Supplement of Atmos. Chem. Phys., 26, 7631–7645, 2026
https://doi.org/10.5194/acp-26-7631-2026-supplement

Supplement of

Machine learning interatomic potentials with accurate long-range interactions for molecular dynamics collision simulations of atmospherically-relevant molecules

Ivo Neefjes et al.

Correspondence to: Ivo Neefjes (ivo.neefjes@chem.au.dk)

Contents

S1 Description of the OPLS-AA force field

The OPLS intramolecular potential consists of harmonic bond and angle terms, as well as a Fourier series for dihedral angles:

$$U_{\text{intra}}^{\text{OPLS}} = \sum_{i=1}^{N_{\text{bonds}}} \frac{k_i^{\text{b}}}{2} \left(r_i - r_i^0\right)^2 + \sum_{j=1}^{N_{\text{angles}}} \frac{k_j^{\theta}}{2} \left(\theta_j - \theta_j^0\right)^2 + \sum_{k=1}^{N_{\text{dihedrals}}} \sum_{n=1}^{4} \frac{V_n}{2} \left[1 + \cos\left(n\phi^k - \phi_n^k\right)\right], \tag{S1}$$

where k_i^{b}, r_i, and r_i^0 are the force constant, instantaneous length, and equilibrium length of bond i; k_j^{θ}, θ_j, and θ_j^0 denote the force constant, instantaneous angle, and equilibrium angle for angle j; and V_n, ϕ_n^k, and ϕ^k represent the Fourier coefficients, phase angles, and instantaneous value of dihedral k.

Intermolecular interactions, along with intramolecular interactions between atoms separated by more than three covalent bonds, are described by Lennard-Jones and Coulomb terms:

$$U_{\text{inter}} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \tag{S2}$$

where r_{ij} is the distance between atoms i and j, ϵ_{ij} and σ_{ij} are the Lennard-Jones energy and distance parameters, q is the partial charge, and ϵ_0 the vacuum permittivity.

We utilized the OPLS parameters from Loukonen et al. (2010). Notably, while the standard OPLS force field scales 1–4 interactions (atoms separated by three bonds) by 0.5, Loukonen et al. (2010) set this scaling factor to zero during parameterization. For consistency, we also neglected these interactions in our simulations.

S2 Hyperparameter tuning plots

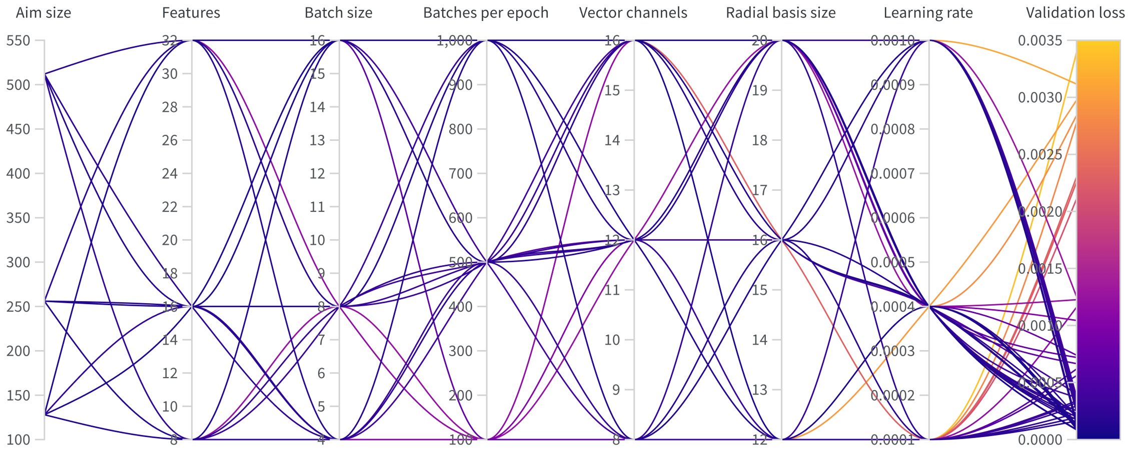

Figure S1: Parallel coordinates plot visualizing the hyperparameter optimization for the AIMNet2 model. The model was trained on 2,000 GFN1-xTB structures from the sulfuric acid–sulfuric acid collision system. Each vertical axis represents a specific hyperparameter, while the final axis displays the resulting validation loss. Each connecting line corresponds to a single training experiment, illustrating how different parameter combinations correlate with model performance.

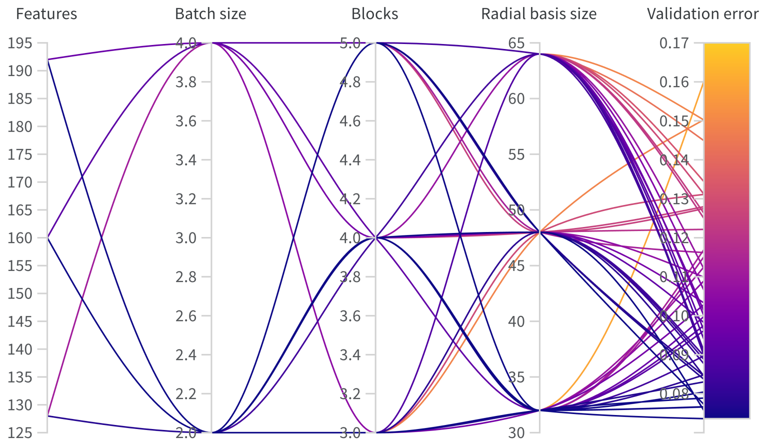

Figure S2: Parallel coordinates plot visualizing the hyperparameter optimization for the PaiNN model. The model was trained on 2,000 GFN1-xTB structures from the sulfuric acid–sulfuric acid collision system. Each vertical axis represents a specific hyperparameter, while the final axis displays the resulting validation loss. Each connecting line corresponds to a single training experiment, illustrating how different parameter combinations correlate with model performance.

S3 List of hyperparameters

This section details the hyperparameters used for training the AIMNet2 and PaiNN models. We also provide example model definition and training configuration files for the AIMNet2 model, specifically for the sulfuric acid–sulfuric acid system trained on GFN1-xTB data. PaiNN calculations were conducted using the JK framework (Kubečka et al., 2024). All hyperparameters not explicitly listed here were maintained at their default values.

AIMNet2

Parameter	Value
AIM size	128
Vector channels (ncomb_v)	16
Features	16
Cutoff radius (rc_s)	5.0 Å
Radial basis size (nshifts_s)	20
Coulombic cutoff	4.6 Å
DFTD3 s6	1.0
DFTD3 s8	2.4
DFTD3 a1	0.63
DFTD3 a2	5.0
Batch size	16
Batches per epoch	4,000
Validation fraction	10%
Charges weight	0.01
Energy weight	0.1
Forces weight	1.0
Learning rate	4×10^{-4}
Weight decay	1×10^{-8}
Epochs	1,000

PaiNN

Parameter	Value
Features	256
Cutoff radius	10.0 Å
Radial basis	32
Batch size	2
Energy weight	0.01
Forces weight	1.0
Validation fraction	10%
Learning rate	1×10^{-4}
Epochs	1,000
Blocks	4

File S1: example of AIMNet2 model definition file

```
class: aimnet.models.AIMNet2
kwargs:
  nfeature: 16
  d2features: true
  ncomb_v: 16
  hidden:
  - - 512
    - 380
  - - 512
    - 380
  - - 512
    - 380
    - 380
  aim_size: 128
  aev:
    rc_s: 5.0
    nshifts_s: 20
  outputs:
    energy_mlp:
      class: aimnet.modules.Output
      kwargs:
        n_in: 128
        n_out: 1
        key_in: aim
        key_out: energy
        mlp:
          activation_fn: torch.nn.GELU
          last_linear: true
          hidden:
          - 128
          - 128
    atomic_shift:
      class: aimnet.modules.AtomicShift
      kwargs:
        key_in: energy
        key_out: energy
    atomic_sum:
      class: aimnet.modules.AtomicSum
      kwargs:
        key_in: energy
        key_out: energy
    lrcoulomb:
      class: aimnet.modules.LRCoulomb
      kwargs:
        rc: 4.6
        key_in: charges
        key_out: energy
    dftd3:
      class: aimnet.modules.DFTD3
      kwargs:
        s6: 1.0
        s8: 2.4
        a1: 0.63
        a2: 5.0
        key_out: energy
```

File S2: example of AIMNet2 training configuration file


```yaml
checkpoint:
  dirname: checkpoints
  filename_prefix: 1sa_1sa_XTB1
  kwargs:
    n_saved: 1
    require_empty: false
data:
  datasets:
    train:
      class: aimnet.data.SizeGroupedDataset
      kwargs: {}
    val:
      class: aimnet.data.SizeGroupedDataset
      kwargs: {}
  ddp_load_full_dataset: false
  loaders:
    train:
      num_workers: 0
      pin_memory: true
    val:
      num_workers: 0
      pin_memory: true
  sae:
    energy:
      file: 1sa_1sa_20K_XTB_training_sae.yaml
      mode: linreg
  samplers:
    train:
      class: aimnet.data.SizeGroupedSampler
      kwargs:
        batch_mode: molecules
        batch_size: 16
        batches_per_epoch: 4000
        shuffle: true
    val:
      class: aimnet.data.SizeGroupedSampler
      kwargs:
        batch_mode: molecules
        batch_size: 16
        batches_per_epoch: -1
        shuffle: false
  separate_val: true
  train: 1sa_1sa_20K_XTB_training.h5
  val:
  val_fraction: 0.1
  x:
  - coord
  - numbers
  - charge
  y:
  - energy
  - forces
  - charges
loss:
  class: aimnet.train.loss.MTLoss
  kwargs:
    components:
      charges:
        fn: aimnet.train.loss.peratom_loss_fn
        kwargs:
          key_pred: charges
          key_true: charges
        weight: 0.01
      energy:
        fn: aimnet.train.loss.energy_loss_fn
        weight: 0.1
      forces:
        fn: aimnet.train.loss.peratom_loss_fn
        kwargs:
          key_pred: forces
          key_true: forces
        weight: 1.0
```

```yaml
metrics:
  class: aimnet.train.metrics.RegMultiMetric
  kwargs:
    cfg:
      charges:
        abbr: q
        peratom: true
      dipole:
        abbr: D
        mult: 3
        scale: 1.0
      energy:
        abbr: E
        scale: 23.06
      forces:
        abbr: F
        mult: 3
        peratom: true
        scale: 23.06
      quadrupole:
        abbr: Q
        mult: 6
        scale: 1.0
      volumes:
        abbr: V
        peratom: true
optimizer:
  class: torch.optim.RAdam
  force_no_train: []
  force_train: []
  kwargs:
    lr: 0.0004
    weight_decay: 1.0e-08
  param_groups:
    shifts:
      re: .*.atomic_shift.shifts.weight$
      weight_decay: 0.0
run_name: grid_search
scheduler:
  class: ignite.handlers.param_scheduler.ReduceLROnPlateauScheduler
  kwargs:
    factor: 0.75
    metric_name: loss
    patience: 12
  terminate_on_low_lr: 1.0e-06
trainer:
  epochs: 1000
  evaluator: aimnet.train.utils.default_evaluator
  trainer: aimnet.train.utils.default_trainer
wandb:
  init:
    mode: offline
    entity: ivo-neefjes
    project: aimnet2_project
  watch_model:
    log: all
    log_freq: 1000
    log_graph: true
```

S4 Removed umbrella sampling simulations

Table S1: List of removed umbrella sampling simulations. Values represent the center-of-mass (COM) distance in Å. The number in parentheses represents the number of failed runs out of the 10 simulations performed per window (e.g., 2.0 (3) means 3 simulations were removed at 2.0 Å).

System	**Data**	**Model**	**Removed COM distances (number)**
H_2SO_4–H_2SO_4	GFN1-xTB	AIMNet2	2.0 (3), 2.2 (2), 2.6 (2), 2.8, 3.8, 4.8, 5.0 (2), 5.2, 5.4
		PaiNN	None
		Δ-PaiNN	None
	ωB97X-3c	AIMNet2	None
		PaiNN	None
		Δ-PaiNN	None
H_2SO_4–$NH(CH_3)_2$	GFN1-xTB	AIMNet2	2.0 (2), 2.2 (2), 3.6
		PaiNN	2.0, 2.2, 2.4 (2)
		Δ-PaiNN	*Not performed*
	ωB97X-3c	AIMNet2	2.0
		PaiNN	2.0 (2), 2.2 (2)
		Δ-PaiNN	None
H_2SO_4–HSO_4^-	GFN1-xTB	AIMNet2	2.2, 2.4, 2.8, 3.2 (2), 13.4, 16.2, 17.4
		PaiNN	2.0
		Δ-PaiNN	*Not performed*
	ωB97X-3c	AIMNet2	2.2, 2.4, 2.8, 3.2 (2), 13.4, 16.2, 17.4
		PaiNN	None
		Δ-PaiNN	2.6, 7.4, 9.4, 10.2 (2), 11.4, 11.8, 12.0 (2), 12.2 (2), 12.4, 12.8, 13.0, 18.4, 19.2

S5 Repulsive region of the H_2SO_4–H_2SO_4 potential of mean force

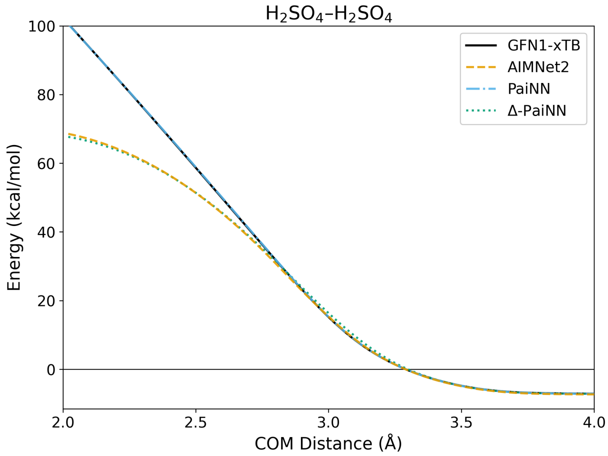

Figure S3: The repulsive region of the potential of mean force along the center-of-mass (COM) distance obtained from umbrella sampling for the H_2SO_4–H_2SO_4 collision system. Results are shown for GFN1-xTB and the AIMNet2, PaiNN, and Δ-PaiNN machine learning models trained on GFN1-xTB.

S6 Collision probability heat maps

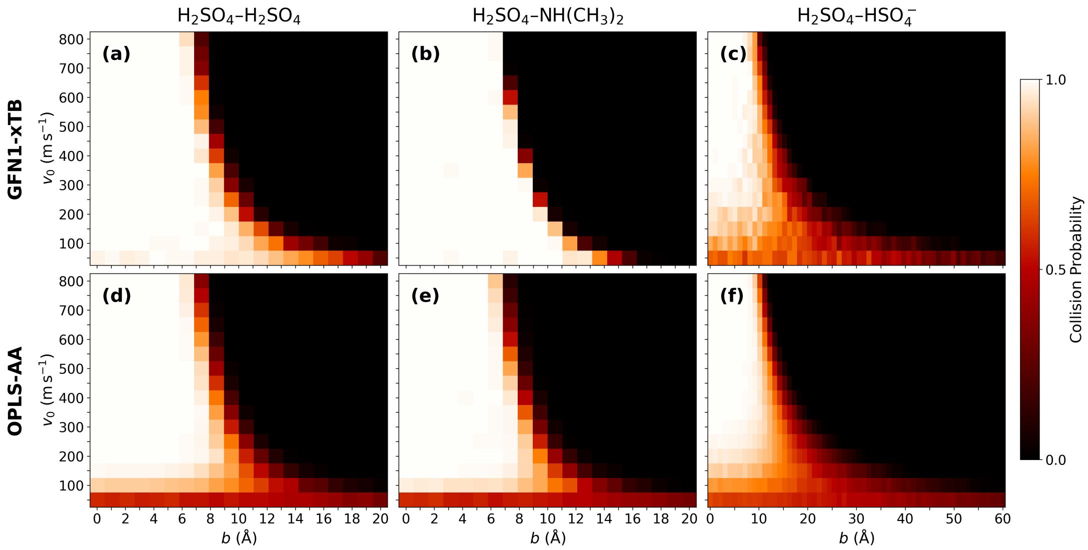

Figure S4: Collision probabilities derived from molecular dynamics simulations for the H_2SO_4–H_2SO_4, H_2SO_4–$NH(CH_3)_2$, and H_2SO_4–HSO_4^- system at 300 K. The heat maps show the probability distribution across impact parameter b and initial relative velocity v_0 for the reference GFN1-xTB method (Grimme et al., 2017) and the classical OPLS-AA force field method (Jorgensen et al., 1996).

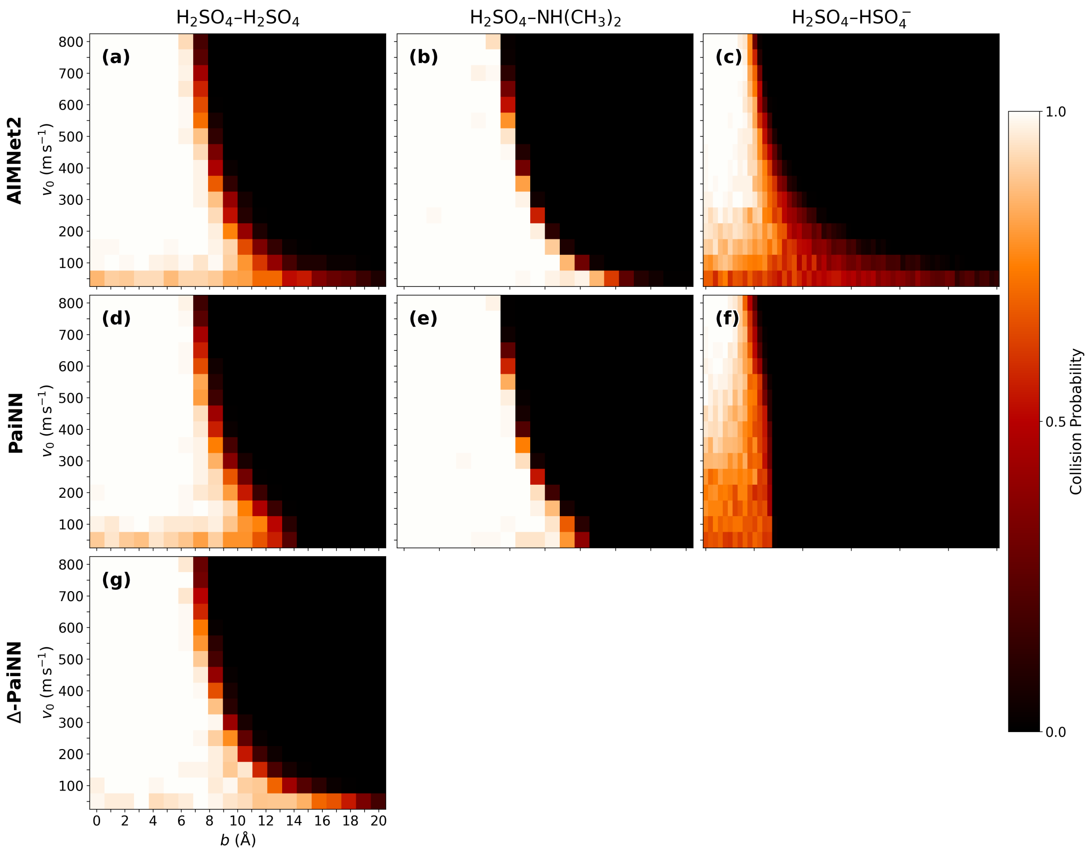

Figure S5: Collision probabilities derived from molecular dynamics simulations for the H_2SO_4–H_2SO_4, H_2SO_4–$NH(CH_3)_2$, and H_2SO_4–HSO_4^- system at 300 K. The heat maps show the probability distribution across impact parameter b and initial relative velocity v_0 for the AIMNet2 (Anstine et al., 2025), PaiNN (Schütt et al., 2021), and Δ-PaiNN machine learning models trained on GFN1-xTB reference data (Grimme et al., 2017).

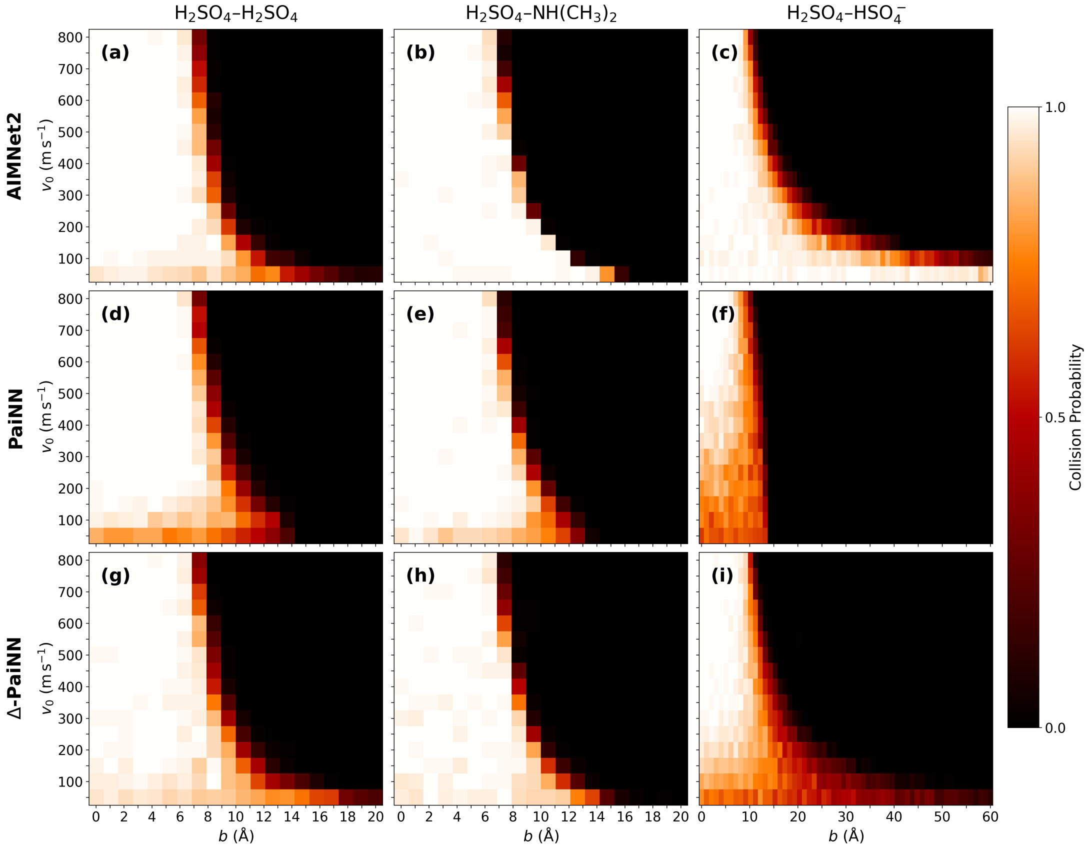

Figure S6: Collision probabilities derived from molecular dynamics simulations for the H_2SO_4–H_2SO_4, H_2SO_4–$NH(CH_3)_2$, and H_2SO_4–HSO_4^- system at 300 K. The heat maps show the probability distribution across impact parameter b and initial relative velocity v_0 for the AIMNet2 (Anstine et al., 2025), PaiNN (Schütt et al., 2021), and Δ-PaiNN machine learning models trained on ωB97X-3c reference data (Müller et al., 2023).

References

Anstine, D. M., Zubatyuk, R., and Isayev, O.: AIMNet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs, Chem. Sci., 16, 10 228–10 244, https://doi.org/10.1039/D4SC08572H, 2025.

Grimme, S., Bannwarth, C., and Shushkov, P.: A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86), J. Chem. Theory Comput., 13, 1989–2009, https://doi.org/10.1021/acs.jctc.7b00118, 2017.

Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J.: Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc., 118, 11 225–11 236, https://doi.org/10.1021/ja9621760, 1996.

Kubečka, J., Ayoubi, D., Tang, Z., Knattrup, Y., Engsvang, M., Wu, H., and Elm, J.: Accurate modeling of the potential energy surface of atmospheric molecular clusters boosted by neural networks, Environ. Sci.: Adv., 3, 1438–1451, https://doi.org/10.1039/D4VA00255E, 2024.

Loukonen, V., Kurtén, T., Ortega, I. K., Vehkamäki, H., Pádua, A. A. H., Sellegri, K., and Kulmala, M.: Enhancing effect of dimethylamine in sulfuric acid nucleation in the presence of water – a computational study, Atmos. Chem. Phys., 10, 4961–4974, https://doi.org/10.5194/acp-10-4961-2010, 2010.

Müller, M., Hansen, A., and Grimme, S.: ωB97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-ζ basis set, J. Chem. Phys., 158, 014 103, https://doi.org/10.1063/5.0133026, 2023.

Schütt, K. T., Unke, O. T., and Gastegger, M.: Equivariant message passing for the prediction of tensorial properties and molecular spectra, arXiv, abs/2102.03150, https://doi.org/10.48550/arXiv.2102.03150, 2021.