Electronic Supplementary Material (ESI) for Digital Discovery.

Supplementary information for: Machine-Learnt Interatomic Potentials for Amorphous Zeolitic Imidazolate Frameworks

Nicolas Castel,[†,‡] Dune André,[†] Connor Edwards,[¶] Jack D. Evans,[*,¶] and François-Xavier Coudert[*,†]

†*Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France*

‡*École des Ponts, 77420 Marne-la-Vallée, France*

¶*School of Physics, Chemistry and Earth Sciences, The University of Adelaide, Adelaide, SA 5005, Australia*

E-mail: j.evans@adelaide.edu.au; fx.coudert@chimieparistech.psl.eu

Generation of *ab initio* training data

System description

Reference data was extracted from *ab initio* molecular dynamics (AIMD) simulations. We considered four AIMD trajectories of 50 ps each performed in the (N, V, T) ensemble of ZIF-4 liquids at four volumes and temperatures of either 1500 K or 1750 K, as reported in Table S1.

Table S1: Temperature of the 50 ps liquid trajectories generated per strain.

Volume change (%)	2	0	-2	-4
Temperature (K)	1500	1500	1750	1750

While (N, V, T) simulations are an efficient way to generate training data at high temperature, we ambition to develop an MLP that could also be used for (N, P, T) simulations, which therefore requires to an accurate reproduction of the stress. An option is to learn from structures simulated at different volumes, by applying deformation to the initial cell of the crystal. We found that it led to better predictions of mechanical properties than the use of a single trajectory without deformation.

We created four (N, V, T) trajectories at 300 K with different deformations. Successive simulations at different volumes were performed, consisting in an instantaneous and isotropic volume change by 2% from the previous volume, followed by an equilibration time larger than 10 ps. These trajectories were taken from a previous work.[1] The final volume deformations were $\{2\%, 0\%, -2\%, -4\%\}$

Once the 300 K crystal models at various volumes were prepared, they were heated up to obtain high temperature trajectories, that should preferably simulate liquid systems. The 0% trajectory was taken from the work by Gaillac et al.,[2] and the other ones were first simulated for this work. The choice of the high temperature value is a trade-off between the need to gather statistics on relatively rare events, and the necessity to preserve the physical consistency of the model. Previous *ab initio* MD simulations of the liquid ZIF-4 have been performed up to 2000 K while preserving the integrity of the imidazolate linkers.[2] Previous

work on the melting of ZIFs[2,3] also showed that a duration of 50 ps is often sufficient to observe melting. We therefore ran multiple (N, V, T) trajectories first at 1500 K, and then at 1750 K if melting was not observed at 1500 K. Each (N, V, T) trajectory was long enough to include 50 ps of liquid.

Detecting liquids

As discussed in,[4] detecting melting is insightful but not straightforward. If a trajectory is sufficiently long, it is possible to identify melting by spotlighting a loss in coordination as Table S2 or a change in the behavior of the MSD as in Figure S1a.

As a complement, ring statistics can provide clear evidence of the onset of the melting transition. By monitoring the number of rings of different sizes as the function of time, as shown on Figure S1b, one can easily identify a transition from the ring statistics of the crystal towards a less ordered system. This had not been used before, but we think it is a powerful descriptor to detect melting.

The choice of the temperature required to observe melting is not straightforward, as we found that it could depend on the deformation. Table S2 and Figure S1a provide evidence that systems with volume deformations of 0% and 2% ended up melting after a few tens of picoseconds at 1500 K, while -2% and -4% did not.

Table S2: Coordination numbers for nitrogen atoms around the zinc cation for 1500 K MD trajectories of ZIF-4 with various deformations. Computed over the last 10 ps of each trajectory.

Volume change (%)	-4	-2	0	2
Average Zn-N coordination number	3.84	3.76	3.59	3.64

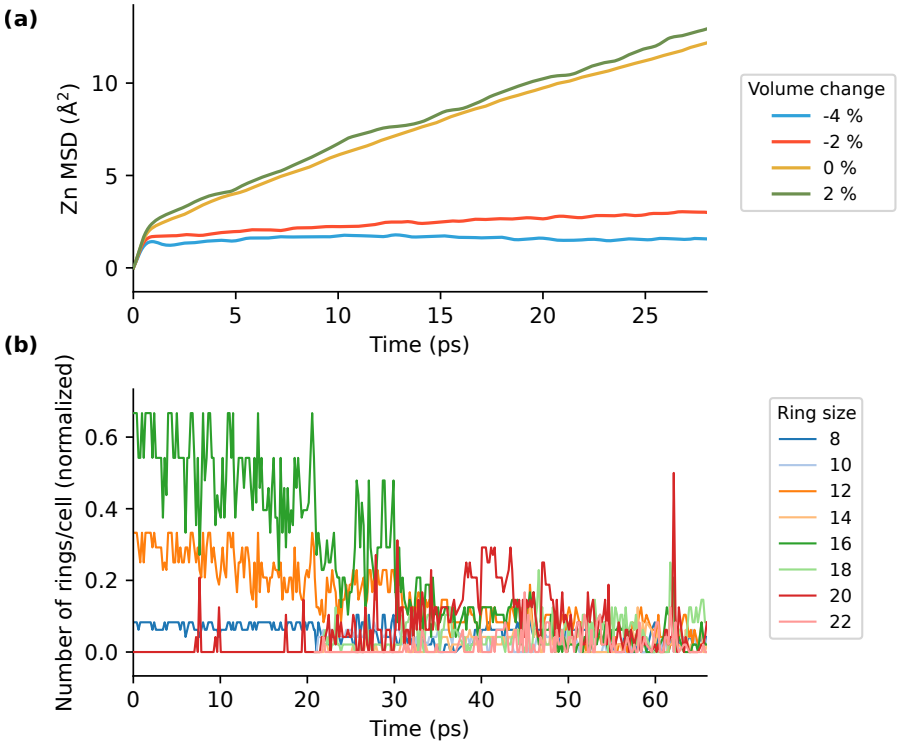

Figure S1: Evidence of melting during 1500 K MD trajectories of ZIF-4. (a) Mean square displacement (MSD) as a function of time for Zn with various deformations. (b) Number of zinc–imidazolate alternate rings of different sizes as the function of time for the trajectory with a deformation of 2%. For clarity, only rings of size comprised between 8 and 22 are shown.

Training NequIP MLPs

Accurately training for stress

Several loss functions were used to simultaneously minimize the errors on the energy (E), forces ($\{\boldsymbol{F}_i\}$) and stress ($\boldsymbol{\sigma}$). The respective contribution of each error in the total loss functions are parameterized by three weights $\lambda_F, \lambda_E, \lambda_\sigma$.

For some MLPs we wanted to bias learning for stress by choosing large λ_σ instead of $\lambda_\sigma = 0$. We tested $\lambda_\sigma \in \{0, 1, 10, 100, 1000\}$ at different stage of development. Here we report the final MLP parameterization, only differing by the value of $\lambda_\sigma \in \{0, 1, 100\}$.

On Table S3, we see that energies and forces are well learned in every case, and any actual learning of the stress occurs only for ($\lambda_\sigma = 100$). We found satisfying final mean absolute errors (MAE) on energy and forces reported in Table S3, compared to previous NequIP works on liquid systems[5] and MOFs.[6]

Table S3: MAEs of the trained models for $\lambda_\sigma \in \{0, 1, 100\}$, evaluated on unseen frames of the liquid trajectories.

λ_σ	0	1	100
Energy MAE (meV/atom)	0.56	0.71	1.7
Forces MAE (meV Å^{-1})	15.1	15.2	16.0
Stress MAE (MPa)	360	300	140

However, Figure S2 clearly evidences that the pressure outputs were singularly different depending on whether the loss function is biased for stress. Strongly biased systems displayed pressure ranges located far from the state of zero pressure, resulting in the extremely low values for the fitted ρ_0 reported in Table S4. On the contrary, the unbiased MLP led to K and ρ_0 values in line with each other and with previous AIMD results. It suggests that training MLPs on energy and forces only is sufficient to reproduce the mechanical properties with reasonable accuracy. We therefore chose ($\lambda_\sigma = 0$) for the production MLP.

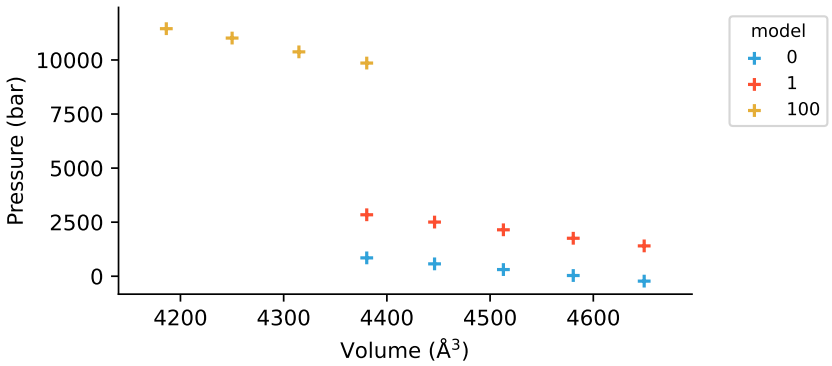

Figure S2: Pressure as the function of volume for the ZIF-4 crystal obtained by the finite strain difference method with multiple λ_σ

Table S4: Bulk modulus K and density at zero pressure ρ_0 for ZIF-4 crystal obtained with the finite strain difference method with multiple λ_σ.

λ_σ	0	1	100
K (GPa)	1.69	1.62	0.80
ρ_0 (g cm^{-3})	1.16	1.06	0.75

MD simulations with NequIP

Table S5: Coordination numbers for nitrogen atoms around the zinc cation for different ZIF-4 phases with the NequIP MLP compared to AIMD.

	Crystal	Liquid	*ab initio* glass
MLP	4.00	3.67	3.94
AIMD	4.00	3.52	3.93

Table S6: Total porous volume (in $\mathrm{cm^3\,kg^{-1}}$) for different ZIF-4 phases with the NequIP MLP compared to AIMD.

	Crystal	Liquid	*ab initio* glass
MLP	53.0	60.2	67.0
AIMD	53.6	52.4	68.1

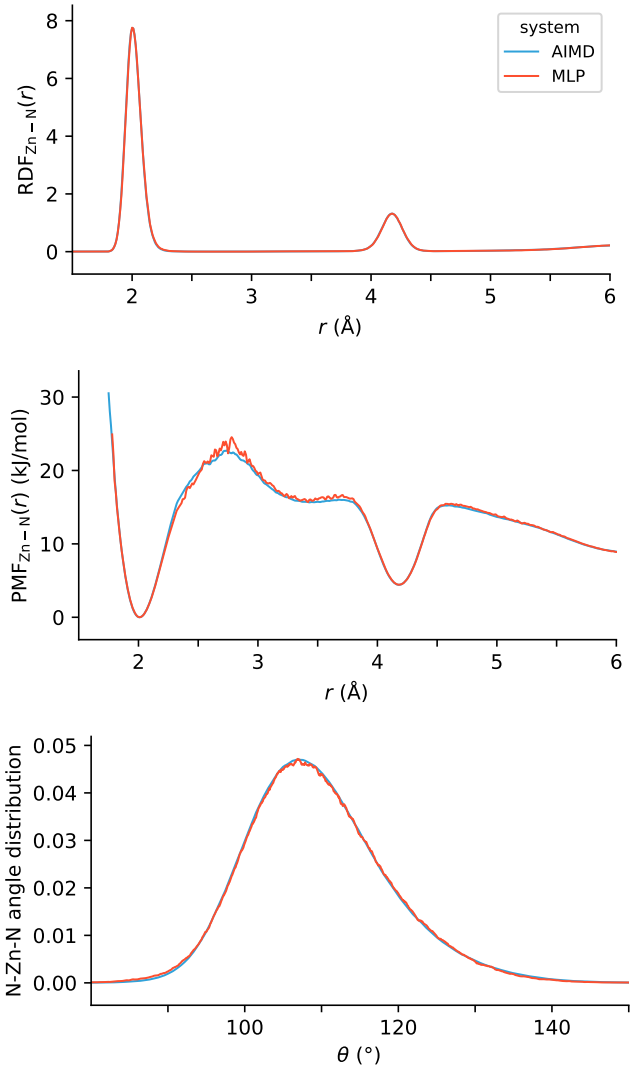

Figure S3: (a) Radial distribution functions (RDF), (b) potentials of mean force (PMF) for the Zn–N atom pairs, and (c) distribution of the N–Zn–N angle for the ZIF-4 *ab initio* glasses with the NequIP MLP compared to AIMD.

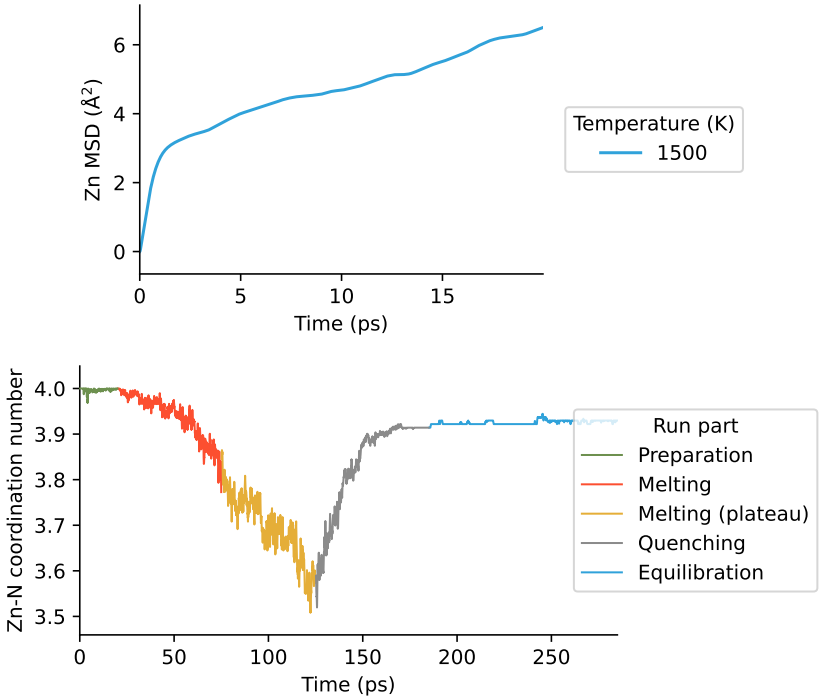

Figure S4: Evidence of melting and glass generation. (a) Mean square displacement (MSD) as a function of time for Zn during the 1500 K plateau. (b) Zn–N coordination number as the function of time with the final value for the glass system distinct from that of the initial crystal.

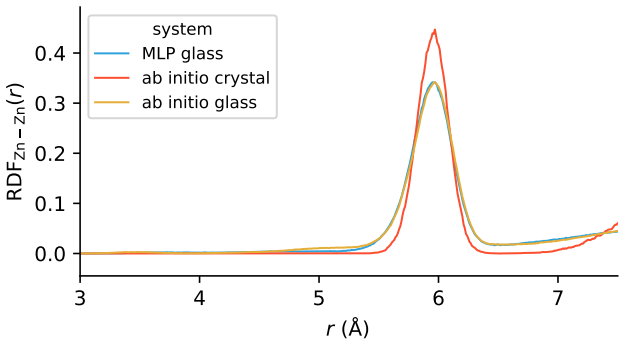

Figure S5: Radial distribution functions (RDF) for the Zn–Zn atom pairs of the MLP glass (blue), *ab initio* glass (orange) and *ab initio* crystal (red).

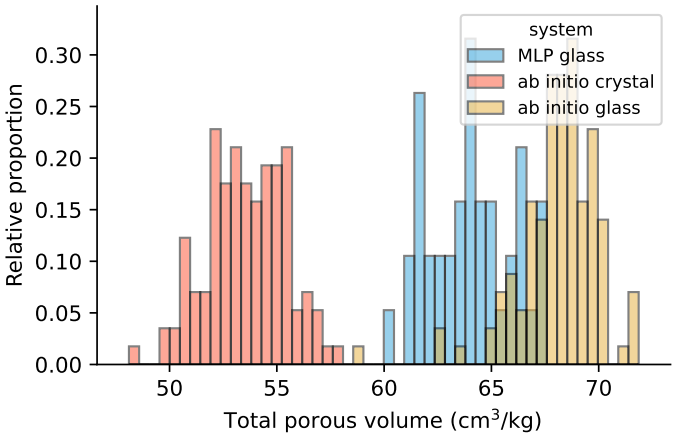

Figure S6: Total porous volume of the MLP glass (blue), *ab initio* glass (orange) and *ab initio* crystal (red).

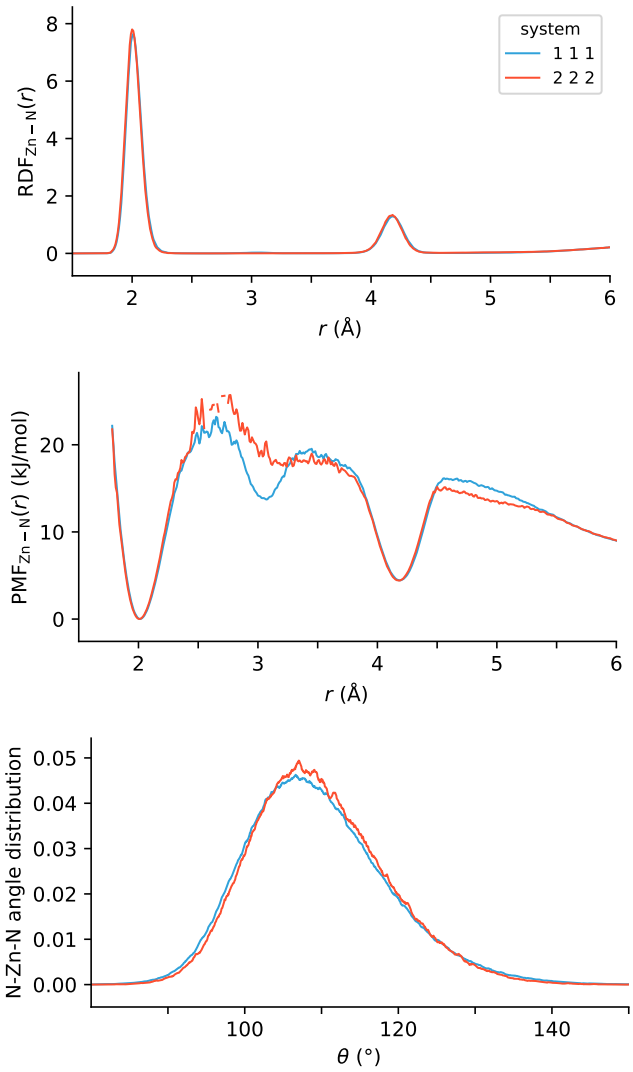

Figure S7: (a) Radial distribution functions (RDF), (b) potentials of mean force (PMF) for the Zn–N atom pairs, and (c) distribution of the N–Zn–N angle for the ZIF-4 the MLP NequIP glass with different system sizes. The single cell properties are averaged over ten glasses.

Mechanical properties with NequIP

System description and methodological details

Finite strain difference method

System description and preparation Eleven systems were prepared following a simplified preparation procedure:

- The same crystallographic model of the crystal used for the *ab initio* simulation of the crystal.[2,7]
- The ten *ab initio* glasses obtained from the work by Gaillac et al.,[7] three of them having been studied with AIMD in a previous work.[1] Properties (K and ρ_0) reported in the article are averaged over the ten glasses.

This preparation procedure consisted in an initial energy minimization, followed by an (N, V, T) run at 300 K for 100 ps with the default temperature damping parameter.

Ten MLP glasses were produced through the same melt-quenching procedure than presented in the main article. The only differences concern the system size (single cell instead of $(2 \times 2 \times 2)$ supercell) and the time spent at the maximal temperature t_{max}. For each glass indexed by $i \in \{1, 2, ..., 10\}$, $t_{max} = 50 + 20 \times i$ (expressed in ps). We checked that every t_{max} was large enough for the crystal to melt. Figure S7 shows that the structural properties are similar, although not identical, between the single cell and $(2 \times 2 \times 2)$ supercell glasses.

Methodological details Successive simulations at different volumes were applied, consisting of a volume deformation by 1.5% and lasting 100 ps from the previous volume followed by an equilibration time of 100 ps. Volume deformation in the (N, V, T) ensemble was a continuous process, during which each dimension of the box changed linearly with time from its initial to final value to achieve a volume reduction by 1.5%.

The value of the pressure for a given volume was taken as the average over the last 100 ps of the simulation.

The well-behaved region of the $P - V$ data for the fit with the second order Birch-Murnaghan EoS was selected to have 5 points at maximum and to include if possible the region of zero pressure. 5 points were found for the crystal, 4.6 ± 0.5 for the *ab initio* glasses and 4.6 ± 0.7 for the MLP glasses.

Strain-fluctuation method

System description For the strain-fluctuation method, the initial systems for the finite strain difference method were taken and were further equilibrated in the (N, P, T) ensemble with a flexible cell until convergence of the elastic constants. This methods was applied for the crystal and the ten *ab initio* glasses.

The equilibration time t_{eq} is of 5 ns for the crystal and 7.5 ns for the *ab initio* glasses.

Methodological details The elastic constants, the volume fluctuations and the volumes were computed over the last 4.5 ns for the crystal and 4.35 ± 0.32 ps for the *ab initio* glasses.

Supplementary figures and tables

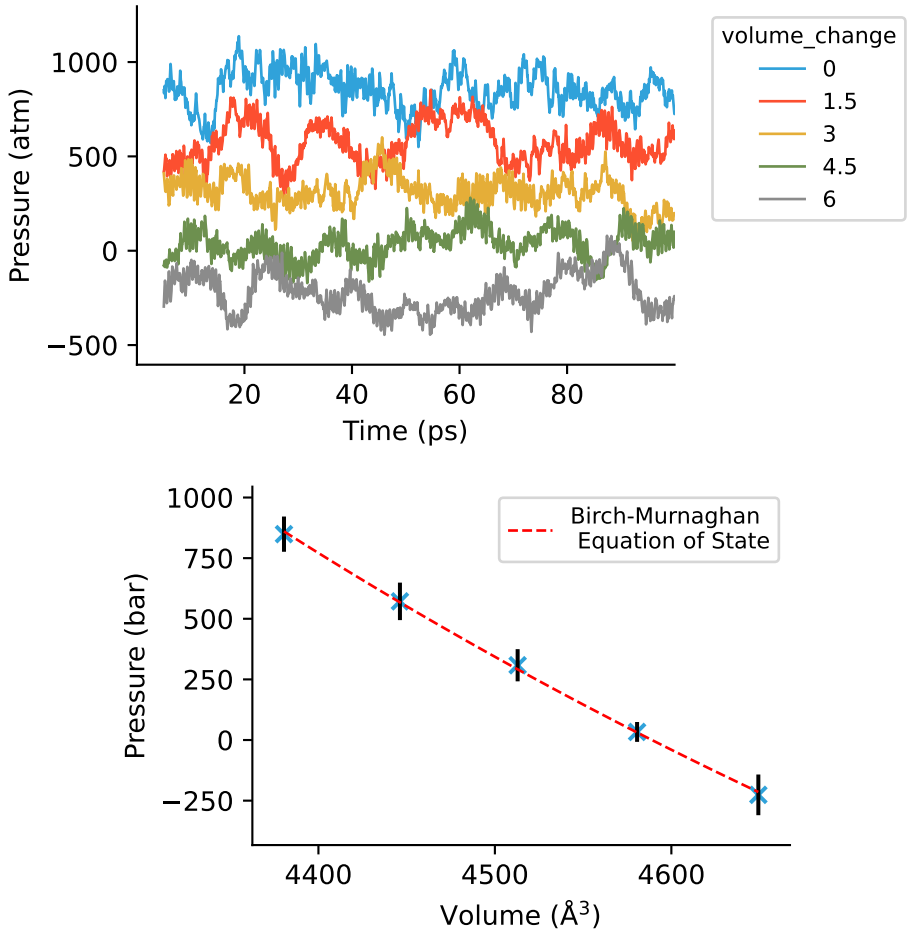

Figure S8: Finite strain difference method for the crystal with the NequIP MLP. (a) Pressure as a function of time for different deformations, shown with a moving average over 5 ps. (b) Pressure as a function of volume fitted with the second order Birch-Murnaghan EoS (red). Each volume corresponds to a (N, V, T) simulation with fixed deformation. Pressure is computed as the average over the last 100 ps of simulations. Error bars are the standard deviations of the pressure averaged with a rolling window of 10 ps.

Table S7: Bulk modulus K and density at zero pressure ρ_0 for ZIF-4 crystal obtained with the finite strain difference method with NequIP MLP and different supercells. Calculations with the $(2 \times 2 \times 2)$ supercell were performed with the shorter deformation and equilibration times of $50\,\mathrm{ps}$, which were sufficient to witness pressure equilibration.

	Single cell	$(2 \times 2 \times 2)$ supercell
K (GPa)	1.69	1.67
ρ_0 ($\mathrm{g\,cm^{-3}}$)	1.16	1.16

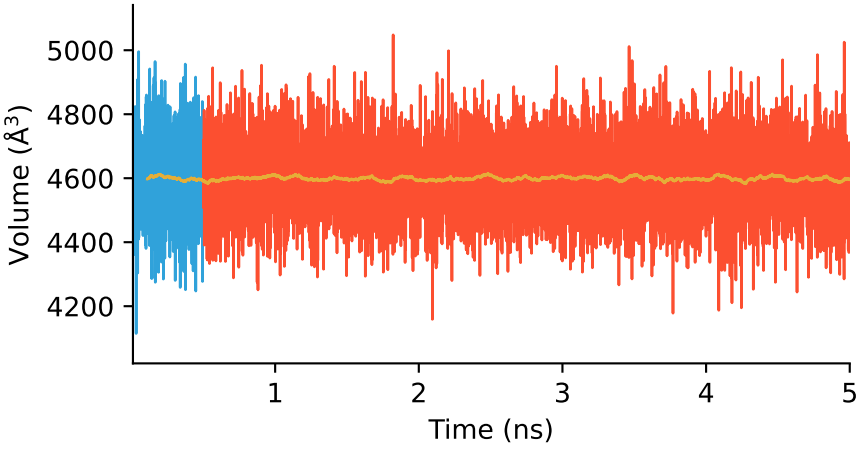

Figure S9: Volume as a function of time during a long 5 ns equilibration in the (N, P, T) ensemble with a flexible cell for the crystal with the NequIP MLP. Elastic constants were computed on the red part of the figure. The orange line shows a moving average over 100 ps.

Table S8: Bulk modulus K and density at zero pressure ρ_0 for ZIF-4 crystal and *ab initio* glasses obtained with two strain-fluctuation methods with the NequIP MLP.

	Crystal	*ab initio* glass
K - volume fluctuations (GPa)	1.59	2.44 ± 1.02
K - elastic constants (GPa)	1.58	2.72 ± 0.97
ρ_0 ($\mathrm{g\,cm^{-3}}$)	1.15	1.30 ± 0.09

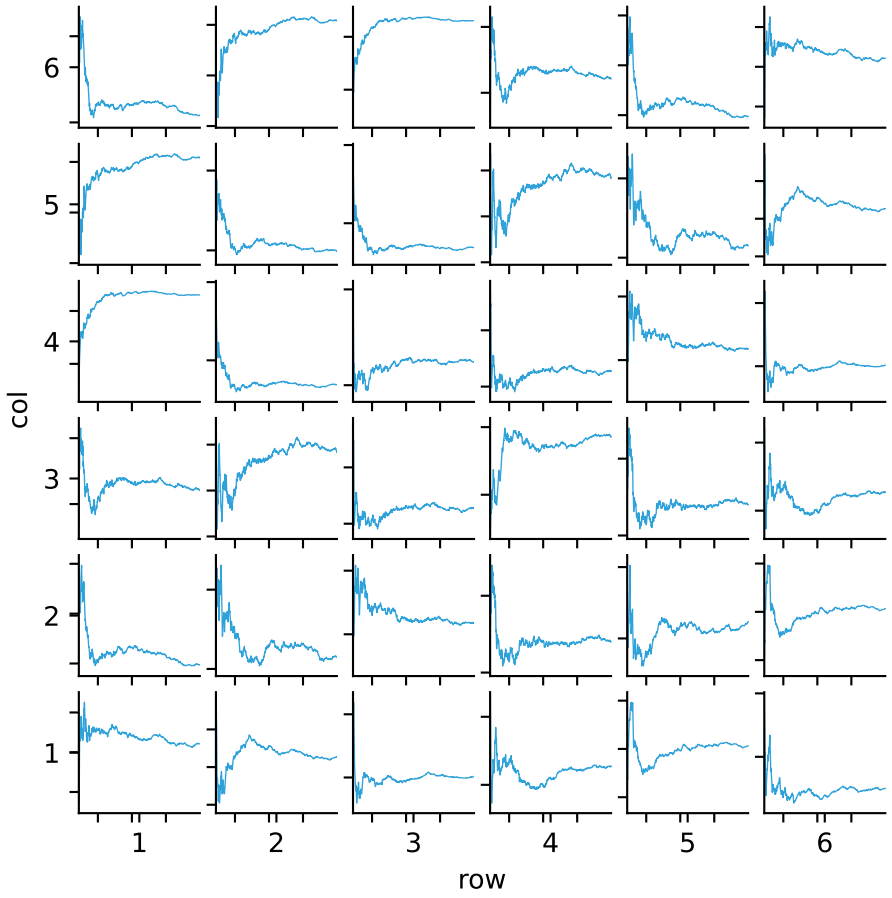

Figure S10: (a) Convergence of the elastic constants $C_{i,j}$ as function of time for the crystal with the NequIP MLP during a long 5 ns equilibration in the (N, P, T) ensemble with a flexible cell. (b) At time t, the represented $C_{i,j}(t)$ corresponds to the elastic constant computed from the 2 ns to (t + 2) ns part of the trajectory. The $C_{i,j}$ used for the computation of the mechanical properties thus correspond to the final $t = 3$ ns.

Allegro MLPs

Training

We conducted experiments starting with the same parameters as the production NequIP model, varying only a number of parameters.

The parameters that were preserved include:

- Cutoff radius R_c: 6 Å
- Maximum rotation order l : 2
- Early stopping lower bounds: Learning Rate (LR): 1.0×10^{-6}

The parameters that were varied include:

- The loss coefficient for stress λ_σ, set to either 0 or 100
- The parity (`parity`) options: `o3_full`, `o3_restricted`, and `so3`. (The available parity options differ from NequIP and exclude the `false` option, so we tested all three provided options.)
- The number of layers (`num_layers`): 1, 2, 3
- The learning rate (`learning_rate`): 0.0005, 0.001, 0.002. (This range of values differs from Allegro compared to NequIP as suggested by the developers.)

Regarding the observations from experiments with varied parameters: The training time is primarily influenced by the number of layers (`num_layers`). For the same result with `num_layers=2`, `o3_full` yields slightly better or identical results compared to `o3_restricted`. As `o3_full` requires only 2% more computing time than `o3_restricted`, it is therefore preferable to retain `o3_full`. Increasing `num_layers` from 2 to 3 doesn't lead to a significant improvement; instead, it often results in overfitting as indicated by a further reduction in training loss without a corresponding improvement in validation loss. Both `num_layers=2` and `num_layers=3` outperform `num_layers=1`. Regarding the impact of the learning rate, there are no significant alterations in the final values observed in the validation dataset.

The chosen parameters are therefore:

- Stress coefficient λ_σ: 0 (same as the NequIP MLP)
- Number of layers: 2 (same as the NequIP MLP)
- Parity option: `o3_full`
- Learning rate: 0.001

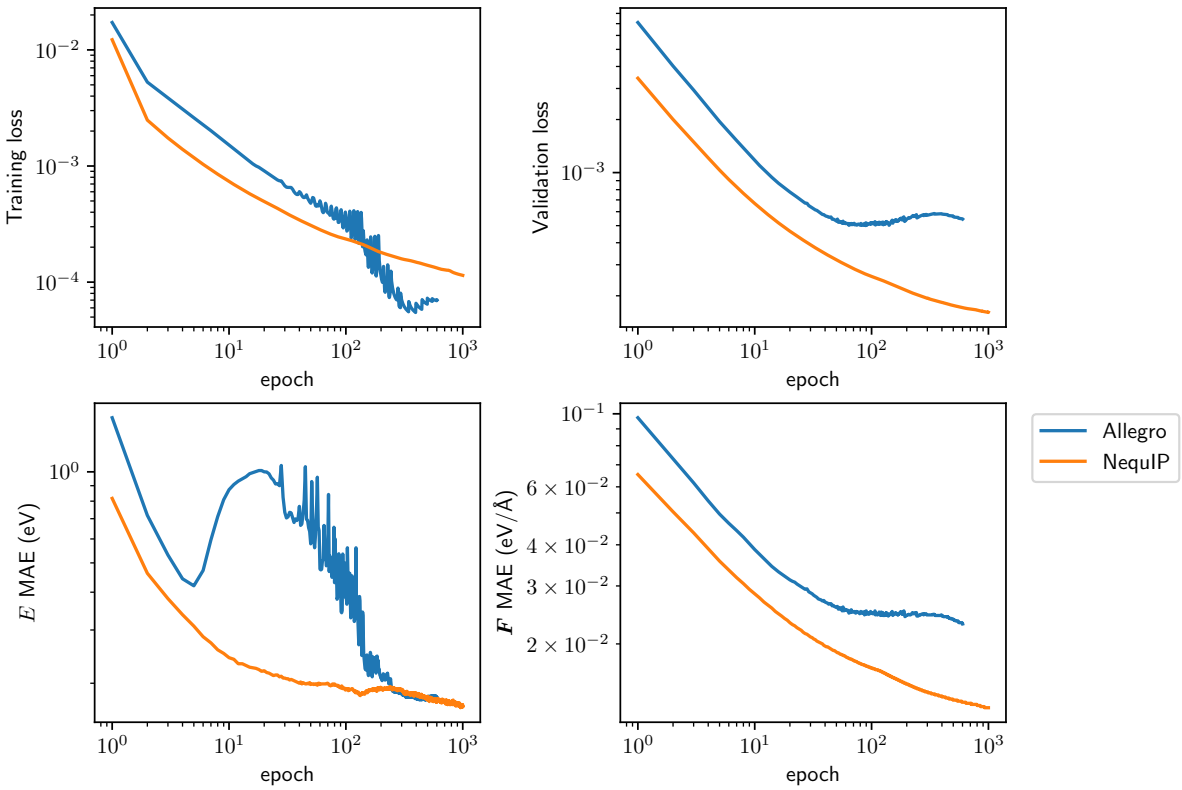

Figure S11: (a) Training and (b) validation loss functions and mean absolute errors (MAE) for (c) energy E and (d) forces $\{\boldsymbol{F}_i\}$ as the function of epoch for the Allegro MLP compared to NequIP.

MD simulations

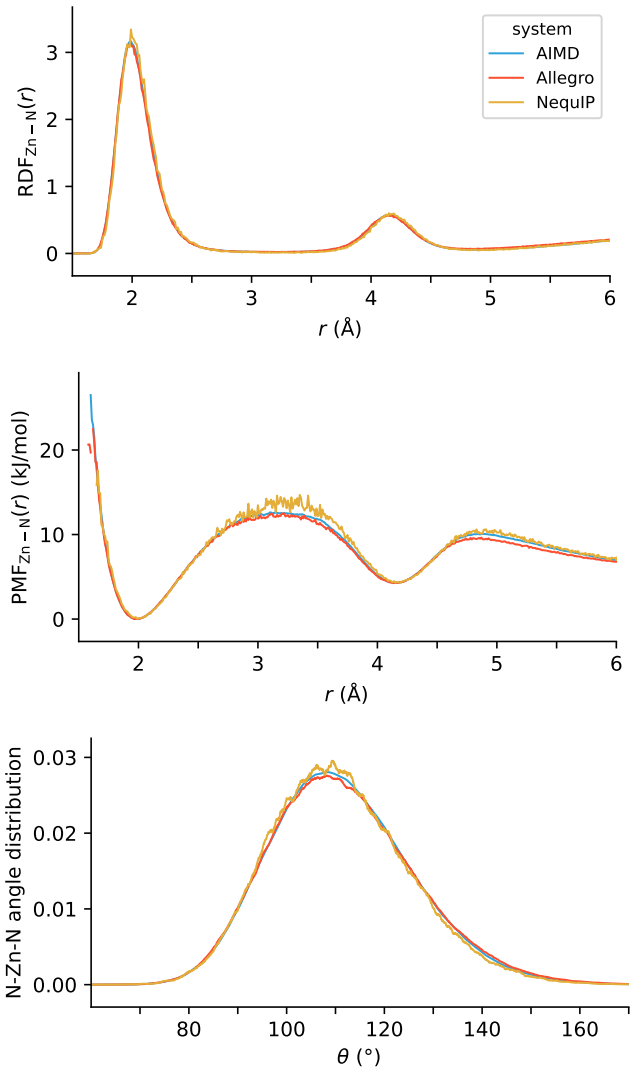

Figure S12: (a) Radial distribution functions (RDF), (b) potentials of mean force (PMF) for the Zn–N atom pairs, and (c) distribution of the N–Zn–N angle for the ZIF-4 liquid at 1500K with Allegro MLP compared to NequIP MLP and AIMD. While the average structural properties are identical, the physical consistency of all imidazolate linkers is not maintained with the Allegro MLP.

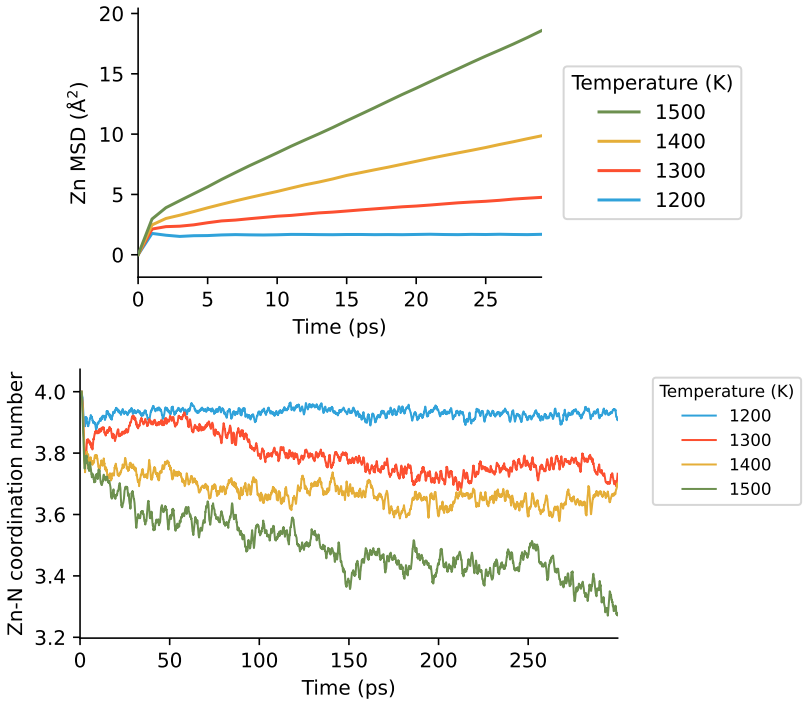

Figure S13: Identification of melting during 300 ps high temperature trajectories with the Allegro MLP. (a) Mean square displacement (MSD) as a function of time for Zn. (b) Zn–N coordination number as the function of time. The use of ring statistics to identify melting is not feasible due to the presence of broken imidazolate.

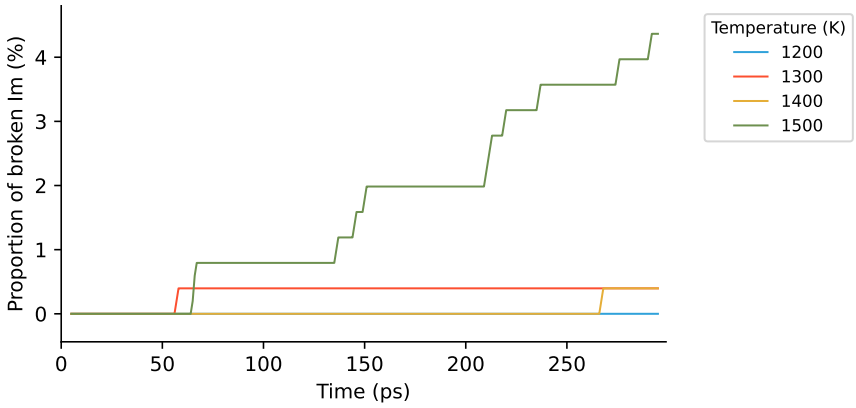

Figure S14: Proportion of broken imidazolate cycles (Im) as a function of time during 300 ps high temperature trajectories with the Allegro MLP. Detected with the Python library `amof`.[4]

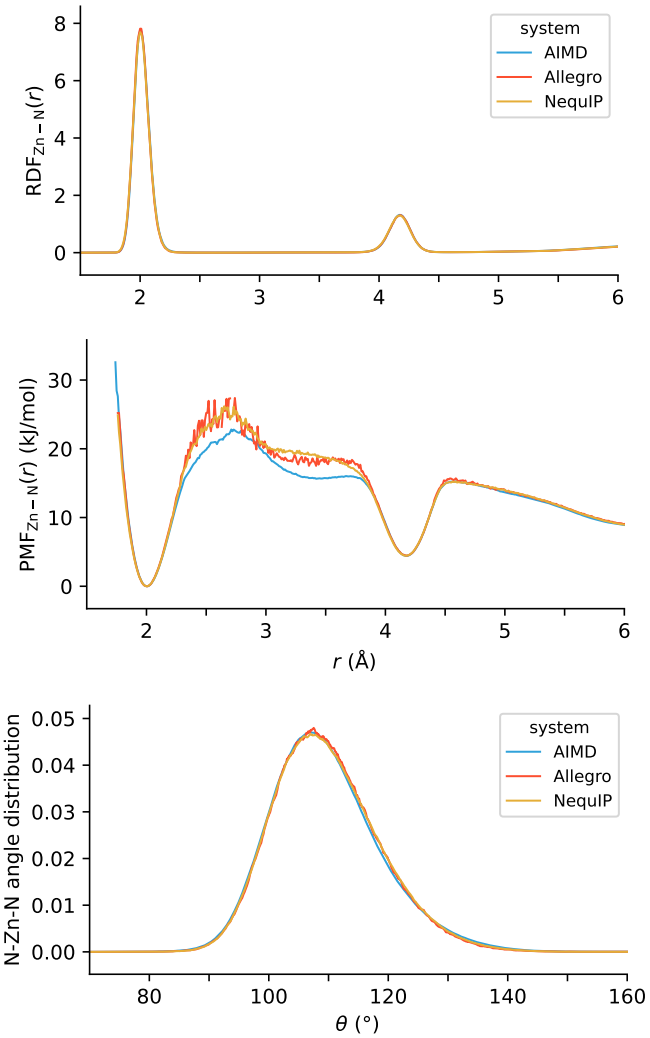

Figure S15: (a) Radial distribution functions (RDF), (b) potentials of mean force (PMF) for the Zn–N atom pairs, and (c) distribution of the N–Zn–N angle for melt-quenched glasses obtained with Allegro (red), NequIP (orange) and AIMD (blue).

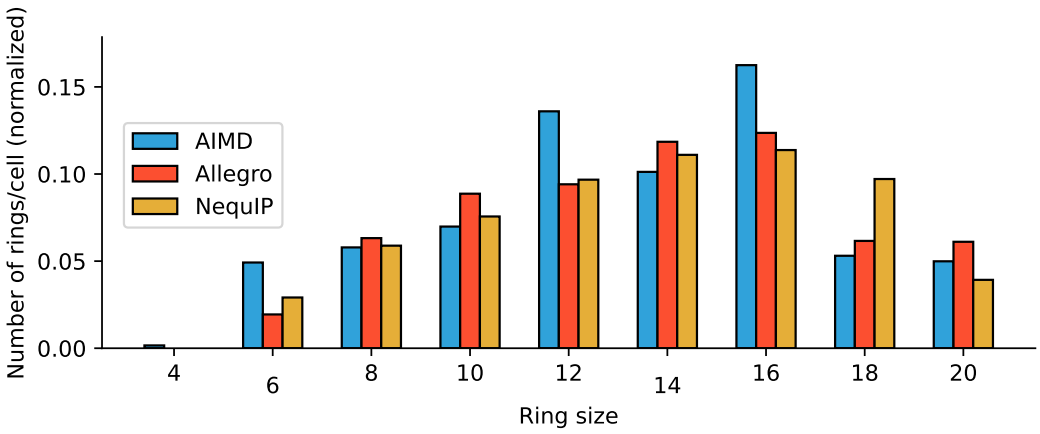

Figure S16: Distributions of size of zinc–imidazolate alternate rings for melt-quenched glasses obtained with Allegro (red), NequIP (orange) and AIMD (blue). The physical consistency of all imidazolate linkers was not maintained through the entire melt-quenching simulations.

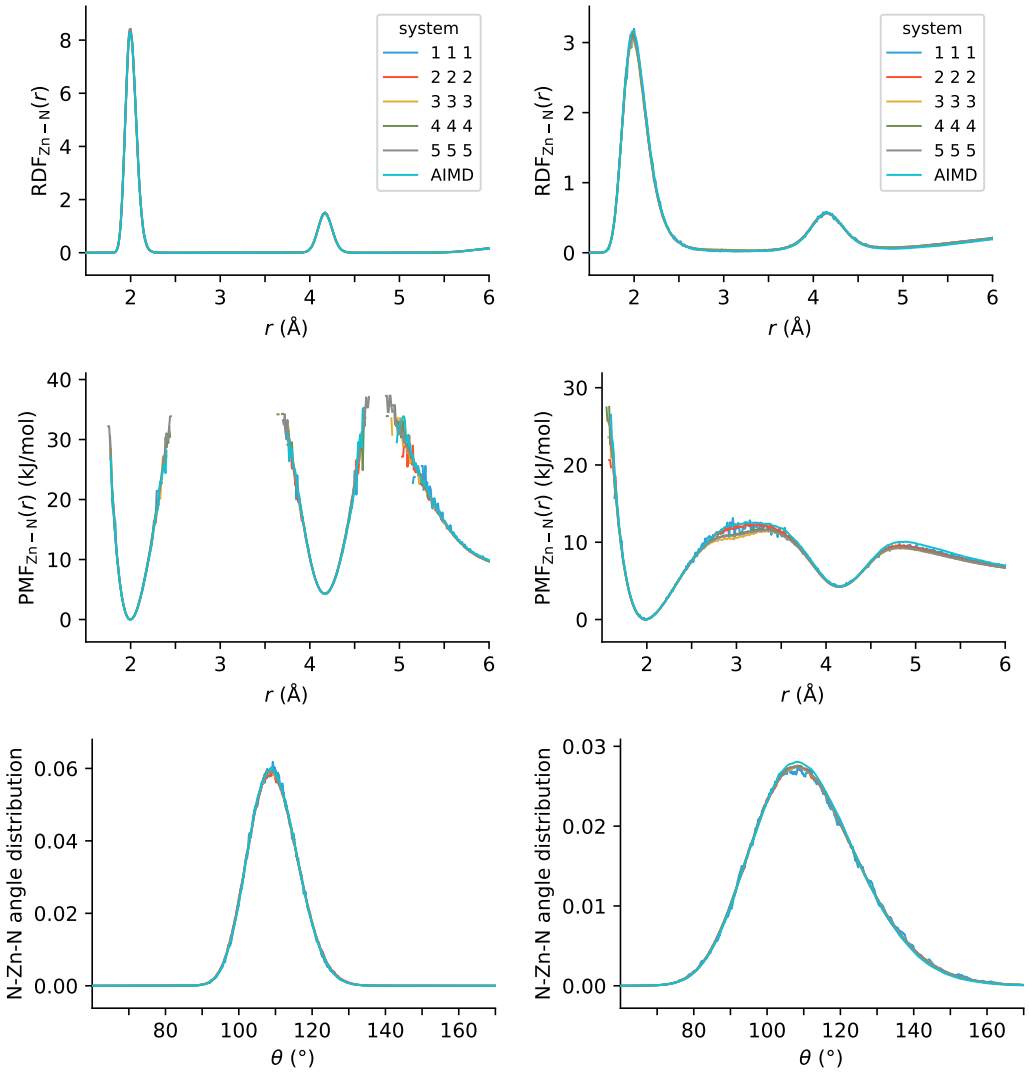

Figure S17: (a) Radial distribution functions (RDF), (b) potentials of mean force (PMF) for the Zn–N atom pairs, and (c) distribution of the N–Zn–N angle for the ZIF-4 crystal (left) and 1500 K liquid (right) with the Allegro MLP and different supercells compared to AIMD.

References

(1) Castel, N.; Coudert, F.-X. Computation of Finite Temperature Mechanical Properties of Zeolitic Imidazolate Framework Glasses by Molecular Dynamics. *Chem. Mater.* **2023**, *35*, 4038–4047.

(2) Gaillac, R.; Pullumbi, P.; Beyer, K. A.; Chapman, K. W.; Keen, D. A.; Bennett, T. D.; Coudert, F.-X. Liquid metal–organic frameworks. *Nature Mater.* **2017**, *16*, 1149–1154.

(3) Gaillac, R.; Pullumbi, P.; Coudert, F.-X. Melting of Zeolitic Imidazolate Frameworks with Different Topologies: Insight from First-Principles Molecular Dynamics. *J. Phys. Chem. C* **2018**, *122*, 6730–6736.

(4) Castel, N.; Coudert, F.-X. Challenges in Molecular Dynamics of Amorphous ZIFs Using Reactive Force Fields. *J. Phys. Chem. C* **2022**, *126*, 19532–19541.

(5) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Commun.* **2022**, *13*, 1.

(6) Vandenhaute, S.; Cools-Ceuppens, M.; DeKeyser, S.; Verstraelen, T.; Van Speybroeck, V. Machine learning potentials for metal-organic frameworks using an incremental learning approach. *npj Comput Mater* **2023**, *9*, 10575.

(7) Gaillac, R.; Pullumbi, P.; Bennett, T. D.; Coudert, F.-X. Structure of Metal–Organic Framework Glasses by *Ab Initio* Molecular Dynamics. *Chem. Mater.* **2020**, *32*, 8004–8011.