Tackling the chemical complexity of atmospheric particle formation by molecular level models

Hanna Vehkamäki

Centre of Excellence VILMA
University of Helsinki, Finland

Contents

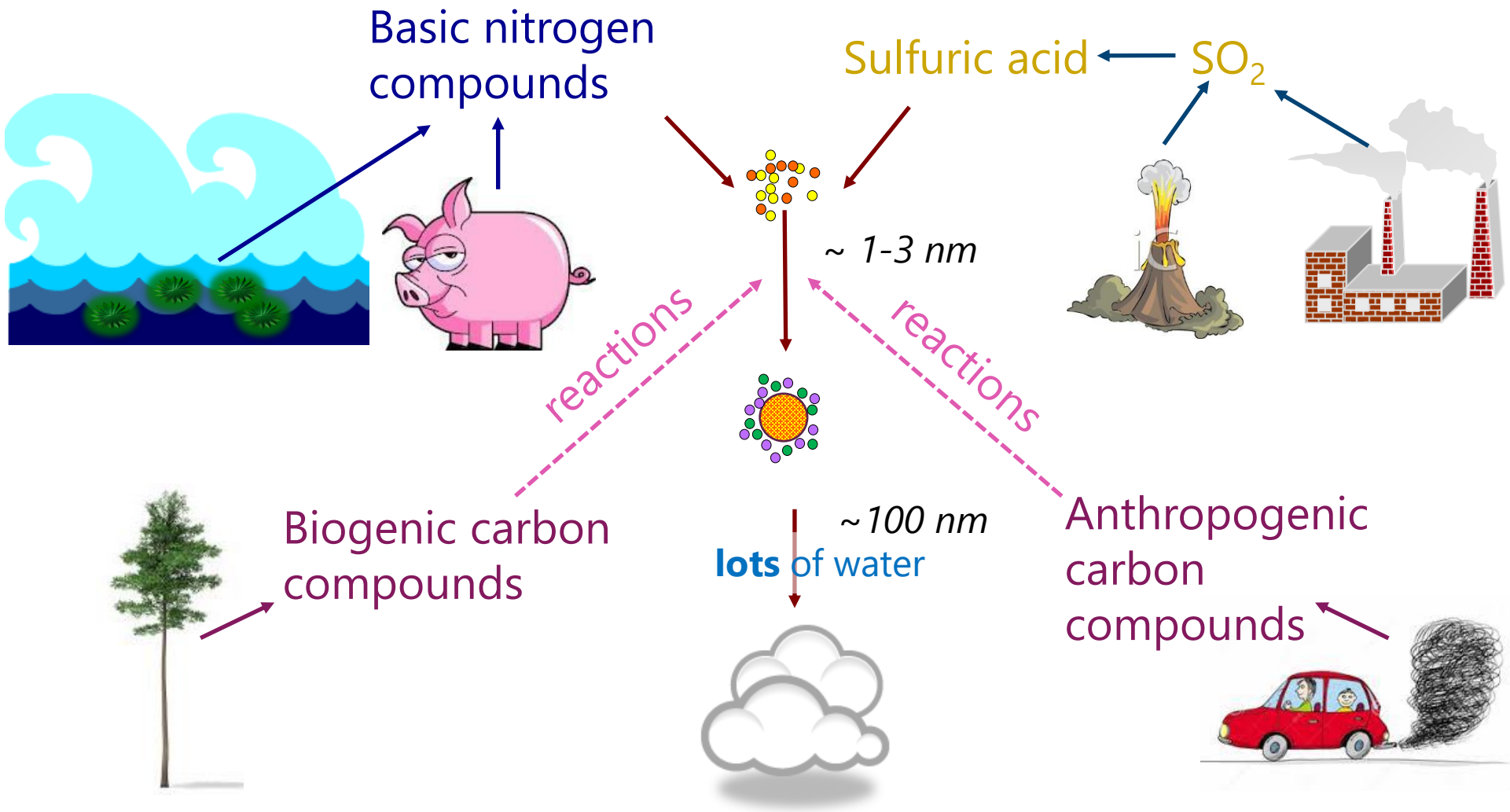
- Aerosol particles matter and they from in the air
- Complex chemistry of forming lowest volatility organics
- Conformer sampling of single molecules
- Predicting saturation vapour pressures
- Conformer sampling of clusters of molecules

- Conclusions

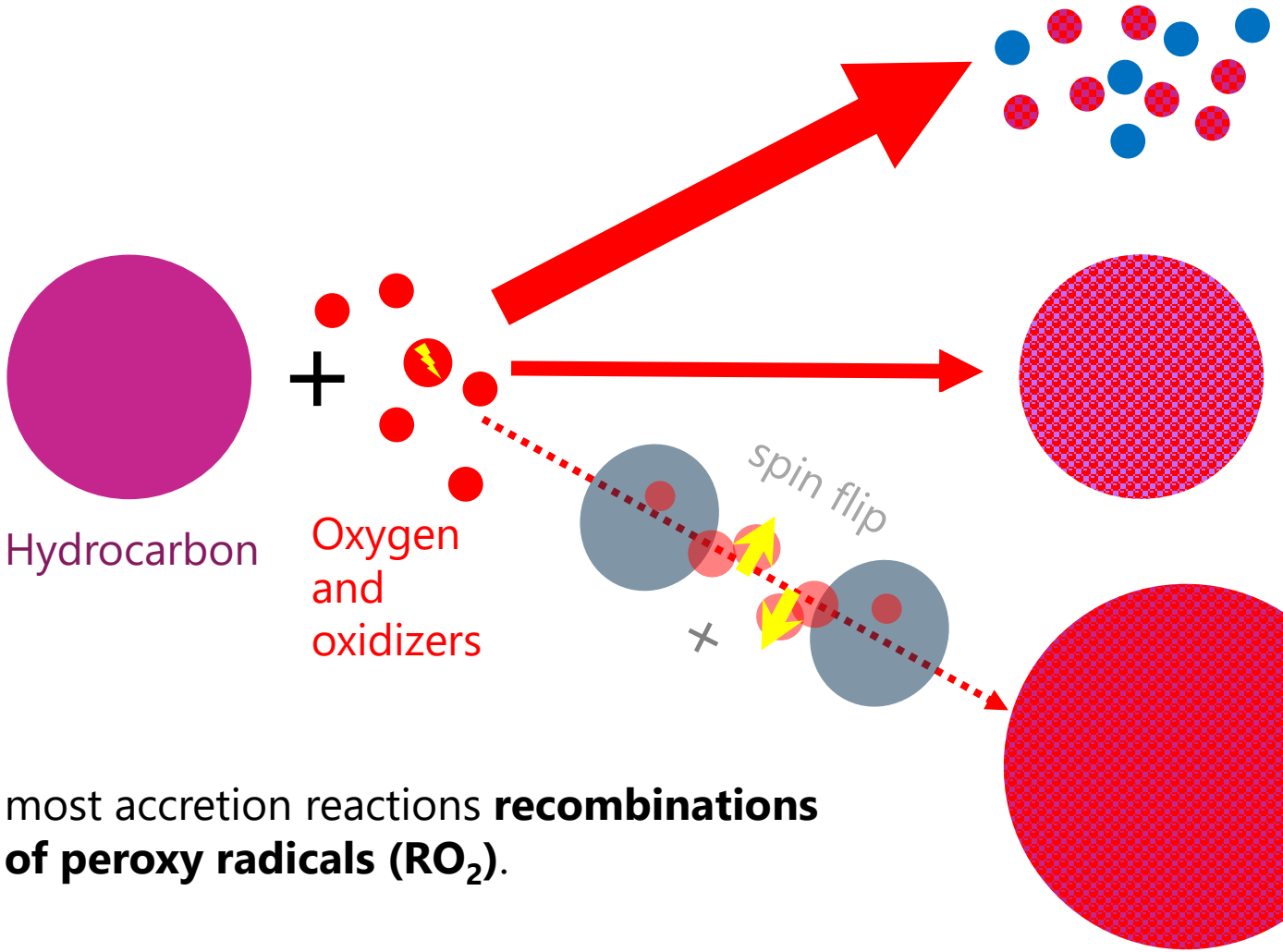
The aerosol paradox

- Aerosols (small particles in the nano-micrometer size range) are needed to form clouds.
- On average, the increase in cloud reflectivity due to human aerosol emissions has slowed global warming – but at the same time aerosols are the main reason for poor air quality.

Formation and growth of atmospheric aerosol



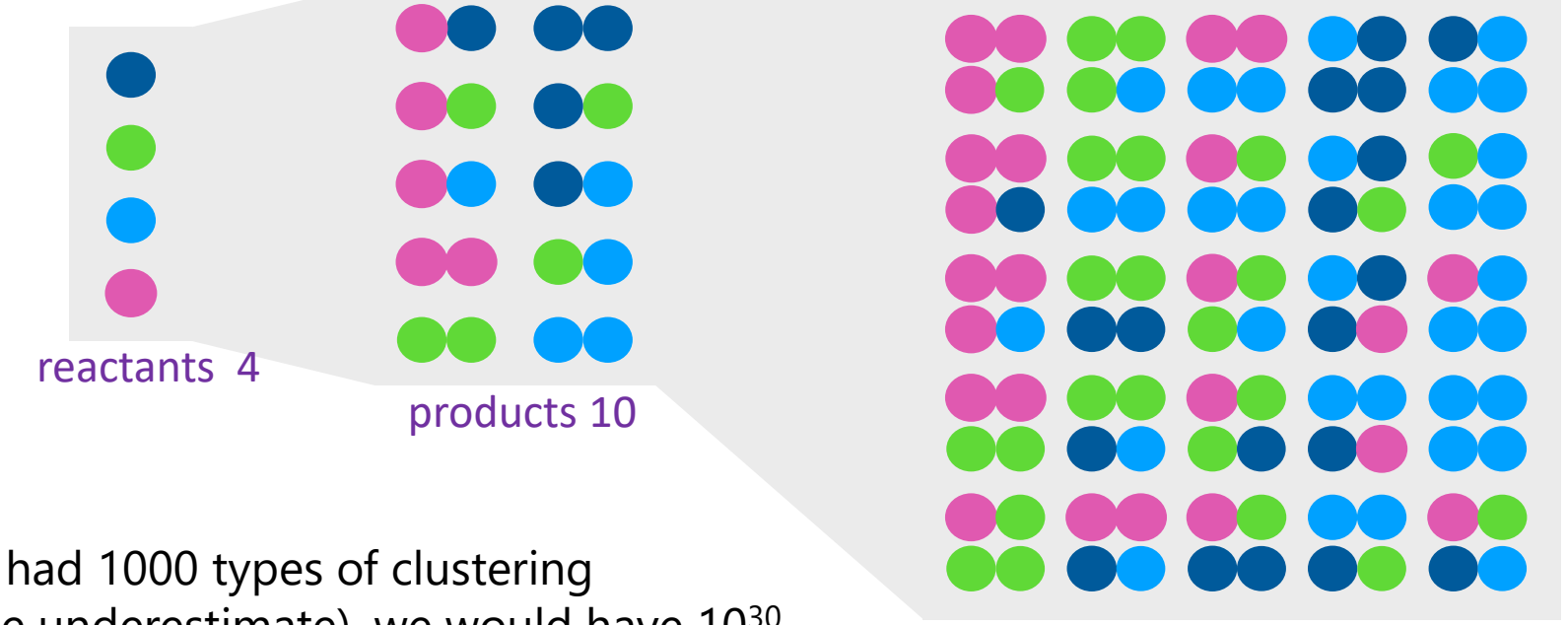
Condensing organic vapours are formed by oxidation



most accretion reactions **recombinations of peroxy radicals (RO_2)**.

Many layers of chemical complexity

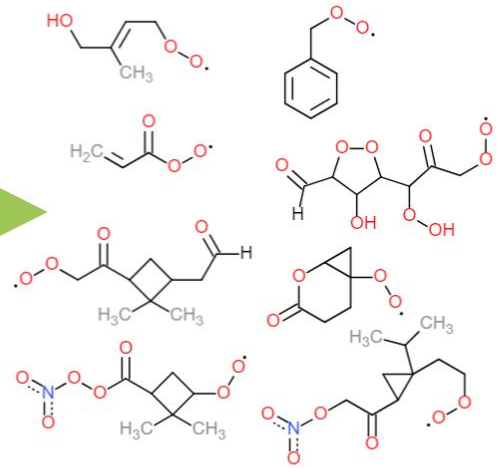
- **Reactants**: large number (> 10^5) of different reactants
  - each has a large number (often > 10^5) of 3-D structures: **conformers**.
  - reactions complicated: extremely difficult/expensive to model (or measure) accurately.
- **Products**: each **combination** of two reactants can lead to one (or more) distinct products (each with even more conformers).
- **Clusters**: molecular clusters formed from very many different combinations of these products (with many possible relative orientations).



Even if we only had 1000 types of clustering molecules (huge underestimate), we would have 10^{30} different 10-molecule clusters…

Tool for qualitative insight (and dataset generation): GECKO-AP

GECKO-A



Peroxy Radicals

GECKO-AP

- Lauri Franzon's modified version of the GECKO reaction mechanism generator: **GECKO-AP.** Includes all known gas phase routes to **accretion products**
  - While the AP module is qualitative, the base GECKO-A chemistry leading to the peroxy radical reactants is semi-quantitative, albeit lacks autoxidation steps.
- Preprint just out: egusphere.copernicus.org/preprints/2024/egusphere-2024-920/

Now we structural formulae for compounds that may form and grow clusters – what next?

- 3D structures for the molecules
  - configurational sampling for single molecules
- Saturation vapour pressures as a first estimate for the relevance of each compounds
  - bulk, flat surface, one component
- Clustering of molecules
  - one or several types of molecules in the same cluster
  - configurational sampling for clusters of molecules

Conformer sampling for single molecules

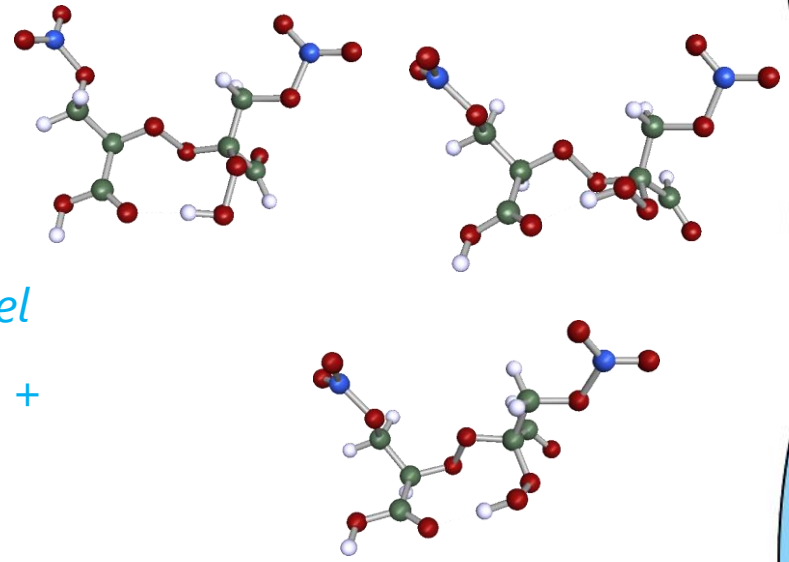
INPUT:
SMILES Strings

C(=O)(O)C(C(ON(=O)(=O)))OOC(OO)(C(ON(=O)(=O)))C(=O)

COSMO*conf*

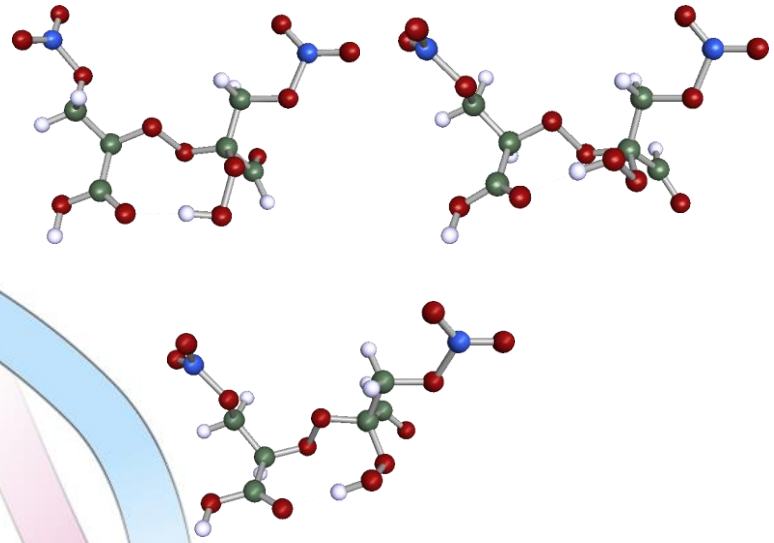
1. Conformer generation (11k)
2. Sorting & data clustering
3. Iterative DFT calculation using *Conductor-like Screening Model for Real Solvents* (COSMO-RS) + Sorting/Clustering

Output:
Conformers



Saturation vapour pressures: COSMO*therm* program

INPUT:
Conformers



COSMO*therm*

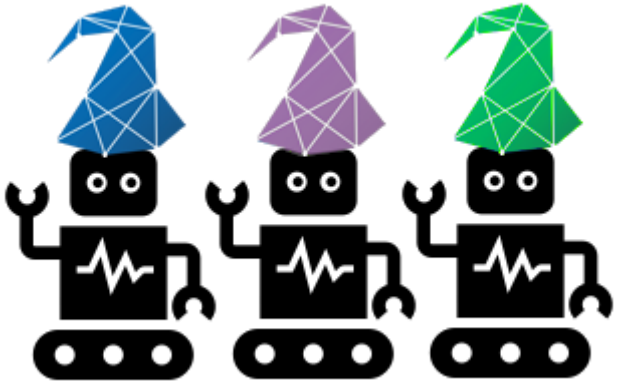
Output:
Saturation Vapor Pressure [mbar]
Chemical potential [kcal/mol]
Free energy [kcal/mol]
Heat of vaporisation [kcal/mol]

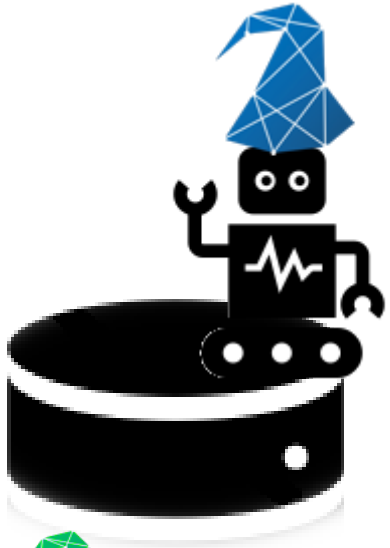
existing saturation vapour pressure prediction methods
(group contribution-based)
unreliable for polyfunctional molecules

Calculation of many molecules: so far 32k molecules, 7M conformers

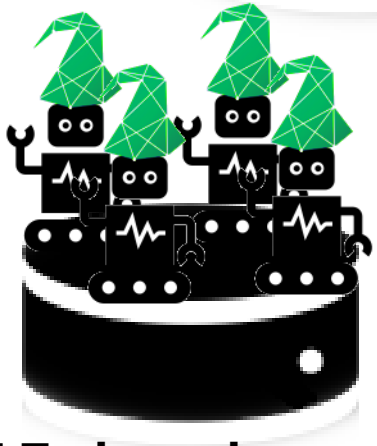
Merlin

Merlin workers for different tasks:

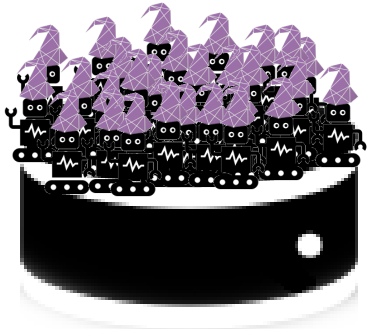




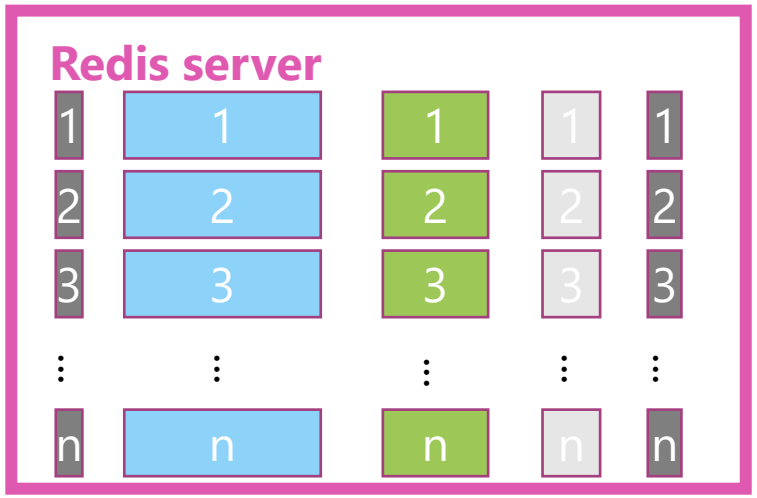
Worker sits on a node and pulls tasks:
2 COSMOconf workers per node (64 cores)



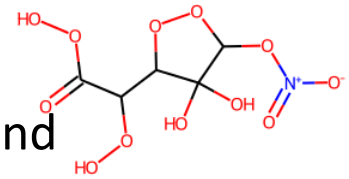
4 Turbomole workers per node (32 cores)



32 Pre-, Post-process & CosmoTherm workers per node (4 cores)



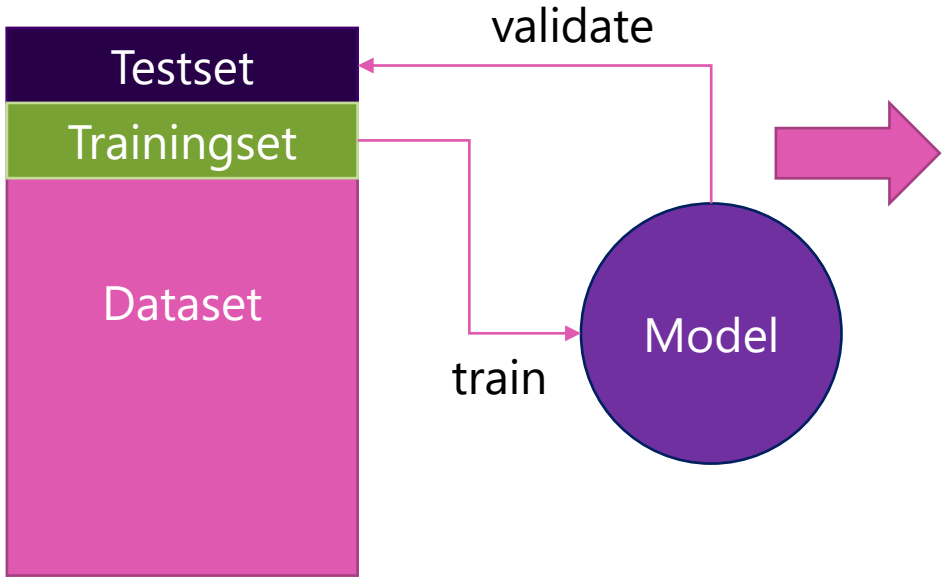
Machine learning saturation vapour pressures

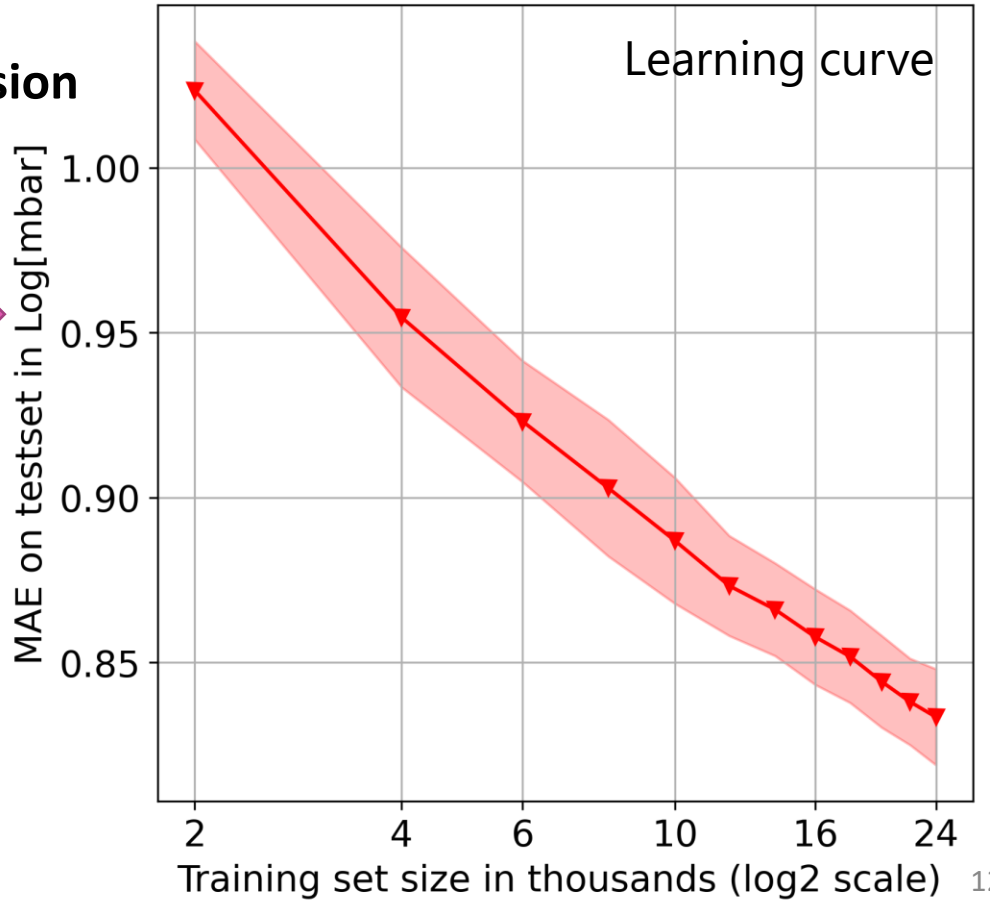


- Computer does not understand

→Molecular descriptor: Topological fingerprint - chosen after extensive testing

- Machine Learning with **Gaussian Process Regression**





Configurational sampling for clusters of molecules: Jammy Key for Configurational Sampling = workflow for acid-base clusters

1. Create large number of guess structures
2. Optimize guess structures at progressively higher levels of theory
3. Filter out unreasonable structures after each optimization

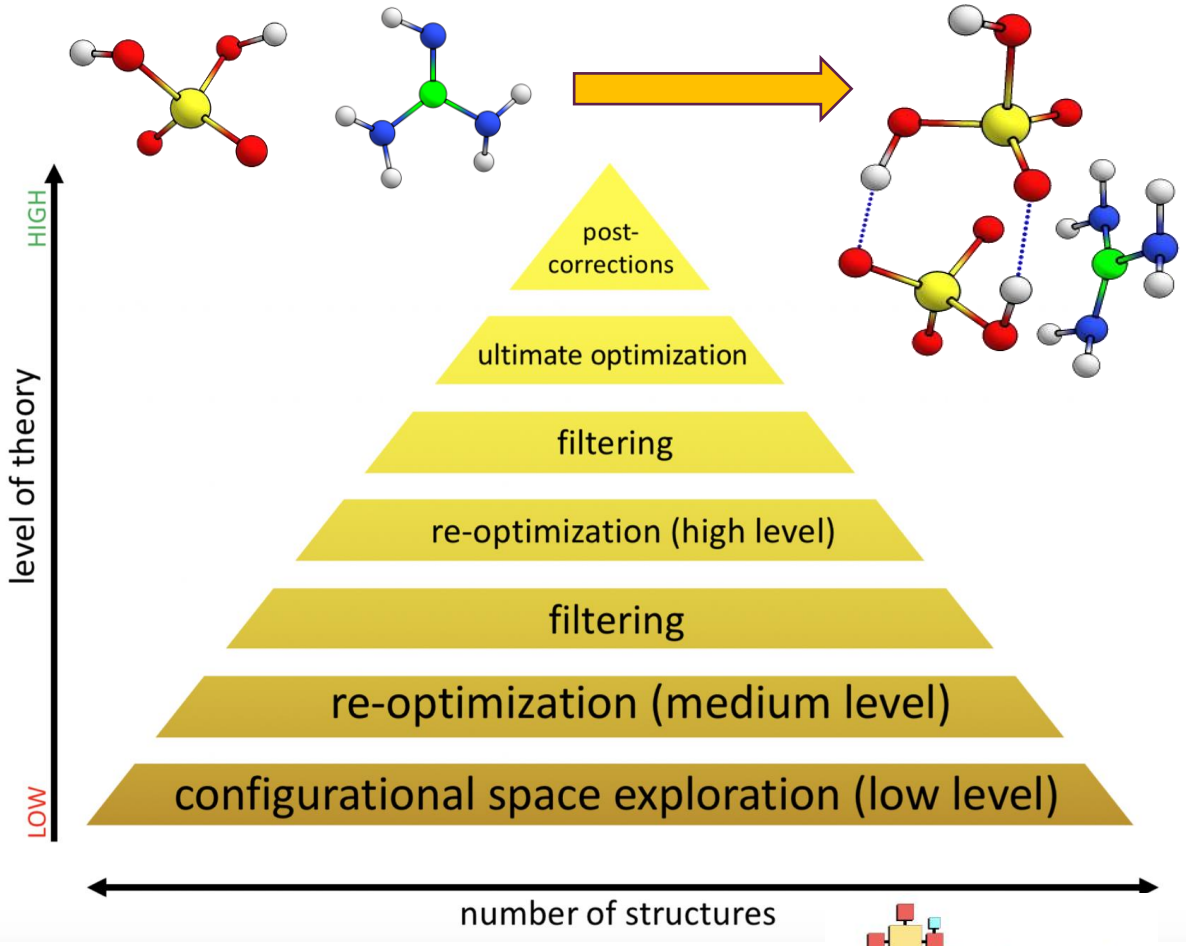
Coupled Cluster

DFT ωB97X-D 6-31G** (max ~100)

XTB semi-empirical (max ~10 000)

[Classical force fields for rigid molecules (max millions)]

ABCluster



Kubečka, J., Besel, V., Kurtén, T. *et al.* (2019) J. Phys. Chem. 123, 28, 6022-6033.

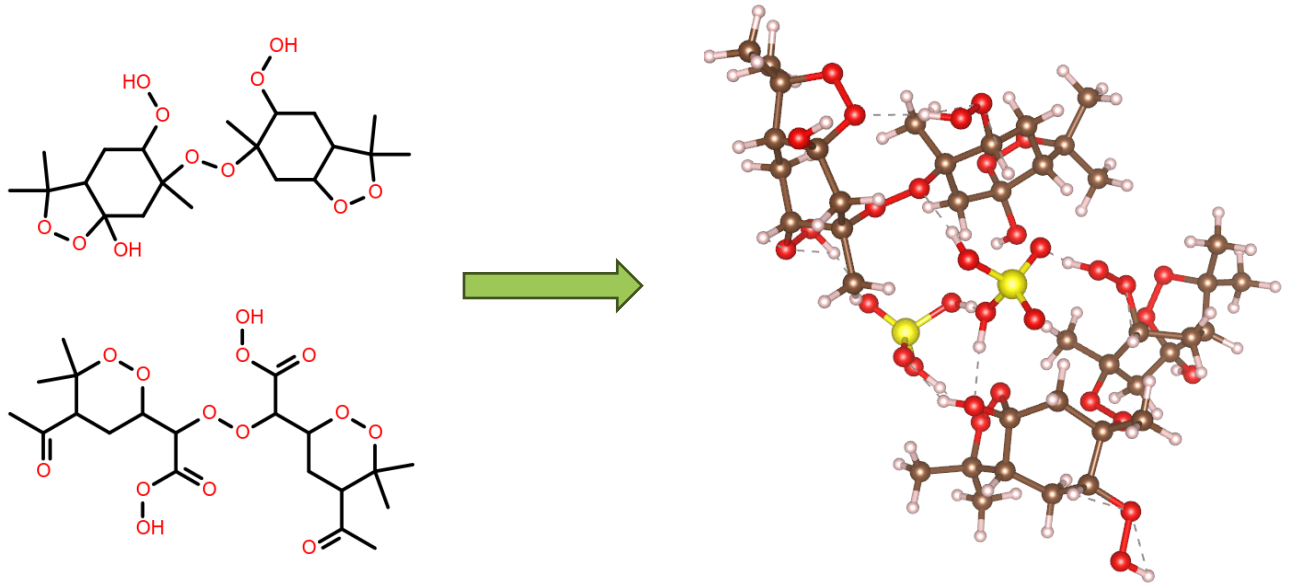
The problem gets much more complex with oxygenated organics

Example: Generate and optimize clusters with 2 oxygenated organics and 0-4 sulphuric acid molecules

- Cheap sampling with ABCluster, start with XTB level simultaneously sample both inter- and intramolecular degrees of freedom

➜ produces 100000 configurations, pick 3000 most reasonable.

- Optimizing one configuration with very cheap DFT B97/def2-SV(P) level takes on average 10 hours

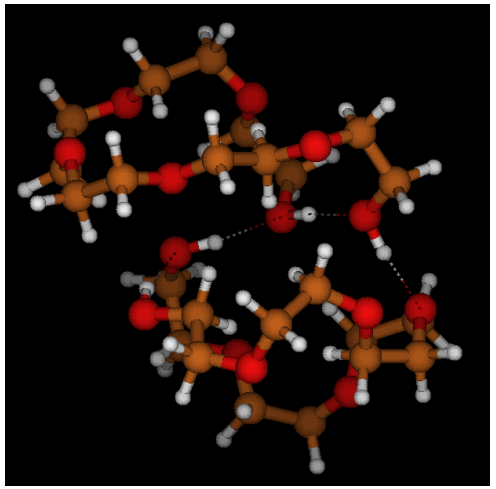


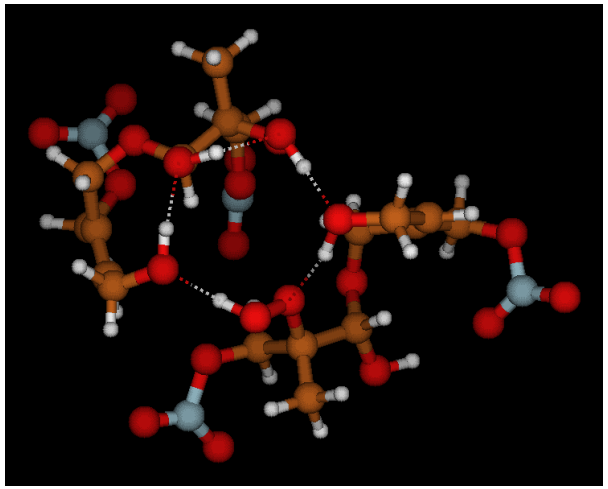
Metadynamics + enforcing H-bonding improves sampling

Assume that the lowest binding energy will be found by maximizing the number of H-bonds between the molecules

Define all possible O··H pairs and their combinations.

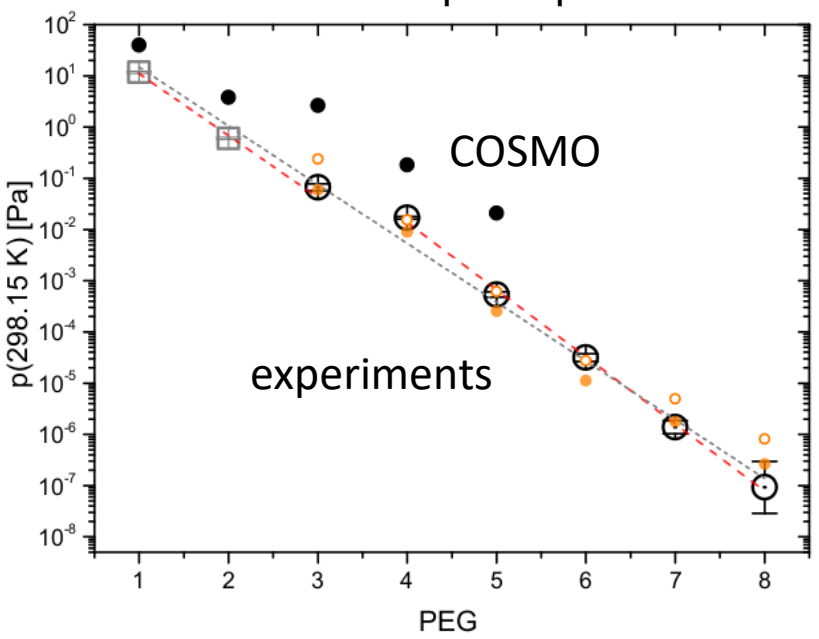
- Force ABCluster (XTB level) to optimize structures where some or all of these combinations are realized.
- Pick the lowest energy structures and run metadynamics on CREST (XTB level) to find further conformers.





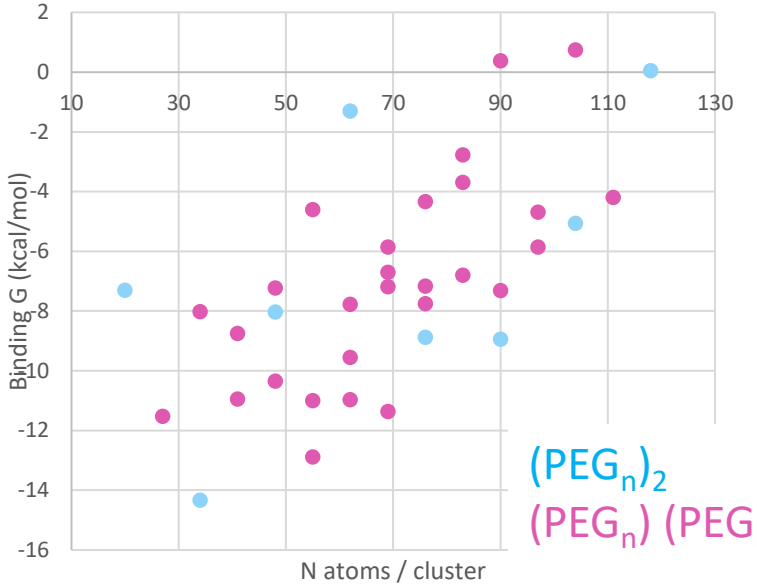
Test calculations for ‘small and simple’ PEGs

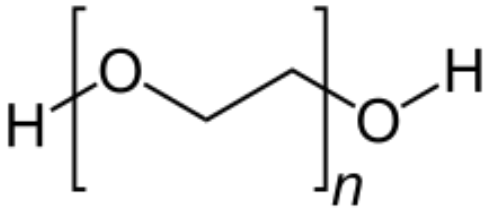
Saturation vapour pressure

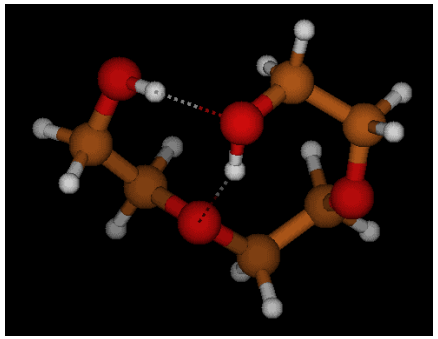


Krieger et al. Atmos. Meas. Tech., 11, 49–63, 2018

Two molecule cluster binding free energy







- saturation vapour pressures drop with size as expected
  - COSMOtherm captures the trend
- two molecule clustering becomes **less** favourable with size
- explained by intramolecular hydrogen bonds in the free monomers (these are stronger in the larger monomers)

Conclusions

- Organic component of atmospheric particle formation probably involves accretion products – formation reactions require very high level of theory to describe
- Qualitative insights wrapped into reaction mechanism generators providing large datasets of realistic candidate compounds
- Developed heavy-duty workflows to predict conformers and saturation vapor pressure
- Modelling organic clustering: work in progress, but strongly indicates that clustering cannot be predicted solely from bulk behaviour
- We still can not answer the burning question: is pure organic particle formation even possible in the Earth's atmosphere, or do we always need inorganic acids?

Thank you!

VILMA

CSC

CENTRES OF EXCELLENCE IN RESEARCH

UNIVERSITY OF HELSINKI

UNIVERSITY OF EASTERN FINLAND

Tampere University

Aalto University