Supplementary Information (SI) for Chemical Communications.
This journal is © The Royal Society of Chemistry 2026

Supplementary Information (SI): Tackling Complexity in Crystal Structure Determination of Metal Organic Compounds Using Machine Learning Interatomic Potentials

Hui Wu,[a,*] Qiang Zhu,[b,c] Wei Zhou[a,*]

[a] *NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, MD, 20899-6102, USA. E-mail: huiwu@nist.gov, wzhou@nist.gov.*

[b] *Department of Mechanical Engineering and Engineering Science, University of North Carolina at Charlotte, Charlotte, NC, 28223, USA.*

[c] *North Carolina Battery Complexity, Autonomous Vehicle and Electrification (BATT CAVE) Research Center, Charlotte, NC, 28223, USA.*

Computational Method: MLIPs-Accelerated Crystal Structure Prediction

1. Global Optimization via Evolutionary Algorithm

Crystal structure prediction (CSP) was performed using the USPEX (v10.5.2) package [1], which utilizes a highly efficient evolutionary algorithm for identifying thermodynamically stable structural motifs. In this work, the search space was explored by treating the organic metal salts as "pseudo-molecular crystals." The metal cations and organic anions were defined as individual structural building blocks, a strategy that significantly reduced the degrees of freedom and accelerated the convergence of the global optimization.

2. Machine Learning Interatomic Potentials (MLIPs)

Structural relaxations were driven by two state-of-the-art universal MLIPs: the UMA (Universal Machine Learning Architecture) model ("uma-s-1p1/omat") developed by Meta FAIR [2], and the Orb-v3 model ("orb_v3_conservative_inf_omat") developed by Orbital Materials [3]. These MLIPs were selected for their capacity to capture complex interactions and their proven scalability. UMA utilizes advanced graph neural networks to model the chemical environment with high directional accuracy, while Orb-3 employs an equivariant transformer architecture to ensure the preservation of physical symmetries during energy evaluation. To further include the DFT-D3 dispersion correction [4] in the predictions of these models, the torch-dftd package [5] was used. A force convergence threshold of 0.01 eV/Å was applied for all structural relaxations.

While a single-stage MLIPs-based relaxation is sufficient for most systems, numerical instabilities can occasionally arise when the models encounter high-energy, non-physical geometries generated during the initial random sampling. In such cases, we adopted a two-stage relaxation protocol: 1) Pre-optimization: A preliminary relaxation using empirical force fields to bring the "raw" structures closer to local energy equilibria, and 2) MLIPs Refinement: High-fidelity relaxation using the UMA or Orb-v3 models.

As USPEX lacks a native interface for these emerging MLIPs frameworks, we developed a custom Python-based interface to facilitate seamless integration. This interface handles the execution of MLIPs-driven crystal structure relaxation and formats the resulting data (structures and energies) for USPEX-compatible ranking and selection.

3. Workflow and Convergence Criteria

For each compound, an initial population of 100 candidate structures per generation was produced. The evolutionary cycle—comprising selection, variation (heredity and mutation), and random sampling—was repeated until the global minimum remained unchanged for 20 consecutive generations, ensuring a thoroughly sampled energy-structure landscape.

4. DFT Validation and Final Refinement

Both UMA and Orb-v3 models yielded nearly identical ground-state structures for all three target compounds. To ensure accuracy, the top 20 candidate structures identified by the MLIPs-CSP search were subjected to final optimization using Density Functional Theory with dispersion corrections (DFT/PBE-D3, with Becke–Johnson damping functions) [4], as implemented in the Quantum Espresso package [6]. A kinetic-energy cut-off of 600 eV was used, and convergence threshold of self-consistency was set to 10^{-6} eV. The structural deviations and energy re-ranking during this final step were found to be small, quantitatively confirming the high fidelity of the MLIPs. All structural parameters reported in this study are derived from these final DFT-D3 refined models.

Table S1 Summary table of the benchmark validation results of our MLIPs-CSP framework on several related compounds. The predicted and experimental crystal structures were quantitatively compared using the "StructureMatcher" module in pymatgen [7] and reasonable matches were obtained in all cases.

The experimental and predicted lattice parameters (all differences < 5%)			
System (with exp values)	UMA	Orb3	DFT/PBE-D3
Phenol (C_6H_5OH), $P2_1$ a = 11.610 Å b = 5.442 Å c = 12.217 Å β = 101.47°	 a = 11.538 Å b = 5.273 Å c = 12.147 Å β = 101.06°	 a = 11.967 Å b = 5.219 Å c = 12.296 Å β = 100.44°	 a = 11.442 Å b = 5.287 Å c = 12.187 Å β = 101.83°
Sodium phenolate (C_6H_5ONa), *Iba*2 a = 10.376 Å b = 19.435 Å c = 5.706 Å	 a = 10.350 Å b = 19.378 Å c = 5.687 Å	 a = 10.362 Å b = 19.400 Å c = 5.690 Å	 a = 10.468 Å b = 19.592 Å c = 5.752 Å
Lithium benzimidazolate ($C_7H_5N_2Li$), *Cmma* a = 5.967 Å b = 10.291 Å c = 11.921 Å	 a = 5.681 Å b = 9.873 Å c = 11.508 Å	 a = 5.650 Å b = 9.968 Å c = 11.503 Å	 a = 5.739 Å b = 9.998 Å c = 11.613 Å
2H-benzimidazol-2-one ($C_7H_4ON_2H_2$), $P2_1/n$ a = 5.296 Å b = 5.077 Å c = 23.444 Å β = 94.50°	 a = 5.280 Å b = 5.065 Å c = 23.353 Å β = 94.31°	 a = 5.298 Å b = 5.080 Å c = 23.455 Å β =94.51°	 a = 5.249 Å b = 5.071 Å c = 23.163 Å β =95.27°

Root Mean Square Deviation (RMSD) for atomic positions (Å)			
System	UMA vs. Exp	Orb3 vs. Exp	DFT-D3 vs. Exp
Phenol	0.082	0.095	0.078
Sodium phenolate	0.047	0.037	0.013
Lithium benzimidazolate	0.065	0.061	0.049
2H-benzimidazol-2-one	0.045	0.039	0.037
Lattice energy difference between the predicted and the experimental structure (kJ/mol)			
System	UMA vs. Exp	Orb3 vs. Exp	DFT-D3 vs. Exp
Phenol	0.86	1.48	0.33
Sodium phenolate	0.69	0.23	1.29
Lithium benzimidazolate	5.76	6.32	2.87
2H-benzimidazol-2-one	0.29	0.12	0.21

Table S2 Crystal symmetries and lattice parameters of the predicted structures reported in this work. Comprehensive crystallographic data are available from the CCDC (deposition numbers 2531939–2531942).

System	Symmetry	Cell Parameters
Lithium phenolate (C_6H_5OLi)	*P2/c*	a = 10.053 Å, b = 4.746 Å, c = 22.594 Å, and β = 97.82°
Sodium cyclohexanolate ($C_6H_{11}ONa$), polymorph 1	C2/m	a = 7.152 Å, b = 6.672 Å, c = 12.434 Å, and β = 101.78°
Sodium cyclohexanolate ($C_6H_{11}ONa$), polymorph 2	*Abm2*	a = 12.352 Å, b = 6.696 Å, c = 7.136 Å
Lithium-benzimidazol-2-one ($C_7H_4ON_2Li_2$)	$P2_1/n$	a = 5.296 Å, b = 5.077 Å, c = 23.444 Å, and β = 94.50°

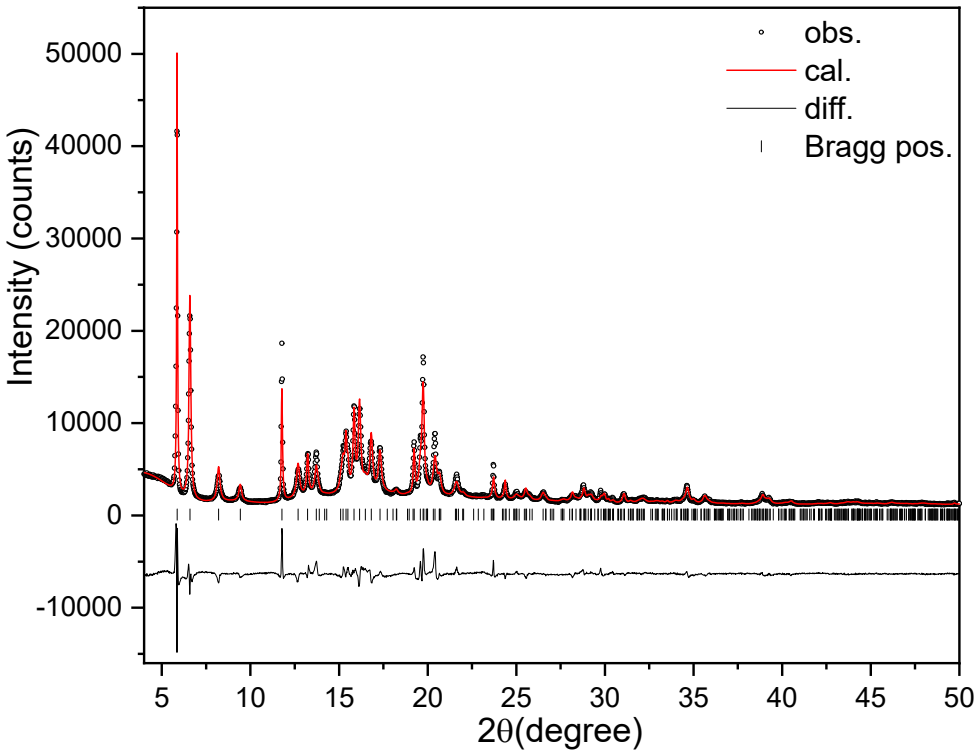

Fig. S1 Rietveld fit of the synchrotron PXRD profile for lithium phenolate (*P2/c* model). Rietveld refinement was performed using GSAS-II [8] with Stephens Model [9] and ellipsoidal domain size model applied to address the anisotropic peak broadening. R_{wp}= 0.0852, reduced χ^2=15.69. Synchrotron data is from reference [10] with measurement wavelength λ=1.14853 Å. The unsatisfactory intensity differences between observed and calculated peaks even after anisotropic peak corrections suggest the multifaceted peak shape complexity from possible shape anisotropy of crystallites and/or texture in the sample used for PXRD measurement.

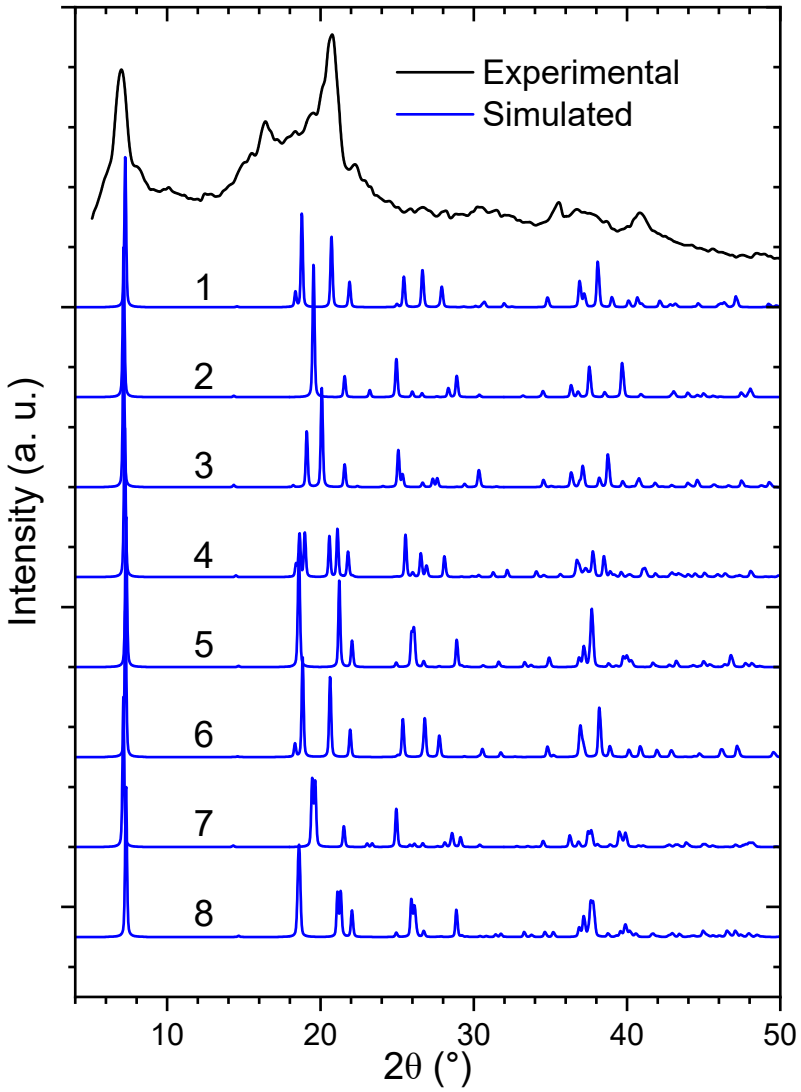

Fig. S2 A comparison between the experimental lab PXRD data of sodium cyclohexanolate (Cu Kα radiation; data from reference [11]) and the calculated patterns of the predicted eight polymorph structures.

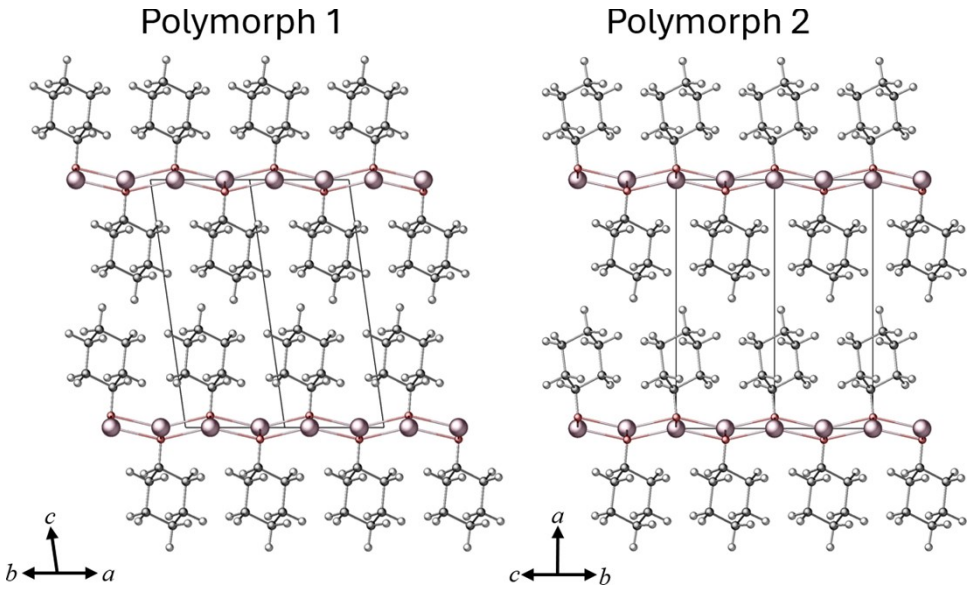

Fig. S3 Representative predicted lowest energy polymorph structures of sodium cyclohexanolate.

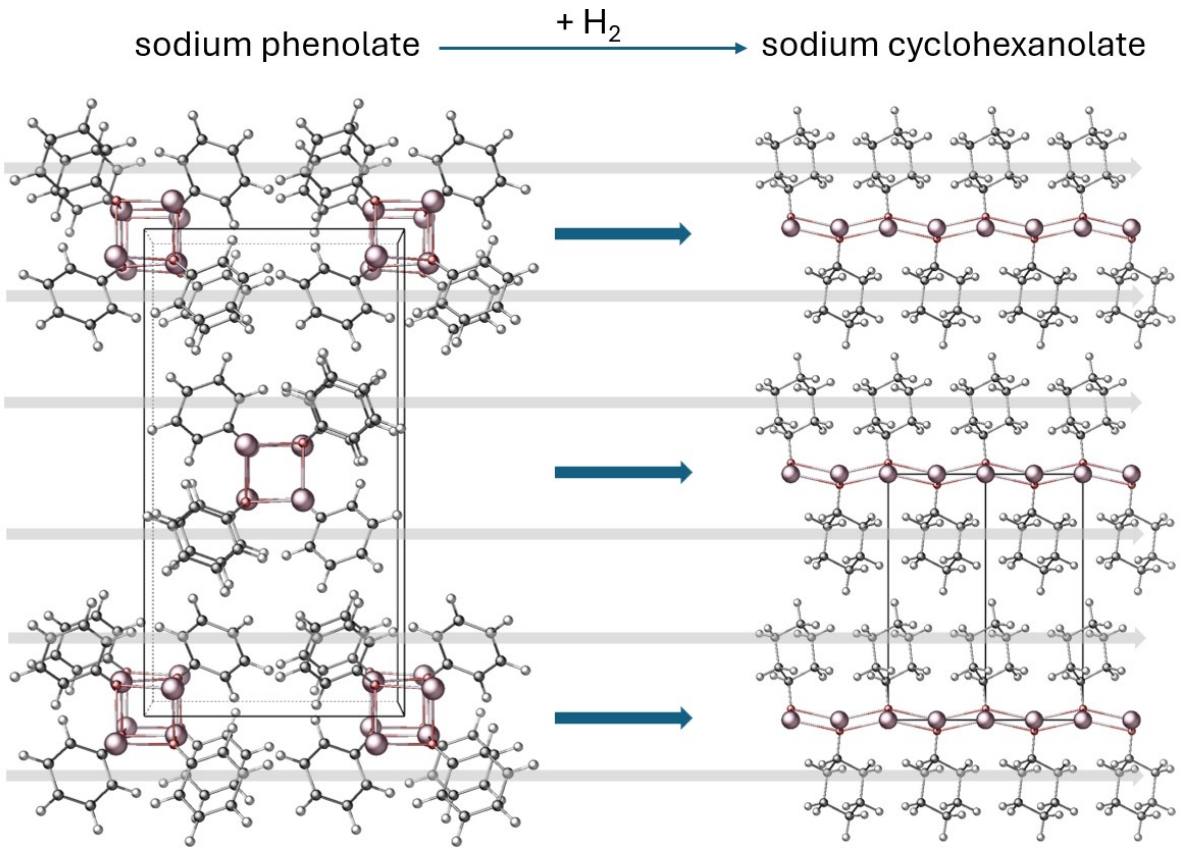

Fig. S4 Structural comparison of sodium phenoxide (left) and sodium cyclohexanolate (right), highlighting the conformational changes in the organic rings and the evolution of the sodium coordination environment upon hydrogenation.

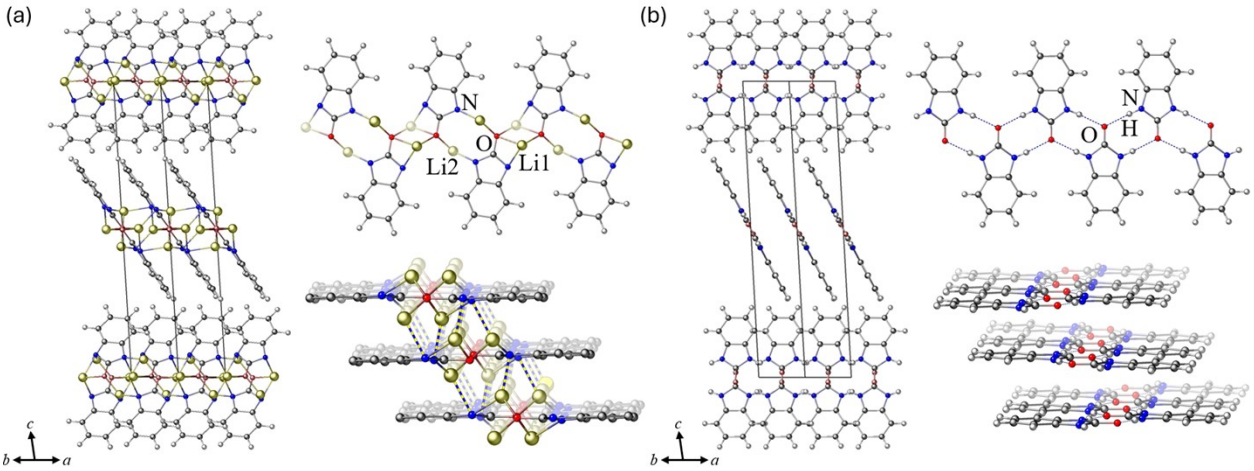

Fig. S5 Structural comparison of (a) lithium-benzimidazol-2-one and (b) 2H-benzimidazol-2-one. In the $[Li_2(C_6H_4N_2O)]_n$ belt, planar $[C_6H_4N_2CO]^{2-}$ anions are connected by the inward- and outward-plane Li^+ cations through Li-N and Li-O bonds (solid lines). These belts are further cross-linked into a 3D framework by inter-belt Li-N bonds (dash lines). In contrast, in the $[C_6H_4(NH)_2O]_n$ belt, planar $[C_6H_4(NH)_2CO]$ molecules are connected by N-H···O hydrogen bonds (dash lines). The inter-belt stabilization is provided by parallel-displaced π-π interaction.

References:

1. (a) A. R. Oganov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 244704. (b) A. O. Lyakhov, A. R. Oganov, H. T. Stokes and Q. Zhu, *Comput. Phys. Commun.*, 2013, **184**, 1172–1182.
2. B. M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J. R. Kitchin, D. S. Levine, K. Michel, A. Sriram, T. Cohen, A. Das, A. Rizvi, S. J. Sahoo, Z. W. Ulissi and C. L. Zitnick, DOI:10.48550/arXiv.250623971.
3. B. Rhodes, S. Vandenhaute, V. Šimkus, J. Gin, J. Godwin, T. Duignan and M. Neumann, DOI:10.48550/arXiv.2504.06231.
4. S. Grimmea, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
5. S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano et al., *Nat. Commun.*, 2022, **13**, 2991.
6. P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio Jr., A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, S. Baroni, *J. Phys. Condens. Matter*, 2017, **29**, 465901.
7. S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Computational Materials Science*, 2013, **68**, 314-319.
8. B. T. Toby, R. B. Von Dreele, GSAS-II: the genesis of a modern open-source all purpose crystallography software package. *J. Appl. Cryst.*, 2013, **46**, 544-549.
9. P. W. Steppens, *J. Appl. Cryst.*, 1999, **32**, 281-289.
10. R. E. Dinnebier, M. Pink, J. Sieler and P. W. Stephens, *Inorg. Chem.*, 1997, **36**, 3398–3401.
11. Y. Yu, T. He, A. Wu, Q. Pei, A. Karkamkar, T. Autrey and P. Chen, *Angewandte Chemie International Edition*, 2019, **58**, 3102–3107.