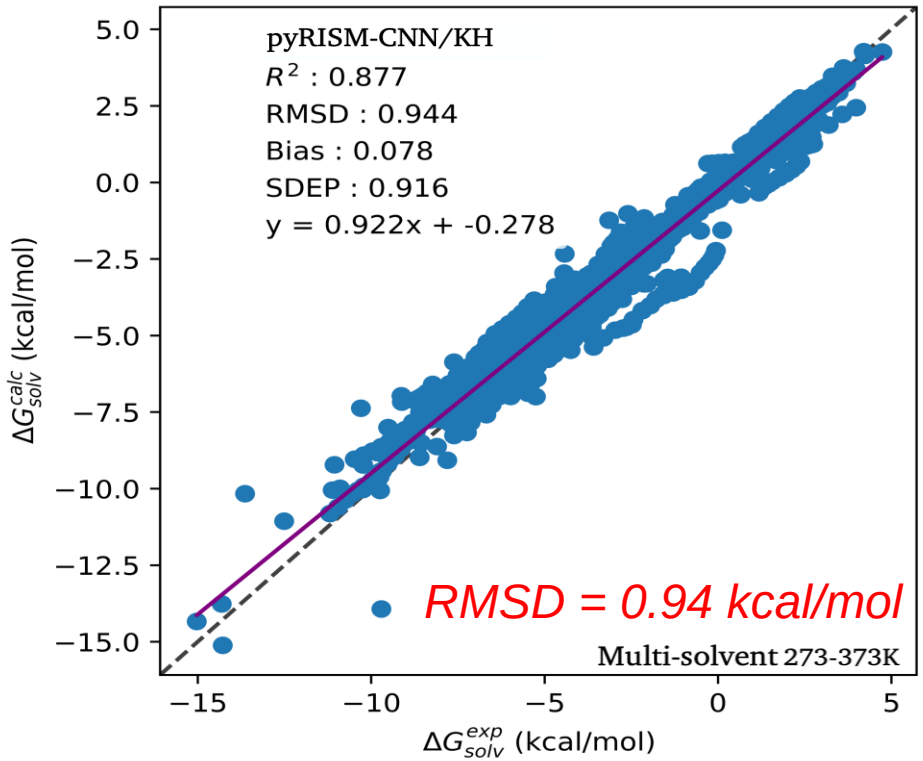


Deep Learning Free Energy Functionals for 1D RISM

Standard 1D RISM theory



pyRISM / CNN

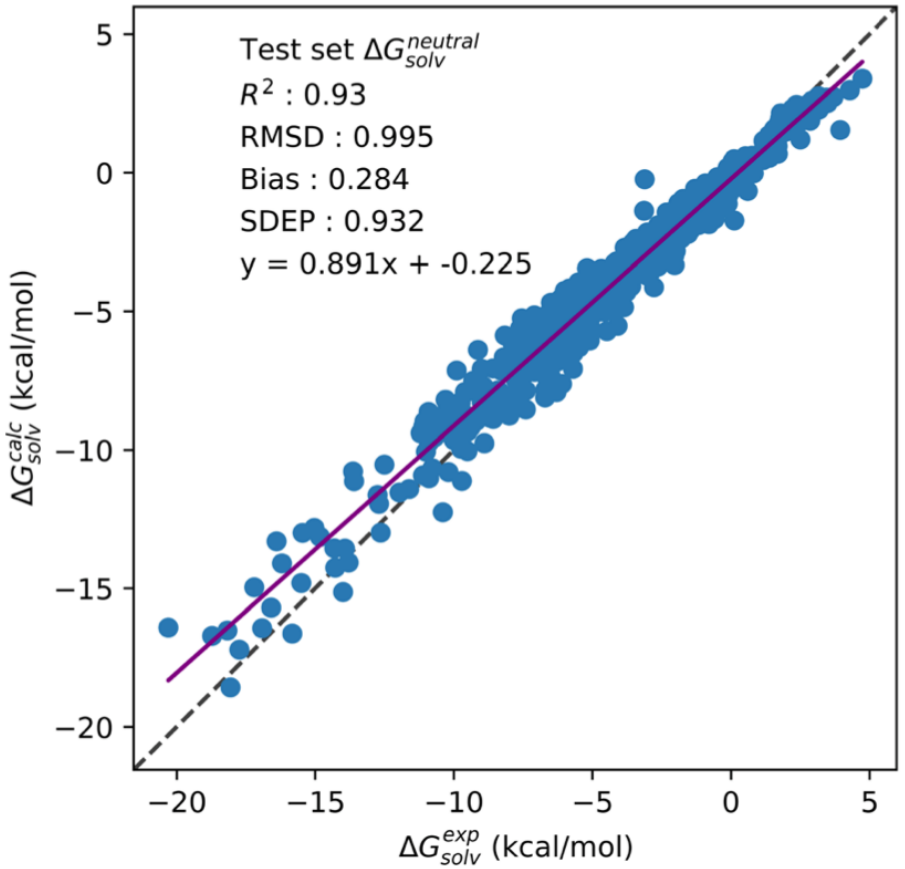


pyRISM-CNN/KH							
Solvent	Temperature	Temp. Descr.	Datapoints	R^2	RMSE	Bias	SDEP
Carbon Tetrachloride	298K	No	79	0.93	0.44	0.06	0.42
Chloroform	298K	No	109	0.92	0.74	0.00	0.72
Water	298K	No	521	0.95	0.91	0.04	0.89
	273-373K	No	3053	0.93	0.66	-0.01	0.65
	273-373K	Yes	3053	0.95	0.55	0.01	0.54
Multi-solvent	298K	No	709	0.95	0.83	-0.01	0.82
	273-373K	No	3241	0.88	0.94	0.08	0.92
	273-373K	Yes	3241	0.87	0.96	0.07	0.94

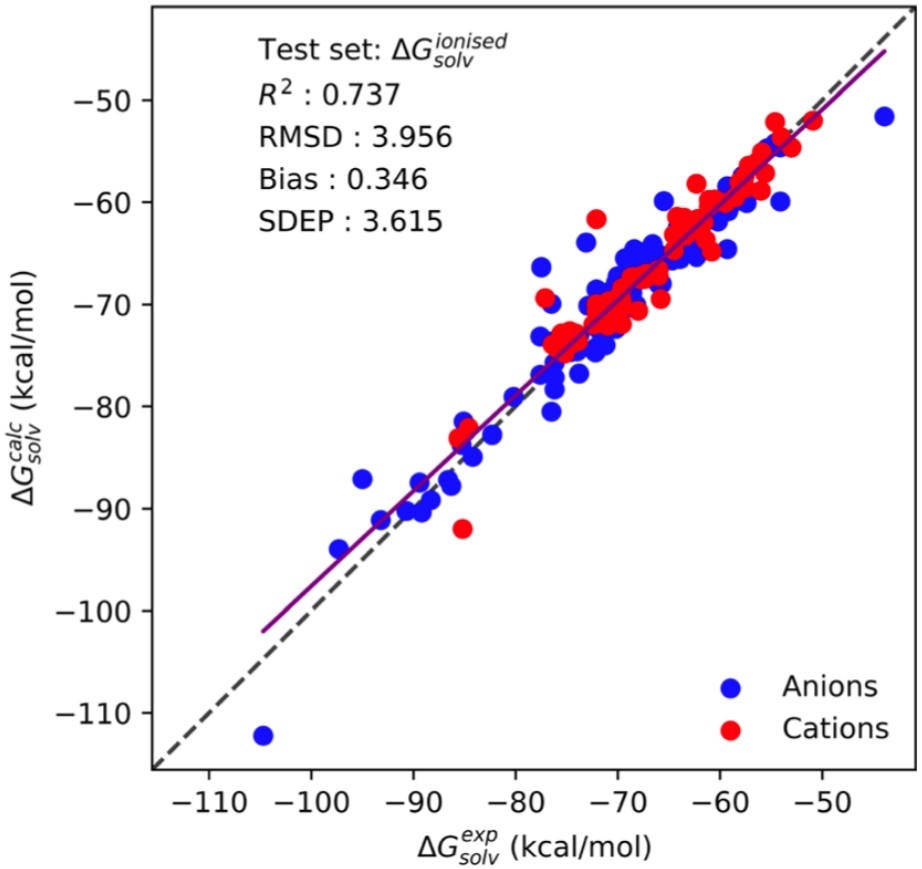
ΔG_{solv} predictions using pyRISM-CNN models trained on 1D-RISM KH SFED

Neutral and Ionised Solutes at 298 K

Neutral Solutes



Ionised Solutes

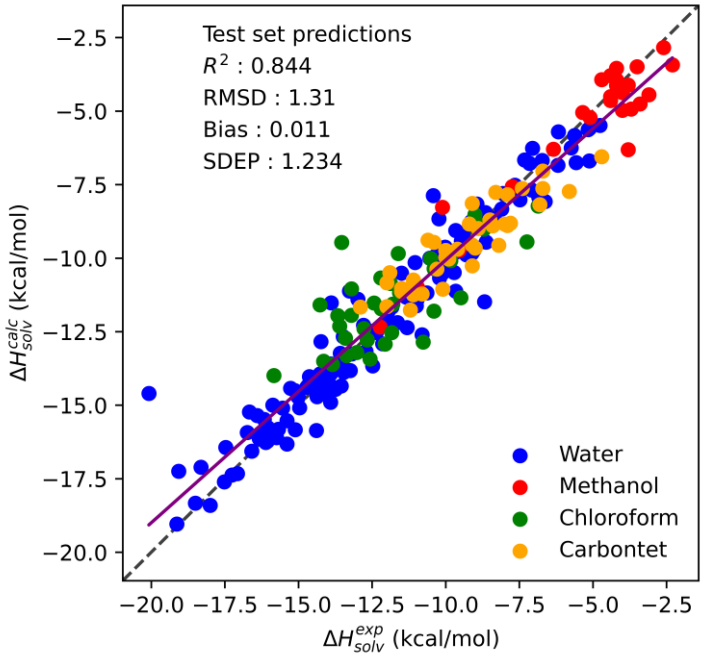


- Models trained separately on neutral and ionized solutes

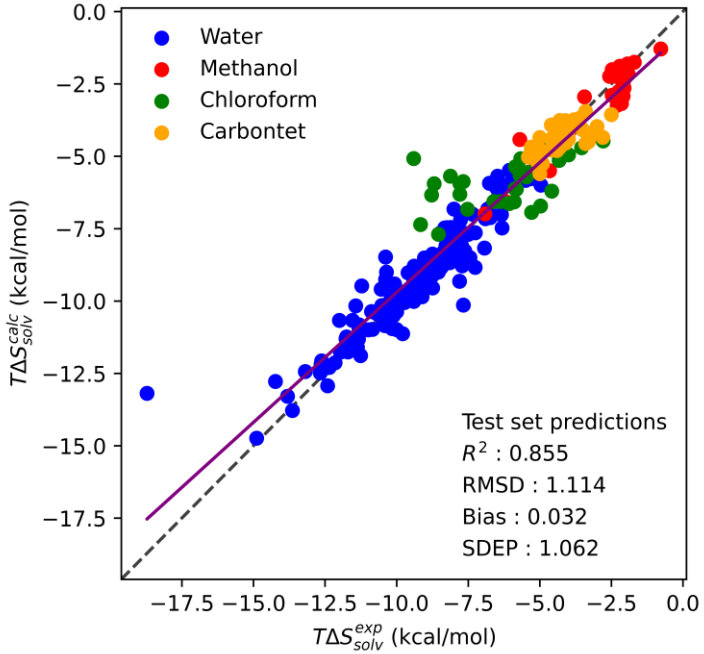
Predicting Entropy, Enthalpy and Free Energy

- *pyRISM-CNN model can be extended to predict entropy, enthalpy, and free energy of solvation together using **multi-task learning** and **transfer learning***
- 238 neutral organic solutes over 4 solvents (water, methanol, chloroform, carbon tetrachloride)

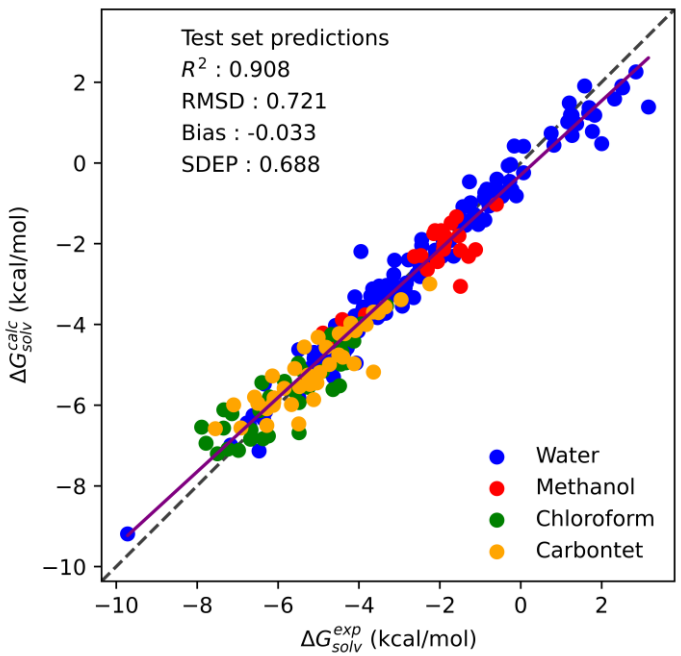
Enthalpy (ΔH)



Entropy (TΔS)



Free Energy (ΔG)



Key Points and Open Questions

Key Points

- single CNN model makes predictions for different solvent systems and varying temperatures
  - grounded in statistical physics
  - solvation free energy densities adapt to changing conditions

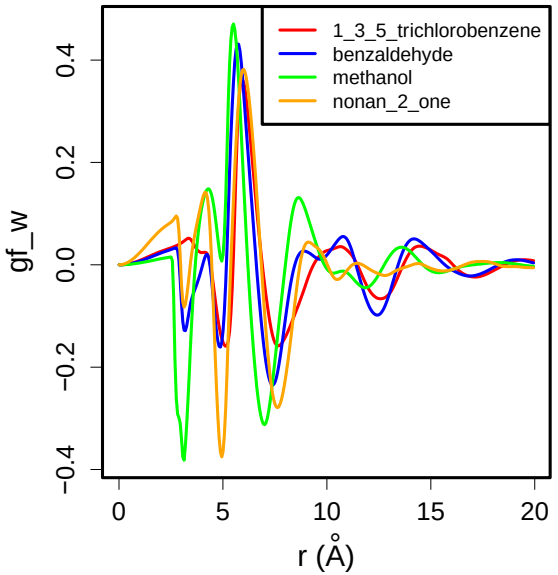
- ~40-fold reduction in predictive error as compared to standard 1D RISM theory

- fast calculation time on single CPU (1-2 mins)
  - suitable for moderate throughput screening

Open Questions

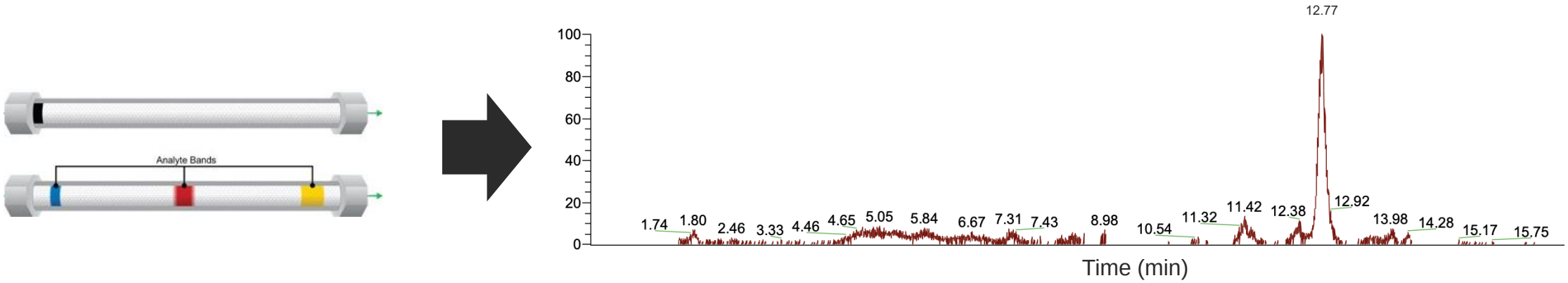
- *Are the solvation descriptors derived from RISM useful for predicting other physchem properties?*

- Potential benefits:
  - tailorable to different solvent environments



Quantitative Structure-Retention Relationships

- In high performance liquid chromatography (HPLC), mixtures are separated based on the varying ratio of affinities that solutes have for a stationary phase (SP) and mobile phase (MP)
- Compounds have a characteristic retention time for a given chromatographic system
- QSRR models, relating retention time to molecular structure, are widely used to identify compounds in HPLC screening experiments, and to guide chromatographic method development[1]

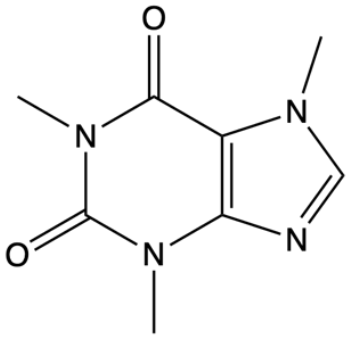


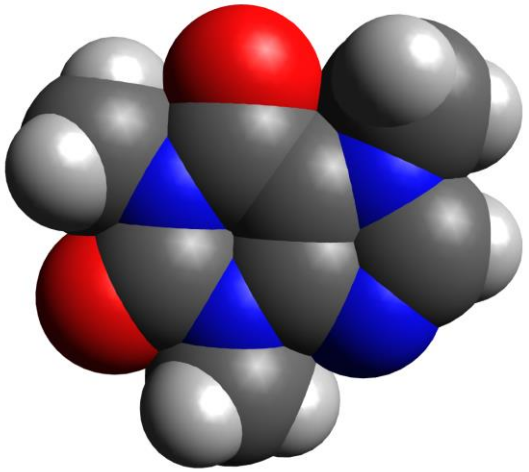
Example MS chromatogram from spiking experiment[2]

[1] P. R. Haddad, M. Taraji and R. Szücs, *Anal. Chem.*, 2021, **93**, 228–256.
[2] M. Cao, K. Fraser, J. Huege, T. Featonby, S. Rasmussen and C. Jones, *Metabolomics*, 2015, **11**, 696–706.

Molecular Descriptors for Quantitative Structure-Retention Relationships

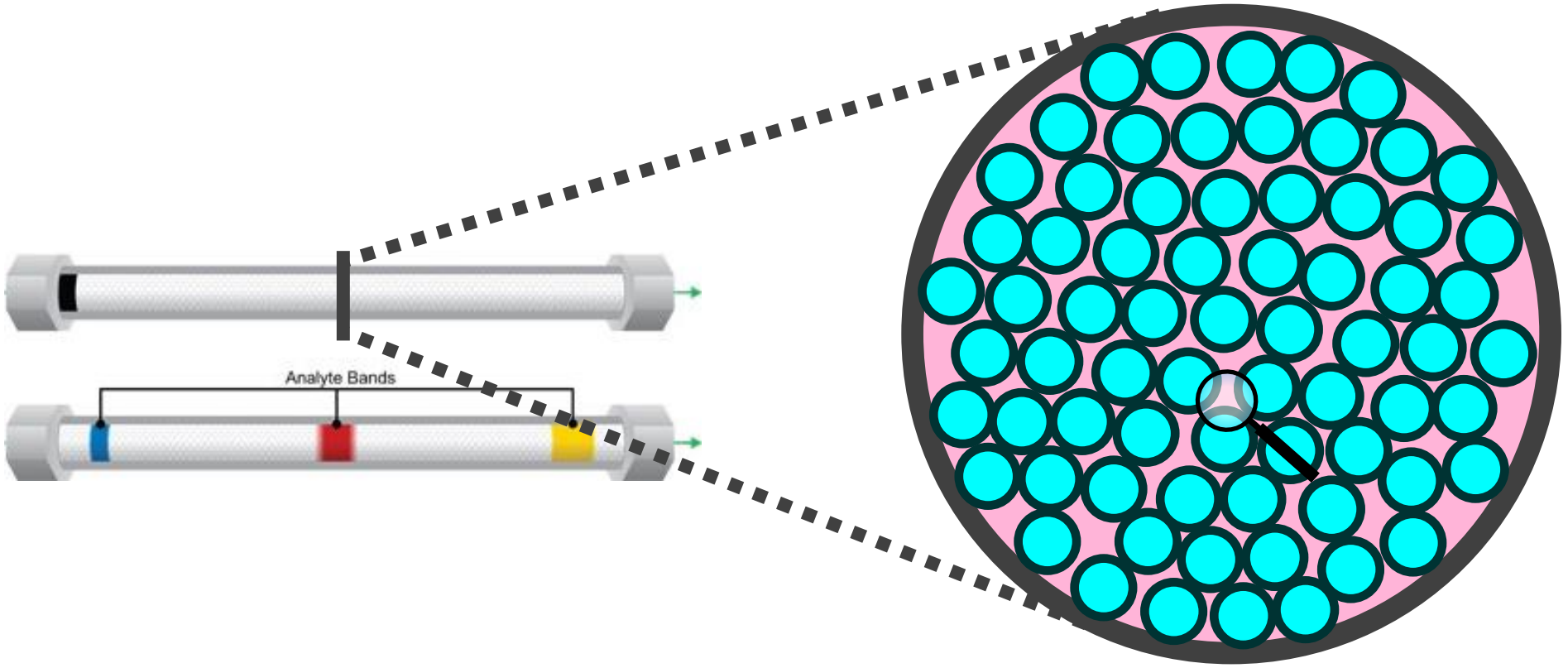
- Their accuracy relies on high quality data, including chosen molecular descriptors
- Typical descriptors are…
  - 2D: MW, atom counts, FG counts, topological indices e.g. Burden matrices
  - 3D: VolSurf, geometric, QM
  - Most are solute centred descriptors
- Retention behavior is governed by solvation/partition interactions
  - Can adding some solvation physics help?





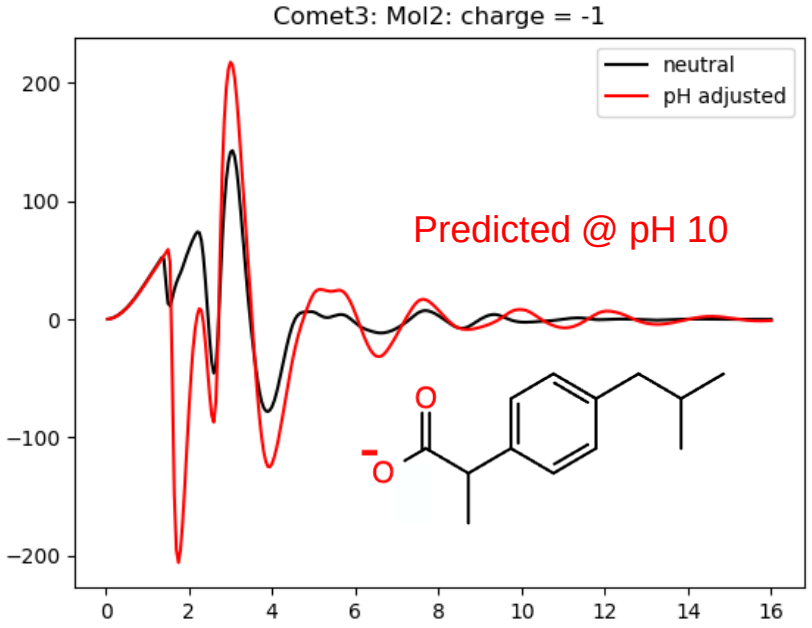
Aim: *Develop chromatography-specific descriptors that represent the chemical environment: partitioning between SP and MP, changing solvent composition (gradient elution), and varying pH.*

Inside a reversed-phase HPLC column...



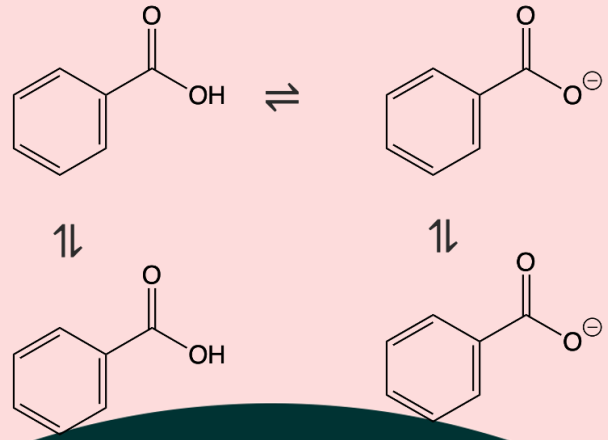
RISM descriptors have been tailored to specific chromatographic conditions

- ✔ Different solvents: water, methanol, and acetonitrile
- ✘ Gradient elution: mixtures
- ✘ Varied pH: buffer ions in solvent
- ✔ Varied pH: charge on solute



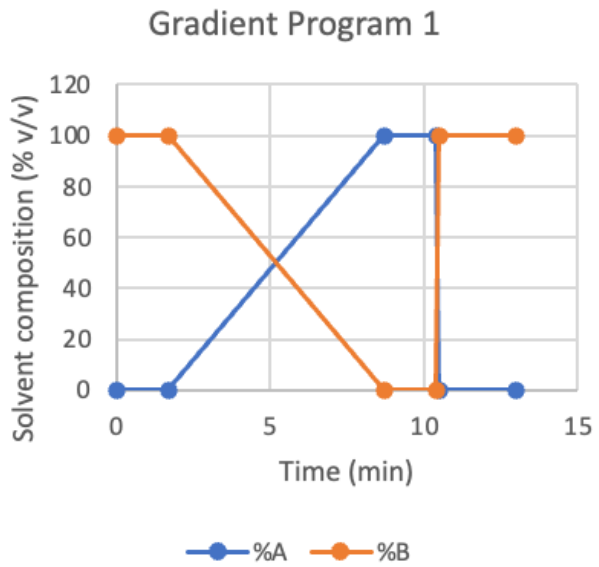
Mobile phase:

- *Gradient: water (weakly eluting) to organic (strongly eluting)*
- *Cosolvent: Acetonitrile or methanol*
- *pH*



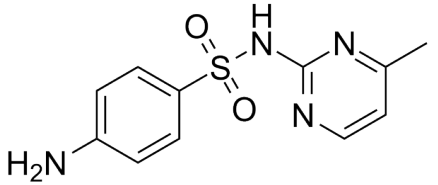
Six datasets with different chromatographic conditions have been used to validate the models

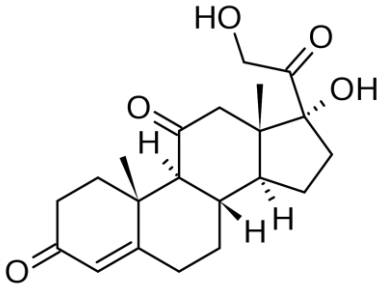
Comet	Mobile phase A	Mobile phase B	pH	Column chemistry	No. of Solutes
1	ACN/H2O (95:5 v/v)	0.1% Formic acid	2.6	C18/carbamate	105
2	MeOH/H2O (95:5 v/v)	10mM Ammonium Acetate	6.8	C18	117
3	MeOH/H2O (95:5 v/v)	0.1% Ammonium Hydroxide	10	phenyl	112
4	ACN/H2O (95:5 v/v)	10mM Ammonium Hydroxide/ACN (95.5 v/v)	10.5	C18/carbamate	102
5	ACN/H2O (95:5 v/v)	10mM Ammonium Acetate	6.8	phenyl	115
6	ACN/H2O (95:5 v/v)	0.1% Formic acid/Acetonitrile (95:5 v/v)	2.6	phenyl	113

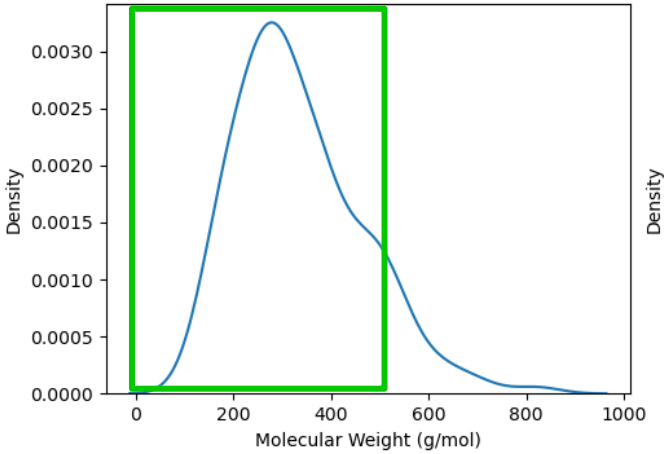


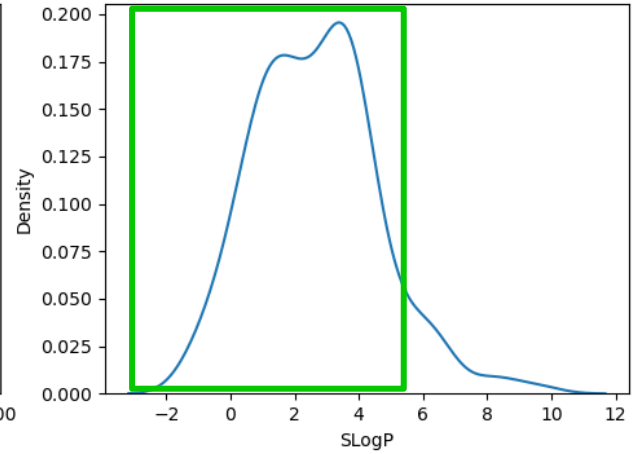
Gradient elution in ~10 mins

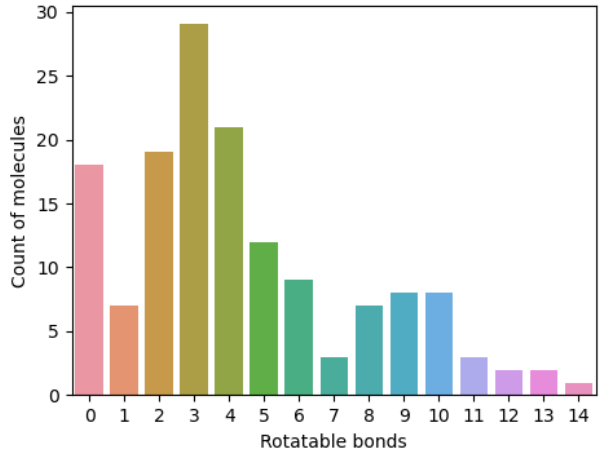
The 149 unique solutes in the six Comet datasets are druglike and chemically diverse

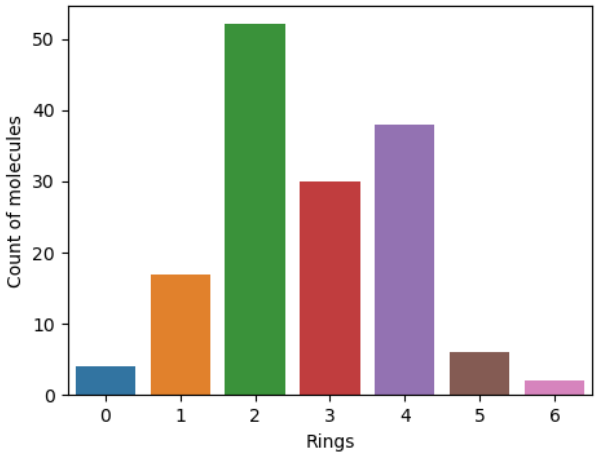


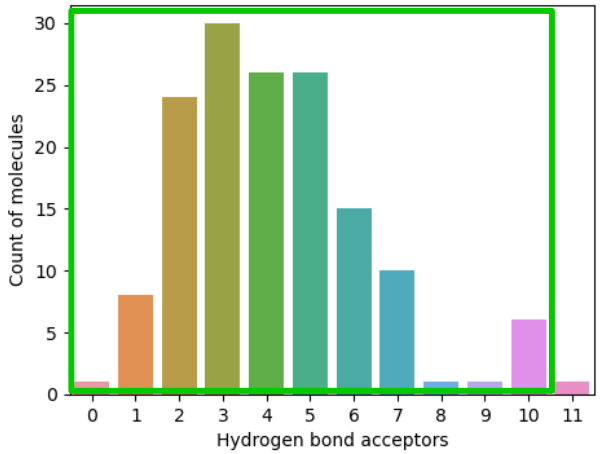


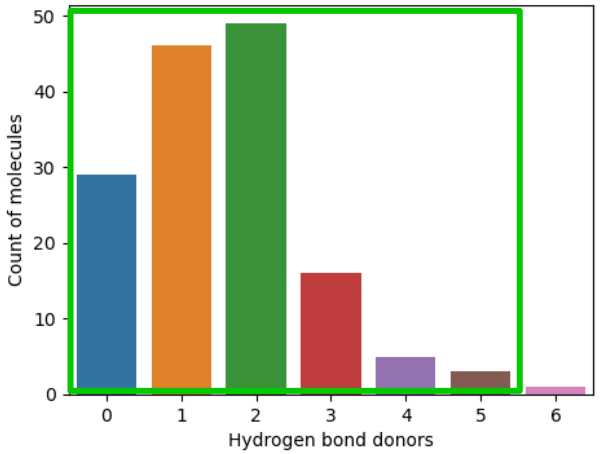


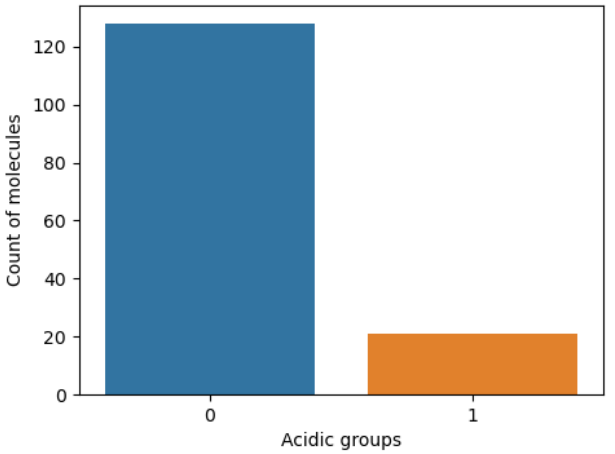


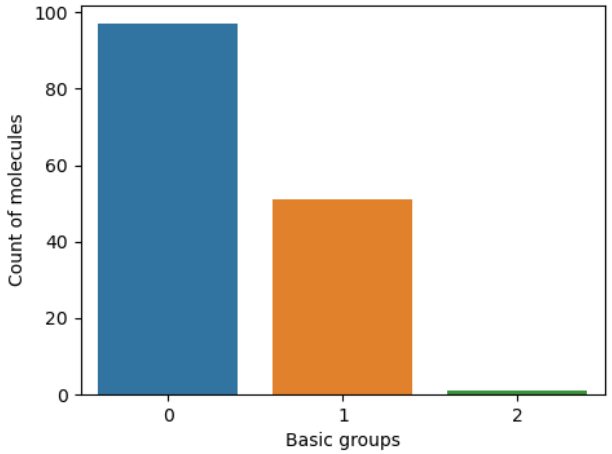






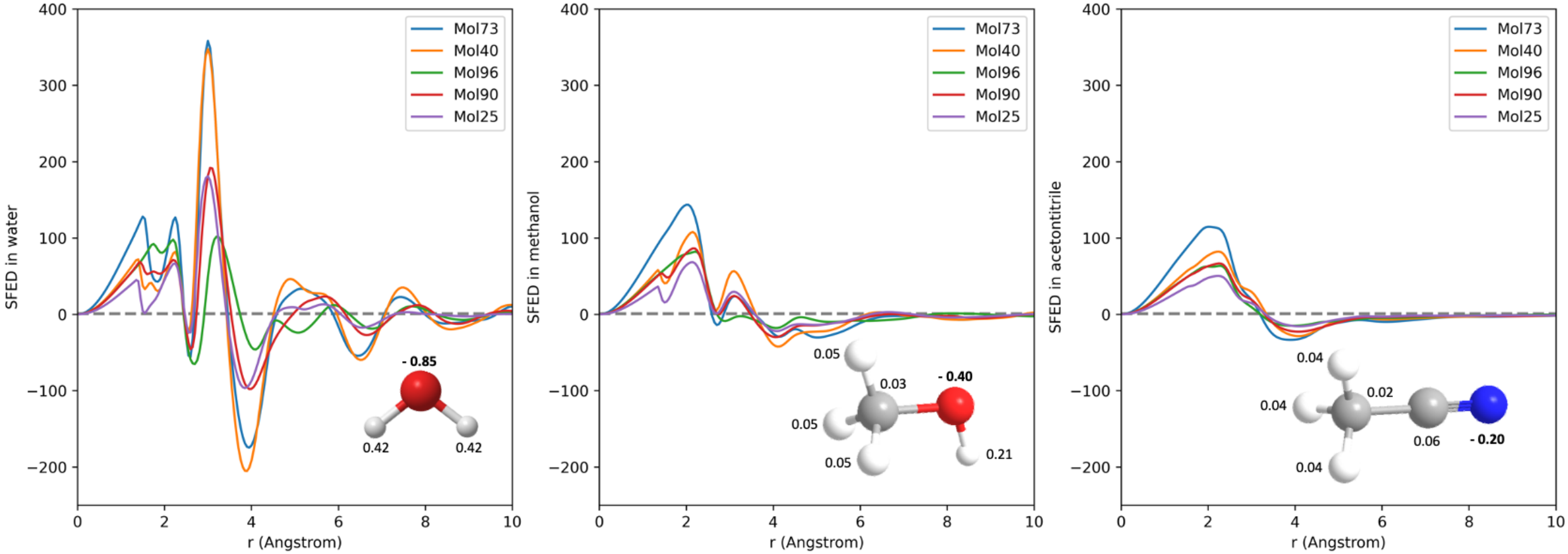




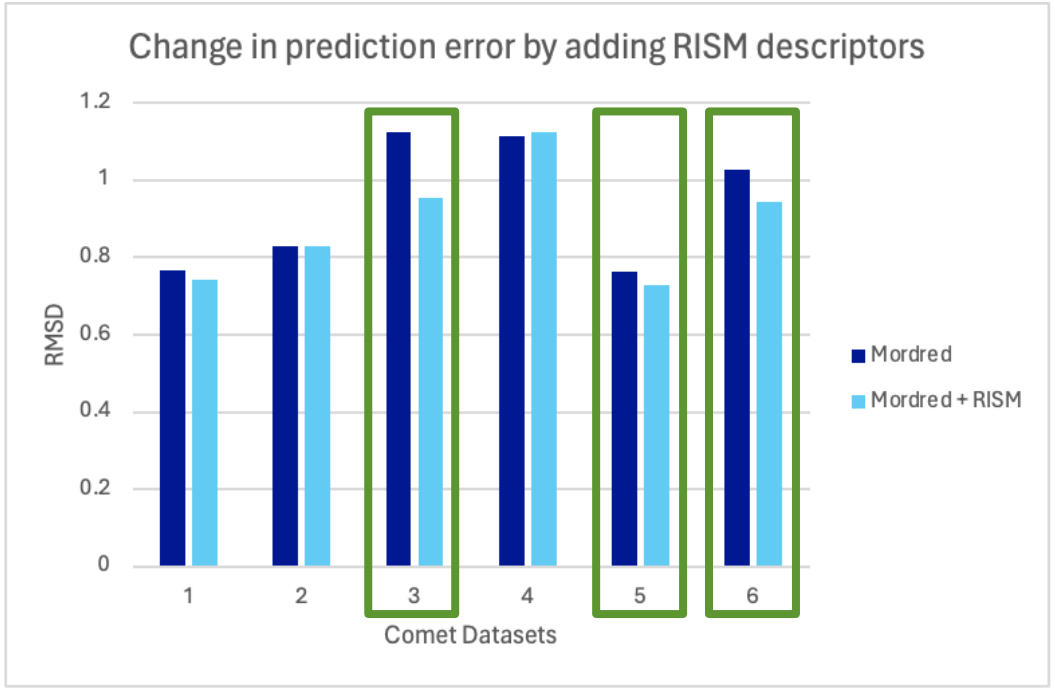


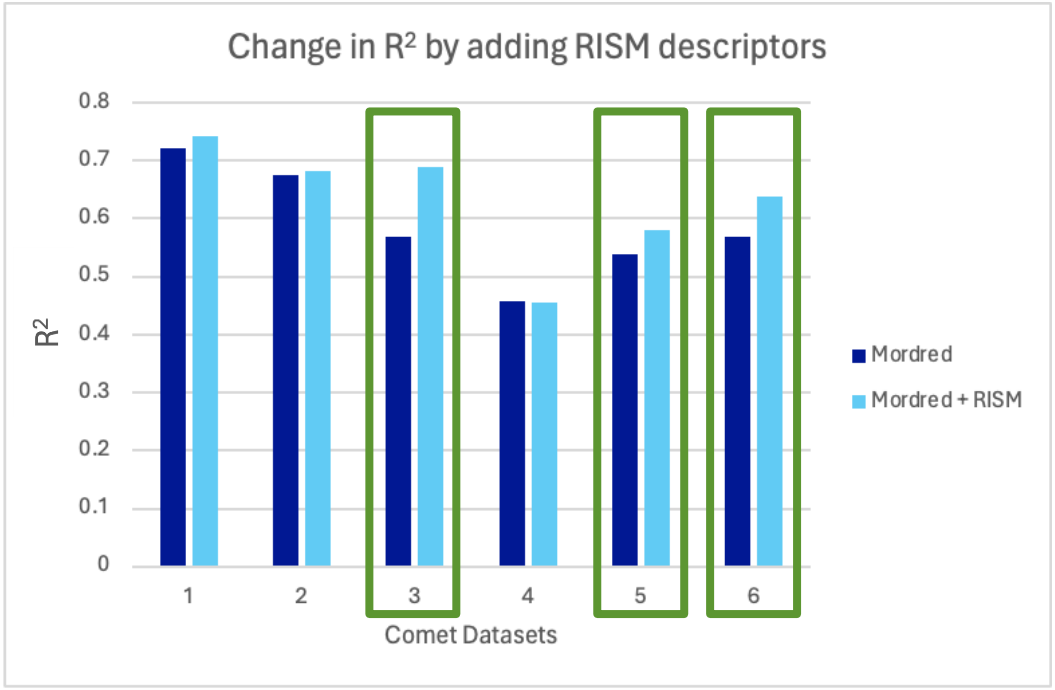
How does solvent choice affect RISM descriptors?

- Degree of structure in the SFED corresponds to the magnitude of partial charges



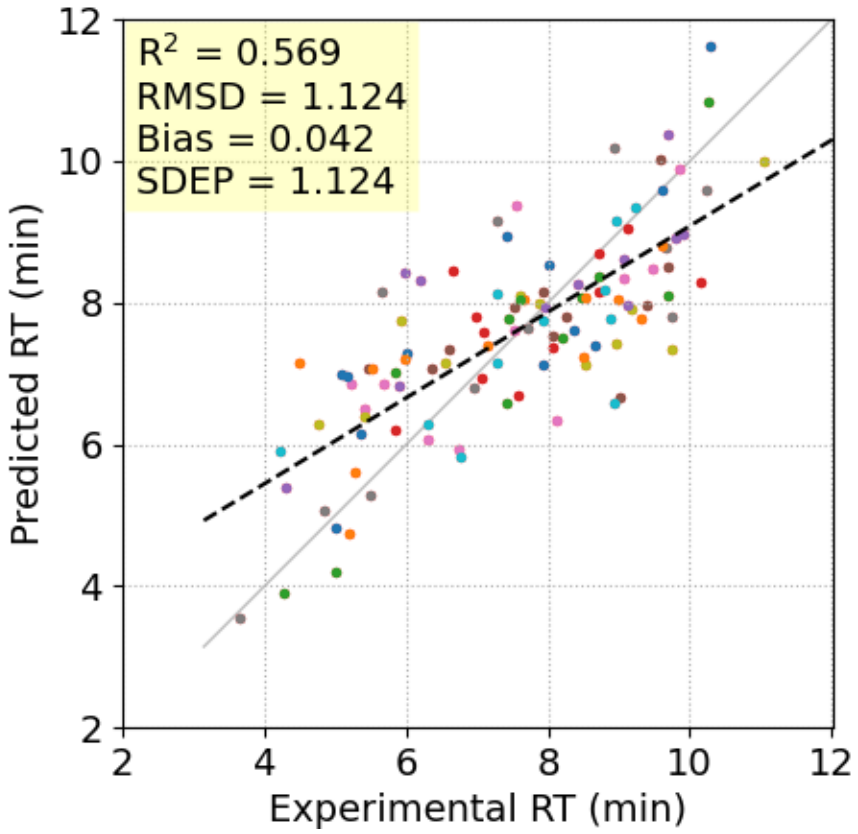
Addition of RISM solvation descriptors calculated for neutral solutes in pure water leads to an improvement in predictive accuracy for some columns



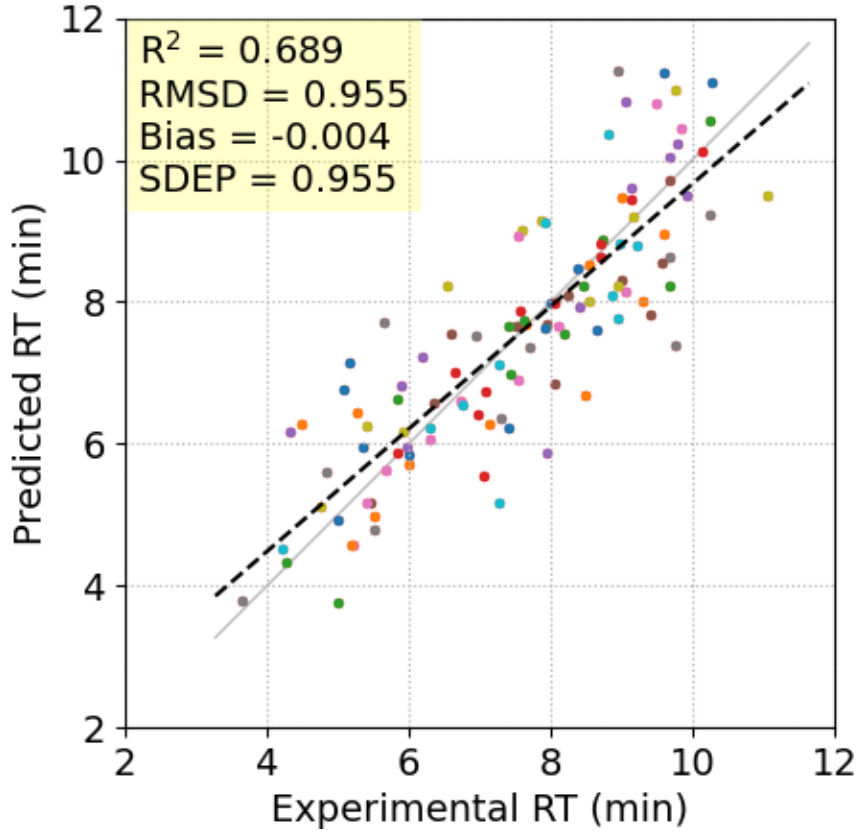


Addition of RISM solvation descriptors calculated for neutral solutes in pure water leads to an improvement in predictive accuracy for Comet 3

PLS: Top 20 Mordred

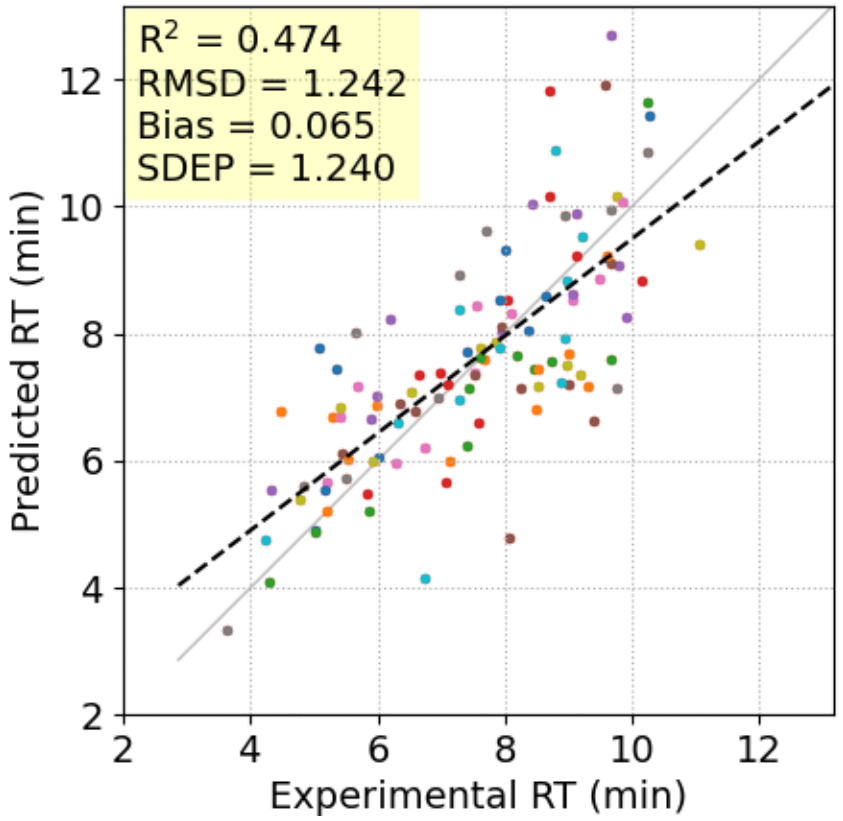


PLS: Top 20 Mordred + RISM

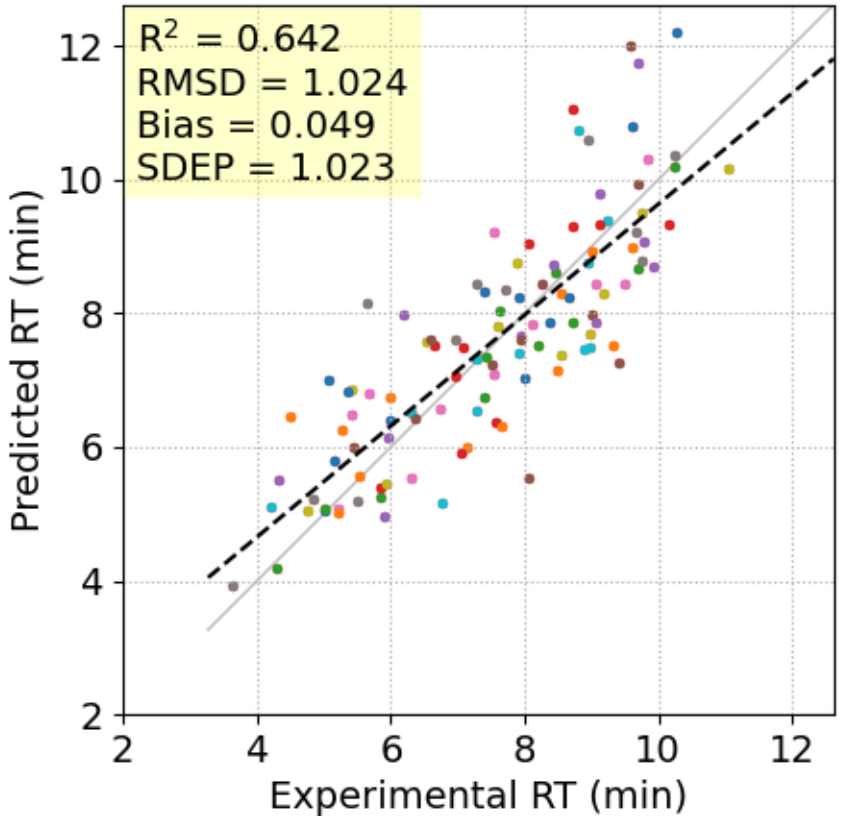


Addition of RISM solvation descriptors calculated for neutral solutes in organic solvent leads to an improvement in predictive accuracy for Comet 3

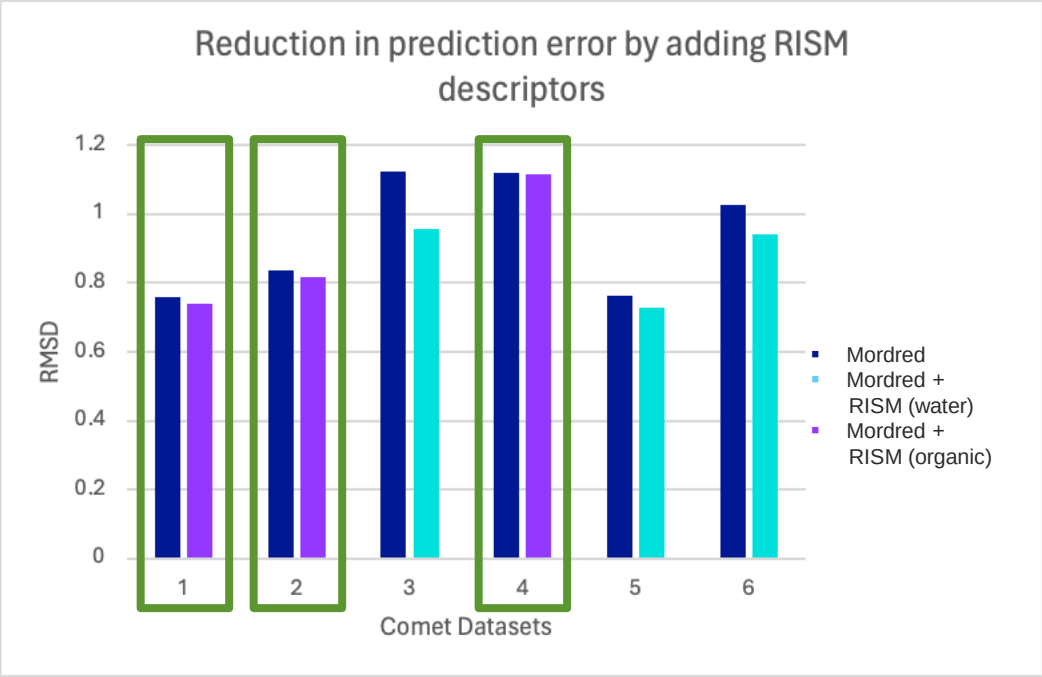
PLS: Top 20 Mordred

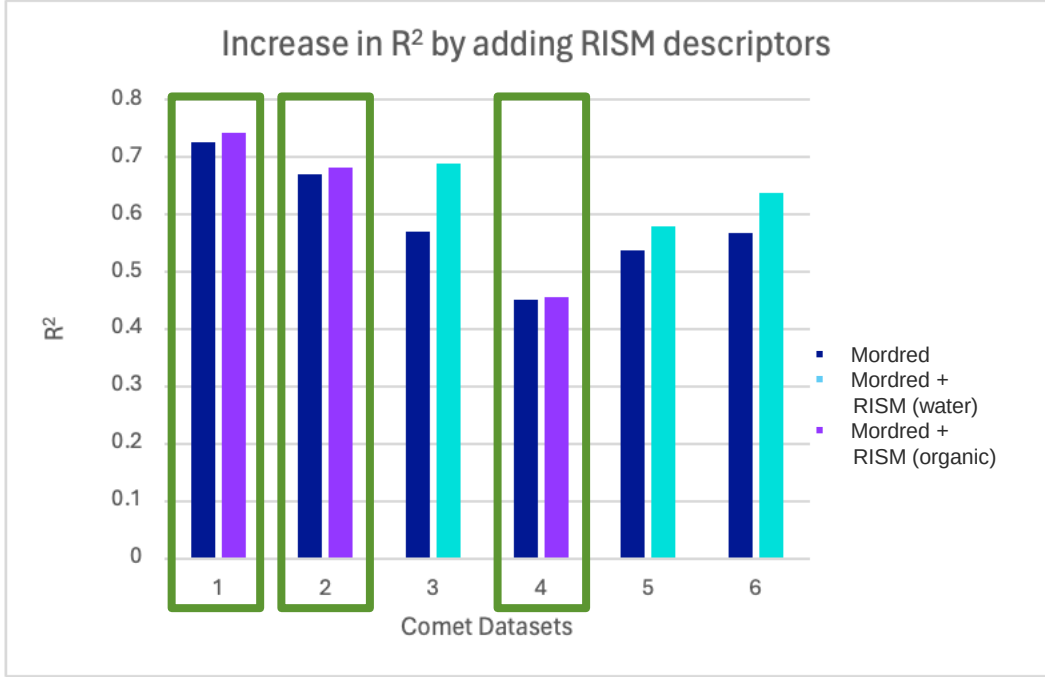


PLS: Top 20 Mordred + RISM



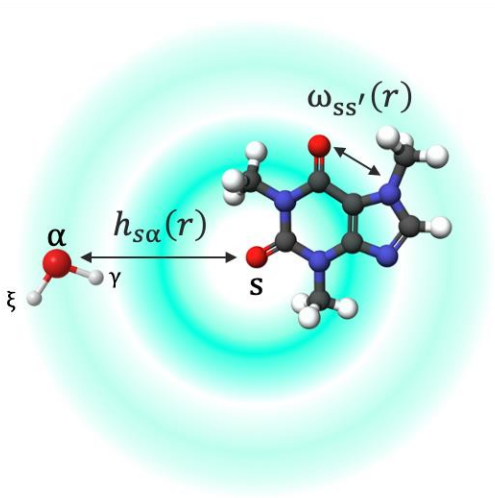
Incorporation of RISM solvation descriptors calculated for neutral solutes in organic solvent leads to a small but consistent improvement in predictive accuracy

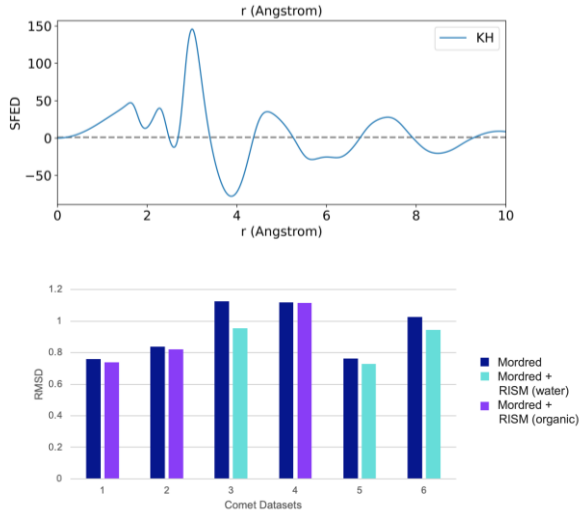


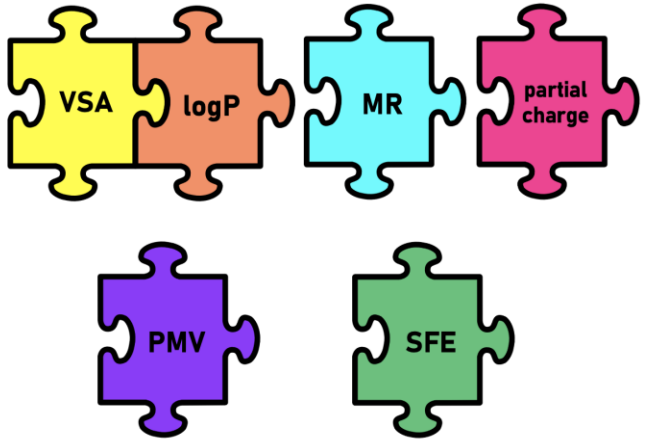


Conclusions

- Physics-based solubility prediction has been demonstrated for druglike molecules.
- A CNN free energy functional for RISM enables accurate predictions of ΔG_{solv}, ΔH_{solv}, ΔS_{solv} for neutral and ionised solutes, and a wide-range of temperatures.
- RISM solvation descriptors have been tailored to QSRR by solving for various solvents + solute states
- A small but consistent increase in accuracy has been achieved with SFED-type RISM solvation descriptors, relative to Mordred descriptors alone







Acknowledgements

- **Madeleine Taylor**, **Abdullah Ahmad**, **Dr Dan Fowles**, and Jonathan Conn.
- Prof Sally Price (UCL), Dr Roman Szücs (Pfizer), Dr Roland Brown (Pfizer), Dr Lucy Morgan (Pfizer) and Dr Jane Kawakami (Pfizer), Dr John Mitchell (St Andrews)
- Thank you to Pfizer for funding via the Community for Analytical Measurement Science (CAMS).
- Results were obtained using the ARCHIE-WeSt High Performance Computer based at the University of Strathclyde (www.archie-west.ac.uk).

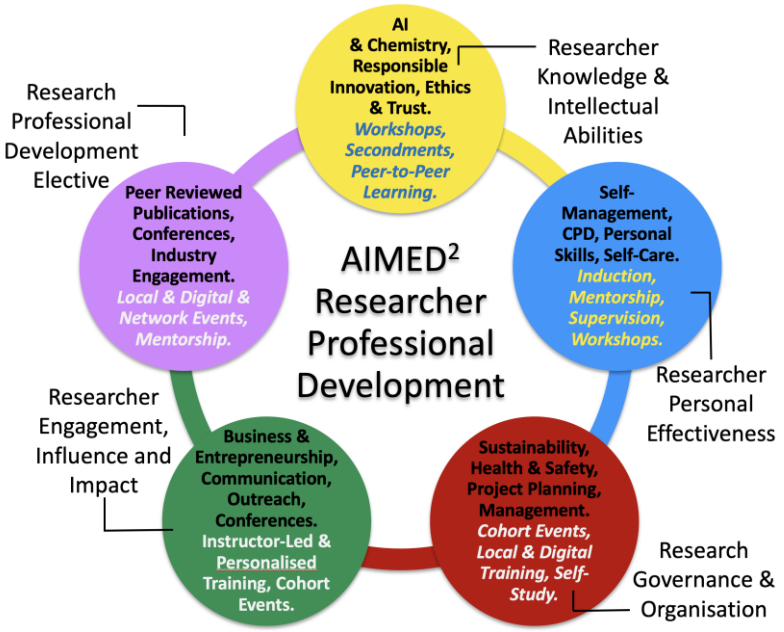
AIMED²

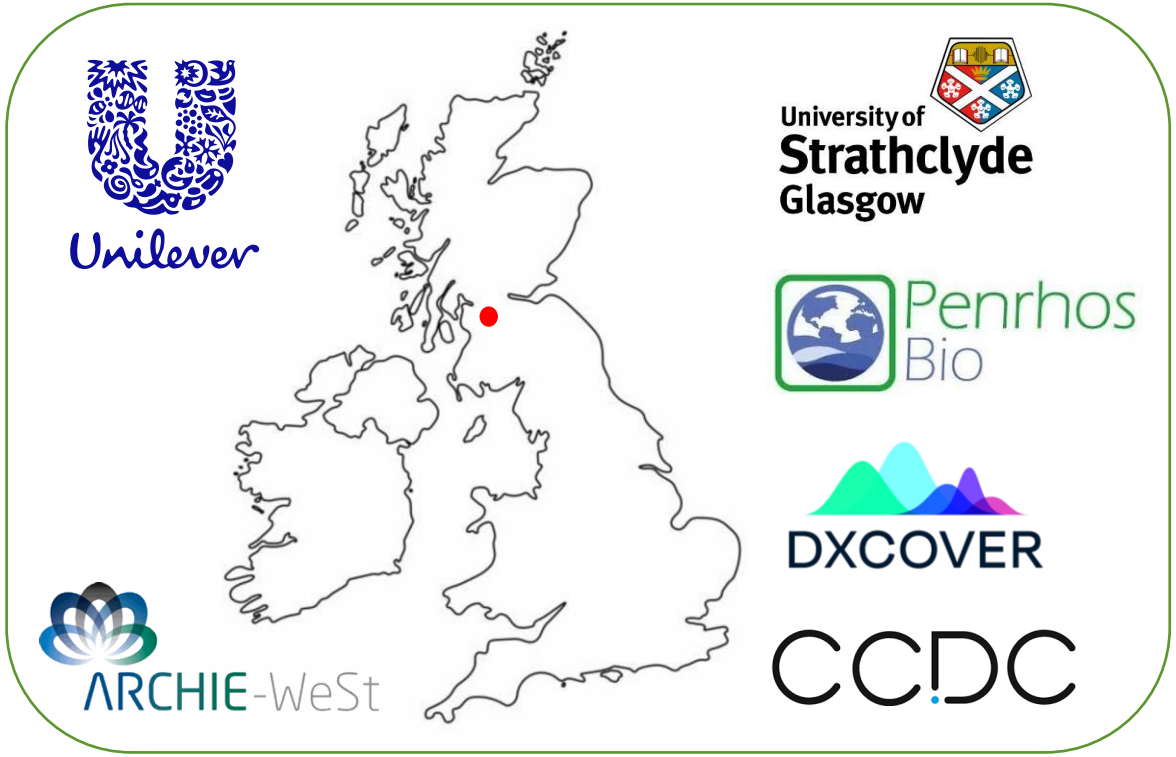
STRATHCLYDE CDT
AI FOR MOLECULAR EXPLORATION, DISCOVERY AND DEVELOPMENT

- First cohort of 5 PhD students beginning in October 2024
  - Academic supervisors from Chemistry, Physics and Computer Science

- Key Research Challenges:
  1. Discovery and Development of Functional Molecules and Materials
  2. Analysis of Complex Chemical Systems
  3. Artificial Chemical Intelligence

- We aim to grow the centre over the coming years!





X THE PLACE OF USEFUL LEARNING

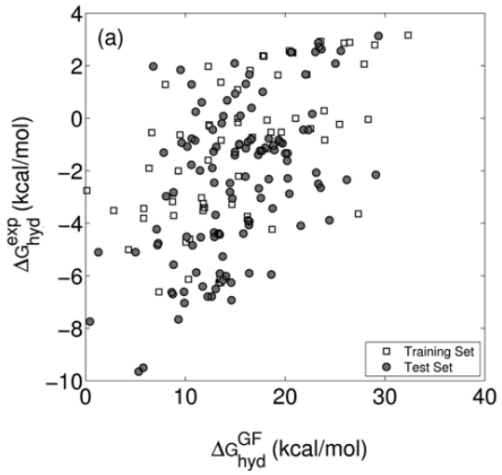
University of Strathclyde

Science

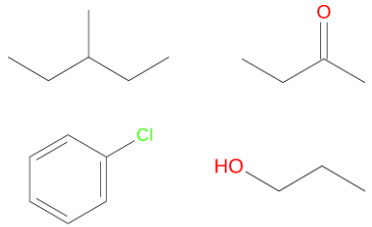
University of Strathclyde Science

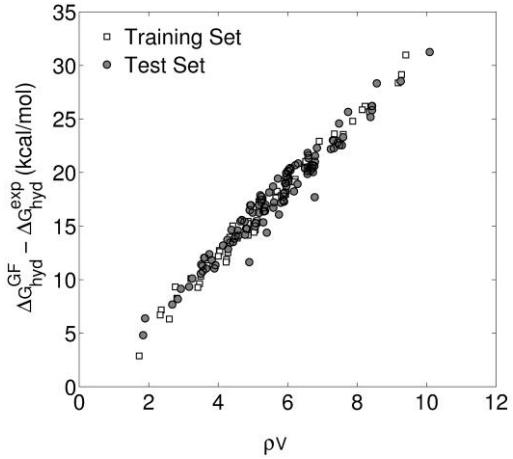
Thank You For Listening!

Solvation free energies from the original 3DRISM theory have enormous errors

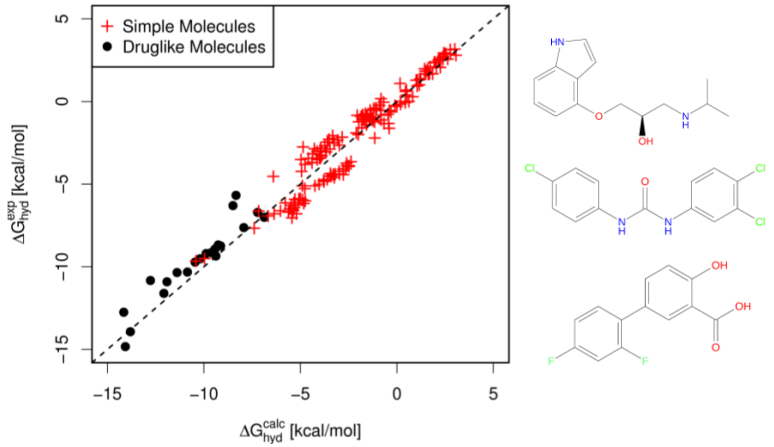


Simple Organic Molecules



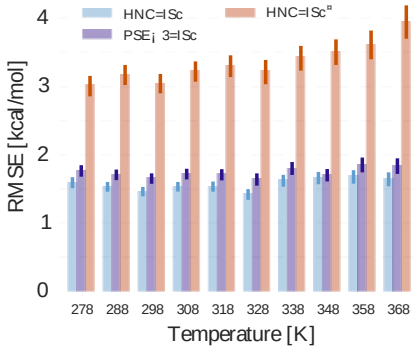


J. Phys. Cond. Matt., **2010**, *22*, 492101



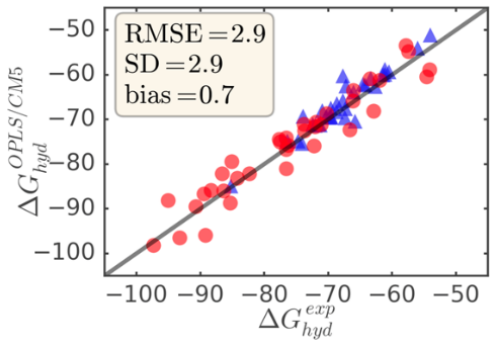
Mol. Pharmaceutics, **2011**, *8*, 1423

Non-ambient temperatures



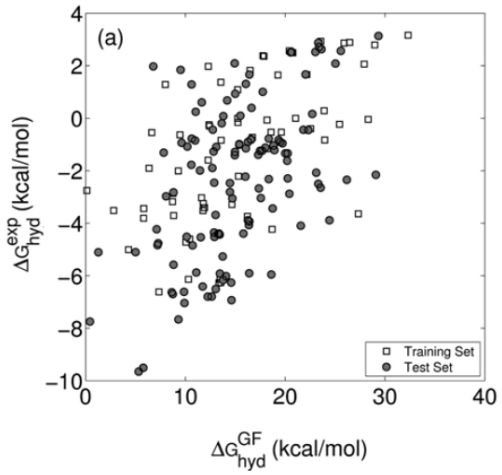
J. Chem. Phys., **2015,** 142, 091105.

Ionized Solutes

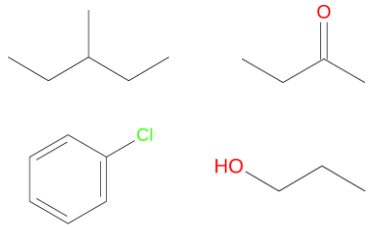


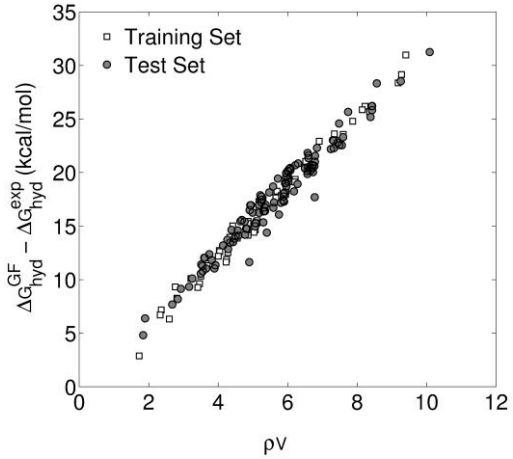
J. Phys. Chem. B. **2016**, *120*, 975–983

Solvation free energies from the original 3DRISM theory have enormous errors



Simple Organic Molecules





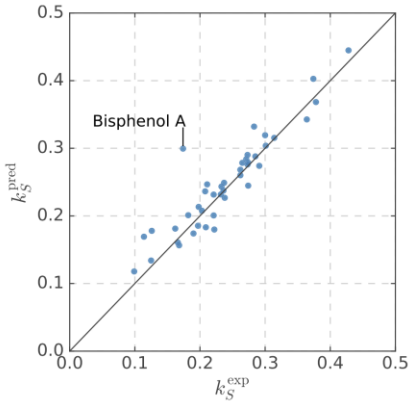
J. Phys. Cond. Matt., **2010**, *22*, 492101

Organic Solvents

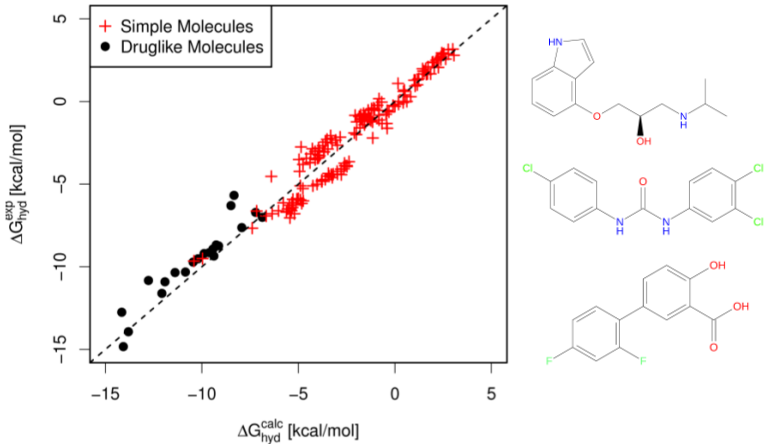
Solvent	N	RMSE	SDE	bias
	Apolar			
1,2-dichloroethane	39	1.16	1.07	0.47
benzene	71	1.28	1.28	0.04
bromobenzene	27	1.17	1.15	-0.23
carbon disulfide	15	0.94	0.89	-0.30
carbon tetrachloride	79	0.85	0.84	-0.11
cyclohexane	103	1.01	0.75	-0.67
isooctane	32	0.98	0.68	-0.70
n-decane	39	1.70	1.23	-1.17
n-decane (4-mer)	39	0.68	0.56	-0.38
n-heptane	67	0.95	0.86	-0.42
n-heptane (3-mer)	67	0.74	0.74	0.05
toluene	51	1.00	0.99	0.08
xylene (mixture)	48	1.00	0.99	-0.10

J. Chem. Phys.
2016, *145*, 194501

Salting-out Constants



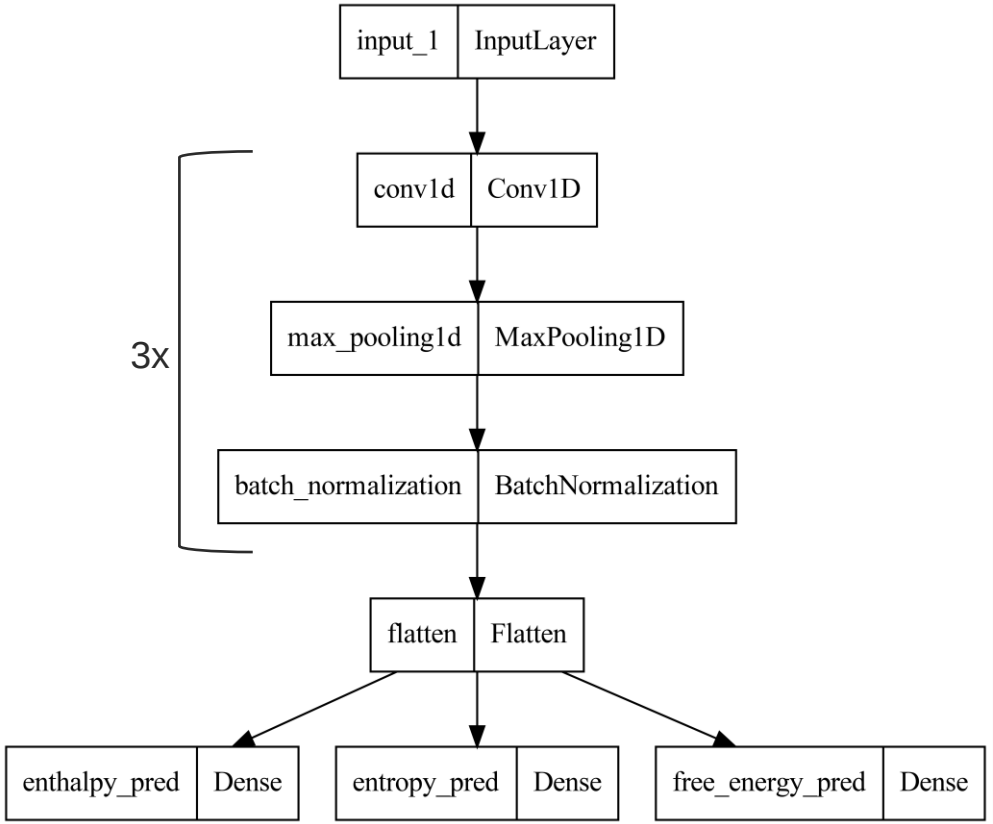
J. Chem. Phys.
2016, *145*, 194501



Mol. Pharmaceutics, **2011**, *8*, 1423

Implemented in the Amber Molecular Dynamics Package (https://ambermd.org/)

Multi-task CNN



Datasets

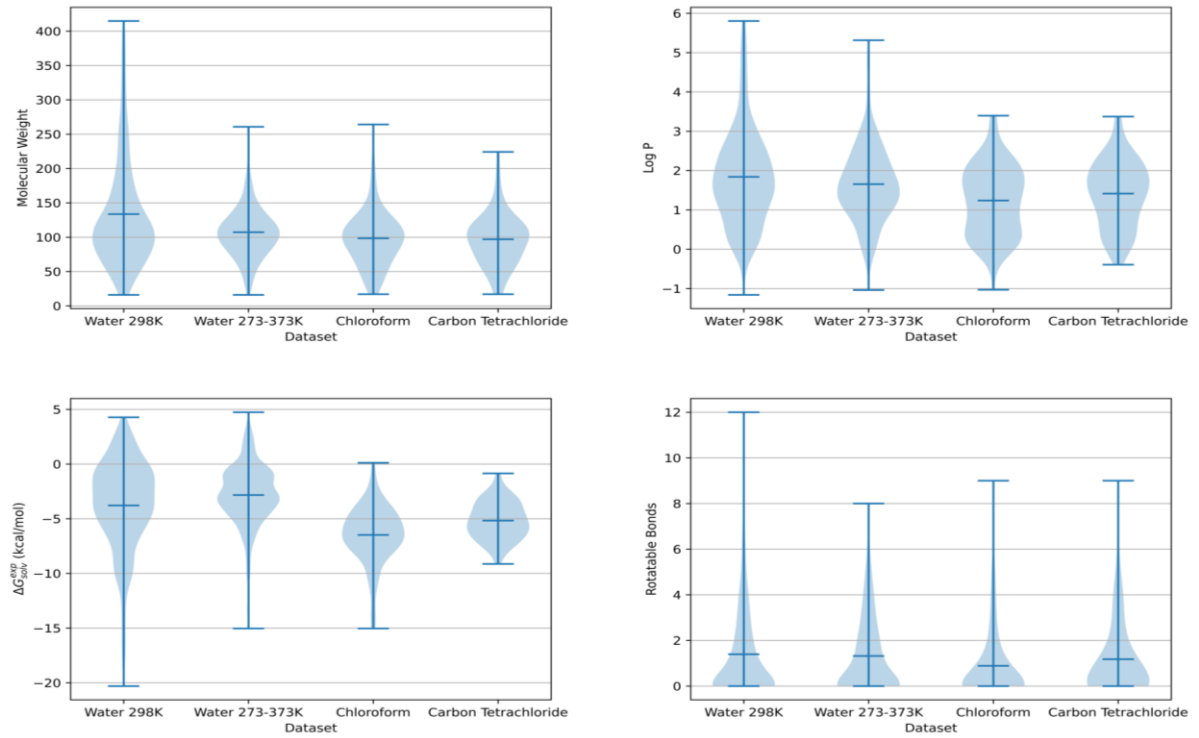
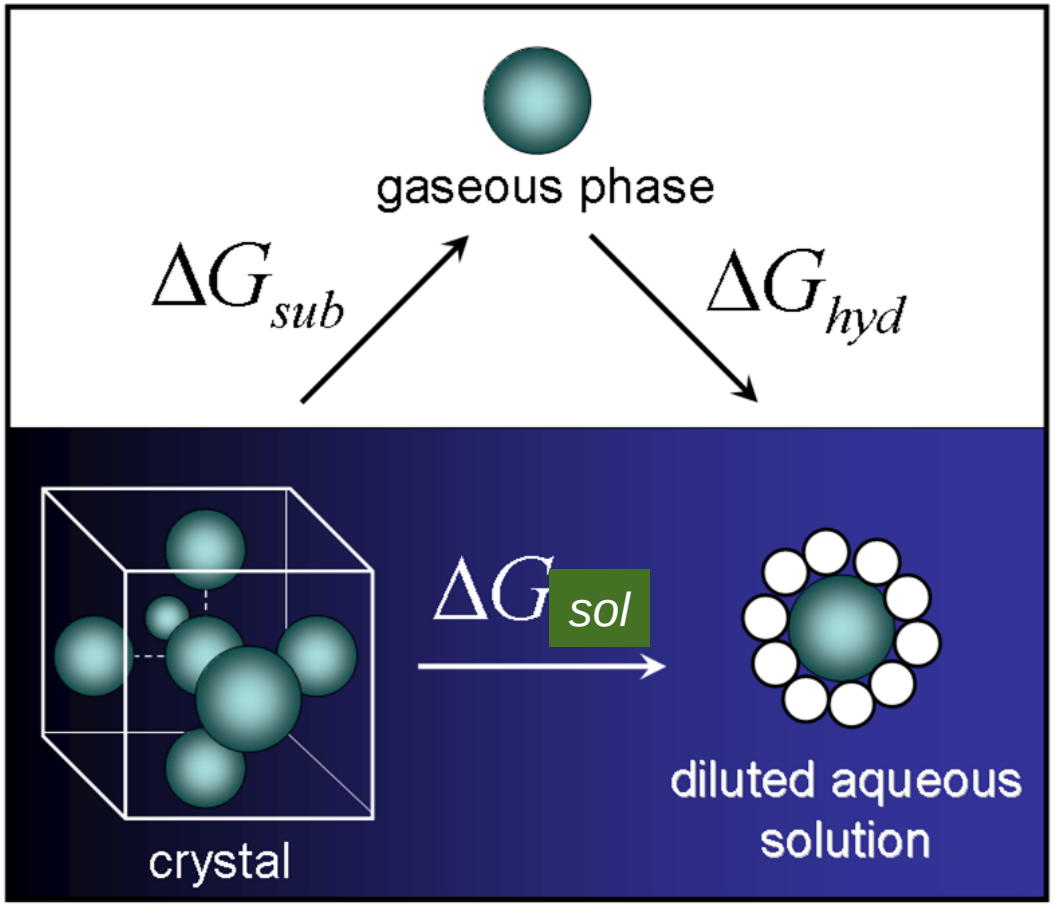


Figure 1: Violin plots showing the distribution of various descriptors for solute molecules within each dataset.

Multi-solvent, multi-temperature dataset

chloroform 298 K (n=109)
carbon tetrachloride 298 K (n=79)
water 273-373 K (n=272, 3053 data points)
water 298 K (n=521)

Hydration Free Energy



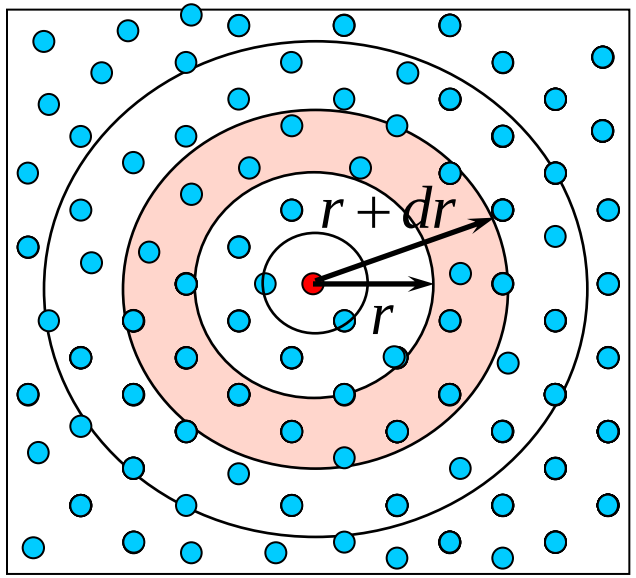
$$\Delta G_{hyd}$$

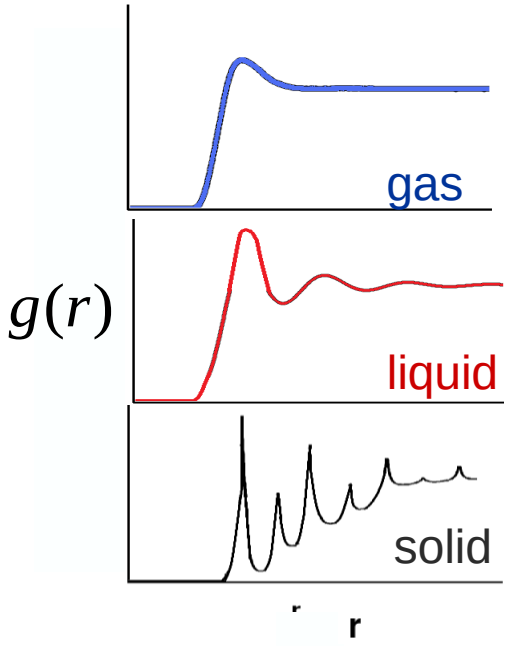
Hydration free energy is the change of the Gibbs free energy that accompanies the transfer of 1 mole of solute from gaseous phase to aqueous solution

$$\Delta G_{hyd} = -RT \ln K = -RT \ln \frac{[X]_{aq}}{[X]_g}$$

Ben-Naim A and Marcus Y 1984 *J. Chem. Phys.* ***81 2016–27***

IET operates with functions that describe the average density distributions of solute and solvent molecules



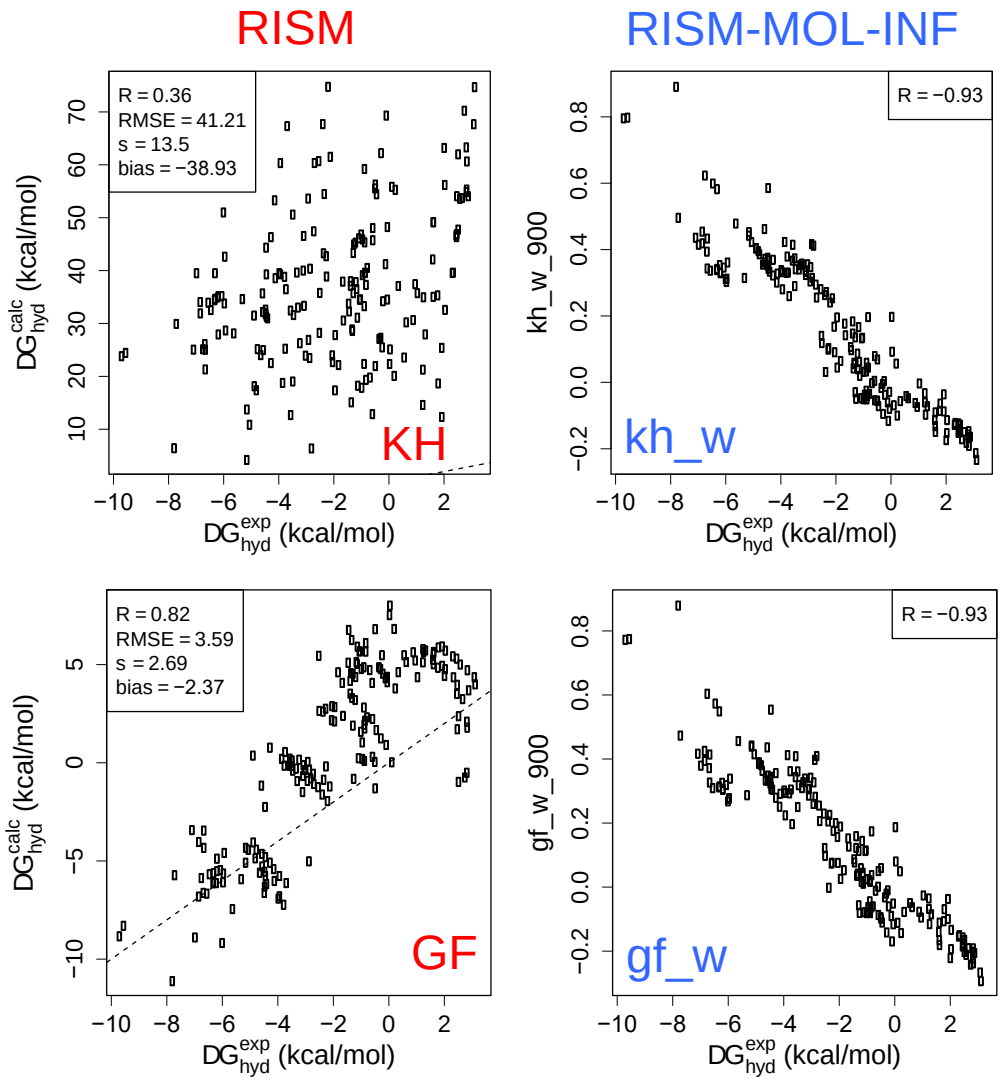


segment number density

$$g(r) = \left\langle \frac{\rho_{segment}(r)}{\rho_{bulk}} \right\rangle$$

bulk number density

$g(r_{12})=g_2(r)$ is known as the **radial distribution function** it is the factor which multiplies the bulk density to give the local density around a particle. If the medium is isotropic then $4\pi r^2 r g(r) dr$ is the number of particles between r and $r+dr$ around the central particle


RISM
RISM-MOL-INF
R = 0.36
RMSE = 41.21
s = 13.5
bias = −38.93
KH
R = −0.93
kh_w
R = 0.82
RMSE = 3.59
s = 2.69
bias = −2.37
GF
R = −0.93
gf_w
DG_hyd^calc (kcal/mol)
DG_hyd^exp (kcal/mol)
kh_w_900
gf_w_900