

A Machine Learned Potential for Investigating Single Crystal to Single Crystal Transformation in Complex Organic Molecular Systems

Chengxi Zhao^{*a,b}, Honglai Liu^a, Da-Hui Qu^a, Andrew I. Cooper^b, Linjiang Chen^{*c,d}

^a Key Laboratory for Advanced Materials and Joint International Research Laboratory of Precision Chemistry and Molecular Engineering, Feringa Nobel Prize Scientist Joint Research Center, Frontiers Science Center for Materiobiology and Dynamic Chemistry, School of Chemistry and Molecular Engineering, East China University of Science and Technology, Shanghai, China

^b Leverhulme Research Centre for Functional Materials Design, Materials Innovation Factory and Department of Chemistry, University of Liverpool, Liverpool, UK.

^c Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China.

^d School of Chemistry, School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK.

Unit cell selection for generating training data

In the training dataset, the $2 \times 1 \times 1$ and part of the $2 \times 1 \times 2$ data were obtained from the initial AIMD process, as the $2 \times 1 \times 1$ size unit cell calculation is faster and can be used to quickly build a base module for active learning. However, the $2 \times 1 \times 1$ unit cell suffers from significant mirror effects. Therefore, the active learning part primarily focuses on the $2 \times 1 \times 2$ data, as the images from this unit cell have a more appropriate size. The $2 \times 2 \times 1$ unit cell was employed in the final few cycles of testing to ensure that the model correctly predicts the behavior of the supercell along the B axis. This was necessary because the dataset did not previously include a supercell along the B axis, despite the original cell being sufficiently large in that direction.

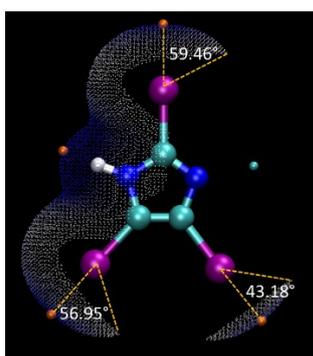


Fig. S1: Electrostatic potential diagrams with only $V_s \geq 0$ is displayed. The σ -hole region could be identified clearly.

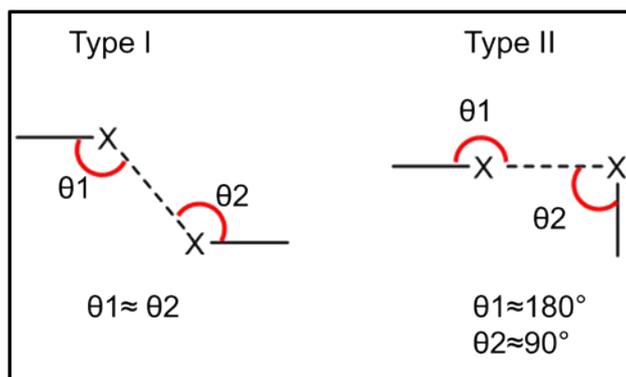


Fig. S2: Schematic diagram of type I and type II halogen-halogen interactions.

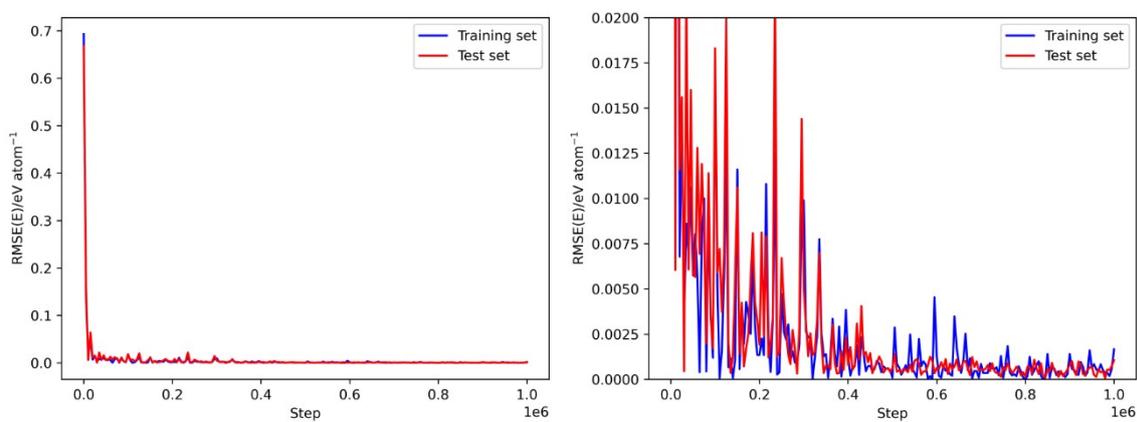


Fig. S3: Energy error of the model throughout the training process. The left and right plots display identical results, with varying Y-axis scales.

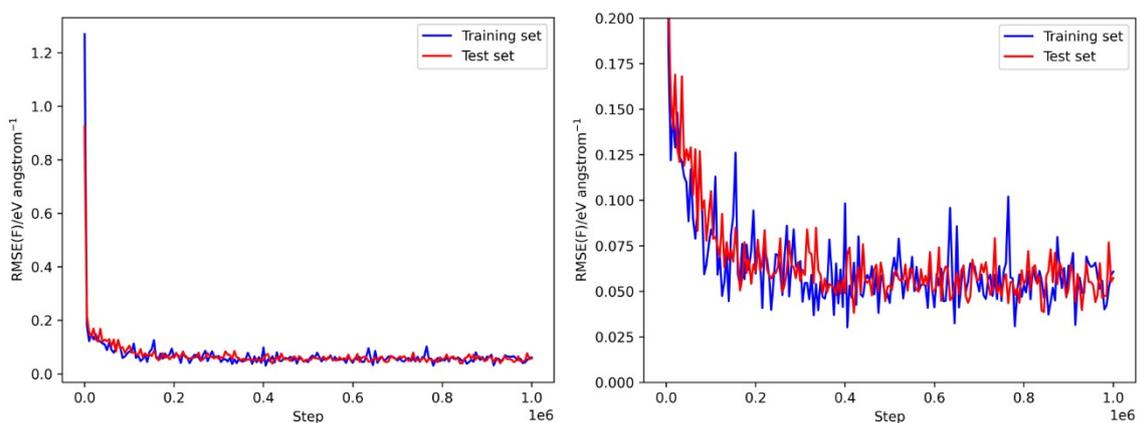


Fig. S4: Force error of the model throughout the training process. The left and right plots display identical results, with varying Y-axis scales.

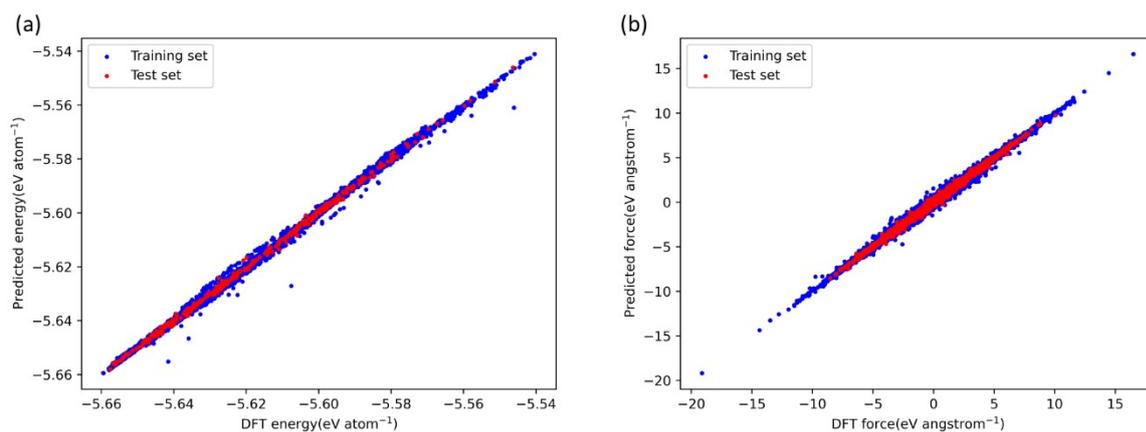


Fig. S5: (a) Energy and (b) (force) errors for the training and test sets.

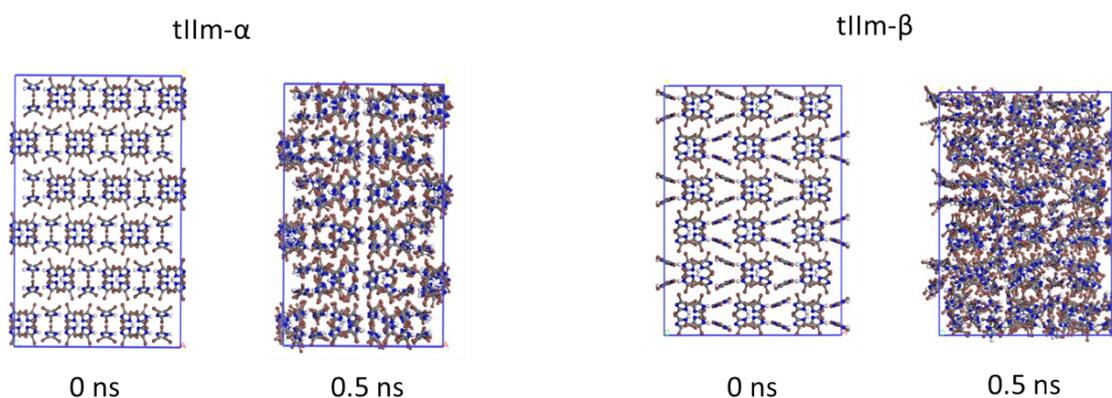


Fig. S6: The snapshots from molecular dynamics simulations using Dreiding force field with a timestep of 0.5 fs. The temperature was controlled using a Nose-Hoover thermostat and the pressure was controlled using a Berendsen barostat. Structures was disintegrated which is not consistent with the experiment result.

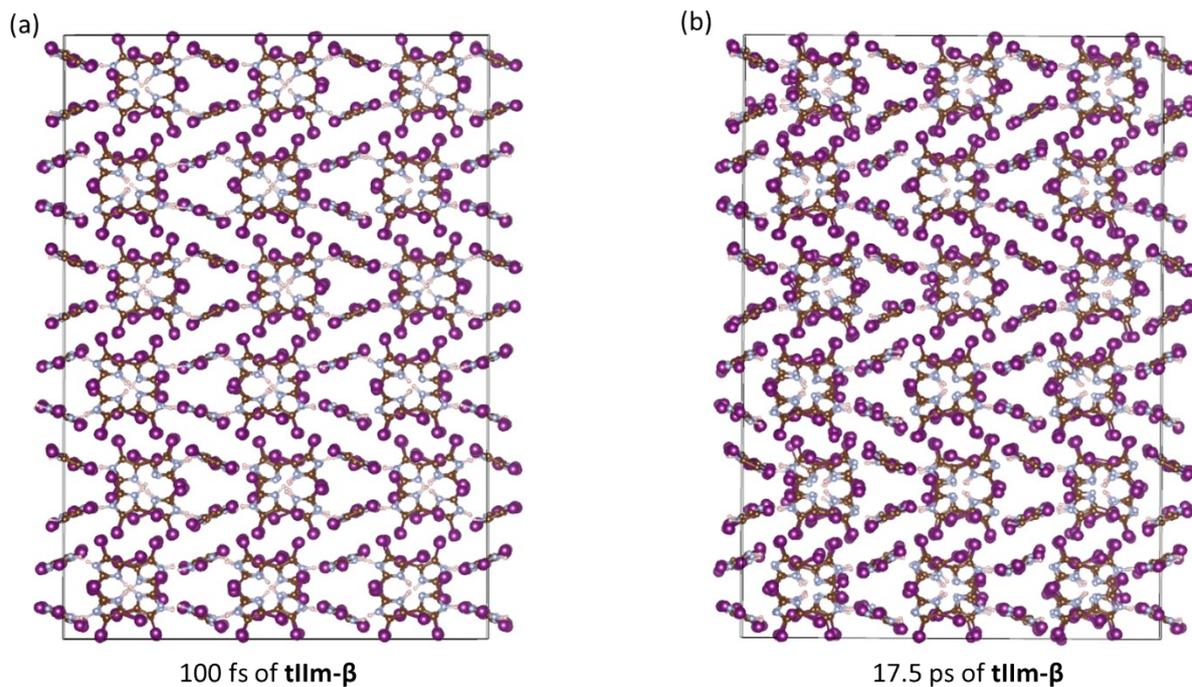


Fig. S7: Snapshot from MD simulations of **tllm- β** that have random initial hydrogen bonding positions.

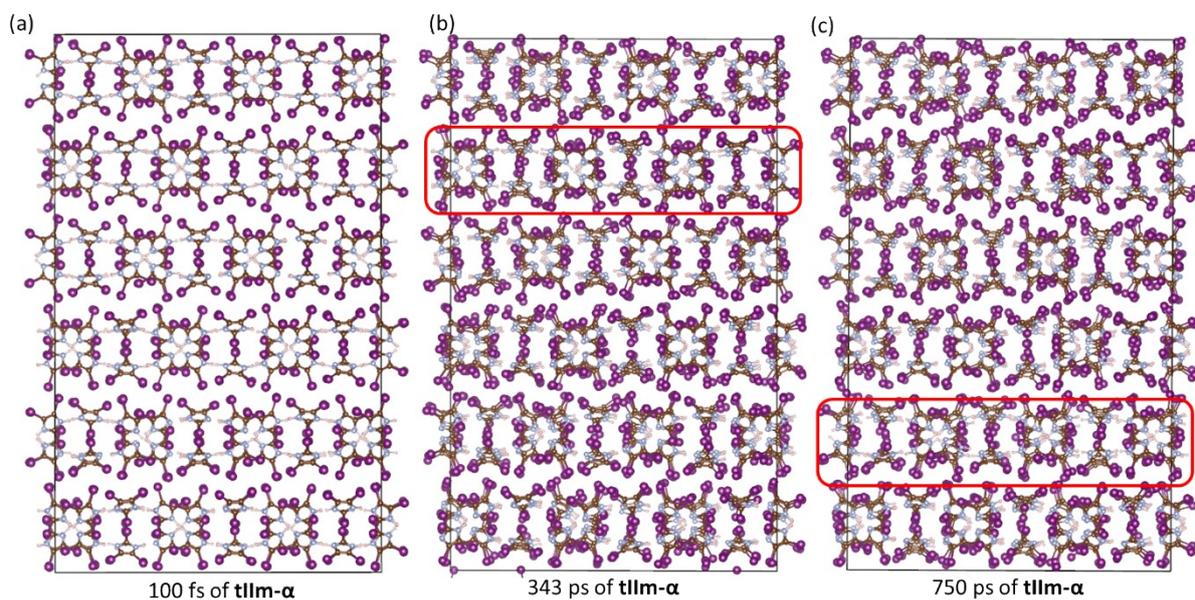


Fig. S8: Snapshot from MD simulations of **tllm- α** that have random initial hydrogen bonding positions.

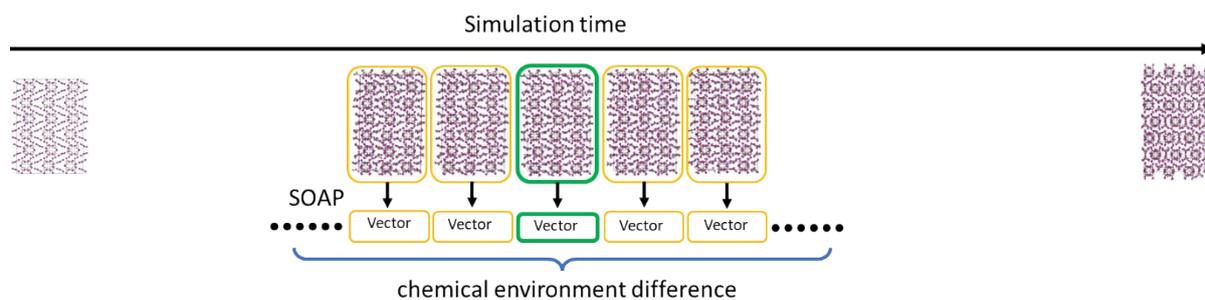


Fig. S9: Diagram of the calculation method for chemical environment difference. Each snapshot will get a float value to quantify its dissimilarity from other snapshots close in time.

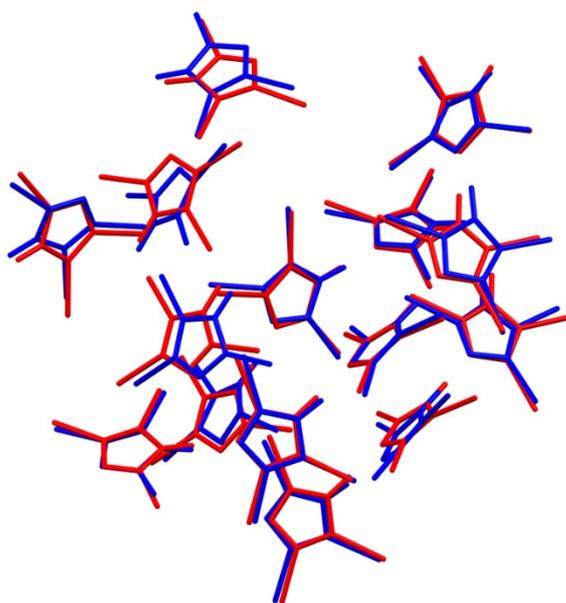


Fig. S10: Crystal packing similarity search performed in Mercury between experimental result of **tllm- α** (red) and a snapshot of **tllm- α** from molecular dynamic simulations (blue). The packing exhibits a 15/15 match with setting of max distance and angle tolerance is 30%.

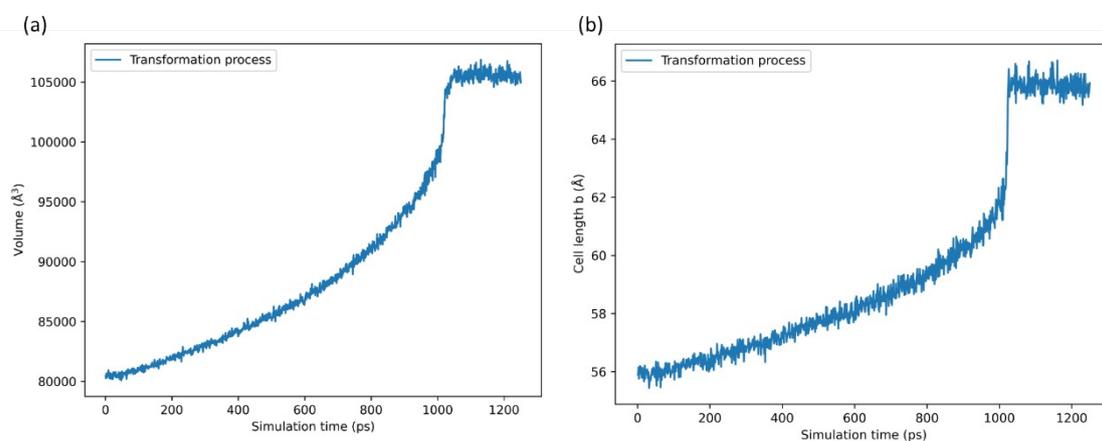


Fig. S11: Simulation trajectory of (a) volume, (b) cell length b as a function of Simulation time for the phase transition from $t11m\text{-}\beta$ to $t11m\text{-}\alpha$. The sudden jump of volume near 1000 ps is due to the happen of phase transition.

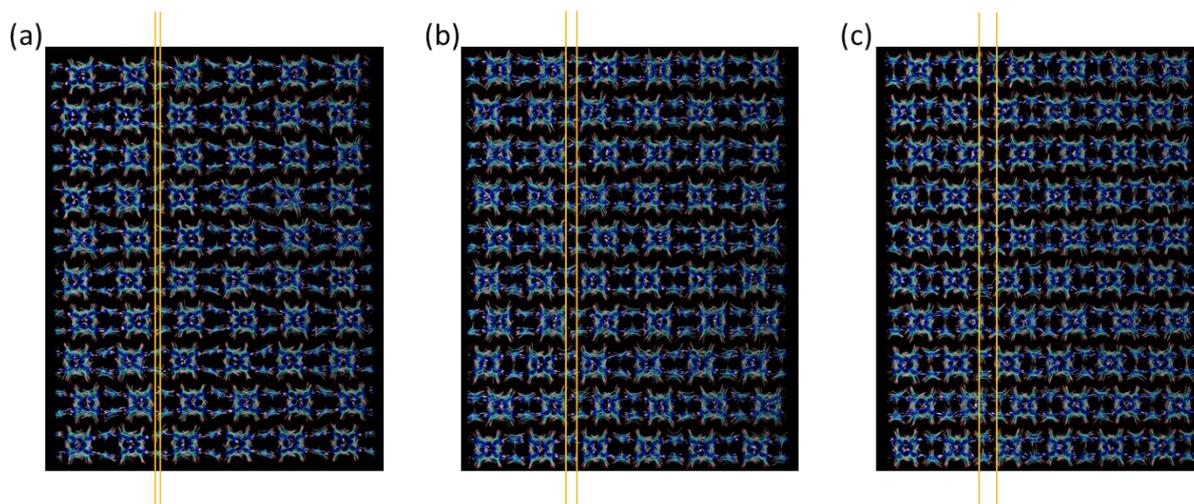


Fig. S12: Simulation trajectory of (a) $t11m\text{-}\beta$, (b) middle of transition process, (c) $t11m\text{-}\alpha$. The simulation is based on a $8 \times 5 \times 5$ super cell. The two yellow lines in the figure represent the isolated molecules in different layer. The rotation of the molecules occurs simultaneously with the layer slides.

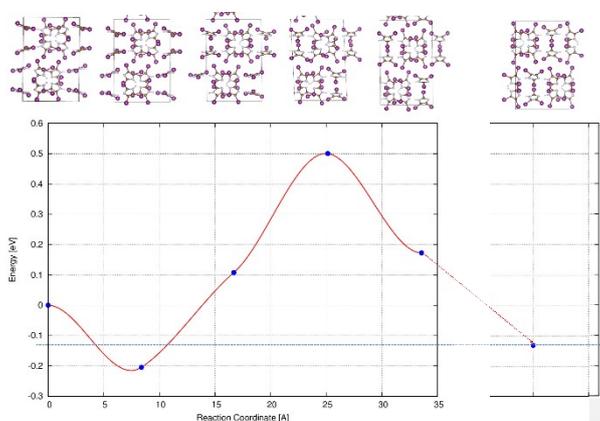


Fig. S13: NEB calculation to provide a rough estimate of the transition barriers. The calculation was carried out as follows: First, a MD simulation of the SCSC transition with the original unit cell was performed until the transition was observed. Next, five images were selected during the MD simulation based on chemical environment difference. The first and last images were optimized and used as the start and end points, while the other images served as the intermediate configurations. Note that here the k -spacing is set as 0.4. We also note that the transition barriers have certain limitations. First, this process involves proton transitions, as shown in the figure. Second, the endpoint does not perfectly match the experimental results due to symmetry breaking caused by thermal fluctuations. We also optimized the perfect α phase, and its energy is shown at the additional position (a separated 6th point in the figure). Third, the pressure factor was not considered in the NEB process. Using images from the MD simulation ensures consistency in the atomic numbering for each image.

Table S1: Dataset details of this research.

Supercell	number of atoms per structure	Structure number
2x1x1	216	3403
2x1x2	432	3639
2x2x1	432	675