



Cite this: DOI: 10.1039/d6cc01457g

 Received 11th March 2026,
Accepted 5th May 2026

DOI: 10.1039/d6cc01457g

rsc.li/chemcomm

Tackling complexity in crystal structure determination of metal organic compounds using machine learning interatomic potentials

 Hui Wu,^{ib}*^a Qiang Zhu,^{ib}^{bc} and Wei Zhou,^{ib}*^a

We integrate *ab initio* crystal structure prediction with universal machine learning interatomic potentials (UMA and Orb-v3) to determine challenging metal organic compound structures. Our framework successfully resolves the previously unknown, technologically significant crystal structures of lithium phenolate, sodium cyclohexanolate, and lithium-benzimidazol-2-one, accelerating materials development.

Structure determination is fundamental to understanding the physicochemical properties and performance of crystalline materials across a broad range of applications. While single-crystal diffraction remains the gold standard, this approach is frequently limited by the inability to synthesize crystals of sufficient size and quality. In such instances, researchers must rely on powder X-ray/neutron diffraction (PXRD/PND). However, because PXRD/PND data are “powder-averaged”, the resulting loss of three-dimensional reciprocal space information could make structure solution a formidable challenge—particularly for systems characterized by low symmetry, poor crystallinity, and/or the presence of persistent impurities.

To overcome these experimental limitations, *ab initio* crystal structure prediction (CSP) has emerged as a powerful computational ally.^{1–4} By identifying the most thermodynamically stable arrangements of atoms or molecules based solely on chemical composition, CSP generates a theoretical library of structural candidates. These candidates can then be cross-referenced with experimental diffraction patterns to facilitate structure determination. Despite its potential, the primary bottleneck in CSP is the vastness of the “energy-structure landscape”. Exhaustive global optimization requires thousands to millions of structural relaxations and energy rankings, which, when performed at the level of

Density Functional Theory (DFT), are computationally prohibitive for complex systems. Machine Learning Interatomic Potentials (MLIPs) are currently revolutionizing this workflow by acting as a high-speed bridge between computationally inexpensive but often inaccurate classical force fields and slow, high-fidelity first-principles methods.⁵ MLIPs offer near-DFT accuracy at a fraction of the computational cost, enable a more thorough exploration of configurational space. We have recently demonstrated that MLIPs-assisted CSP exhibits excellent performance in predicting the structures of porous organic crystals;^{6,7} however, its application to the multifaceted interactions within metal organic compounds (in particular, organic metal salts) remains an open frontier.

In this work, we employ an MLIPs-accelerated CSP framework to tackle the structure determination of several technologically significant organic metal salts whose structures have remained elusive. These materials, composed of metal cations and organic anions, are critical to the development of next-generation energy storage and conversion technologies. Researchers have previously been forced to rely on empirical assumptions regarding working mechanisms due to a lack of definitive structural data. The reliable models provided in this work enable researchers to conduct in-depth mechanistic studies and apply structure–property relationships to the design of improved materials. We focus on three key cases:

- (1) Lithium phenolate (C₆H₅OLi): a foundational reagent in chemistry⁸ and a recently identified dynamic intermediate in sustainable, lithium-mediated ammonia electrosynthesis.⁹
- (2) Sodium cyclohexanolate (C₆H₁₁ONa): A promising material for high-capacity, on-board hydrogen storage.¹⁰
- (3) Lithium-benzimidazol-2-one (C₇H₄ON₂Li₂): A critical compound for mitigating irreversible Li⁺ loss and extending the cycle life of high-energy-density lithium-ion batteries.¹¹

We first present our computational methodology briefly (more details provided in the SI). We utilized two state-of-the-art universal MLIPs: the UMA model¹² and the Orb-v3 model¹³. These MLIPs were selected for their capacity to capture complex interactions and their proven scalability. The crystal structure

^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



prediction was conducted within a global optimization framework. We used the USPEX code for the random structure generation and structure evolution.^{14,15} In each cycle, the structural candidates were first relaxed using MLIPs, and then processed by USPEX for energy ranking, followed by the generation of the next batch of candidates through an evolutionary algorithm as well as additional random sampling. This process was iterated until no further structures with lower energies could be identified.

To establish the reliability of the MLIPs-CSP framework for organic metal salts, we first performed a benchmark test on a set of known crystalline structures within this family, including phenol (C_6H_5OH), sodium phenolate (C_6H_5ONa), lithium benzimidazole ($C_7H_5N_2Li$) and 2H-benzimidazol-2-one ($C_7H_4ON_2H_2$), which are closely related to the target materials. We found that the MLIPs-CSP consistently identified the experimental structures as the global energy minimum, confirming that UMA and Orb-3, augmented with the D3 dispersion correction, can reliably capture strong ionic M–O/M–N bonding as well as subtle intermolecular van der Waals (vdW) forces. Table S1 summarizes the benchmark results, including experimental and predicted lattice parameters, root-mean-square deviations (RMSD) for atomic positions, and the lattice energy differences between the predicted and experimental structures. Following this successful benchmarking, the MLIPs-CSP method was applied to the three target salts.

We start on lithium phenolate, which is a versatile reagent in organic synthesis and a recently recognized important dynamic intermediate in green chemistry.⁹ Attempt on its crystal structure determination was made back in 1996. The researchers used high-resolution synchrotron PXRD and successfully determined the crystal structures of several alkali metal phenolates, including K, Rb, and Cs.⁸ However, the crystal structure of lithium phenolate could not be solved. The authors cited several reasons for this “structural elusive” nature, including poor crystallinity of the sample and anisotropic diffraction line broadening. Nevertheless, they were able to successfully index the powder pattern of lithium phenolate to a monoclinic $P2_1/a$ crystal system.

Knowing the crystal symmetry and unit cell can dramatically reduce the CSP search space. Therefore, in our CSP work of this system, we constrained the lattice parameters to the indexed experimental values and restricted the space groups to $P2_1/a$, $P2_1/c$, and their subgroups ($P2_1$, $C2$, and Pc). The MLIPs-CSP calculations quickly converged (within ~ 30 iterations), and the ground-state structure (Fig. 1a and Table S2) was identified. Fig. 1d shows a comparison between the experimental PXRD data and the calculated pattern of the predicted lowest energy structure. Overall, the agreement is very well, except that there are some minor intensity mismatches. In particular, the experimental (00 l) reflections exhibit notably higher intensities than the predicted, suggesting possible shape anisotropy of the crystallites and/or texture (preferred orientation) in the sample used for the PXRD measurement. This is also consistent with the observation that the (00 l) reflections in the experimental data have notably narrower line width than others. We attempted Rietveld refinements by incorporating crystallite domain size and micro-strain broadening corrections for this anisotropy, yet the results were not satisfactory due to the aforementioned multifaceted

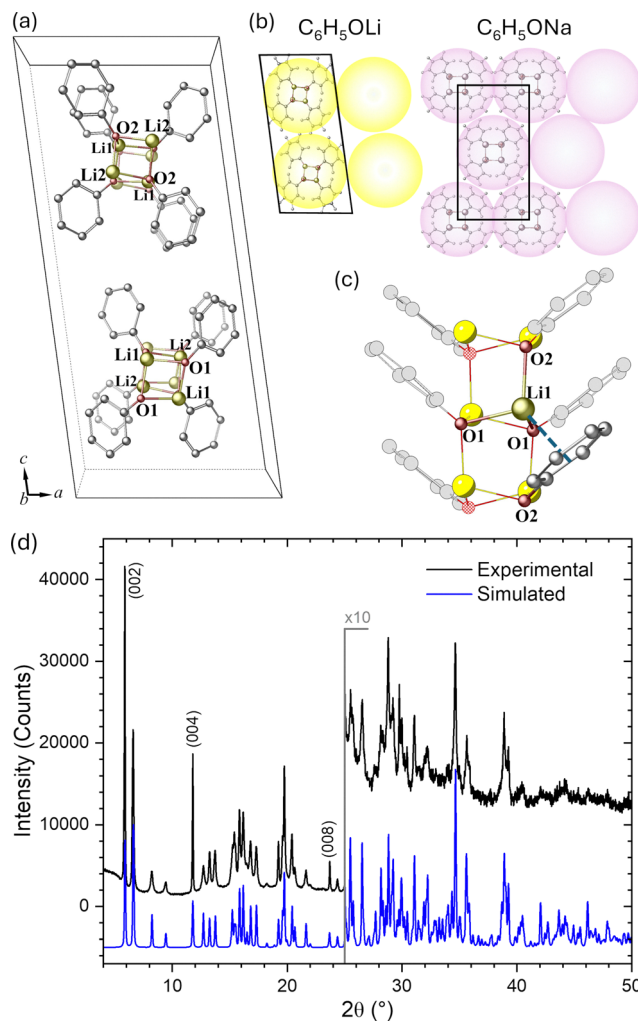


Fig. 1 (a) The predicted crystal structure of lithium phenolate. H atoms are omitted for clarity. (b) Different packing of the $[MOPh]_n$ column clusters in lithium phenolate and sodium phenolate. (c) The tetrahedral coordination of Li^+ in lithium phenolate. (d) The experimental synchrotron PXRD data ($\lambda = 1.14853 \text{ \AA}$; from ref. 8) in comparison with the calculated pattern of the predicted structure.

complexity (see Fig. S1). Nevertheless, Fig. 1d convincingly shows that the predicted structure is a near-perfect representation of lithium phenolate crystal.

The structure of lithium phenolate is composed of polymeric $[LiOPh]_n$ columns propagating along the $[010]$ direction. Within the column, every two Li^+ cations and two transverse OPh^- anions are bonded to form a Li_2O_2 four-membered ring. The two Li^+ cations in a given Li_2O_2 ring are also coordinated with the O atoms of two OPh^- anions from the adjacent Li_2O_2 ring directly above, and interact with the C–C edges of the phenyl rings from a neighbouring Li_2O_2 unit right below. This results a tetrahedral coordination for each Li^+ cation (Fig. 1c), with the Li_2O_2 rings stacking along the $[010]$ direction forming $[LiOPh]_n$ column-like cluster. These $[LiOPh]_n$ columns then cohere together *via* weak inter-molecular interactions to form the 3D crystal. Notably, the tetrahedral Li coordination and the construction of the $[LiOPh]_n$ columns are quite similar to those for Na in sodium phenolate.



The primary difference between the two lies in the assembling of the $[\text{MOPh}]_n$ columns. As shown in Fig. 1b, the $[\text{LiOPh}]_n$ columns in lithium phenolate adopt a pseudo-square packing, whereas the $[\text{NaOPh}]_n$ columns in sodium phenolate exhibit a pseudo-hexagonal packing.

Next, we discuss sodium cyclohexanolate. In 2019, the sodium phenolate and sodium cyclohexanolate pair was identified as a highly promising material system for H_2 storage,¹⁰ with sodium cyclohexanolate serving as the hydrogenated form. Characterizing the structural transformation between these states is crucial for understanding the H_2 storage mechanism. While the starting sodium phenolate material is highly crystalline with a well-documented structure (as discussed above), the material lost much of its crystallinity upon hydrogenation, presumably due to substantial crystal volume expansion. The experimental PXRD data is dominated by two broad features centred at $2\theta \sim 7.1^\circ$ and 20.8° (see Fig. S2), which precludes unit cell indexing and conventional structure solution. Without any experimental inputs, we had to perform a thorough MLIPSP-CSP calculation on sodium cyclohexanolate. We considered all space groups and various Z' values (number of formula units in the asymmetric unit: 1, 2, 3, 4 and 8), evaluating about ~ 100 K candidate structures in total before achieving convergence. We found that the energy landscape of this material is relatively flat. Rather than a single lowest energy structure, we identified eight conceptually similar but crystallographically distinct structures with nearly identical energies ($< 1 \text{ kJ mol}^{-1}$ in difference). Note that the existence of multiple polymorphs, a common cause for structure disorder, is entirely consistent with the broad PXRD features observed. The structures of representative predicted polymorphs are shown in Fig. 2 and Fig. S3 (also see Table S2). Their corresponding calculated PXRD patterns are shown in Fig. S2 along with the experimental pattern. Although it is impossible to do a one-to-one peak match in this case, the general agreement does support that the predicted structures are largely reasonable.

Different from sodium phenolate, where the Na and O atoms form columns of Na_2O_2 four-membered rings, in sodium cyclohexanolate they form a quasi-planar motif (Fig. 2b), with each Na coordinated to O in a pseudo square planar geometry. The cyclohexanolate rings extend from these O atoms, alternating

above and below the Na–O plane to form 2D layers. The 3D crystal structure can be viewed as a stacking of these layers *via* vdW interactions. Note that the cyclohexanolate rings can assume different orientations and the interlayer stacking mode can also vary, hence a diverse array of polymorphs can form with only small energy differences. Because these polymorphs are energetically nearly identical and structurally quite similar, their bulk thermodynamic properties (such as the enthalpy of hydrogen release) are expected to be very similar. A side-by-side structural comparison between sodium phenolate and sodium cyclohexanolate is shown in Fig. S4, illustrating the significant hydrogenation-induced structural evolution of this system.

Our third case study focuses on lithium-benzimidazol-2-one. It was reported in 2025 that this compound can function as a high-performance Li^+ compensation agent, significantly enhancing the cycling stability of high-energy-density lithium-ion batteries.¹¹ Despite reported laboratory PXRD data, its crystal structure remained unsolved. Lithium-benzimidazol-2-one was synthesized by reacting 2H-benzimidazol-2-one with lithium methoxide. We noted that the reported PXRD data of the 2H-benzimidazol-2-one precursor indicate that the crystallites possess anisotropic shapes and preferred orientation in the sample. The PXRD data from the lithium-benzimidazol-2-one sample are also complicated by these same factors, mirroring the challenges encountered with lithium phenolate. We first indexed the PXRD pattern and determined the crystal symmetry to be monoclinic, while the lattice parameters could not be uniquely identified. In our CSP work, we then restricted the structure search to monoclinic space groups and considered various Z' values. In total, ~ 20 K candidate structures were evaluated before convergence was achieved. Fig. 3a depicts the predicted structure of lithium-benzimidazol-2-one (also see Table S2). The corresponding calculated PXRD pattern is shown in Fig. 3b alongside the experimental pattern. Despite the complications of sample texture and crystallite shape anisotropy, a one-to-one peak correspondence is apparent, strongly supporting the validity of the predicted structure.

The structure of lithium-benzimidazol-2-one exhibits great similarity to that of its precursor, 2H-benzimidazol-2-one. A side-by-side comparison is provided in Fig. S5. Before metalation, the planar $\text{C}_6\text{H}_4(\text{NH})_2\text{CO}$ molecules are oriented with their imidazolinone moieties toward one another and are linked *via* $\text{O} \cdots \text{H}-\text{N}$ hydrogen bonds. This forms a $[\text{C}_6\text{H}_4(\text{NH})_2\text{CO}]_n$ flat belt extending along the $\langle 110 \rangle$ directions. These aromatic ring-containing belts then stack through the parallel-displaced $\pi-\pi$ interaction, forming arrays of $[\text{C}_6\text{H}_4(\text{NH})_2\text{CO}]_n$ belts. The 3D crystal structure of 2H-benzimidazol-2-one is composed of such stacks of parallel $[\text{C}_6\text{H}_4(\text{NH})_2\text{CO}]_n$ belts alternating along the $[001]$ direction with their stacking direction orthogonal to each other. Upon Li substitution, the arrangement of the $[\text{C}_6\text{H}_4\text{N}_2\text{CO}]^{2-}$ anions remains similar to that in the precursor structure, however, the planar $[\text{C}_6\text{H}_4\text{N}_2\text{CO}]^{2-}$ anions within each flat belt are now interconnected by Li–O and Li–N bonds. Unlike the covalently bonded in-plane H (N) atoms in the $\text{C}_6\text{H}_4(\text{NH})_2\text{CO}$ molecule, the Li^+ cations are situated in out-of-plane positions (Fig. 3a). As these $\text{Li}_2[\text{C}_6\text{H}_4\text{N}_2\text{CO}]$ belts stack, the out-of-plane Li^+ cations can further

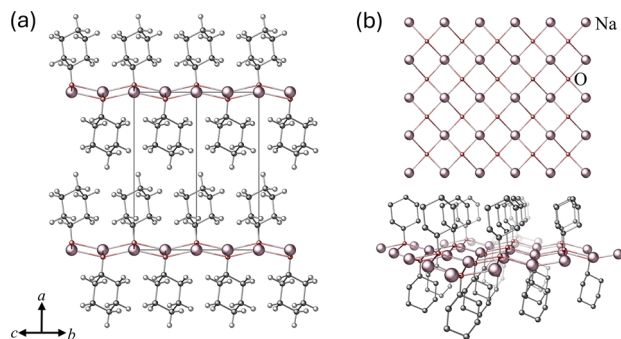


Fig. 2 (a) Representative predicted lowest energy structure of sodium cyclohexanolate. (b) The pseudo-square planar coordination of Na^+ .



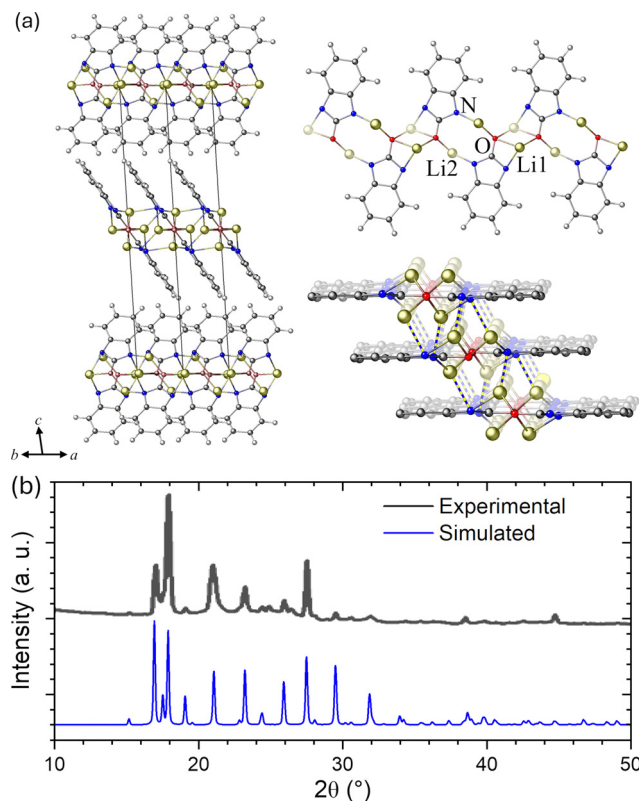


Fig. 3 (a) Predicted crystal structure of lithium-benzimidazol-2-one. The Li coordination environment is shown in detail. (b) The experimental lab PXRD data (Cu K_{α} radiation; data from ref. 11) along with the calculated pattern of the predicted structure.

interact with the $[\text{C}_6\text{H}_4\text{N}_2\text{CO}]^{2-}$ anions in adjacent parallel belts through Li–N bonding, effectively knitting a stack of $\text{Li}_2[\text{C}_6\text{H}_4\text{N}_2\text{CO}]$ flat belts altogether (Fig. 3a). The knitted stacks of $\text{Li}_2[\text{C}_6\text{H}_4\text{N}_2\text{CO}]$ flat belts then orthogonally pack along the [001] direction *via* weak vdW interactions. Notably, the dual Li–N/O coordination identified herein provides a fundamental physical basis for the high stability and favorable lithium-release thermodynamics of this important Li compensation material.

In summary, we have presented an *ab initio* crystal structure prediction framework accelerated by universal machine learning interatomic potentials. Using this method, we successfully resolved the structures of several organic metal salts that were previously unknown and beyond the reach of conventional crystallography. While these computational structures await future confirmation by higher-quality experimental diffraction data, the structural insights reported here provide essential information for understanding the properties of these compounds in related energy storage and conversion applications, facilitating the continued development of these materials. We expect that the MLIPs-accelerated CSP method will be broadly adopted across various materials in the near future.

Conflicts of interest

There are no conflicts to declare.

Disclaimer

Certain commercial suppliers are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Data availability

All data of this study are included in this published article and its supplementary information (SI). Supplementary information: details of the computational method, supporting tables and figures, cif files of the predicted eight sodium cyclohexanolate polymorphs, and representative USPEX input files. CCDC 2531939–2531942 contain the supplementary crystallographic data for this paper. See DOI: <https://doi.org/10.1039/d6cc01457g>.

Acknowledgements

The authors thank Prof. Peter W. Stephens (Stony Brook University) for providing the original experimental synchrotron PXRD data of lithium phenolate and for valuable scientific discussions. Q. Z. acknowledges the NSF (DMR-2410178) for financial support. Unless otherwise noted, NIST work was funded solely by the United States Government.

References

- G. M. Day, *Crystallogr. Rev.*, 2011, **17**, 3–52.
- S. L. Price, *Chem. Soc. Rev.*, 2014, **43**, 2098–2111.
- A. R. Oganov, C. J. Pickard, Q. Zhu and R. J. Needs, *Nat. Rev. Mater.*, 2019, **4**, 331–348.
- Q. Zhu and S. Hattori, *J. Mater. Res.*, 2023, **38**, 19–36.
- R. Jacobs, D. Morgan, S. Attarian, J. Meng, C. Shen, Z. Wu, C. Y. Xie, J. H. Yang, N. Artrith, B. Blaiszik, G. Ceder, K. Choudhary, G. Csanyi, E. D. Cubuk, B. Deng, R. Drautz, X. Fu, J. Godwin, V. Honavar, O. Isayev, A. Johansson, B. Kozinsky, S. Martiniani, S. P. Ong, I. Poltavsky, K. Schmidt, S. Takamoto, A. P. Thompson, J. Westermayr and B. M. Wood, *Curr. Opin. Solid State Mater. Sci.*, 2025, **35**, 101214.
- Q. Zhu and S. Hattori, *Digital Discovery*, 2025, **4**, 120–134.
- M. M. Mukta, R. Perriot, S. Hattori, W. Zhou and Q. Zhu, *RSC Adv.*, 2026, **16**, 7221–7229.
- R. E. Dinnebier, M. Pink, J. Sieler and P. W. Stephens, *Inorg. Chem.*, 1997, **36**, 3398–3401.
- X. Fu, A. Xu, J. B. Pedersen, S. Li, R. Sažinas, Y. Zhou, S. Z. Andersen, M. Saccoccio, N. H. Deissler, J. B. V. Mygind, J. Kibsgaard, P. C. K. Vesborg, J. K. Nørskov and I. Chorkendorff, *Nat. Commun.*, 2024, **15**, 2417.
- Y. Yu, T. He, A. Wu, Q. Pei, A. Karkamkar, T. Autrey and P. Chen, *Angew. Chem., Int. Ed.*, 2019, **58**, 3102–3107.
- Z. Kang, S. Wang, G. Wu, S. Chen, Z. Zheng, W. Wang, X. Du, H. Li, M. Zhu, H. Peng and Y. Gao, *J. Am. Chem. Soc.*, 2025, **147**, 30591–30598.
- 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, *arXiv*, 2025, preprint, arXiv:2506.23971, DOI: [10.48550/arXiv.2506.23971](https://doi.org/10.48550/arXiv.2506.23971).
- B. Rhodes, S. Vandenhoute, V. Simkus, J. Gin, J. Godwin, T. Duignan and M. Neumann, *arXiv*, 2025, preprint, arXiv:2504.06231, DOI: [10.48550/arXiv.2504.06231](https://doi.org/10.48550/arXiv.2504.06231).
- A. R. Oganov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 244704.
- A. O. Lyakhov, A. R. Oganov, H. T. Stokes and Q. Zhu, *Comput. Phys. Commun.*, 2013, **184**, 1172–1182.

