Probing the limits of the Universal Models for Atoms: energetic and structural analysis of polyoxometalates

Hugo Salazar-Lozas[1], Jordi buils[1], Carles Bo[1], Albert Solé Daura[1]
[1]Institute of Chemical Research of Catalonia, Av. Països Catalans, 16, Tarragona, 43007, Spain
hsalazar@iciq.es

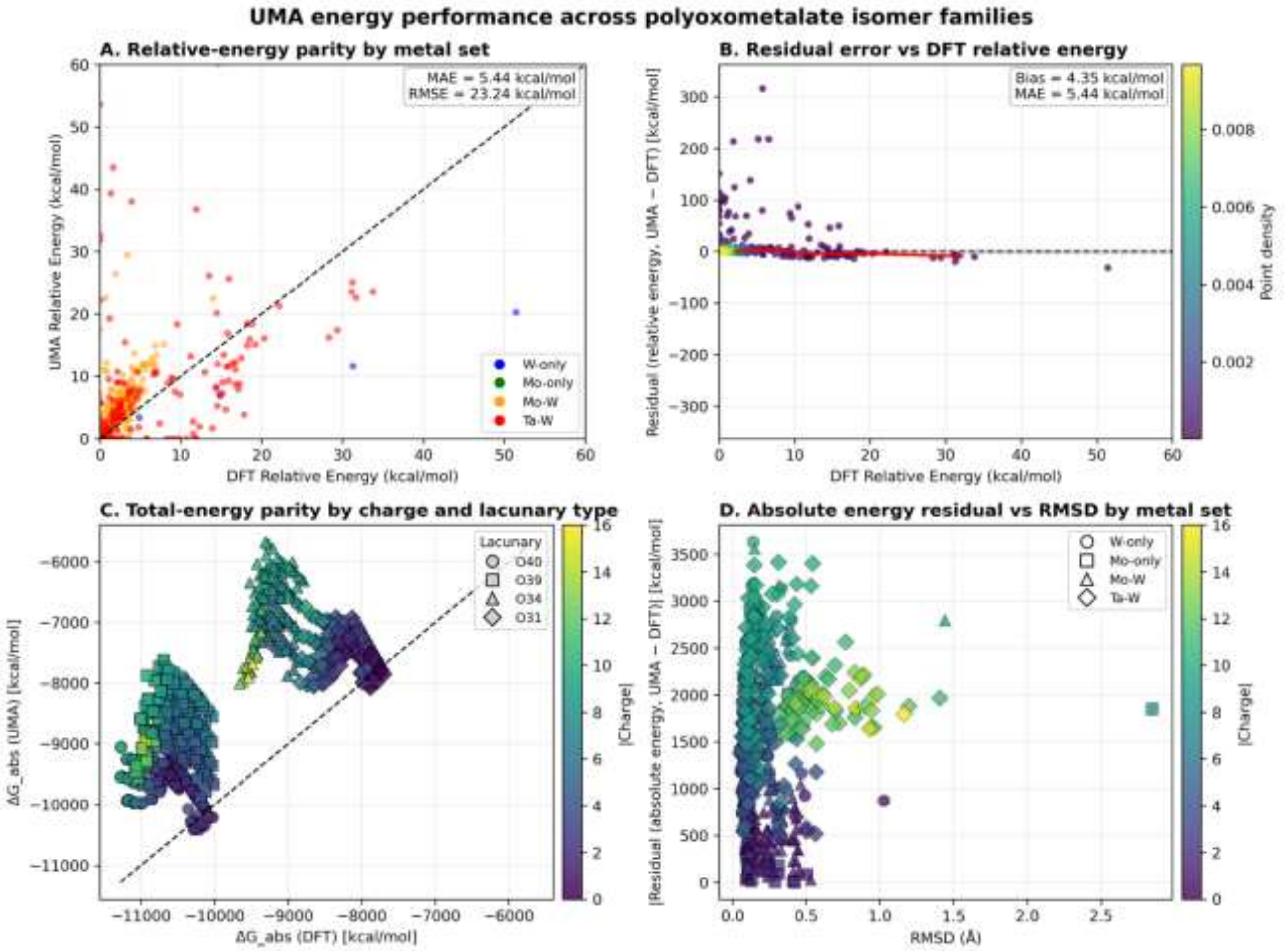
Machine-learning interatomic potentials (MLIPs) have rapidly emerged as powerful tools for accelerating atomistic simulations while retaining near–first-principles accuracy [1–4]; however, their reliability in chemically complex systems that lie outside the dominant training distribution remains an open challenge [4]. In this work, we evaluate the performance of the Universal Model for Atoms (UMA) potential [6], a universal MLIP, in polyoxometalates (POMs) [5], a challenging class of highly charged metal–oxo clusters characterized by structural diversity, strong electrostatic interactions, and complex electronic structure. We first benchmark UMA against density functional theory (DFT) calculations performed with ADF and Gaussian16 for the five canonical Keggin isomers (α, β, γ, δ, and ε), comparing total energies, relative energies, structural deviations (RMSD), and computational cost. The results show that UMA accurately reproduces structural features and captures the main energetic trends while reducing computational time by orders of magnitude. We then extend the analysis to a large dataset of POMs including different central atoms (As, P, Si, and V), metal compositions (W, Mo, Ta, and mixed-metal systems), and structural motifs (Keggin, mono-lacunary, and tri-lacunary clusters). Despite the limited representation of these elements and motifs in the UMA training dataset, the model preserves the relative-energy ordering of isomers across most POM families, successfully identifying low-energy structures in a large fraction of cases. In contrast, deviations in absolute energies increase for highly charged systems and mixed-metal compositions, reflecting the importance of electrostatic effects and implicit solvation present in the DFT reference but absent in UMA. Overall, our results demonstrate that UMA remains a robust and efficient MLIP for ranking relative energies in complex inorganic systems, even in out-of-distribution regimes, providing new insights into the applicability limits of universal machine-learning potentials and highlighting their potential for accelerating the exploration of large polyoxometalate configurational spaces.

References

[1] Behler, J.; Parrinello, M., Phys. Rev. Lett., 98 (2007) 146401.
[2] Bartók, A. P. et al., Phys. Rev. Lett., 104 (2010) 136403.
[3] Batzner, S. et al., Nat. Commun., 13 (2022) 2453.
[4] Unke, O. T. et al., Chem. Rev., 121 (2021) 10142–10186.
[5] Long, D.-L.; Burkholder, E.; Cronin, L., Angew. Chem. Int. Ed., 49 (2010) 1736–1758.
[6] Universal Models for Atoms, arXiv:2506.23971v2 (2025).
[7] C. Bo, G. Dias Da Silva, F. Hosseini, A. R. Mullukkandy, H. Salazar-Lozas, A. Sole-Daura, and Y. Wang, DOI: 10.26434/chemrxiv.15001161/v1 (2026).

Figures

Figure 1. Performance of UMA for polyoxometalate isomers.
(A) Relative-energy parity plot showing that UMA preserves the energetic ordering across different metal compositions.
(B) Residual error versus DFT relative energy, highlighting systematic deviations.
(C) Absolute-energy parity colored by cluster charge and lacunary type, revealing increasing errors for highly charged systems.
(D) Absolute energy residual versus RMSD, indicating that energy errors are dominated by charge rather than structural deviations.


UMA energy performance across polyoxometalate isomer families
A. Relative-energy parity by metal set
MAE = 5.44 kcal/mol
RMSE = 23.24 kcal/mol
UMA Relative Energy (kcal/mol)
DFT Relative Energy (kcal/mol)
0
10
20
30
40
50
60
W-only
Mo-only
Mo-W
Ta-W
B. Residual error vs DFT relative energy
Bias = 4.35 kcal/mol
MAE = 5.44 kcal/mol
Residual (relative energy, UMA − DFT) [kcal/mol]
DFT Relative Energy (kcal/mol)
300
200
100
0
−100
−200
−300
Point density
0.008
0.006
0.004
0.002
C. Total-energy parity by charge and lacunary type
Lacunary
O40
O39
O34
O31
ΔG_abs (UMA) [kcal/mol]
ΔG_abs (DFT) [kcal/mol]
−6000
−7000
−8000
−9000
−10000
−11000
|Charge|
16
14
12
10
8
6
4
2
0
D. Absolute energy residual vs RMSD by metal set
W-only
Mo-only
Mo-W
Ta-W
|Residual (absolute energy, UMA − DFT)| [kcal/mol]
RMSD (Å)
3500
3000
2500
2000
1500
1000
500
0.0
0.5
1.0
1.5
2.0
2.5
|Charge|