

# The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models

Daniel S. Levine<sup>1,\*</sup>, Muhammed Shuaibi<sup>1,\*</sup>, Evan Walter Clark Spotte-Smith<sup>2</sup>, Michael G. Taylor<sup>3</sup>, Muhammad R. Hasyim<sup>4</sup>, Kyle Michel<sup>1</sup>, Ilyes Batatia<sup>5</sup>, Gábor Csányi<sup>5</sup>, Misko Dzamba<sup>1</sup>, Peter Eastman<sup>6</sup>, Nathan C. Frey<sup>7</sup>, Xiang Fu<sup>1</sup>, Vahe Gharakhanyan<sup>1</sup>, Aditi S. Krishnapriyan<sup>8,9,10</sup>, Joshua A. Rackers<sup>7</sup>, Sanjeev Raja<sup>8</sup>, Ammar Rizvi<sup>1</sup>, Andrew S. Rosen<sup>11</sup>, Zachary Ulissi<sup>1</sup>, Santiago Vargas<sup>12,13</sup>, C. Lawrence Zitnick<sup>1,†</sup>, Samuel M. Blau<sup>13,†</sup>, Brandon M. Wood<sup>1</sup>,

<sup>1</sup>FAIR at Meta, <sup>2</sup>Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, <sup>3</sup>Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA, <sup>4</sup>Simons Center for Computational Physical Chemistry, New York University, New York, NY, USA,

<sup>5</sup>Engineering Laboratory, University of Cambridge, Cambridge, UK, <sup>6</sup>Department of Chemistry, Stanford University, Stanford, CA, USA, <sup>7</sup>Prescient Design, Genentech, New York, NY, USA,

<sup>8</sup>Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA, <sup>9</sup>Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA, <sup>10</sup>Applied Mathematics and Computational Research, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, <sup>11</sup>Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, USA, <sup>12</sup>Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, <sup>13</sup>Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

\*Co-first Author, †Co-corresponding Author

Machine learning (ML) models hold the promise of transforming atomic simulations by delivering quantum chemical accuracy at a fraction of the computational cost. Realization of this potential would enable high-throughput, high-accuracy molecular screening campaigns to explore vast regions of chemical space and facilitate *ab initio*-level simulations at sizes and time scales that were previously inaccessible. However, a fundamental challenge to creating ML models that perform well across molecular chemistry is the lack of comprehensive data for training. Despite substantial efforts in data generation, no large-scale molecular dataset exists that combines broad chemical diversity with a high level of accuracy. To address this gap, Meta FAIR introduces Open Molecules 2025 (OMol25), a large-scale dataset composed of more than 100 million density functional theory (DFT) calculations at the *ωB97M-V/def2-TZVPD* level of theory, representing billions of CPU core-hours of compute. OMol25 uniquely blends elemental, chemical, and structural diversity including: 83 elements, a wide-range of intra- and intermolecular interactions, explicit solvation, variable charge/spin, conformers, and reactive structures. There are ~83M unique molecular systems in OMol25 covering small molecules, biomolecules, metal complexes, and electrolytes, including structures obtained from existing datasets. OMol25 also greatly expands on the size of systems typically included in DFT datasets, with systems of up to 350 atoms. In addition to the public release of the data, we provide baseline models and a comprehensive set of model evaluations to encourage community engagement in developing the next-generation ML models for molecular chemistry.

**Dataset:** <https://huggingface.co/facebook/OMol25>

**Models:** <https://huggingface.co/facebook/OMol25>

**Code:** <https://github.com/facebookresearch/fairchem>

**Correspondence:** C.L.Z. ([zitnick@meta.com](mailto:zitnick@meta.com)), S.M.B. ([smbbau@lbl.gov](mailto:smbbau@lbl.gov)), B.M.W. ([bmwood@meta.com](mailto:bmwood@meta.com))

## 1 Introduction

Molecular chemistry is a cornerstone of modern society, driving innovation in areas such as medicine [1, 2], energy production [3, 4] and storage [5, 6], advanced computing [7, 8], agriculture [9, 10], and more. Further progress in these areas hinges on the design and discovery of molecular systems with improved or novel

properties [11]. Computational chemistry has become an essential tool to aid in this pursuit [12]. It offers a means to efficiently explore the molecular design space and provide mechanistic insights, both of which are difficult to achieve with experimentation alone [13–17].

Over the past few decades, Density Functional Theory (DFT) has become the prevailing quantum chemistry modeling tool [18–20] due to its ability to model complex atomic interactions very generally with a reasonable compromise between accuracy and computational efficiency. However, its computational complexity remains a major limitation, which scales approximately cubically with the number of electrons. This inhibits routine, large-scale screening campaigns and severely limits its use for long time-scale simulations or structures containing more than a few hundred atoms.

Recently, Machine Learning Interatomic Potentials (MLIPs) that act as DFT surrogates have emerged as a means to approach the accuracy of DFT at a small fraction of the required computation [21, 22]. This progress has been driven in part by modeling innovation, moving from small descriptor-based neural networks or kernel methods to much larger graph neural networks [23–30], which currently represent the state of the art. As models increase in size and accuracy, further advances in learning general representations depend on having access to large-scale, high-quality data.

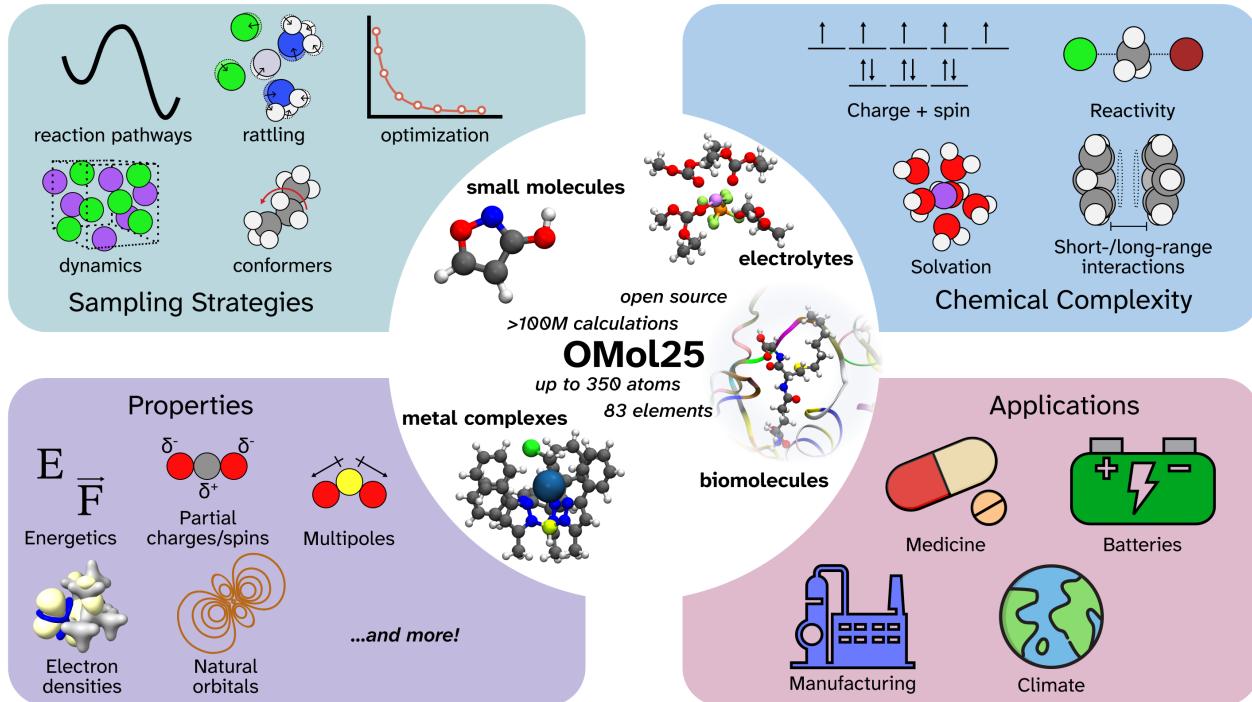
Small-scale ( $< \mathcal{O}(1M)$ ) datasets, such as MD-17 [31] and QM9 [32], helped launch the field but are limited to a few atom types (e.g., C, H, O, N, and F) and have narrow chemical diversity. More recently, there have been a number of  $\mathcal{O}(1 - 10M)$  dataset efforts that expand chemical and structural diversity and, to a lesser extent, elemental diversity [33–40]. However, the vast majority of calculations involve only charge neutral, isolated organic molecules with a relatively small number of atoms ( $N < 50$ ). An ideal molecular dataset would contain a mix of system sizes that have high elemental, chemical, and structural diversity with awareness of charge and spin [41]. The reason no such large-scale  $\mathcal{O}(100M)$  molecular dataset exists is the prohibitive computational expense of DFT.

Beyond models and training data, building useful evaluations is another important component of driving improvements in MLIPs. In the past, evaluations have focused on total energy, forces, and structure metrics such as the mean absolute error (MAE) computed on random in-distribution splits of the data. While these metrics can be informative, they fail to capture whether a model is actually useful for practical chemistry applications. Additional metrics are needed to determine whether models respect certain physical properties such as energy conservation [30, 42, 43], as well as generalize to out-of-distribution data. Thorough and well-motivated assessments of MLIPs on practically-relevant, domain-informed tasks would help to clarify for the community where current deficiencies lie and where sufficient accuracy has already been achieved. There have been recent efforts in this area for molecular systems [44], but more complete efforts are needed.

In this paper, we present the Open Molecules 2025 (OMol25) Dataset, a large-scale resource for training molecular chemistry machine learning models. OMol25 comprises over 100 million DFT single-point calculations containing up to 350 atoms at a high level of DFT theory ( $\omega$ B97M-V/def2-TZVPD) [45–47]. OMol25 draws from diverse chemistry disciplines including biochemistry, electrochemistry, and organic and inorganic chemistry with all of the first 83 elements represented. The dataset provides a wide sampling of chemical complexity, including systems with varying charge and spin states, explicit solvation, reactivity, and various intermolecular interactions. Structural diversity is incorporated through a variety of sampling techniques, such as classical and MLIP-based molecular dynamics (MD) and conformer sampling. Additionally, we recompute a number of existing datasets to ensure a consistent level of DFT theory across previous efforts. Beyond the dataset, we introduce a series of evaluation tasks that represent common objectives in computational chemistry, such as conformer ranking, ionization energies, spin gaps, and distance scaling. We evaluate baseline models trained on OMol25 on these tasks to provide insight on where future model development should focus. We provide all of the OMol25 data with a CC BY 4.0 license, and model weights with a commercially permissive license (with some geographic and acceptable use restrictions). In the near future, we will also release a public leaderboard to inspire innovation on ML models for molecular chemistry.

## 2 Open Molecules 2025 Dataset

For ML models to act as DFT surrogates across the broad field of molecular chemistry, we must ensure the training dataset encapsulates the behavior of atoms across the field’s numerous individual subdisciplines. To



**Figure 1** Overview of OMol25, including chemical scope, sampling strategies used to construct structures, chemical phenomena we seek to capture, properties available for each datapoint, and envisioned application areas.

accomplish this, we explore three major domains within the Open Molecule 2025 dataset: **biomolecules**, **metal complexes**, and **electrolytes**; we additionally re-evaluate existing **community** datasets and derivatives thereof. Biomolecules are focused on proteins, DNA, and RNA. Metal complexes feature monometallic transition metal (TM), main group metal, and lanthanide systems with diverse ligands. The electrolytes domain includes collections of multiple molecules, often charged, and their interactions with solvent molecules. Finally, the community domain includes various datasets and derivatives that are generally focused on organic chemistry.

Within each of these domains, we sought to ensure coverage of several important chemical concepts: the effect of charge and spin, the role of molecular conformations, and the propensity for reactivity. As a result of these different domains and cross-cutting considerations, there are no fewer than 12 conceptually distinct methods used to generate input data. A thorough explanation of all of the methods employed is given in the Appendix and the code used for generation will be made available on GitHub. In this section, we shall only summarize the methods and describe the resulting dataset.

## 2.1 Biomolecules

Accurately capturing the interactions of biomolecules is critical for applications in drug design, both of small molecules and biologics, and biochemistry research generally (*e.g.*, computationally studying a protein’s mechanism of action). To enable these diverse applications, we include a broad set of protein–ligand, protein–protein, nucleic acid–nucleic acid, and protein–nucleic acid interactions.

Annotations for ligands (both small molecules and metal ions), DNA, and RNA bound to proteins were taken from the BioLiP2 database of protein-ligand interactions [48]. To obtain inputs of suitable size for DFT, fragments of these large macromolecular systems are extracted. This is accomplished by pulling out the immediate protein pocket environment of the experimental PDB structure around a small-molecule ligand, metal, or nucleic acid residue (and potentially additional nearby residues). To these extracted pockets, hydrogens are added to generate various protonation/tautomeric states of the protein and ligands, and protein and nucleic acid residues are appropriately capped. Molecular dynamics (MD) simulation of the extracted system is then performed with restraints on the protein and nucleic acid backbones to prevent the fragment from falling apart now that they are not held together by the rest of the protein or nucleic acid structure.

This workflow utilizes the Schrödinger Software Suite [49] for manipulating and elaborating these biomolecular systems.

Additional protein-ligand structures were generated by docking random drug-like molecules from the GEOM [39], ChEMBL [50], and ZINC20 [51] databases into the above-extracted protein pockets using smina [52] and simulating them by the same MD procedure as above. Only the two closest chains to the ligand are retained, as opposed to the entire pocket.

Protein–protein interactions are sampled by extracting the environments around buried residues (as determined by the per residue relative solvent accessible surface area [53]) and residues at the interfaces of protein subunits (according to the DIPS-Plus database [54]). These extracted protein–protein fragments were similarly prepared, capped, and run through the MD protocol described above. Nucleic acid–nucleic acid interactions were probed by extracting collections of nearby residues in various topologies with an analogous procedure. In addition to the protein–nucleic acid structures we obtain from BioLiP2, the Nucleic Acid Knowledge Base [55] was mined for non-traditional nucleic acid structures, such as A-form and Z-form DNA, triplex systems, and Holliday junctions [56].

## 2.2 Metal Complexes

Metal complexes, including organometallic species, Werner coordination complexes, and any well-defined molecular system in which a set of ligands stabilizes one or more metal centers, are critically important for catalysis, energy harvesting, and manufacturing [57]. The metals in these structures, which span the periodic table, present several challenges relative to organic chemistry. Metals have a much wider array of configurational flexibility, from strongly enforced bond angles to loose, complexly-varying arrangements around the metal center. Metal–ligand bonds tend to break more easily than main-group bonds, which allows for variability in the number of bonded ligands. Finally, metals have much more flexibility when it comes to electronic structure. Many metals support multiple oxidation states with each potentially having profoundly different behaviors and chemistries.

### 2.2.1 Ground State Structures

In order to sample the diverse space of metal complexes, we employed the Architector package [58]. Architector takes in a specification of the metal center (*e.g.*, element, oxidation state, coordination number) and a set of ligands and generates 3D structures of the requested metal center coordinated by the ligands in multiple geometries and conformations. We leveraged a curated list of experimentally used ligands [59] and metal–oxidation state pairs from the mendeleev package [60]. In this way, an extremely large number of complexes can be generated by randomly assembling collections of ligands to attach to randomly chosen metal centers. The structures which are output by Architector have been optimized by xTB [61] during and after generation and can be used directly in DFT calculations. The spin state for each system was set to the maximum that would be reasonable for the given metal oxidation state (*e.g.*, a  $d^8$  center would be a triplet). Optimizations with no more than five steps were carried out on these complexes. The TM complex inputs were also run as single points in the lowest possible spin state (singlets for even-number electron systems, doublets of odd-number) and a subset of these were also optimized with a maximum of five steps.

### 2.2.2 Metal Reactivity

Diverse samples of reactive metal complex structures were generated by taking existing datasets of metal complex reactivity MOR41 [62], ROST61 [63], and MOBH35 [64] and swapping the identity of metals and ligands using Architector. The atoms of these reactant and product complexes were renumbered to bring them into correspondence and a reaction path was generated using the Artificial Force Induced Reaction (AFIR) scheme [65, 66]. In AFIR, the atoms in bonds that are breaking are pushed apart from each other while the atoms in the bonds that are forming are pushed toward each other with a fictitious, constant force. A geometry optimization is run with progressively higher force constants until the reaction occurs. The points along this optimization path form a reasonable guess for the minimum energy path from reactant to product. This path was subsampled to generate DFT inputs of metal complexes in reactive geometries using the highest spin configuration.

## 2.3 Electrolytes

Electrolytes, solutions containing ions and other additives, are vital components of batteries, play a central role in biological and geochemical processes, and are essential to electrochemical manufacturing. In this work, we consider electrolytes broadly defined to include aqueous and non-aqueous solutions, ionic liquids, and molten salts. The solvation of molecular and ionic electrolytes, and the presence of other electrolyte species, profoundly affects molecular stability and structure, particularly by stabilizing highly charged groups. This creates a challenging problem for MLIPs seeking to predict the subtle forces governing intermolecular interactions, particularly due to the presence of varying local charge.

### 2.3.1 Molecular Dynamics-Based Sampling

We employ MD to simulate a diverse sampling of electrolyte structures in large periodic boxes. The initial structures are created using the Desmond MD package [67] along with Schrödinger's Disordered System Builder. After equilibration, the structures were simulated with NPT MD using the OPLS4 [68] force field at a range of temperatures and concentrations. From these simulations, a set of frames were collected and the environment around every ion was extracted. The ion's environment included all molecules where any atom of that molecule was within some fixed cutoff radius of the atom(s) of the central ion. In this work, we used both 3Å and 5Å for the cutoffs, which corresponds roughly to the primary and secondary coordination shells, respectively. These extracted clusters are subsampled to obtain a set of clusters that vary significantly in composition and geometry. A similar procedure is employed around solvent molecules, except that we require no solute to be present in the extracted solvent-centered conformers, i.e. they are solvent-only clusters. Clusters from the 3Å cutoff were optimized for up to five steps, while clusters from the 5Å cutoff were run as single-points.

As electrolytes are used in batteries and subject to high electric fields, electrons are frequently pulled out of or pushed onto clusters of the kinds described here. We therefore also compute a random sample of systems with an either an electron added or removed from the cluster.

Different intermolecular interactions have different scaling behavior with distance and these effects must be captured by an MLIP. A random sample of clusters were dilated by random amounts (changing all intermolecular distances, but keeping intramolecular distances fixed), spreading the molecules apart or contracting them together.

In purely classical MD, many possible configurations may not be sampled. To account for this, we utilize Ring Polymer Molecular Dynamics (RPMD) that includes the use of nuclear quantum effects (NQE) to increase the diversity of the configurations. We used OpenMM's [69] RPMD implementation to simulate another batch of electrolytes which were skewed to contain more light atoms that are more affected by NQE. The same extraction procedure as above was employed to obtain DFT input structures for single point calculations.

At interfaces, solvents can form different structures than in bulk systems due to the absence of, for example, hydrogen bonding partners on the other side of the interface. We apply a spherical restraining potential around a large droplet of electrolyte to create a gas-solvent interface. This droplet is sampled via the same strategy as above to create a set of interfacial structures for DFT single points.

Given water's importance as a solvent, additional large clusters of up to 70 water molecules were also included. Water was simulated with the AMOEBA [70–72] force field and snapshots were used as DFT single point inputs.

### 2.3.2 Small Molecule Generation

Electrolyte research often involves the development of new ions and additives that may stabilize battery chemistry or prevent degradation. In order to sample from diverse electrolyte-like molecules, an array of 77 electrolyte “core” structures and 240 functional groups were curated and these cores and functional groups were randomly chained together to create novel small molecules. Ions and solvent molecules were placed around these new species with the Architect package [58]. These systems were optimized for up to five steps.

### 2.3.3 Electrolyte Reactivity

Electrolyte reactions were taken from previous work on reaction networks and mechanistic studies of electrolyte decomposition and solid-electrolyte interphase formation [14, 73–76]. Coordinates for all of these reaction templates were generated if they did not already exist and random metals were swapped with the metals present in the templates to generate a library of reactants and products undergoing electrolyte-type reactions in the presence of various ions. The Popcornn [77] method was used to generate reaction paths between reactant and product; these paths were subsampled to generate DFT inputs.

## 2.4 Community

### 2.4.1 Existing Community Datasets

Numerous molecular datasets have been previously released [31–38, 78] with varying levels of theory. As part of Open Molecules 2025, we have recomputed several of the most widely used datasets to upgrade them to a consistent and higher level of theory and to fill in missing data, such as forces. The datasets calculated were ANI-2X [33], Transition-1X [34], ANI-1xBB [35], Orbnnet Denali [36], SPICE2 [37, 38], and Solvated Protein Fragments [78]. We also recomputed approximately 30% of the GEOM [39] dataset. The GEOM systems were optimized, with a fraction having their initial positions randomly perturbed. The Transition-1X dataset is a database of reactive trajectories and was recomputed in the UKS formalism.

### 2.4.2 Interpolated Reactivity Datasets

To increase the number of reactivity samples within OMol25, we utilize several reactivity datasets. Robustly sampling the reactivity of even simple systems can be quite challenging. Feasible reactions often require atoms to be carefully arranged, while remote moieties can strongly influence energetics. Moreover, the vast majority of reactions in the chemical literature describe overall chemical transformations which consist of several steps, rather than the elementary reaction steps that trace the movement of atoms. To sample structures containing reactivity, we use the RGD1 dataset [79] of reactant-transition state-product triples, and the PMechDB [80] and RMechDB [81] databases of elementary reaction steps of polar and radical reactions, respectively. For RGD1, we carry out an interpolation in internal coordinate space from reactant to transition state to product, creating a series of frames that are likely near the minimum energy path. For RMechDB and PMechDB, we generate 3D structures, assign equivalent atom-numbering with the Schrödinger Suite, and then use the AFIR procedure as described in Section 2.2.2. All of these DFT inputs were run in the UKS formalism.

## 2.5 ML-Based MD

In order to increase the structural diversity of the dataset and discover areas where the ML models may need additional data to avoid pathological behavior, we undertook ML-based MD of the types of inputs used in the first three domains above. Three EquiformerV2 models [28] were trained on approximately half of the data from each of the three domains. Additional Architector metal complexes, periodic boxes of electrolytes, and protein–protein interface clusters were prepared as described above. The EqV2 models and the MACE-MP0 model [82] were used to simulate short MD trajectories which were randomly subsampled to create ML-MD-based DFT inputs. For metal complexes, a small amount of data was obtained by rattling the atomic positions according to a Boltzmann distribution as was done in Open Materials 2024 [83].

## 2.6 Calculation Details

In selecting the level of theory for Open Molecules 2025, we sought to create a long-lasting dataset using the highest quality settings that were possible given the computational resources available. The DFT functional selected, the range-separated hybrid meta-GGA  $\omega$ B97M-V [45], has consistently been shown by various authors to be one of, if not the most, accurate functional for a broad array of quantum chemistry tasks [20, 84]. Only double-hybrid functionals have been consistently shown to outperform it, at prohibitive computational cost. Because the dataset contains anions, a basis set with diffuse functions was required; the triple-zeta def2-TZVPD basis set [85] was thus selected as the Ahlrichs basis sets are well-optimized for use with DFT, with effective core potentials (ECPs) [86] that allow support for elements 1–83. We computed all singlet

systems which contain transition metal and lanthanide complexes or where bonds are expected to be breaking in the Unrestricted Kohn-Sham (UKS) formalism and rotate by 20° between the HOMO and LUMO in the  $\beta$  space in order to break spin symmetry in the initial guess. Non-singlet systems were also run in UKS.

Calculations were carried out with the ORCA 6.0.0 DFT package [87, 88], using both RI-J [89] and COSX [90] integral acceleration techniques and tight convergence settings. Benchmarking of grid settings (both the exchange-correlation grid and the COSX grid) indicated that typical grids led to small numerical inconsistencies between energy and forces due to grid incompleteness. These errors were significant on the scale of errors with state-of-the-art MLIPs. In order to achieve sufficiently tight consistency, ORCA’s DEFGRID3 offered the best trade-off of convergence and cost. Calculations were only considered in OMol25 if they met several quality control and error checks; details are described in Appendix A.1.

The released dataset required 6 billion CPU core-hours in total. Nearly all calculations were run on Elastic Compute within Meta’s private cloud [91], utilizing a heterogeneous pool of servers that can be preempted at any time to accommodate increased demand on higher-priority workloads. Although this presents challenges at the software level, especially around handling of partially complete calculations when hosts are reclaimed, it allowed for the creation of this dataset on servers that would otherwise have been sitting idle.

## 2.7 Dataset Profile

The final Open Molecules 2025 dataset contains more than 100 million DFT calculations. The number of atoms ranges from 2 to 350, with 50 on average. Charges vary from -10 to +10 and spin multiplicity varies from 1 to 11. Owing to the wide range of system sizes in the dataset, we report the breakdowns of each domain by the total number of atoms rather than the total number of systems. There are approximately 1.2B atoms in each of Biomolecules and Metal Complexes, 1.4B in Community, and 2.0B in Electrolytes. The breakdown of each domain into the subdomains discussed above is given in Figure 3. In addition to energy and force data, the dataset includes a variety of partial charge and spin schemes, orbital energies, Fock matrices, densities, and more as described in Appendix A.2.

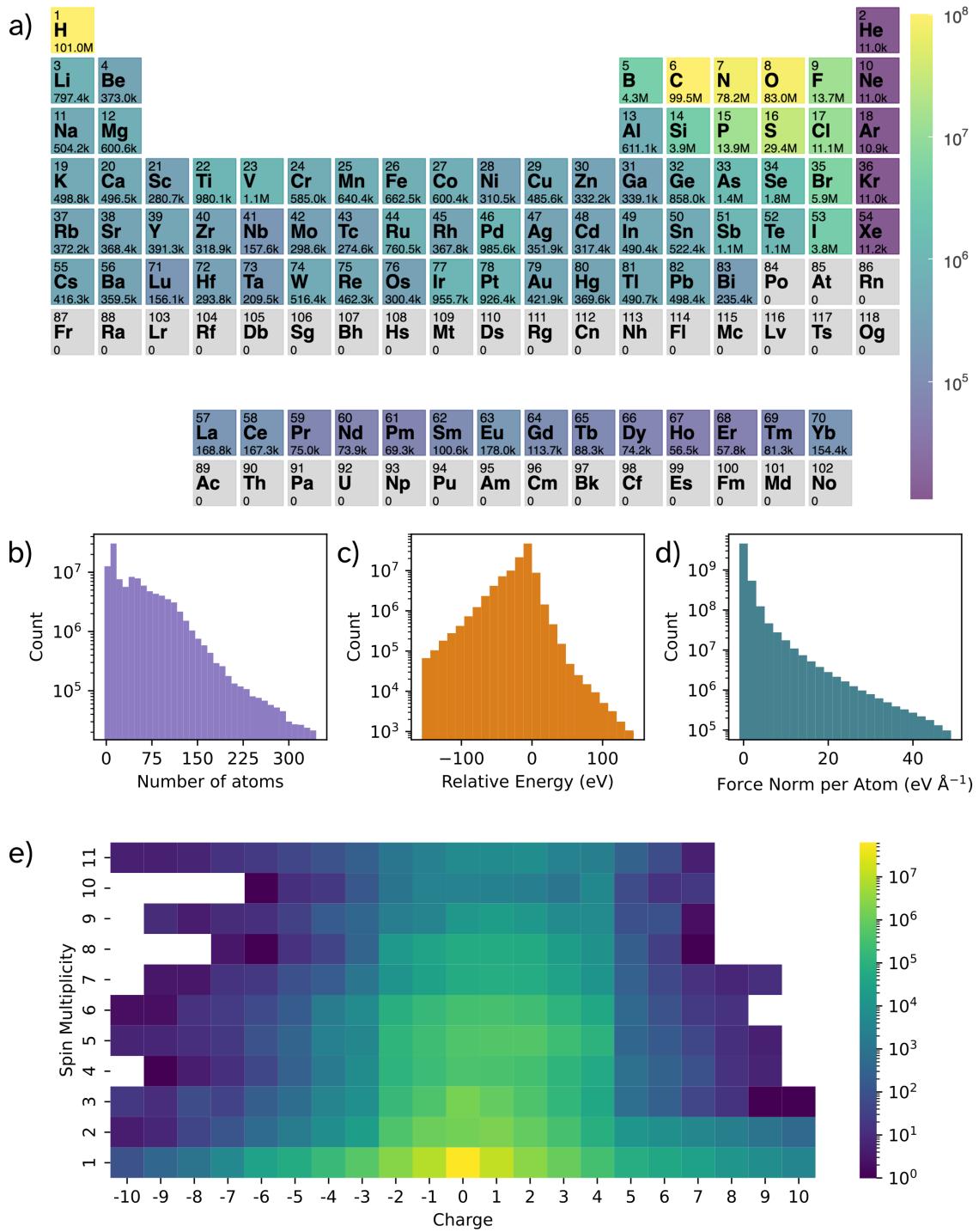
**Table 1** Size of the OMol25 train, validation, and test splits.

	Split	Size	Description
Train	All	101,666,280	Full training set
	4M	3,986,754	Uniform ~4M subset
	Neutral	34,335,828	Charge neutral, singlet subset
Val	Comp	2,762,021	Out-of-distribution compositions
Test	Comp	2,805,046	Out-of-distribution compositions
	M-Lig	39,615	Unique metal-ligand pairs
	PDB-TM	26,614	Metal-containing protein structures
	Reactivity	64,898	Held-out organic and metal-complex reactions
	COD	76,156	Experimental crystal structures
	Anions	16,488	Unique anion structures

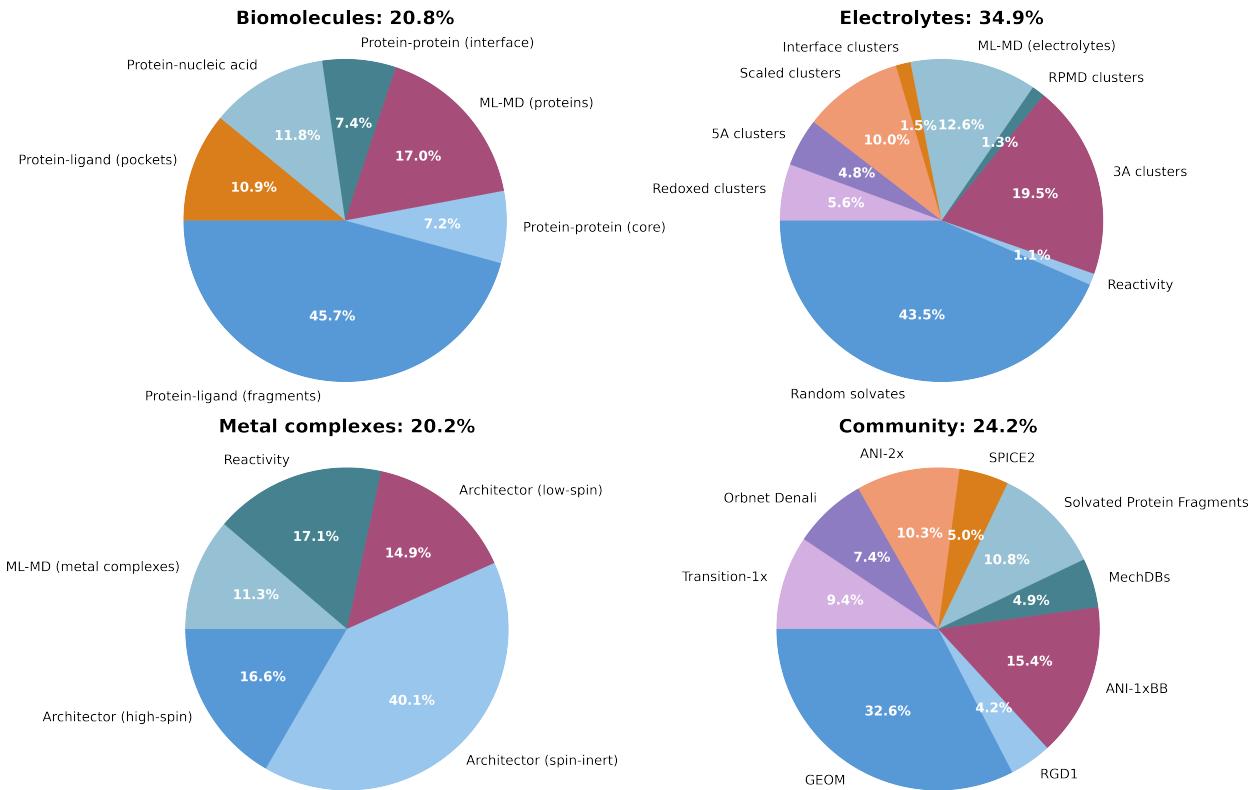
## 2.8 OMol25 Training, Validation, and Test Splits

The OMol25 dataset is divided into training, validation, and test splits to ensure consistent evaluations within the community. Splits are created based on compositions (*i.e.*, molecular formula). After enumerating all unique compositions among the computed data, we hold out ~2.5% each for a validation and test set. We create three dataset sizes - the full OMol25 dataset (“All”), a much smaller ~4M split (“4M”), and a charge-neutral, singlet split (“Neutral”). The 4M split is sampled uniformly across the electrolytes, metal complexes, biomolecules, reactivity (RMechDB, PMechDB, and ANI-1xBB), and community domains. The Neutral split corresponds to only the charge-neutral singlets from a subset of the community domain - ANI-2X, Orbnnet Denali, SPICE2, GEOM, Transition-1X, and RGD1. This split is intended to measure the performance of models on datasets the community is familiar with, without worrying about the complexity of charge and spin. All training and validations splits are made publicly available.

To test generalizability beyond just out-of-distribution (OOD) compositions, we also generate several explicit OOD splits. Dataset splits are summarized in Table 1.



**Figure 2** OMol25 dataset composition. a) Periodic table heat map, showing the number of snapshots in the training set containing a given element. b) Histogram in log scale for the number of training set snapshots with a given number of atoms. c) Histogram in log scale of training set snapshots with a given energy value relative to atomic references. d) Histogram in log scale of atomic force norms in the training set. e) Heat map for number of training set snapshots of different charge:spin.



**Figure 3** OMol25 breakdown by domain and sampling strategy by number of atoms. The percentage next to each domain indicates its share of the total dataset’s number of atoms.

### 2.8.1 Out-of-Distribution Splits

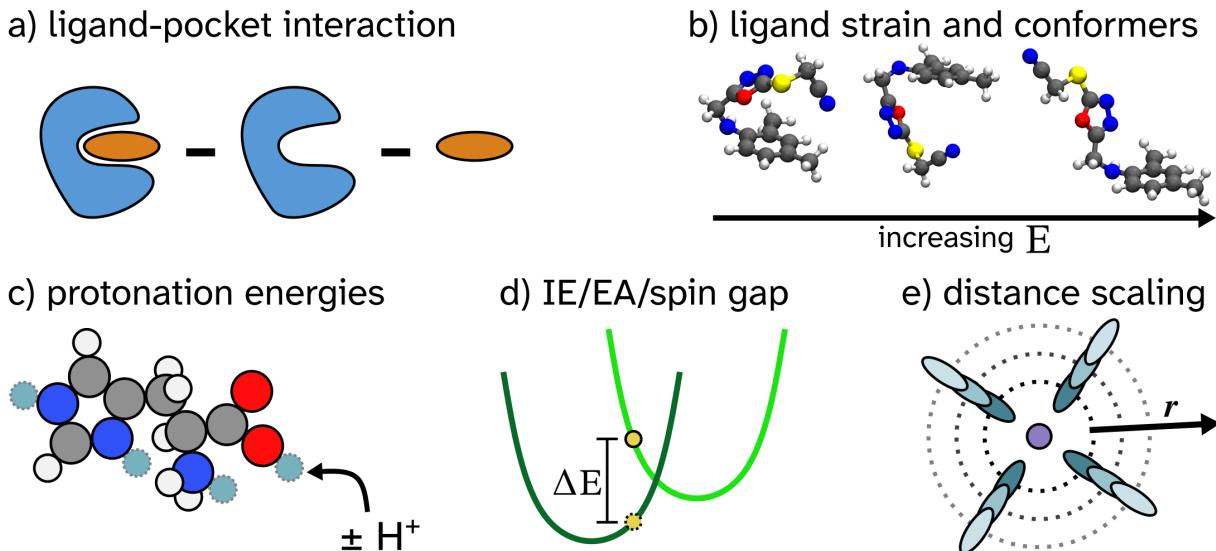
**Metal-Ligand Bonds** As a test of the generalizability of metal-ligand interactions, we randomly sampled 50 metal-ligand (M-L) bond combination (e.g. Pt-OAc) to hold out from training. This validation set consists of 39,615 complexes which contain one or more of the OOD M-L bonds.

**PDB-TM** OMol25 explicitly excludes metal-containing protein structures from the main training split so that it can be used to test the learning transfer from metal complex and electrolytes data. These include single ions coordinated to amino acids (e.g. the Zn of a zinc finger), metal complexes (e.g. the Fe of a heme group), and multimetallic clusters (e.g. 4Fe-4S clusters).

**Reactivity** Out-of-distribution organic reactivity data is taken from the work of Grambow, Pattanaik, and Green [92]. More specifically, 1782 of their reactions are not found in either Transition-1x or RGD1. These reactant, transition state, product triples were elaborated into reaction paths as was done in Section 2.4.2. Additionally, several of the reaction templates in Section 2.2.2 were incompatible with the automated ligand swapping methodology, and thus were held out of training data generation. After metal swapping, these templates yielded 1436 reactions, which were then subjected to the AFIR procedure described in Section 2.2.2.

**COD** To assess the sufficiency of OMol25’s generated metal complexes for describing real-world systems, we draw from the Crystallography Open Database (COD) [93] of experimental crystal structures of metal complexes. This replicates the typical starting point of real-world computational projects involving metal complexes. In this case, both the highest and lowest reasonable spin states were used for first-row TMs, only low-spin states for second and third row TMs, and only high-spin states for lanthanide complexes.

**Anions** Battery research often involves the development of new electrolytes. To assess the ability of models to describe new types of electrolyte systems, two additional electrolyte anions were simulated according the same procedure described in Section D.1. While additional non-OOD anions, cations, and solvents appear in the resulting clusters, all of them also contain one or more of these new anions.



**Figure 4** OMol25 evaluations. a) Ligand-pocket interaction energy/force, as defined by the energy/force difference between the ligand-pocket complex and the isolated ligand and isolated pocket. b) Ligand strain energy and conformer optimization/ordering, both of which involve a global optimization where many conformers of a structure are all subjected to a tightly converged geometry optimization. c) Relative protonation energies, as defined by the energy difference between optimized structures of different protonation states. d) Unoptimized ionization energy (IE) / electron affinity (EA) / spin gap, as defined by the energy difference between static structures of varying charge and/or spin multiplicity. e) Distance scaling, which seeks to predict the energy difference between instances of the same structure containing multiple molecular components, with the inter-component distance scaled by some factor.

Additional splits including out-of-distribution cations/solvents, TorsionNet500 [94], and the Wiggle150 [44] benchmark are described in the Appendix I.4.

### 3 Evaluations

In order to evaluate the capacity for models trained on OMol25 to capture the breadth of chemistry that the training data seeks to cover, we developed a range of evaluation tasks, as summarized in Figure 4. Each task is carried out on at least 1,000 different structures to achieve robust statistics.

#### 3.1 Protein-ligand Interaction Energy and Forces

Protein-ligand binding is central to biological processes and is the mechanistic underpinning of many life-saving treatments. It is highly desirable that models trained on OMol25 be able to accurately capture the physics of protein-ligand binding in realistic environments. Since calculating a true binding free energy is beyond the scope of an MLIP evaluation, requiring an appropriate sampling procedure, we use ligand-pocket interaction energy ( $E_{\text{ligand+pocket}} - E_{\text{ligand}} - E_{\text{pocket}}$ ) MAE and ligand-pocket interaction forces ( $\vec{F}_{\text{ligand+pocket}} - \vec{F}_{\text{ligand}} - \vec{F}_{\text{pocket}}$ ) MAE as a proxy for the binding free energy task. A full free energy procedure would be required to accurately predict interaction energies at each time step (Fig. 4a). Further details are provided in Appendix H.1.

#### 3.2 Ligand Strain

Another recently established evaluation task for atomistic simulation of protein-ligand binding is the ligand strain energy [95], defined as the energy difference between the local minimum corresponding to the bioactive conformation of a ligand and the ligand’s global minimum energy. Details of how we obtain the local minimum of the bioactive geometry are provided in Appendix H.2. The global minimum is identified via tightly converged geometry optimizations on structures generated by CREST [96], RDKit [97], and MacroModel [98–100] run

on the original bioactive conformation. Evaluation metrics are then the strain energy MAE and the RMSD between the DFT global minimum structure and the MLIP global minimum structure.

### 3.3 Conformers

Molecules with many torsional degrees of freedom and configurationally flexible metal complexes can have a large number of local minima on their potential energy surface, and it is frequently of critical importance to recover the lowest energy conformer(s) whose properties dominate the thermal ensemble (Fig. 4b). To evaluate model capacity in this context, families of conformers are optimized to very tight convergence with both DFT and the MLIP. We then map the DFT-optimized to the MLIP-optimized conformers via RMSD-based linear sum assignment, which finds the correspondence between DFT and MLIP structures with lowest total RMSD. Using this mapping, the total RMSD and a Boltzmann-weighted version are used as evaluation metrics (see Appendix H.3 for more details). We further compare the  $\Delta E$  between each conformer and the global minimum conformer versus the equivalent value from the MLIP, yielding a  $\Delta E$  MAE evaluation metric.

Additionally, we use the MLIP to re-optimize DFT-optimized conformer structures to probe how far MLIP minima are from DFT minima, yielding a reoptimization RMSD evaluation metric. We also calculate MLIP  $\Delta E$  values between each conformer and the DFT global minimum conformer at the MLIP-reoptimized structures, which can be compared to the DFT  $\Delta E$  values at the DFT-optimized structures, yielding a  $\Delta E_{reopt}$  MAE evaluation metric.

### 3.4 Protonation Energies

Protonation state plays a central role in many chemical transformations in biological, environmental, and industrial processes. Since a rigorous treatment of acidity/basicity in solution (*i.e.*, pKa prediction) is beyond the scope of MLIP evaluation, we opt for a proxy evaluation task of protonation energies (Fig. 4c). These protonation energies are used as part of DFT-based pKa prediction workflows [101, 102]. We perform geometry optimization on pairs of structures that differ by one proton and then calculate the  $\Delta E$  between the two structures, yielding  $\Delta E$  MAE, and the RMSD between MLIP-optimized and ORCA-optimized structures as evaluation metrics. As with conformers, we further re-optimize DFT-optimized structures, yielding reoptimization RMSD and  $\Delta E_{reopt}$  MAE evaluation metrics. Note that this evaluation is only possible for MLIPs which can simulate species of different total charge. Further details are provided in Appendix H.4.

### 3.5 Unoptimized IE/EA and Spin Gap

The addition, removal, or transfer of electrons are central to enzymatic, catalytic, electrochemical, and other redox processes. An MLIP which faithfully describes such processes would be able to predict energy and force differences between two electronic states of the same atomic configuration (Fig. 4d). The  $\Delta E$  between two such states are the well-known vertical ionization energy (IE) and electron affinity (EA), while  $\Delta F$  captures the differences in the potential energy surface between the two states. We therefore evaluate  $\Delta E$  MAE,  $\Delta \vec{F}$  MAE, and  $\Delta \vec{F}$  cosine similarity between principal and charge-modified structures. Further details are provided in Appendix H.5.

Of similar importance is a model’s ability to accurately predict the energy and force on different spin surfaces. Accurately simulating systems containing open-shell d-block metals often requires determining which is the lowest energy spin state, which may change in the course of a reaction, and the energy gap between spin states can play a critical role in properties of molecular optical devices and photoactive catalysts [103–105]. Our evaluation metrics are thus  $\Delta E$  MAE,  $\Delta \vec{F}$  MAE, and  $\Delta \vec{F}$  cosine similarity between structures simulated at their highest viable spin multiplicity and each possible lower spin multiplicity. Further details are provided in Appendix H.5.

### 3.6 Distance Scaling: Short-range and Long-range Interactions

Different intermolecular interactions have distinct scaling behavior with distance (*e.g.*,  $1/r$  for charge-charge interactions,  $1/r^3$  for dipole-dipole interactions,  $1/r^6$  for dispersion forces). It is important that an MLIP smoothly and accurately captures both the correct magnitude and scaling behavior to make valid predictions

on observables such as phase changes, ion/mass transport, density, viscosity, etc. We probe this feature by scaling the distance between components of molecular clusters and complexes. Many MLIP models employ a distance cutoff beyond which two atoms do not communicate directly. We therefore split this evaluation into a “short-range” and “long-range” regime around 6Å of separation (typical of many model’s cutoff radius). We thus compute a distance scaling scan and evaluate  $\Delta E$  and  $\Delta \vec{F}$  between the points on that scan and the lowest energy structure along the scan that lies in the short-range regime. Further details are provided in Appendix H.6.

**Table 2** Summary of metrics for each evaluation task.

Evaluation Task	Metrics
Protein-ligand	Interaction Energy MAE
	Interaction Forces MAE
Ligand Strain	Strain Energy MAE
	Global Minimum RMSD
Conformers	Ensemble RMSD
	Boltzmann-weighted ensemble RMSD
	$\Delta E$ MAE
	Reoptimization RMSD
Protonation	$\Delta E_{reopt}$ MAE
	RMSD
	$\Delta E$ MAE
	Reoptimization RMSD
IE/EA	$\Delta E_{reopt}$ MAE
	$\Delta E$ MAE
	$\Delta \vec{F}$ MAE
Spin Gap	Cosine Similarity
	$\Delta E$ MAE
	$\Delta \vec{F}$ MAE
Distance Scaling	Cosine Similarity
	SR and LR $\Delta E$
	SR and LR $\Delta \vec{F}$

## 4 Baseline Models

We present a set of baseline results to aid the community in evaluating the performance of current state-of-the-art models on the OMol25 validation and test sets as well as the OMol25 evaluations. The current<sup>1</sup> set of baseline models includes eSEN [30], GemNet-OC [106], and MACE [27]. While these models are by no means an exhaustive survey of the field, they serve to provide a representative overview of the capabilities of current equivariant and invariant models.

All the baseline MLIPs are message-passing graph neural networks where nodes represent atoms and edges represent interactions with neighboring atoms. Typically, these models take atom types (*i.e.*, elements) and 3D atom positions as input and return energy as an output. Conserving models compute per-atom forces by taking the gradient of the energy with respect to the positions, while direct models predict the per-atom forces as an additional output of the model.

One necessary modification to baseline models is enabling the use of total charge and spin as inputs. Given the extent to which molecules with different charge and spin exist in OMol25 (see Figure 2) charge- and spin-aware models are a requirement when training on the full dataset. For example, there are numerous examples in the dataset where the same 3D structure has a different charge or spin. If the model is not given this information, it would incorrectly predict these structures to have the same energy. To accomplish this in baseline models, we introduce a simple combined charge and spin embedding based on the input total charge and spin, which is added to the node embeddings (more details can be found in Appendix I). We recognize that this may not be sufficient to describe how charge and spin might localize in a system and leave more complex schemes for incorporating charge and spin for future work.

Additionally, we train eSEN models of different sizes to understand how performance changes with increasing

<sup>1</sup>Additional baseline models will be included in the future

model capacity. Due to the scale of OMol25, we utilize a multi-step training procedure to improve the efficiency of medium and large models. The procedure involves first pre-training direct force models with lower precision [30], followed by a second stage at FP32. Small models are trained in both a direct and conserving configuration using full precision. GemNet-OC was trained using its default cutoff radius of 12Å, but also a configuration with 6Å, consistent with eSEN. MACE was trained only on the neutral split due to the publicly available model not supporting charge and spin. For training, energies are referenced using a modified heat of formation followed by a linear reference. Full training details and model hyperparameters can be found in Appendix I.

## 5 Results

### 5.1 Test

For all splits, we evaluate our models' performance on predicting a structure's energy and per-atom forces. The per-atom mean absolute error (MAE) is used as the primary evaluation metric, similar to previous work [83, 107, 108].

**Table 3** Structure to energy and force results across the different **test** splits. **Energy per atom** and force mean absolute error metrics are reported across the two different training splits - All and 4M. Total energy metrics are provided in Appendix I.2.

Dataset	Model	# of params	Test													
			Comp		M-Lig		PDB-TM		Reactivity		COD		Anions		Average	
			Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
All	eSEN-sm-d.	6.3M	1.55	9.09	1.18	17.15	1.06	15.31	3.63	52.21	3.50	23.45	1.01	11.37	1.99	21.43
	eSEN-sm-cons.	6.3M	1.35	7.39	1.02	14.66	0.83	12.72	2.85	44.25	2.51	20.56	0.55	9.23	1.52	18.14
	eSEN-md-d.	50.7M	0.85	4.29	0.68	8.79	0.61	9.06	2.28	32.21	2.13	14.30	0.63	5.40	<b>1.20</b>	<b>12.34</b>
	GemNet-OC	39.1M	0.57	5.85	0.74	12.55	0.66	11.38	2.82	41.98	2.47	18.42	0.61	7.20	1.31	16.23
	eSEN-lg-d.	690.4M	training in progress													
4M	eSEN-sm-d.	6.3M	2.23	11.86	1.62	21.63	1.67	18.70	4.84	64.56	5.02	30.18	1.68	15.83	2.84	27.13
	eSEN-sm-cons.	6.3M	2.04	9.89	1.28	18.86	1.13	15.93	3.59	54.17	3.47	27.10	1.02	12.76	2.09	23.12
	eSEN-md-d.	50.7M	1.33	5.99	0.99	12.02	1.05	12.21	3.34	43.88	3.11	19.01	1.01	8.23	<b>1.81</b>	<b>16.89</b>
	GemNet-OC-r6	39.1M	1.31	8.85	1.15	17.17	1.35	15.62	4.26	57.06	4.76	26.24	1.07	11.71	2.32	22.78
	GemNet-OC	39.1M	0.99	8.00	1.08	16.45	1.16	15.71	4.17	55.57	4.38	24.36	0.87	10.99	2.11	21.85

Energy(meV/atom), Forces(meV/Å)

Results for eSEN and GemNet-OC are given in Table 3 across the different test splits. Conserving models (cons.) outperform their direct (d.) counterparts across all splits and metrics. Larger models, such as eSEN-md, also outperform their smaller variants. GemNet-OC outperforms eSEN-sm across all splits, and is comparable to eSEN-md across most metrics, outperforming it only on the out-of-distribution compositions ("Comp") split. When we compare a GemNet-OC model trained with a smaller cutoff radius of 6Å (the same as used in eSEN models), we generally observe diminishing performance. Models trained on the full dataset exhibit 50% to 100% improved performance versus models trained on the 4M subset. Model trends between the All and 4M datasets, however, correlate very well, suggesting that model development can be confidently done on the subset. Of the test splits, Comp and metal ligand pairs ("M-Lig") have the lowest errors; with reactivity and COD the highest. Averaged across all splits, eSEN-md performs the best with an energy MAE of 1.20 meV/atom and 12.34 meV/Å forces MAE on the All set, and 1.81 meV/atom and 16.89 meV/Å, respectively, on the 4M set.

To measure model performance across the different domains, we evaluate the respective subsets within the out-of-distribution Comp test split. Results are given in Table 4. Across all models, biomolecules have the lowest per-atom energy errors, followed by electrolytes, neutral organics, and metal complexes being the most difficult. Similar trends are observed with forces, with biomolecules and electrolytes having lower errors than neutral organics and metal complexes. Total energy test results and validation results can be found in Appendix I.2 and I.3, respectively.

**Table 4** Out-of-distribution composition **test** results broken across biomolecules, electrolytes, metal complexes, and neutral organics. **Energy per atom** and force mean absolute error metrics are reported across the two different training splits - All and 4M. Total energy metrics are provided in Appendix 1.2.

		Test-Comp							
Dataset	Model	Biomolecules		Electrolytes		Metal Complexes		Neutral Organics	
		Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
All	eSEN-sm-d.	0.69	6.39	1.54	10.41	2.24	31.70	1.93	18.71
	eSEN-sm-cons.	0.60	4.64	1.34	9.09	1.96	27.23	1.36	16.41
	eSEN-md-d.	0.34	2.62	0.90	5.16	1.42	18.71	0.97	7.66
	GemNet-OC	0.15	3.95	0.64	6.53	1.54	23.98	1.50	14.52
	eSEN-lg-d.					training in progress			
4M	eSEN-sm-d.	0.91	8.25	2.39	13.84	3.04	38.77	3.31	27.43
	eSEN-sm-cons.	0.87	6.23	2.14	12.34	2.40	33.31	2.46	25.46
	eSEN-md-d.	0.47	3.40	1.54	7.54	2.13	25.46	1.95	12.52
	GemNet-OC-r6	0.40	5.92	1.71	10.29	2.40	31.98	3.15	23.85
	GemNet-OC	0.26	5.29	1.19	9.16	2.27	31.14	2.85	22.70

Energy(meV/atom), Forces(meV/Å)

To gauge model performance on a task more common in the literature, we evaluate eSEN and MACE on OMol25’s charge neutral, singlet split. Results are provided in Table 5, with metrics reported on the neutral subset of the validation set. eSEN outperforms MACE on both energy and force metrics.

**Table 5** Results trained and evaluated on OMol25’s **Neutral** split. Total energy, energy per atom, and force mean absolute error metrics are reported for eSEN-sm-d. and MACE. Although evaluated for a fixed step count, both models continue to converge, with numbers expected to continue improving.

Model	Neutral		
	Energy [meV]	Energy [meV/atom]	Forces [meV/Å]
eSEN-sm-d.	29.80	0.88	10.11
MACE	54.09	1.55	16.83

## 5.2 Evaluations

Evaluations are broken into two categories: optimization and single-point based tasks. Optimization results (ligand-strain, conformers, and protonation) are reported in Table 6. Single-point results (protein-ligand interaction, IE/EA, spin gap, and distance scaling) are reported in Table 7. Evaluations requiring an optimization used Sella [109] to ensure consistency with the underlying DFT calculated.

**Ligand strain** evaluation are in very good agreement with DFT for all of the methods investigated here, with the largest errors of only 7.68 meV for the eSEN-sm-d down to 3.69 meV for eSEN-md-d. Models trained on the 4M dataset were only worse by a few meV. RMSDs were less tightly converged to DFT, but still performed fairly well. The eSEN-sm-cons and eSEN-md-d performed similarly, with an RMSD of 0.26 and 0.27 Å, respectively.

**Conformers** RMSDs were essentially the same for Boltzmann-weighted vs. ensemble metrics, indicating that lower energy structures were treated equally well as higher energy ones in the evaluation. Energy differences were also all well within chemical accuracy ( $\sim 1$  kcal/mol or 43 meV); the best performing models were eSEN-sm-cons and eSEN-md-d, which produced errors in energy gaps of 5.6 and 5.1 meV, respectively. While there was significant erosion in performance for models trained on the 4M dataset, they also were well within chemical accuracy. When starting from converged DFT minima, RMSD and  $\Delta E$  errors were even smaller, as is to be expected. For models trained on the All split, essentially no structural change is observed (the max RMSD, for the eSEN-sm-d and GemNet-OC models, was 0.06) and  $\Delta E$  error is less than 3.5 meV in all cases except the eSEN-sm-d model, which produced a slightly larger error of 5.69 meV.

**Table 6 Optimization Evaluations** including ligand-strain, conformer prediction, and protonation states. Results are reported across a variety of energy and structure based metrics for each task.

Dataset	Model	Ligand strain			Conformers				Protonation			
		Strain energy MAE /meV/ $\downarrow$	RMSD min./ $\text{\AA}$ / $\downarrow$	RMSD ensemble / $\text{\AA}$ / $\downarrow$	RMSD boltz./ $\text{\AA}$ / $\downarrow$	$\Delta$ Energy MAE /meV/ $\downarrow$	RMSD reopt./ $\text{\AA}$ / $\downarrow$	$\Delta$ Energy reopt./meV/ $\downarrow$	RMSD $\text{\AA}$ / $\downarrow$	$\Delta$ Energy MAE /meV/ $\downarrow$	RMSD reopt./ $\text{\AA}$ / $\downarrow$	$\Delta$ Energy reopt./meV/ $\downarrow$
All	eSEN-sm-d.	7.68	0.50	0.15	0.13	12.83	0.06	5.69	0.25	50.81	0.08	23.31
	eSEN-sm-cons.	4.52	0.26	0.06	0.05	5.60	0.02	3.08	0.12	51.77	0.03	15.02
	eSEN-md-d.	3.69	0.27	0.05	0.05	5.10	0.01	2.13	0.11	26.14	0.02	9.31
	GemNet-OC	4.93	0.39	0.14	0.12	9.73	0.06	3.31	0.20	38.81	0.07	11.01
	eSEN-lg-d.					in progress						
4M	eSEN-sm-d.	10.92	0.59	0.20	0.18	18.34	0.11	9.11	0.31	74.82	0.13	41.48
	eSEN-sm-cons.	6.87	0.35	0.08	0.07	8.35	0.04	4.74	0.17	59.07	0.06	31.99
	eSEN-md-d.	4.95	0.31	0.07	0.06	6.80	0.02	3.32	0.15	33.12	0.03	17.94
	GemNet-OC-r6	10.80	0.48	0.19	0.17	15.09	0.09	5.42	0.28	52.24	0.10	21.59
	GemNet-OC	6.68	0.45	0.17	0.15	13.36	0.08	4.87	0.25	55.76	0.09	18.43

**Protonation** proved to be the hardest of the optimization-based evaluations. RMSD errors were about twice as large in most cases as with conformers, though they remain 0.25 Å or better which indicates that the optimized structures are a very close match between the MLIP and DFT optimizations. Energy differences, on the other hand, ranged from 26 meV for eSEN-md-d to 51.77 meV for eSEN-sm-cons, in other words 5-10x worse than for conformers. While these results are still likely useful for assessing relative protonation state stability, it is clear this is more challenging than relative conformer stability. Evaluations that begin with the DFT optimized minima are again more closely aligned with the DFT results.

**Table 7 Singlepoint Evaluations** including protein-ligand, ionization energies/electron affinity (IE/EA), spin gap, and distance scaling. Results are reported across a variety of energy and force based metrics for each task.

Dataset	Model	Protein-ligand			IE/EA			Spin gap			Distance scaling		
		Ixn Energy MAE	Ixn Forces MAE	$\Delta$ Energy MAE	$\Delta$ Forces MAE	$\Delta$ Forces cosine sim.	$\Delta$ Energy MAE	$\Delta$ Forces MAE	$\Delta$ Forces cosine sim.	$\Delta$ Energy MAE	$\Delta$ Forces (SR) MAE	$\Delta$ Energy (LR) MAE	$\Delta$ Forces (LR) MAE
All	eSEN-sm-d.	144.94	4.64	385.37	79.05	0.75	596.88	106.75	0.70	28.62	3.05	206.68	2.73
	eSEN-sm-cons.	147.26	4.32	315.23	72.54	0.79	552.71	101.36	0.73	28.59	3.07	268.58	4.40
	eSEN-md-d.	64.25	2.16	264.08	56.44	0.82	463.92	84.17	0.77	19.00	2.08	167.35	2.54
	GemNet-OC	19.37	2.67	253.62	61.60	0.81	516.49	92.07	0.75	11.83	2.20	143.42	1.80
	eSEN-lg-d.					in progress							
4M	eSEN-sm-d.	196.52	5.67	491.63	94.02	0.69	786.65	121.17	0.64	35.69	3.71	205.19	3.05
	eSEN-sm-cons.	243.88	5.36	383.69	85.16	0.74	642.66	112.75	0.69	35.39	3.79	324.37	4.92
	eSEN-md-d.	98.34	2.57	378.71	72.62	0.76	623.74	103.07	0.71	26.44	2.75	166.41	2.87
	GemNet-OC-r6	96.04	4.35	442.75	81.67	0.73	720.77	112.37	0.67	25.57	3.13	127.16	2.93
	GemNet-OC	36.22	3.25	335.75	77.36	0.75	675.01	109.52	0.68	17.38	2.80	169.67	2.06

**Protein-ligand Interaction** evaluation metrics improve dramatically between eSEN-sm-d (144.9 meV interaction energy MAE, 4.64 meV/Å interaction forces MAE) and eSEN-md-d (64.3 meV interaction energy MAE, 2.16 meV/Å interaction forces MAE). Still further improvements in interaction energy are seen with GemNet-OC (19.4 meV). These improvements may imply that higher angular momentum representations are important for accurately capturing the interactions. While interaction forces MAE is comparable with biomolecules force MAE, the interaction energy MAE is significantly greater than biomolecules energy MAE, demonstrating the additional challenge inherent in specifically capturing intermolecular protein-ligand interactions.

**IE/EA** and **Spin Gap** evaluation metrics demonstrate the severe challenge presented by energy and force differences of metal-containing complexes with different charge and/or spin states. eSEN-md-d achieves 264.1 meV energy on IE/EA and 463.9 meV  $\Delta E$  MAE on spin gaps. Force errors were also notably higher by a large margin than any test set or evaluation metric. While  $\Delta E$  and  $\Delta \vec{F}$  errors do improve by nearly 50% on IE/EA between eSEN-sm-d and eSEN-md-d, they only improve by around 25% on spin gap, suggesting that more than just model scaling will be required to achieve sufficiently low error for scientific utility on these tasks.

**Distance Scaling** evaluation metrics demonstrate that capturing non-covalent interactions over different length scales was a significant challenge for current baseline models. While  $\Delta E$  and  $\Delta F$  errors are fairly small inside the cutoff distance (SR), they dramatically increase once systems are passed that cutoff (LR). While this is to be expected since none of these models contain any correction for long range interactions as are sometimes included in other models [110, 111], the fact that these models can display significant discontinuities in the PES at the cutoff is a serious issue that further work should address.

## 6 Outlook and future directions

OMol25 is the first dataset of its kind to span the breadth of major chemistry domains (inorganic, organic, biochemistry) at a high level of theory. However, we recognize that gaps in the coverage of chemical space still exist. For example, the OMol25 dataset does not include any of the radioactive elements after bismuth (*e.g.*, actinides), metal complexes with intermediate spin, or structures which might arise in polymer materials. Furthermore, the coverage of certain classes of materials such as lanthanides complexes, multimetallic structures, and solvated protonated organic molecules and metal complexes are relatively limited. Although baseline models trained on OMol25 may still be suitable for these applications, it has yet to be demonstrated. We invite the community to explore its capabilities to help inform future dataset development.

Baseline models trained on OMol25 have demonstrated great potential for MLIPs to be highly accurate across a wide range of chemical tasks. With models like eSEN-md achieving average energy and force errors as low as 1.2 meV/atom (71 meV) and 12.3 meV/Å, respectively, across all test splits; OMol25's combination of scale, diversity, quality, and complexity represents a significant leap in molecular DFT datasets for training MLIPs.

While baseline models have shown strong performance, approaching chemical accuracy ( $\sim 1\text{kcal/mol}$ ) especially in domains such as biomolecules and neutral organics; OMol25's evaluation tasks reveal significant gaps that need to be addressed. Notably, ionization energies/electron affinity, spin-gap, and long range scaling have errors as high as 200-500 meV. Architectural improvements around charge, spin, and long-range interactions are especially critical here. Partial charge and spin data were included in OMol25 to assist in these efforts. In the future, we hope to extend our evaluations to include free energy and reactivity tasks (for example, transition state or reaction path optimization). Such evaluations would require accurate geometry second derivatives (*i.e.*, Hessians), tests of which are currently absent from our MLIP evaluations, and may reveal new challenges to be overcome. Other properties computed in this dataset, such as multipole moments and electron densities, may enable new models which can also predict spectroscopic observables.

In order to foster community engagement and motivate rapid method development, we intend to release a public leaderboard in the near future to track the community's efforts on OMol25's evaluation tasks. We hope the OMol25 dataset will be a valuable resource to the community as we all seek to advance the state of the art. Whether OMol25 is leveraged for pre-training, specialized domain training (*e.g.* biomolecules), or fine-tuning, we are eager to see what challenges the community will address with this dataset.

## 7 Acknowledgments

M.G.T. acknowledges the support of the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (BES), Heavy Element Chemistry Program (KC0302031) under contract number E3M2. M.G.T. acknowledges support from the Laboratory Directed Research and Development program through the Institute for Materials Science (IMS) of Los Alamos National Laboratory (LANL) for his work on development of sampling methods for metal complexes. LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (contract no. 89233218CNA000001).

M.R.H. acknowledges support by a grant from the Simons Foundation (Grant 839534, MET) and the NYU IT High Performance Computing resources, services, and staff expertise.

I.B. and G.C. acknowledge the use of the AI computing and storage resources by GENCI at IDRIS thanks to the grant 2024-GC010815458 on the supercomputer Jean Zay's H100 partition.

P.E. acknowledges support from the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM140090.

A. S. K. and S. R. acknowledge support from the U.S. Department of Energy, Office of Science, Energy Earthshot initiatives as part of the Center for Ionomer-based Water Electrolysis at Lawrence Berkeley National Laboratory under Award Number DE-AC02-05CH11231.

A.S.R. acknowledges support from the Wilke 1989 School of Engineering and Applied Science Innovation Fund at Princeton University.

S.V. was supported by the DOE's National Nuclear Security Administration's Office of Defense Nuclear Nonproliferation Research and Development (NA-22) as part of the NextGen Nonproliferation Leadership Development Program.

S.M.B. acknowledges support from the Laboratory Directed Research and Development program of Lawrence Berkeley National Laboratory (LBNL), supported by the Office of Science, Office of Basic Energy Sciences (BES), of the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231. For his contributions to the electrolyte modeling portion of the dataset, S.M.B. acknowledges support from the Energy Storage Research Alliance (ESRA) (DE-AC02-06CH11357), an Energy Innovation Hub funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences.

## References

- [1] Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Medicinal Research Reviews* **1996**, *16*, 3–50.
- [2] Lombardino, J. G.; Lowe, J. A. The role of the medicinal chemist in drug discovery — then and now. *Nature Reviews Drug Discovery* **2004**, *3*, 853–862.
- [3] Huber, G. W.; Iborra, S.; Corma, A. Synthesis of Transportation Fuels from Biomass: Chemistry, Catalysts, and Engineering. *Chemical Reviews* **2006**, *106*, 4044–4098, PMID: 16967928.
- [4] Hagfeldt, A.; Grätzel, M. Molecular Photovoltaics. *Accounts of Chemical Research* **2000**, *33*, 269–277, PMID: 10813871.
- [5] Placke, T.; Kloepsch, R.; Dühnen, S.; Winter, M. Lithium ion, lithium metal, and alternative rechargeable battery technologies: the odyssey for high energy density. *Journal of Solid State Electrochemistry* **2017**, *21*, 1939–1964.
- [6] Miller, M. A.; Petrasch, J.; Randhir, K.; Rahmatian, N.; Klausner, J. In *Thermal, Mechanical, and Hybrid Chemical Energy Storage Systems*; Brun, K., Allison, T., Dennis, R., Eds.; Academic Press, 2021; pp 249–292.
- [7] Liu, L.; Liu, P.; Ga, L.; Ai, J. Advances in Applications of Molecular Logic Gates. *ACS Omega* **2021**, *6*, 30189–30204, PMID: 34805654.
- [8] Hinsberg, W. D.; Houle, F. A.; Sanchez, M. I.; Wallraff, G. M. Chemical and physical aspects of the post-exposure baking process used for positive-tone chemically amplified resists. *IBM Journal of Research and Development* **2001**, *45*, 667–682.
- [9] Yahaya, S. M.; Mahmud, A. A.; Abdullahi, M.; Haruna, A. Recent advances in the chemistry of nitrogen, phosphorus and potassium as fertilizers in soil: A review. *Pedosphere* **2023**, *33*, 385–406.
- [10] Leigh, G. J. In *Catalysts for Nitrogen Fixation: Nitrogenases, Relevant Chemical Models and Commercial Processes*; Smith, B. E., Richards, R. L., Newton, W. E., Eds.; Springer Netherlands: Dordrecht, 2004; pp 33–54.
- [11] Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365.
- [12] Head-Gordon, M.; Artacho, E. Chemistry on the computer. *Physics Today* **2008**, *61*, 58–63.
- [13] Garza, A. J.; Bell, A. T.; Head-Gordon, M. Mechanism of CO<sub>2</sub> Reduction at Copper Surfaces: Pathways to C<sub>2</sub> Products. *ACS Catalysis* **2018**, *8*, 1490–1499.
- [14] Spotte-Smith, E. W. C.; Kam, R. L.; Barter, D.; Xie, X.; Hou, T.; Dwaraknath, S.; Blau, S. M.; Persson, K. A. Toward a Mechanistic Model of Solid-Electrolyte Interphase Formation and Evolution in Lithium-Ion Batteries. *ACS Energy Letters* **2022**, *7*, 1446–1453.
- [15] Vaissier Welborn, V.; Head-Gordon, T. Computational Design of Synthetic Enzymes. *Chemical Reviews* **2019**, *119*, 6613–6630, PMID: 30277066.
- [16] Gómez-Bombarelli, R. et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials* **2016**, *15*, 1120–1127.
- [17] Qu, X.; Jain, A.; Rajput, N. N.; Cheng, L.; Zhang, Y.; Ong, S. P.; Brafman, M.; Maginn, E.; Curtiss, L. A.; Persson, K. A. The Electrolyte Genome project: A big data approach in battery materials discovery. *Computational Materials Science* **2015**, *103*, 56–67.
- [18] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- [19] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- [20] Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics* **2017**, *115*, 2315–2372, Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/00268976.2017.1333644>.
- [21] Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203.

- [22] Yao, K.; Herr, J. E.; Toth, D.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **2018**, *9*, 2261–2269.
- [23] Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- [24] Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **2017**, *30*.
- [25] Gasteiger, J.; Becker, F.; Günemann, S. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems* **2021**, *34*, 6790–6802.
- [26] 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 communications* **2022**, *13*, 1–11.
- [27] Batatia, I.; Kovacs, D. P.; Simm, G. N. C.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in Neural Information Processing Systems*. 2022.
- [28] Liao, Y.-L.; Wood, B.; Das, A.; Smidt, T. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. 2024; <http://arxiv.org/abs/2306.12059>, arXiv:2306.12059 [cs].
- [29] Qu, E.; Krishnapriyan, A. S. The Importance of Being Scalable: Improving the Speed and Accuracy of Neural Network Interatomic Potentials Across Chemical Domains. *Advances in Neural Information Processing Systems*. 2024; pp 139030–139053.
- [30] Fu, X.; Wood, B. M.; Barroso-Luque, L.; Levine, D. S.; Gao, M.; Dzamba, M.; Zitnick, C. L. Learning Smooth and Expressive Interatomic Potentials for Physical Property Prediction. 2025; <https://arxiv.org/abs/2502.12147>.
- [31] Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Science Advances* **2017**, *3*, e1603015.
- [32] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data* **2014**, *1*, 140022.
- [33] Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *Journal of Chemical Theory and Computation* **2020**, *16*, 4192–4202, Publisher: American Chemical Society.
- [34] Schreiner, M.; Bhowmik, A.; Vegge, T.; Busk, J.; Winther, O. Transition1x - a dataset for building generalizable reactive machine learning potentials. *Scientific Data* **2022**, *9*, 779, Publisher: Nature Publishing Group.
- [35] Zhang, S.; Zubatyuk, R.; Yang, Y.; Roitberg, A.; Isayev, O. ANI-1xBB: An ANI-Based Reactive Potential for Small Organic Molecules. *Journal of Chemical Theory and Computation* **2025**, Publisher: American Chemical Society.
- [36] Christensen, A. S.; Sirumalla, S. K.; Qiao, Z.; O'Connor, M. B.; Smith, D. G. A.; Ding, F.; Bygrave, P. J.; Anandkumar, A.; Welborn, M.; Manby, F. R.; Miller, T. F., III OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy. *The Journal of Chemical Physics* **2021**, *155*, 204103.
- [37] Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y.; De Fabritiis, G.; Markland, T. E. SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials. *Scientific Data* **2023**, *10*.
- [38] Eastman, P.; Pritchard, B. P.; Chodera, J. D.; Markland, T. E. Nutmeg and SPICE: Models and Data for Biomolecular Machine Learning. *Journal of Chemical Theory and Computation* **2024**, *20*, 8583–8593, Publisher: American Chemical Society.
- [39] Axelrod, S.; Gómez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data* **2022**, *9*, 185, Publisher: Nature Publishing Group.
- [40] Ganscha, S.; Unke, O. T.; Ahlin, D.; Maennel, H.; Kashubin, S.; Müller, K.-R. The QCML dataset, Quantum chemistry reference data from 33.5M DFT and 14.7B semi-empirical calculations. *Scientific Data* **2025**, *12*, 406.

- [41] Yuan, E. C.-Y.; Liu, Y.; Chen, J.; Zhong, P.; Raja, S.; Kreiman, T.; Vargas, S.; Xu, W.; Head-Gordon, M.; Yang, C.; others Foundation Models for Atomistic Simulation of Chemistry and Materials. *arXiv preprint arXiv:2503.10538* **2025**,
- [42] Bigi, F.; Langer, M.; Ceriotti, M. The dark side of the forces: assessing non-conservative force models for atomistic machine learning. 2025; <https://arxiv.org/abs/2412.11569>.
- [43] Amin, I.; Raja, S.; Krishnapriyan, A. S. Towards Fast, Specialized Machine Learning Force Fields: Distilling Foundation Models via Energy Hessians. The Thirteenth International Conference on Learning Representations. 2025.
- [44] Brew, R. R.; Nelson, I. A.; Binayeva, M.; Nayak, A. S.; Simmons, W. J.; Gair, J. J.; Wagen, C. C. Wiggle150: Benchmarking Density Functionals and Neural Network Potentials on Highly Strained Conformers. *Journal of Chemical Theory and Computation* **2025**, *21*, 3922–3929, Publisher: American Chemical Society.
- [45] Mardirossian, N.; Head-Gordon, M. omegaB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of Chemical Physics* **2016**, *144*, 214110.
- [46] Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of Chemical Physics* **2010**, *133*, 134105.
- [47] Hellweg, A.; Rappoport, D. Development of new auxiliary basis functions of the Karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC calculations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 1010–1017.
- [48] Zhang, C.; Zhang, X.; Freddolino, L.; Zhang, Y. BioLiP2: an updated structure database for biologically relevant ligand–protein interactions. *Nucleic Acids Research* **2023**, *52*, D404–D412.
- [49] Schrödinger Schrödinger Release 2024-2: Maestro. Schrödinger, LLC: New York, NY, 2024.
- [50] Mendez, D. et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic acids research* **2019**, *47*, D930–D940.
- [51] Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of Chemical Information and Modeling* **2020**, *60*, 6065–6073, Publisher: American Chemical Society.
- [52] Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. *Journal of Chemical Information and Modeling* **2013**, *53*, 1893–1904, Publisher: American Chemical Society.
- [53] Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. Interior and surface of monomeric proteins. *Journal of Molecular Biology* **1987**, *196*, 641–656.
- [54] Morehead, A.; Chen, C.; Sedova, A.; Cheng, J. DIPS-Plus: The enhanced database of interacting protein structures for interface prediction. *Scientific Data* **2023**, *10*, 509, Publisher: Nature Publishing Group.
- [55] Lawson, C. L.; Berman, H.; Chen, L.; Vallat, B.; Zirbel, C. The Nucleic Acid Knowledgebase: a new portal for 3D structural information about nucleic acids. *Nucleic Acids Research* **2024**, *52*, D245–D254.
- [56] Ortiz-Lombardía, M.; González, A.; Eritja, R.; Aymamí, J.; Azorín, F.; Coll, M. Crystal structure of a DNA Holliday junction. *Nature Structural Biology* **1999**, *6*, 913–917, Publisher: Nature Publishing Group.
- [57] (a) Nandy, A.; Duan, C.; Taylor, M. G.; Liu, F.; Steeves, A. H.; Kulik, H. J. Computational Discovery of Transition-metal Complexes: From High-throughput Screening to Machine Learning. *Chemical Reviews* **2021**, *121*, 9927–10000, Publisher: American Chemical Society; (b) Kinzel, N. W.; Werlé, C.; Leitner, W. Transition Metal Complexes as Catalysts for the Electroconversion of CO<sub>2</sub>: An Organometallic Perspective. *Angewandte Chemie International Edition* **2021**, *60*, 11628–11686, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202006988>; (c) Larsen, C. B.; Wenger, O. S. Photoredox Catalysis with Metal Complexes Made from Earth-Abundant Elements. *Chemistry – A European Journal* **2018**, *24*, 2039–2058, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/chem.201703602>; (d) Sun, L.; Yuan, G.; Gao, L.; Yang, J.; Chhowalla, M.; Gharahcheshmeh, M. H.; Gleason, K. K.; Choi, Y. S.; Hong, B. H.; Liu, Z. Chemical vapour deposition. *Nature Reviews Methods Primers* **2021**, *1*, 1–20, Publisher: Nature Publishing Group; (e) George, S. M. Atomic Layer Deposition: An Overview. *Chemical Reviews* **2010**, *110*, 111–131, Publisher: American Chemical Society.

- [58] Taylor, M. G.; Burrill, D. J.; Janssen, J.; Batista, E. R.; Perez, D.; Yang, P. Architector for high-throughput cross-periodic table 3D complex building. *Nature Communications* **2023**, *14*, 2786, Publisher: Nature Publishing Group.
- [59] Taylor, M. G.; Burrill, D. J.; Janssen, J.; Batista, E. R.; Perez, D.; Yang, P. Supplemental Data for Architector for high- throughput cross-periodic table 3D complex building. 2023; <https://doi.org/10.5281/zenodo.7764697>.
- [60] Mentel, L. mendeleev - A Python package with properties of chemical elements, ions, isotopes and methods to manipulate and visualize periodic table. 2021; <https://github.com/lmmentel/mendeleev>.
- [61] Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation* **2019**, *15*, 1652–1671, Publisher: American Chemical Society.
- [62] Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. *Journal of Chemical Theory and Computation* **2018**, *14*, 2596–2608.
- [63] Maurer, L. R.; Bursch, M.; Grimme, S.; Hansen, A. Assessing Density Functional Theory for Chemically Relevant Open-Shell Transition Metal Reactions. *Journal of Chemical Theory and Computation* **2021**, *17*, 6134–6151.
- [64] Semidalas, E.; Martin, J. M. The MOBH35 Metal–Organic Barrier Heights Reconsidered: Performance of Local-Orbital Coupled Cluster Approaches in Different Static Correlation Regimes. *Journal of Chemical Theory and Computation* **2022**, *18*, 883–898.
- [65] Maeda, S.; Harabuchi, Y.; Takagi, M.; Taketsugu, T.; Morokuma, K. Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces. *The Chemical Record* **2016**, *16*, 2232–2248.
- [66] Levine, D. S.; Jacobson, L. D.; Bochevarov, A. D. Large Computational Survey of Intrinsic Reactivity of Aromatic Carbon Atoms with Respect to a Model Aldehyde Oxidase. *Journal of Chemical Theory and Computation* **2023**, *19*, 9302–9317, Publisher: American Chemical Society.
- [67] SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing. 2006.
- [68] Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *Journal of Chemical Theory and Computation* **2021**, *17*, 4291–4300, Publisher: American Chemical Society.
- [69] Eastman, P. et al. OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials. *The Journal of Physical Chemistry B* **2023**, *128*, 109–116.
- [70] Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* **2013**, *9*, 4046–4063.
- [71] Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010**, *114*, 2549–2564.
- [72] Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B* **2003**, *107*, 5933–5947.
- [73] Xie, X.; Clark Spotte-Smith, E. W.; Wen, M.; Patel, H. D.; Blau, S. M.; Persson, K. A. Data-Driven Prediction of Formation Mechanisms of Lithium Ethylene Monocarbonate with an Automated Reaction Network. *Journal of the American Chemical Society* **2021**, *143*, 13245–13258.
- [74] Barter, D.; Clark Spotte-Smith, E. W.; Redkar, N. S.; Khanwale, A.; Dwaraknath, S.; Persson, K. A.; Blau, S. M. Predictive stochastic analysis of massive filter-based electrochemical reaction networks. *Digital Discovery* **2023**, *2*, 123–137.
- [75] Spotte-Smith, E. W. C.; Petrocelli, T. B.; Patel, H. D.; Blau, S. M.; Persson, K. A. Elementary Decomposition Mechanisms of Lithium Hexafluorophosphate in Battery Electrolytes and Interphases. *ACS Energy Letters* **2022**, *8*, 347–355.
- [76] Spotte-Smith, E. W. C.; Blau, S. M.; Barter, D.; Leon, N. J.; Hahn, N. T.; Redkar, N. S.; Zavadil, K. R.; Liao, C.; Persson, K. A. Chemical Reaction Networks Explain Gas Evolution Mechanisms in Mg-Ion Batteries. *Journal of the American Chemical Society* **2023**, *145*, 12181–12192.

- [77] Hegazy, K.; Kumar, A.; Yuan, .; Blau, S. Path Optimization with a Continuous Representation Neural Network for reaction path with machine learning potentials. <https://github.com/khegazy/Popcornn>, 2025.
- [78] Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *Journal of Chemical Theory and Computation* **2019**, *15*, 3678–3693, Publisher: American Chemical Society.
- [79] Zhao, Q.; Vaddadi, S. M.; Woulfe, M.; Ogunfowora, L. A.; Garimella, S. S.; Isayev, O.; Savoie, B. M. Comprehensive exploration of graphically defined reaction spaces. *Scientific Data* **2023**, *10*, 145, Publisher: Nature Publishing Group.
- [80] Tavakoli, M.; Miller, R. J.; Angel, M. C.; Pfeiffer, M. A.; Gutman, E. S.; Mood, A. D.; Van Vranken, D.; Baldi, P. PMechDB: A Public Database of Elementary Polar Reaction Steps. *Journal of Chemical Information and Modeling* **2024**, *64*, 1975–1983, Publisher: American Chemical Society.
- [81] Tavakoli, M.; Chiu, Y. T. T.; Baldi, P.; Carlton, A. M.; Van Vranken, D. RMechDB: A Public Database of Elementary Radical Reaction Steps. *Journal of Chemical Information and Modeling* **2023**, *63*, 1114–1123, Publisher: American Chemical Society.
- [82] Batatia, I. et al. A foundation model for atomistic materials chemistry. 2024; <https://arxiv.org/abs/2401.00096>.
- [83] Barroso-Luque, L.; Shuaibi, M.; Fu, X.; Wood, B. M.; Dzamba, M.; Gao, M.; Rizvi, A.; Zitnick, C. L.; Ulissi, Z. W. Open Materials 2024 (OMat24) Inorganic Materials Dataset and Models. 2024; <http://arxiv.org/abs/2410.12771>, arXiv:2410.12771 [cond-mat].
- [84] (a) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics* **2017**, *19*, 32184–32215, Publisher: The Royal Society of Chemistry; (b) Najibi, A.; Goerigk, L. The Nonlocal Kernel in van der Waals Density Functionals as an Additive Correction: An Extensive Analysis with Special Emphasis on the B97M-V and  $\omega$ B97M-V Approaches. *Journal of Chemical Theory and Computation* **2018**, *14*, 5725–5738, Publisher: American Chemical Society; (c) Santra, G.; Sylvetsky, N.; Martin, J. M. L. Minimally Empirical Double-Hybrid Functionals Trained against the GMTKN55 Database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *The Journal of Physical Chemistry A* **2019**, *123*, 5129–5143, Publisher: American Chemical Society; (d) Garrison, A. G.; Heras-Domingo, J.; Kitchin, J. R.; dos Passos Gomes, G.; Ulissi, Z. W.; Blau, S. M. Applying Large Graph Neural Networks to Predict Transition Metal Complex Energies Using the tmQM\_wB97MV Data Set. *Journal of Chemical Information and Modeling* **2023**, *63*, 7642–7654.
- [85] (a) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297; (b) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105; (c) Rappoport, D. Property-optimized Gaussian basis sets for lanthanides. *The Journal of Chemical Physics* **2021**, *155*, 124102.
- [86] (a) Dolg, M.; Stoll, H.; Savin, A.; Preuss, H. Energy-adjusted pseudopotentials for the rare earth elements. *Theor. Chim. Acta* **1989**, *75*, 173–194; (b) Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. *Theor. Chim. Acta* **1990**, *77*, 123–141; (c) Kaupp, M.; Schleyer, P. v. R.; Stoll, H.; Preuss, H. Pseudopotential approaches to Ca, Sr, and Ba hydrides. Why are some alkaline earth MX<sub>2</sub>compounds bent? *J. Chem. Phys.* **1991**, *94*, 1360–1366; (d) Dolg, M.; Stoll, H.; Preuss, H. A combination of quasirelativistic pseudopotential and ligand field calculations for lanthanoid compounds. *Theor. Chim. Acta* **1993**, *85*, 441–450; (e) Leininger, T.; Nicklass, A.; Küchle, W.; Stoll, H.; Dolg, M.; Bergner, A. The accuracy of the pseudopotential approximation: non-frozen-core effects for spectroscopic constants of alkali fluorides XF (X = K, Rb, Cs). *Chem. Phys. Lett.* **1996**, *255*, 274–280; (f) Metz, B.; Stoll, H.; Dolg, M. Small-core multiconfiguration-Dirac-Hartree-Fock-adjusted pseudopotentials for post-d main group elements: Application to PbH and PbO. *J. Chem. Phys.* **2000**, *113*, 2563–2569; (g) Metz, B.; Schweizer, M.; Stoll, H.; Dolg, M.; Liu, W. A small-core multiconfiguration Dirac-Hartree-Fock-adjusted pseudopotential for Tl - application to Tl X ( X = F, Cl, Br, I). *Theor. Chim. Acta* **2000**, *104*, 22–28; (h) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16-18 elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.
- [87] Neese, F. Software update: the ORCA program system, version 5.0. *WIREs Comput. Molec. Sci.* **2022**, *12*, e1606.

- [88] Neese, F. Software Update: The ORCA Program System—Version 6.0. *WIREs Computational Molecular Science* **2025**, *15*, e70019, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.70019>.
- [89] Neese, F. An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix. *J. Comp. Chem.* **2003**, *24*, 1740–1747.
- [90] Helmich-Paris, B.; de Souza, B.; Neese, F.; Izsák, R. An improved chain of spheres for exchange algorithm. *J. Chem. Phys.* **2021**, *155*, 104109.
- [91] Gupta, N. et al. Dynamic Idle Resource Leasing To Safely Oversubscribe Capacity At Meta. Proceedings of the 2024 ACM Symposium on Cloud Computing. New York, NY, USA, 2024; p 792–810.
- [92] Grambow, C. A.; Pattanaik, L.; Green, W. H. Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry. *Scientific Data* **2020**, *7*, 137, Publisher: Nature Publishing Group.
- [93] (a) Vaitkus, A.; Merkys, A.; Sander, T.; Quirós, M.; Thiessen, P. A.; Bolton, E. E.; Gražulis, S. A workflow for deriving chemical entities from crystallographic data and its application to the Crystallography Open Database. *Journal of Cheminformatics* **2023**, *15*, Publisher: Springer Science and Business Media LLC; (b) Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research* **2012**, *40*, D420–D427, \_eprint: <https://nar.oxfordjournals.org/content/40/D1/D420.full.pdf+html>; (c) Merkys, A.; Vaitkus, A.; Grybauskas, A.; Konovalovas, A.; Quirós, M.; Gražulis, S. Graph isomorphism-based algorithm for cross-checking chemical and crystallographic descriptions. *Journal of Cheminformatics* **2023**, *15*, Publisher: Springer Science and Business Media LLC; (d) Vaitkus, A.; Merkys, A.; Gražulis, S. Validation of the Crystallography Open Database using the Crystallographic Information Framework. *Journal of Applied Crystallography* **2021**, *54*, 661–672; (e) Merkys, A.; Vaitkus, A.; Butkus, J.; Okulič-Kazarinas, M.; Kairys, V.; Gražulis, S. COD::CIF::Parser: an error-correcting CIF parser for the Perl language. *Journal of Applied Crystallography* **2016**, *49*; (f) Gražulis, S.; Merkys, A.; Vaitkus, A.; Okulič-Kazarinas, M. Computing stoichiometric molecular composition from crystal structures. *Journal of Applied Crystallography* **2015**, *48*, 85–91.
- [94] Rai, B. K.; Sresht, V.; Yang, Q.; Unwalla, R.; Tu, M.; Mathiowitz, A. M.; Bakken, G. A. TorsionNet: A Deep Neural Network to Rapidly Predict Small-Molecule Torsional Energy Profiles with the Accuracy of Quantum Mechanics. *Journal of Chemical Information and Modeling* **2022**, *62*, 785–800, Publisher: American Chemical Society.
- [95] Wallace, E. R. S.; Frey, N. C.; Rackers, J. A. Strain Problems got you in a Twist? Try StrainRelief: A Quantum-Accurate Tool for Ligand Strain Calculations. 2025; <https://arxiv.org/abs/2503.13352>.
- [96] Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. CREST—A program for the exploration of low-energy molecular chemical space. *The Journal of Chemical Physics* **2024**, *160*, 114110.
- [97] Landrum, G. et al. rdkit/rdkit: 2025\_03\_2 (Q1 2025) Release. 2025; <https://zenodo.org/records/15286010>.
- [98] Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of Computational Chemistry* **1990**, *11*, 440–467.
- [99] Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. Macrocycle Conformational Sampling with MacroModel. *Journal of Chemical Information and Modeling* **2014**, *54*, 2680–2696.
- [100] Schrödinger, L. MacroModel.
- [101] (a) Cao, Y. et al. Quantum chemical package Jaguar: A survey of recent developments and unique features. *The Journal of Chemical Physics* **2024**, *161*, 052502; (b) Tang, H. et al. Discovery of a Novel Class of d-Amino Acid Oxidase Inhibitors Using the Schrödinger Computational Platform. *Journal of Medicinal Chemistry* **2022**, *65*, 6775–6802, Publisher: American Chemical Society.
- [102] Johnston, R. C.; Yao, K.; Kaplan, Z.; Chelliah, M.; Leswing, K.; Seekins, S.; Watts, S.; Calkins, D.; Chief Elk, J.; Jerome, S. V.; Repasky, M. P.; Shelley, J. C. Epik: pKa and Protonation State Prediction through Machine Learning. *Journal of Chemical Theory and Computation* **2023**, *19*, 2380–2388, Publisher: American Chemical Society.
- [103] Kumar, K. S.; Ruben, M. Sublimable Spin-Crossover Complexes: From Spin-State Switching to Molecular Devices. *Angewandte Chemie International Edition* **2021**, *60*, 7502–7521.

- [104] Sun, Z.; Lin, L.; He, J.; Ding, D.; Wang, T.; Li, J.; Li, M.; Liu, Y.; Li, Y.; Yuan, M.; Huang, B.; Li, H.; Sun, G. Regulating the Spin State of FeIII Enhances the Magnetic Effect of the Molecular Catalysis Mechanism. *Journal of the American Chemical Society* **2022**, *144*, 8204–8213, PMID: 35471968.
- [105] Senthil Kumar, K.; Ruben, M. Emerging trends in spin crossover (SCO) based functional materials and devices. *Coordination Chemistry Reviews* **2017**, *346*, 176–205, SI: 42 iccc, Brest– by invitation.
- [106] Gasteiger, J.; Shuaibi, M.; Sriram, A.; Günnemann, S.; Ulissi, Z.; Zitnick, C. L.; Das, A. GemNet-OC: Developing Graph Neural Networks for Large and Diverse Molecular Simulation Datasets. 2022; <https://arxiv.org/abs/2204.02782>.
- [107] Chanussot, L.; Das, A.; Goyal, S.; Lavril, T.; Shuaibi, M.; Riviere, M.; Tran, K.; Heras-Domingo, J.; Ho, C.; Hu, W.; others Open catalyst 2020 (OC20) dataset and community challenges. *Acs Catalysis* **2021**, *11*, 6059–6072.
- [108] Sriram, A.; Choi, S.; Yu, X.; Brabson, L. M.; Das, A.; Ulissi, Z.; Uyttendaele, M.; Medford, A. J.; Sholl, D. S. The Open DAC 2023 dataset and challenges for sorbent discovery in direct air capture. 2024.
- [109] Hermes, E. D.; Sargsyan, K.; Najm, H. N.; Zádor, J. Sella, an open-source automation-friendly molecular saddle point optimizer. *Journal of Chemical Theory and Computation* **2022**, *18*, 6974–6988.
- [110] Gong, S. et al. A predictive machine learning force field framework for liquid electrolyte development. 2025; <http://arxiv.org/abs/2404.07181>, arXiv:2404.07181 [cond-mat].
- [111] Anstine, D. M.; Zubatyuk, R.; Isayev, O. AIMNet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs. *Chemical Science* **2025**, Publisher: The Royal Society of Chemistry.
- [112] Kovács, D. P.; Moore, J. H.; Browning, N. J.; Batatia, I.; Horton, J. T.; Kapil, V.; Witt, W. C.; Magdău, I.-B.; Cole, D. J.; Csányi, G. MACE-OFF23: Transferable machine learning force fields for organic molecules. *arXiv preprint arXiv:2312.15211* **2023**,
- [113] Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **2013**, *27*, 221–234, Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer Netherlands.
- [114] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **1988**, *28*, 31–36.
- [115] Kneiding, H.; Nova, A.; Balcells, D. Directional multiobjective optimization of metal complexes at the billion-system scale. *Nature Computational Science* **2024**, *4*, 263–273.
- [116] Chen, S.-S.; Meyer, Z.; Jensen, B.; Kraus, A.; Lambert, A.; Ess, D. H. ReaLigands: A Ligand Library Cultivated from Experiment and Intended for Molecular Computational Catalyst Design. *Journal of Chemical Information and Modeling* **2023**, *63*, 7412–7422, PMID: 37987743.
- [117] Nandy, A.; Taylor, M. G.; Kulik, H. J. Identifying Underexplored and Untapped Regions in the Chemical Space of Transition Metal Complexes. *The Journal of Physical Chemistry Letters* **2023**, *14*, 5798–5804, PMID: 37338110.
- [118] Gilbert Taylor, M. coordcomplexsampling. 2025; <https://github.com/lanl/coordcomplexsampling>.
- [119] Jang, S. S.; Molinero, V.; Çağın, T.; Goddard, W. A. Nanophase-segregation and transport in Nafion 117 from molecular dynamics simulations: effect of monomeric sequence. *The Journal of Physical Chemistry B* **2004**, *108*, 3149–3157.
- [120] Sengupta, A.; Li, Z.; Song, L. F.; Li, P.; Merz, K. M. J. Parameterization of Monovalent Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *Journal of Chemical Information and Modeling* **2021**, *61*, 869–880, Publisher: American Chemical Society.
- [121] Gupta, S.; Wai, N.; Lim, T. M.; Mushrif, S. H. Force-field parameters for vanadium ions (+ 2, + 3, + 4, + 5) to investigate their interactions within the vanadium redox flow battery electrolyte solution. *Journal of Molecular Liquids* **2016**, *215*, 596–602.
- [122] Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences* **2005**, *102*, 6665–6670.
- [123] Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14\* CM1A-LBCC: localized bond-charge corrected CM1A charges for condensed-phase simulations. *The Journal of Physical Chemistry B* **2017**, *121*, 3864–3870.

- [124] Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research* **2017**, *45*, W331–W336.
- [125] Li, Z.; Song, L. F.; Li, P.; Merz Jr, K. M. Parametrization of trivalent and tetravalent metal ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *Journal of Chemical Theory and Computation* **2021**, *17*, 2342–2354.
- [126] Li, Z.; Song, L. F.; Li, P.; Merz Jr, K. M. Systematic parametrization of divalent metal ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *Journal of Chemical Theory and Computation* **2020**, *16*, 4429–4442.
- [127] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general AMBER force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.
- [128] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Erratum: “Development and testing of a general AMBER force field,” *Journal of Computational Chemistry* (2004) 25(9) 1157–1174. *Journal of Computational Chemistry* **2005**, *26*, 114.
- [129] Mamatkulov, S.; Polák, J.; Razzokov, J.; Tomanik, L.; Slavicek, P.; Dzubiella, J.; Kanduc, M.; Heyda, J. Unveiling the Borohydride Ion through Force-Field Development. *Journal of Chemical Theory and Computation* **2024**, *20*, 1263–1273.
- [130] Ishida, T.; Nishikawa, K.; Shirota, H. Atom substitution effects of [XF<sub>6</sub>]- in ionic liquids. 2. Theoretical study. *The Journal of Physical Chemistry B* **2009**, *113*, 9840–9851.
- [131] Sambasivarao, S. V.; Acevedo, O. Development of OPLS-AA force field parameters for 68 unique ionic liquids. *Journal of Chemical Theory and Computation* **2009**, *5*, 1038–1050.
- [132] Doherty, B.; Zhong, X.; Gathiaka, S.; Li, B.; Acevedo, O. Revisiting OPLS force field parameters for ionic liquid simulations. *Journal of Chemical Theory and Computation* **2017**, *13*, 6131–6145.
- [133] Wang, Y.-L.; Shah, F. U.; Glavatskikh, S.; Antzutkin, O. N.; Laaksonen, A. Atomistic insight into orthoborate-based ionic liquids: force field development and evaluation. *The Journal of Physical Chemistry B* **2014**, *118*, 8711–8723.
- [134] González, M. A.; Abascal, J. L. A flexible model for water based on TIP4P/2005. *The Journal of Chemical Physics* **2011**, *135*.
- [135] Adil, M.; Sarkar, A.; Roy, A.; Panda, M. R.; Nagendra, A.; Mitra, S. Practical Aqueous Calcium-Ion Battery Full-Cells for Future Stationary Storage. *ACS Applied Materials and Interfaces* **2020**, *12*, 11489–11503.
- [136] Adil, M.; Ghosh, A.; Mitra, S. Water-in-Salt Electrolyte-Based Extended Voltage Range, Safe, and Long-Cycle-Life Aqueous Calcium-Ion Cells. *ACS Applied Materials and Interfaces* **2022**, *14*, 25501–25515.
- [137] Aiken, C. P.; Logan, E. R.; ElDesoky, A.; Hebecker, H.; Oxner, J. M.; Harlow, J. E.; Metzger, M.; Dahn, J. R. Li[Ni<sub>0.5</sub>Mn<sub>0.3</sub>Co<sub>0.2</sub>]O<sub>2</sub> as a superior alternative to LiFePO<sub>4</sub> for long-lived low voltage Li-ion cells. *Journal of The Electrochemical Society* **2022**, *169*, 050512.
- [138] Aurbach, D.; Skaletsky, R.; Gofer, Y. The electrochemical behavior of calcium electrodes in a few organic electrolytes. *Journal of The Electrochemical Society* **1991**, *138*, 3536–3545.
- [139] Børresen, B.; Haarberg, G.; Tunold, R. Electrodeposition of magnesium from halide melts—charge transfer and diffusion kinetics. *Electrochimica Acta* **1997**, *42*, 1613–1622.
- [140] Bhide, A.; Hofmann, J.; Katharina Dürr, A.; Janek, J.; Adelhelm, P. Electrochemical stability of non-aqueous electrolytes for sodium-ion batteries and their compatibility with Na<sub>0.7</sub>CoO<sub>2</sub>. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1987–1998.
- [141] Bialik, M.; Sedin, P.; Theliander, H. Boiling Point Rise Calculations in Sodium Salt Solutions. *Industrial and Engineering Chemistry Research* **2008**, *47*, 1283–1287.
- [142] Biria, S.; Pathreker, S.; Genier, F. S.; Li, H.; Hosein, I. D. Plating and stripping calcium at room temperature in an ionic-liquid electrolyte. *ACS Applied Energy Materials* **2020**, *3*, 2310–2314.
- [143] Carbone, L.; Munoz, S.; Gobet, M.; Devany, M.; Greenbaum, S.; Hassoun, J. Characteristics of glyme electrolytes for sodium battery: nuclear magnetic resonance and electrochemical study. *Electrochimica Acta* **2017**, *231*, 223–229.
- [144] Casteel, J. F.; Amis, E. S. Specific conductance of concentrated solutions of magnesium salts in water-ethanol system. *Journal of Chemical and Engineering Data* **1972**, *17*, 55–59.

- [145] Chang, G.; Liu, S.; Fu, Y.; Hao, X.; Jin, W.; Ji, X.; Hu, J. Inhibition role of trace metal ion additives on zinc dendrites during plating and striping processes. *Advanced Materials Interfaces* **2019**, *6*.
- [146] Chen, L.; Bao, J. L.; Dong, X.; Truhlar, D. G.; Wang, Y.; Wang, C.; Xia, Y. Aqueous Mg-ion battery based on polyimide anode and prussian blue cathode. *ACS Energy Letters* **2017**, *2*, 1115–1121.
- [147] Chen, S.; Zheng, J.; Yu, L.; Ren, X.; Engelhard, M. H.; Niu, C.; Lee, H.; Xu, W.; Xiao, J.; Liu, J.; Zhang, J.-G. High-efficiency lithium metal batteries with fire-retardant electrolytes. *Joule* **2018**, *2*, 1548–1558.
- [148] Cheng, Y.; Stolley, R. M.; Han, K. S.; Shao, Y.; Arey, B. W.; Washton, N. M.; Mueller, K. T.; Helm, M. L.; Sprenkle, V. L.; Liu, J.; Li, G. Highly active electrolytes for rechargeable Mg batteries based on a  $[\text{Mg}_2(\mu\text{-Cl})_2]^{2+}$  cation complex in dimethoxyethane. *Physical Chemistry Chemical Physics* **2015**, *17*, 13307–13314.
- [149] Deng, L.; Zhang, Y.; Wang, R.; Feng, M.; Niu, X.; Tan, L.; Zhu, Y. Influence of  $\text{KPF}_6$  and KFSI on the performance of anode materials for potassium-ion batteries: a case study of  $\text{MoS}_2$ . *ACS Applied Materials and Interfaces* **2019**, *11*, 22449–22456.
- [150] Ding, C.; Nohira, T.; Kuroda, K.; Hagiwara, R.; Fukunaga, A.; Sakai, S.; Nitta, K.; Inazawa, S. NaFSA- $\text{C}_1\text{C}_3\text{pyrFSA}$  ionic liquids for sodium secondary battery operating over a wide temperature range. *Journal of Power Sources* **2013**, *238*, 296–300.
- [151] Doi, T.; Shimizu, Y.; Hashinokuchi, M.; Inaba, M. Dilution of highly concentrated  $\text{LiBF}_4$ /propylene carbonate electrolyte solution with fluoroalkyl ethers for 5-V  $\text{LiNi}_{0.5}\text{Mn}_{1.5}\text{O}_4$  positive electrodes. *Journal of The Electrochemical Society* **2017**, *164*, A6412–A6416.
- [152] Eiberweiser, A.; Nazet, A.; Heftner, G.; Buchner, R. Ion hydration and association in aqueous potassium phosphate solutions. *The Journal of Physical Chemistry B* **2015**, *119*, 5270–5281.
- [153] Etman, A. S.; Carboni, M.; Sun, J.; Younesi, R. Acetonitrile-based electrolytes for rechargeable zinc batteries. *Energy Technology* **2020**, *8*.
- [154] Geng, L.; Meng, J.; Wang, X.; Han, C.; Han, K.; Xiao, Z.; Huang, M.; Xu, P.; Zhang, L.; Zhou, L.; Mai, L. Eutectic electrolyte with unique solvation structure for high-performance zinc-ion batteries. *Angewandte Chemie International Edition* **2022**, *61*.
- [155] Hosaka, T.; Kubota, K.; Kojima, H.; Komaba, S. Highly concentrated electrolyte solutions for 4 V class potassium-ion batteries. *Chemical Communications* **2018**, *54*, 8387–8390.
- [156] Hou, S.; Ji, X.; Gaskell, K.; Wang, P.-f.; Wang, L.; Xu, J.; Sun, R.; Borodin, O.; Wang, C. Solvation sheath reorganization enables divalent metal batteries with fast interfacial charge transfer kinetics. *Science* **2021**, *374*, 172–178.
- [157] Hou, T.; Fong, K. D.; Wang, J.; Persson, K. A. The solvation structure, transport properties and reduction behavior of carbonate-based electrolytes of lithium-ion batteries. *Chemical Science* **2021**, *12*, 14740–14751.
- [158] Jiang, L.; Lu, Y.; Zhao, C.; Liu, L.; Zhang, J.; Zhang, Q.; Shen, X.; Zhao, J.; Yu, X.; Li, H.; Huang, X.; Chen, L.; Hu, Y.-S. Building aqueous K-ion batteries for energy storage. *Nature Energy* **2019**, *4*, 495–503.
- [159] Jiao, H.; Wang, C.; Tu, J.; Tian, D.; Jiao, S. A rechargeable Al-ion battery: Al/molten  $\text{AlCl}_3$ -urea/graphite. *Chemical Communications* **2017**, *53*, 2331–2334.
- [160] Kühnel, R.-S.; Reber, D.; Battaglia, C. A high-voltage aqueous electrolyte for sodium-ion batteries. *ACS Energy Letters* **2017**, *2*, 2005–2006.
- [161] Kao, Y.-C.; Tu, C.-H. Solubility, density, viscosity, refractive index, and electrical conductivity for potassium nitrate-water-2-propanol at (298.15 and 313.15) K. *Journal of Chemical and Engineering Data* **2009**, *54*, 1927–1931.
- [162] Keyzer, E. N.; Glass, H. F. J.; Liu, Z.; Bayley, P. M.; Dutton, S. E.; Grey, C. P.; Wright, D. S. Mg( $\text{PF}_6$ )<sub>2</sub>-based electrolyte systems: understanding electrolyte-electrode interactions for the development of Mg-ion batteries. *Journal of the American Chemical Society* **2016**, *138*, 8682–8685.
- [163] Ko, S.; Yamada, Y.; Yamada, A. A 62 m K-ion aqueous electrolyte. *Electrochemistry Communications* **2020**, *116*, 106764.
- [164] Komaba, S.; Murata, W.; Ishikawa, T.; Yabuuchi, N.; Ozeki, T.; Nakayama, T.; Ogata, A.; Gotoh, K.; Fujiwara, K. Electrochemical Na insertion and solid electrolyte interphase for hard-carbon electrodes and application to Na-ion batteries. *Advanced Functional Materials* **2011**, *21*, 3859–3867.

- [165] Kubota, K.; Nohira, T.; Goto, T.; Hagiwara, R. Novel inorganic ionic liquids possessing low melting temperatures and wide electrochemical windows: binary mixtures of alkali bis(fluorosulfonyl)amides. *Electrochemistry Communications* **2008**, *10*, 1886–1888.
- [166] Kubota, K.; Nohira, T.; Hagiwara, R. New inorganic ionic liquids possessing low melting temperatures and wide electrochemical windows: ternary mixtures of alkali bis(fluorosulfonyl)amides. *Electrochimica Acta* **2012**, *66*, 320–324.
- [167] Lazouski, N.; Schiffer, Z. J.; Williams, K.; Manthiram, K. Understanding continuous lithium-mediated electrochemical nitrogen reduction. *Joule* **2019**, *3*, 1127–1139.
- [168] Leonard, D. P.; Wei, Z.; Chen, G.; Du, F.; Ji, X. Water-in-salt electrolyte for potassium-ion batteries. *ACS Energy Letters* **2018**, *3*, 373–374.
- [169] Li, J.; Han, C.; Ou, X.; Tang, Y. Concentrated electrolyte for high-performance Ca-ion battery based on organic anode and graphite cathode. *Angewandte Chemie* **2022**, *134*.
- [170] Lin, M.-C.; Gong, M.; Lu, B.; Wu, Y.; Wang, D.-Y.; Guan, M.; Angell, M.; Chen, C.; Yang, J.; Hwang, B.-J.; Dai, H. An ultrafast rechargeable aluminium-ion battery. *Nature* **2015**, *520*, 324–328.
- [171] Liu, G.; Cao, Z.; Zhou, L.; Zhang, J.; Sun, Q.; Hwang, J.; Cavallo, L.; Wang, L.; Sun, Y.; Ming, J. Additives engineered nonflammable electrolyte for safer potassium ion batteries. *Advanced Functional Materials* **2020**, *30*.
- [172] Liu, S.; Mao, J.; Zhang, Q.; Wang, Z.; Pang, W. K.; Zhang, L.; Du, A.; Sencadas, V.; Zhang, W.; Guo, Z. An intrinsically non-flammable electrolyte for high-performance potassium batteries. *Angewandte Chemie International Edition* **2020**, *59*, 3638–3644.
- [173] Ma, L.; Chen, S.; Li, N.; Liu, Z.; Tang, Z.; Zapien, J. A.; Chen, S.; Fan, J.; Zhi, C. Hydrogen-free and dendrite-free all-solid-state Zn-ion batteries. *Advanced Materials* **2020**, *32*.
- [174] Marczewski, M. J.; Stanje, B.; Hanzu, I.; Wilkening, M.; Johansson, P. "Ionic liquids-in-salt" – a promising electrolyte concept for high-temperature lithium batteries? *Phys. Chem. Chem. Phys.* **2014**, *16*, 12341–12349.
- [175] Morishita, M.; Koyama, K.; Mori, Y. Inhibition of anodic dissolution of zinc-plated steel by electro-deposition of magnesium from a molten salt. *ISIJ International* **1997**, *37*, 55–58.
- [176] Naveed, A.; Yang, H.; Shao, Y.; Yang, J.; Yanna, N.; Liu, J.; Shi, S.; Zhang, L.; Ye, A.; He, B.; Wang, J. A highly reversible Zn anode with intrinsically safe organic electrolyte for long-cycle-life batteries. *Advanced Materials* **2019**, *31*.
- [177] Nguyen, D.-T.; Eng, A. Y. S.; Ng, M.-F.; Kumar, V.; Sofer, Z.; Handoko, A. D.; Subramanian, G. S.; Seh, Z. W. A high-performance magnesium triflate-based electrolyte for rechargeable magnesium batteries. *Cell Reports Physical Science* **2020**, *1*, 100265.
- [178] NuLi, Y.; Yang, J.; Wang, J.; Xu, J.; Wang, P. Electrochemical magnesium deposition and dissolution with high efficiency in ionic liquid. *Electrochemical and Solid-State Letters* **2005**, *8*, C166.
- [179] Pan, H.; Shao, Y.; Yan, P.; Cheng, Y.; Han, K. S.; Nie, Z.; Wang, C.; Yang, J.; Li, X.; Bhattacharya, P.; Mueller, K. T.; Liu, J. Reversible aqueous zinc/manganese oxide energy storage from conversion reactions. *Nature Energy* **2016**, *1*.
- [180] Pluhařová, E.; Mason, P. E.; Jungwirth, P. Ion pairing in aqueous lithium salt solutions with monovalent and divalent counter-anions. *The Journal of Physical Chemistry A* **2013**, *117*, 11766–11773.
- [181] Ponrouch, A.; Frontera, C.; Bardé, F.; Palacín, M. R. Towards a calcium-based rechargeable battery. *Nature Materials* **2015**, *15*, 169–172.
- [182] Qian, J.; Henderson, W. A.; Xu, W.; Bhattacharya, P.; Engelhard, M.; Borodin, O.; Zhang, J.-G. High rate and stable cycling of lithium metal anode. *Nature Communications* **2015**, *6*.
- [183] Ren, X. et al. Localized high-concentration sulfone electrolytes for high-efficiency lithium-metal batteries. *Chem* **2018**, *4*, 1877–1892.
- [184] Shi, P.; Fang, S.; Huang, J.; Luo, D.; Yang, L.; Hirano, S.-i. A novel mixture of lithium bis(oxalato)borate, gamma-butyrolactone and non-flammable hydrofluoroether as a safe electrolyte for advanced lithium ion batteries. *Journal of Materials Chemistry A* **2017**, *5*, 19982–19990.

- [185] Shterenberg, I.; Salama, M.; Yoo, H. D.; Gofer, Y.; Park, J.-B.; Sun, Y.-K.; Aurbach, D. Evaluation of (CF<sub>3</sub>SO<sub>2</sub>)<sub>2</sub>N-(TFSI) Based Electrolyte Solutions for Mg Batteries. *Journal of The Electrochemical Society* **2015**, *162*, A7118–A7128.
- [186] Son, S.-B.; Gao, T.; Harvey, S. P.; Steirer, K. X.; Stokes, A.; Norman, A.; Wang, C.; Cresce, A.; Xu, K.; Ban, C. An artificial interphase enables reversible magnesium chemistry in carbonate electrolytes. *Nature Chemistry* **2018**, *10*, 532–539.
- [187] Song, Y.; Hu, J.; Tang, J.; Gu, W.; He, L.; Ji, X. Real-time X-ray imaging reveals interfacial growth, suppression, and dissolution of zinc dendrites dependent on anions of ionic liquid additives for rechargeable battery applications. *ACS Applied Materials and Interfaces* **2016**, *8*, 32031–32040.
- [188] Soundharajan, V.; Sambandam, B.; Kim, S.; Mathew, V.; Jo, J.; Kim, S.; Lee, J.; Islam, S.; Kim, K.; Sun, Y.-K.; Kim, J. Aqueous magnesium zinc hybrid battery: an advanced high-voltage and high-energy MgMn<sub>2</sub>O<sub>4</sub> cathode. *ACS Energy Letters* **2018**, *3*, 1998–2004.
- [189] Su, D.; McDonagh, A.; Qiao, S.; Wang, G. High-capacity aqueous potassium-ion batteries for large-scale energy storage. *Advanced Materials* **2016**, *29*.
- [190] Suo, L.; Borodin, O.; Gao, T.; Olguin, M.; Ho, J.; Fan, X.; Luo, C.; Wang, C.; Xu, K. "Water-in-salt" electrolyte enables high-voltage aqueous lithium-ion chemistries. *Science* **2015**, *350*, 938–943.
- [191] Suo, L.; Borodin, O.; Wang, Y.; Rong, X.; Sun, W.; Fan, X.; Xu, S.; Schroeder, M. A.; Cresce, A. V.; Wang, F.; Yang, C.; Hu, Y.; Xu, K.; Wang, C. "Water-in-salt" electrolyte makes aqueous sodium-ion battery safe, green, and long-lasting. *Advanced Energy Materials* **2017**, *7*.
- [192] Tsuneto, A.; Kudo, A.; Sakata, T. Lithium-mediated electrochemical reduction of high pressure N<sub>2</sub> to NH<sub>3</sub>. *Journal of Electroanalytical Chemistry* **1994**, *367*, 183–188.
- [193] Vidal-Abarca, C.; Lavela, P.; Tirado, J.; Chadwick, A.; Alfredsson, M.; Kelder, E. Improving the cyclability of sodium-ion cathodes by selection of electrolyte solvent. *Journal of Power Sources* **2012**, *197*, 314–318.
- [194] Wang, J.; Yamada, Y.; Sodeyama, K.; Chiang, C. H.; Tateyama, Y.; Yamada, A. Superconcentrated electrolytes for a high-voltage lithium-ion battery. *Nature Communications* **2016**, *7*.
- [195] Wang, D.; Gao, X.; Chen, Y.; Jin, L.; Kuss, C.; Bruce, P. G. Plating and stripping calcium in an organic electrolyte. *Nature Materials* **2017**, *17*, 16–20.
- [196] Wang, F.; Fan, X.; Gao, T.; Sun, W.; Ma, Z.; Yang, C.; Han, F.; Xu, K.; Wang, C. High-voltage aqueous magnesium ion batteries. *ACS Central Science* **2017**, *3*, 1121–1128.
- [197] Wang, M.; Jiang, C.; Zhang, S.; Song, X.; Tang, Y.; Cheng, H.-M. Reversible calcium alloying enables a practical room-temperature rechargeable calcium-ion battery with a high discharge voltage. *Nature Chemistry* **2018**, *10*, 667–672.
- [198] Wang, H. et al. Reversible electrochemical interface of Mg metal and conventional electrolyte enabled by intermediate adsorption. *ACS Energy Letters* **2019**, *5*, 200–206.
- [199] Wang, Z.; Diao, J.; Burrow, J. N.; Reimund, K. K.; Katyal, N.; Henkelman, G.; Mullins, C. B. Urea-modified ternary aqueous electrolyte with tuned intermolecular interactions and confined water activity for high-stability and high-voltage zinc-ion batteries. *Advanced Functional Materials* **2023**, *33*.
- [200] Watanabe, Y.; Ugata, Y.; Ueno, K.; Watanabe, M.; Dokko, K. Does Li-ion transport occur rapidly in localized high-concentration electrolytes? *Physical Chemistry Chemical Physics* **2023**, *25*, 3092–3099.
- [201] Watarai, A.; Kubota, K.; Yamagata, M.; Goto, T.; Nohira, T.; Hagiwara, R.; Ui, K.; Kumagai, N. A rechargeable lithium metal battery operating at intermediate temperatures using molten alkali bis(trifluoromethylsulfonyl)amide mixture as an electrolyte. *Journal of Power Sources* **2008**, *183*, 724–729.
- [202] Weng, G.-M.; Li, Z.; Cong, G.; Zhou, Y.; Lu, Y.-C. Unlocking the capacity of iodide for high-energy-density zinc/polyiodide and lithium/polyiodide redox flow batteries. *Energy and Environmental Science* **2017**, *10*, 735–741.
- [203] Xiao, N.; McCulloch, W. D.; Wu, Y. Reversible dendrite-free potassium plating and stripping electrochemistry for potassium secondary batteries. *Journal of the American Chemical Society* **2017**, *139*, 9475–9478.

- [204] Xie, J.; Li, X.; Lai, H.; Zhao, Z.; Li, J.; Zhang, W.; Xie, W.; Liu, Y.; Mai, W. A robust solid electrolyte interphase layer augments the ion storage capacity of bimetallic-sulfide-containing potassium-ion batteries. *Angewandte Chemie* **2019**, *131*, 14882–14889.
- [205] Yamamoto, T.; Matsumoto, K.; Hagiwara, R.; Nohira, T. Physicochemical and electrochemical properties of  $K[N(SO_2F)_2]-[N\text{-methyl-N-propyl}]\text{pyrrolidinium}[N(SO_2F)_2]$  ionic liquids for potassium-ion batteries. *The Journal of Physical Chemistry C* **2017**, *121*, 18450–18458.
- [206] Yang, H.; Hwang, J.; Tonouchi, Y.; Matsumoto, K.; Hagiwara, R. Sodium difluorophosphate: facile synthesis, structure, and electrochemical behavior as an additive for sodium-ion batteries. *Journal of Materials Chemistry A* **2021**, *9*, 3637–3647.
- [207] Yoo, D.-J.; Liu, Q.; Cohen, O.; Kim, M.; Persson, K. A.; Zhang, Z. Understanding the role of SEI layer in low-temperature performance of lithium-ion batteries. *ACS Applied Materials and Interfaces* **2022**, *14*, 11910–11918.
- [208] Yoon, H.; Zhu, H.; Hervault, A.; Armand, M.; MacFarlane, D. R.; Forsyth, M. Physicochemical properties of  $N\text{-propyl-N-methyl}\text{pyrrolidinium bis(fluorosulfonyl)imide}$  for sodium metal battery applications. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12350–12355.
- [209] Zhang, N.; Cheng, F.; Liu, Y.; Zhao, Q.; Lei, K.; Chen, C.; Liu, X.; Chen, J. Cation-deficient spinel  $ZnMn_2O_4$  cathode in  $Zn(CF_3SO_3)_2$  electrolyte for rechargeable aqueous Zn-ion battery. *Journal of the American Chemical Society* **2016**, *138*, 12894–12901.
- [210] Zhang, R.; Bao, J.; Pan, Y.; Sun, C.-F. Highly reversible potassium-ion intercalation in tungsten disulfide. *Chemical Science* **2019**, *10*, 2604–2612.
- [211] Zhang, X.-Q.; Chen, X.; Hou, L.-P.; Li, B.-Q.; Cheng, X.-B.; Huang, J.-Q.; Zhang, Q. Regulating anions in the solvation sheath of lithium ions for stable lithium metal batteries. *ACS Energy Letters* **2019**, *4*, 411–416.
- [212] Zhao, J.; Zou, X.; Zhu, Y.; Xu, Y.; Wang, C. Electrochemical intercalation of potassium into graphite. *Advanced Functional Materials* **2016**, *26*, 8103–8110.
- [213] Zhou, J.; Shan, L.; Wu, Z.; Guo, X.; Fang, G.; Liang, S. Investigation of  $V_2O_5$  as a low-cost rechargeable aqueous zinc ion battery cathode. *Chemical Communications* **2018**, *54*, 4457–4460.
- [214] Zhu, Y.; Yin, J.; Zheng, X.; Emwas, A.-H.; Lei, Y.; Mohammed, O. F.; Cui, Y.; Alshareef, H. N. Concentrated dual-cation electrolyte strategy for aqueous zinc-ion batteries. *Energy and Environmental Science* **2021**, *14*, 4463–4473.
- [215] Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; others Scalable algorithms for molecular dynamics simulations on commodity clusters. Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. 2006; pp 84–es.
- [216] D. E. Shaw Research; Schrödinger Schrödinger Release 2024-2: Desmond Molecular Dynamics System; Maestro-Desmond Interoperability Tools. D. E. Shaw Research and Schrödinger, LLC: New York, NY, 2024; Desmond Molecular Dynamics System (2024); Maestro-Desmond Interoperability Tools (2024).
- [217] Habershon, S.; Manolopoulos, D. E.; Markland, T. E.; Miller III, T. F. Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space. *Annual review of physical chemistry* **2013**, *64*, 387–413.
- [218] Markland, T. E.; Manolopoulos, D. E. An efficient ring polymer contraction scheme for imaginary time path integral simulations. *The Journal of Chemical Physics* **2008**, *129*.
- [219] Erkut, E. The discrete p-dispersion problem. *European Journal of Operational Research* **1990**, *46*, 48–60.
- [220] Ghosh, J. B. Computational aspects of the maximum diversity problem. *Operations Research Letters* **1996**, *19*, 175–181.
- [221] Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The Journal of Physical Chemistry B* **2015**, *119*, 5113–5123.
- [222] Liao, Y.-L.; Wood, B.; Das\*, A.; Smidt\*, T. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. International Conference on Learning Representations (ICLR). 2024.
- [223] Zhu, X.; Thompson, K. C.; Martínez, T. J. Geodesic interpolation for reaction pathways. *The Journal of Chemical Physics* **2019**, *150*, 164103.

- [224] Sameera, W. M. C.; Maeda, S.; Morokuma, K. Computational Catalysis Using the Artificial Force Induced Reaction Method. *Accounts of Chemical Research* **2016**, *49*, 763–773.
- [225] Unke, O. T.; Meuwly, M. Solvated protein fragments. 2019; <https://zenodo.org/record/2605372>.
- [226] Riebesell, J.; Goodall, R. E.; Benner, P.; Chiang, Y.; Deng, B.; Lee, A. A.; Jain, A.; Persson, K. A. Matbench Discovery—A framework to evaluate machine learning crystal stability predictions. *arXiv preprint arXiv:2308.14920* **2023**,
- [227] Tran, R.; Lan, J.; Shuaibi, M.; Wood, B. M.; Goyal, S.; Das, A.; Heras-Domingo, J.; Kolluru, A.; Rizvi, A.; Shoghi, N.; others The Open Catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis* **2023**, *13*, 3066–3084.

# Appendix

## Appendix Table of Contents

<b>A Calculation Details</b>	<b>32</b>
A.1 Calculation Quality Filters . . . . .	32
A.2 Computed Properties . . . . .	32
<b>B Biomolecules</b>	<b>33</b>
B.1 Protein Pocket + Ligand . . . . .	33
B.2 Fragmented Protein Pockets + Ligand . . . . .	35
B.3 Protein + Protein: Core . . . . .	35
B.4 Protein + Protein: Interface . . . . .	35
B.5 DNA/RNA . . . . .	35
<b>C Metal Complexes</b>	<b>36</b>
C.1 Complexes from Architect . . . . .	36
C.2 Complexes from the Crystallography Open Database (COD) . . . . .	38
<b>D Electrolytes</b>	<b>40</b>
D.1 Bulk Electrolytes . . . . .	40
D.2 Small Molecule Generation . . . . .	47
D.3 Molecular Clusters . . . . .	48
D.4 Water Clusters . . . . .	48
D.5 Noble Gases . . . . .	48
<b>E Molecular Dynamics with MLIPs</b>	<b>49</b>
<b>F Reactivity</b>	<b>49</b>
F.1 RGD1 . . . . .	49
F.2 RMechDB and PMechDB . . . . .	49
F.3 Metal Reactivity . . . . .	51
F.4 Electrolyte Reactivity . . . . .	52
<b>G Community Datasets</b>	<b>53</b>
<b>H Evaluation Details</b>	<b>53</b>
H.1 Protein-ligand Interactions . . . . .	53
H.2 Ligand Strain . . . . .	53
H.3 Conformers . . . . .	53
H.4 Protonation Energies . . . . .	54
H.5 Unoptimized IE/EA and Spin Gap . . . . .	54
H.6 Distance Scaling: Short-range and Long-range Interactions . . . . .	54
<b>I Baseline Models and Results</b>	<b>55</b>
I.1 Training . . . . .	55
I.2 Total Energy Test Results . . . . .	56
I.3 Validation Results . . . . .	56
I.4 Additional Results . . . . .	56
I.5 Wiggle150 . . . . .	58
I.6 Hyperparameters . . . . .	59

## A Calculation Details

All calculations in OMol25 were carried out with the ORCA 6.0.0 DFT package [87, 88]. ORCA supports various integral acceleration techniques, including RI-J and COSX, which dramatically improve the computational cost of these calculations at a very small cost in error, and these were used here. Experimentation with integral threshold settings indicated that the best trade-off between robust convergence and computational cost was to set the integral threshold (`thresh` in ORCA) to 1e-12 and the primitive batch threshold (`tcut` in ORCA) to 1e-13; these values were subsequently adopted by the ORCA package to be defaults in future versions of ORCA. ORCA’s tight convergence settings were employed. Extensive benchmarking with various combinations of grid settings (both the exchange-correlation grid and the COSX grid) indicated that typical grids led to small numerical inconsistencies between energy and forces. In other words, there were discrepancies between the derivative of the energy with respect to coordinates and the computed forces due to grid incompleteness, and these errors were significant on the scale of errors with state-of-the-art MLIPs). In order to achieve sufficiently tight consistency, ORCA’s DEFGRID3 offered the best trade-off of convergence and cost. This corresponds to a pruned grid with 590 angular points for exchange-correlation and 302 for the final COSX grid.

As described in Section 2.6, we typically compute systems where bonds may be breaking in UKS. We could not find mention of whether Transition-1X was originally run in RKS or UKS and so have computed the dataset both ways, though we advocate for the use of the UKS version. In the end, this is a small consideration as <5% of the Transition-1X data in UKS had an expectation value of  $S^2$  greater than 0.001, indicating that, for more than 95% of the dataset, the RKS and UKS SCF solutions are the same.

### A.1 Calculation Quality Filters

Given the scale and diversity of OMol25, adequate quality and error checks are required before considering a calculation valid, particularly for ML training. We enforce several quality checks on the resulting DFT calculations before considering them in the final dataset:

- Referenced energies shall not exceed  $\pm 150$  eV. This removes highly unreasonable configurations;
- Max per atom force shall not exceed 50eV/Å. This also removes highly unreasonable configurations;
- $S^2 < 0.5$  for open-shell metal-containing systems,  $S^2 < 1.1$  otherwise. We enforce tighter constraints on metal-systems to avoid incorrect SCF solutions but allow full spin-unpairing which may occur in organic reactivity;
- Enforce alpha, beta, and total electron consistency with the integrated densities. This indicates insufficient grid density;
- ORCA errors where the final COSX exchange deviates considerably. This indicate convergence errors;
- Non-negative HOMO-LUMO gaps.

Furthermore, using a trained model on an early snapshot of OMol25, we perform an error analysis similar to the work of MACE-OFF on SPICE [112]. Energy errors were evaluated for a held-out test set of OMol25, followed by manual inspection on some of the worst offenders. This allowed us to identify and filter systematically problematic inputs, such as isolated metal centers far away from the rest of the structure resulting in convergence to excited SCF minima.

### A.2 Computed Properties

The properties computed for each point in OMol25 are the following:

1. Total energy (in eV)
2. Forces (in eV/Å)
3. charge
4. spin

5. Number of atoms
6. Number of electrons
7. Number of ECP electrons
8. Number of basis functions
9. Unrestricted vs. Restricted
10. Number of SCF steps
11. Energy computed by VV10
12.  $S^2$  expectation value
13. Deviation of  $S^2$  from ideal
14. Integrated density (should be very close to the total number of electrons)
15. HOMO energy (in eV),  $\alpha$  and  $\beta$  for unrestricted
16. HOMO-LUMO gap (in eV),  $\alpha$  and  $\beta$  for unrestricted
17. Maximum force magnitude for a given atom in a given direction (fmax)
18. Mulliken charges (and spins if unrestricted)
19. Loewdin charges (and spins if unrestricted)
20. NBO charges (and spins if unrestricted) if the total number of atom  $<= 70$
21. Any ORCA warnings that are generated
22. ORCA .gbw files and densities will be made available in the near future

Additional properties which have been computed, such as multipole moments, Fock matrices, and orbital energies, will be released in future versions of the dataset.

## B Biomolecules

Biological application represent one of the largest current uses of computational chemistry. Annual spending by pharmaceutical companies on structure-based drug design, both of small molecules and biologics is on the order of a billion dollars. Methods such as free energy perturbation require extremely high accuracy in the potential used. Due to the large size of biological systems and the need to simulate often for tens of nanoseconds, atomistic simulations nearly universally employ classical molecular dynamics force field. To achieve—and surpass—state-of-the-art classical forcefields, high quality quantum chemistry data which covers the breadth of biological interactions is needed.

To this end, we developed several focused subdatasets intended to probe a variety of applications which would reasonably be expected to arise in biochemistry research. All structures taken from the Protein Data Bank (PDB) were first prepared with Schrödinger’s Protein Preparation Wizard [113]. Molecular Dynamics (MD) calculations were performed with Schrödinger’s Desmond molecular dynamics engine with the OPLS4 forcefield [68]. Desmond/OPLS4 automatically assigns forcefield parameters to an extremely broad amount of chemical matter automatically, which facilitated this high-throughput process. Protonation and tautomeric states of ligands were sampled with Schrödinger’s Epik program [102, 113].

### B.1 Protein Pocket + Ligand

Ligand-protein interactions are some of the most important for computational chemistry, and it is critical to obtain a broad and diverse sampling of ligand chemistries and geometries. Moreover, often neglected from ligand-protein interaction is the effect of protonation state and tautomerism on both ligands and proteins. The BioLiP2 database contains nearly 9.5M entries detailing ligands (broadly defined, including small molecules, metal ions, peptides, DNA, and RNA) and their corresponding receptor residues. All non-peptide, non-nucleic

acid ligand structures in the BioLiP2 database as of May 31, 2024 were extracted and processed. In order to obtain unique ligand binding environments, we removed the following entries as duplicates:

1. if the same ligand id appeared with identical receptor residues (including residue numbering); we do not require PDB IDs to be identical as closely related proteins often have conserved binding sites
2. if the same ligand ID appeared in the same PDB structure with receptor residues that were largely the same (one being a proper subset of the other), we discard the larger binding site, as those marginal residues do not represent key interactions

This resulted in 258,467 unique ligand-pocket pairs. We further elaborated those pairs using Schrödinger's Epik in conjunction with the Protein Preparation Wizard to add up to the 10 lowest energy protonation/tautomer states and optimize the protonation of receptor residues in their presence. The ligand and the residues identified by the BioLiP2 entry were extracted along with any residues with any atom within 2.5 Å of the ligand (these residues include such things as waters coordinated to metal centers and ligands which may consist of multiple covalently attached positions). All protein residues were capped with ACE or NMA at the N- and C-termini, respectively, and these caps were rotated to match the C $\alpha$  positions of the native protein structure. If two residues were extracted that had a single residue in between that was not extracted, then these capping groups would overlap and instead of capping, a glycine is placed between the two residues with C $\alpha$  at the same location as the residue in the native protein.

Various cleanup steps were applied to ensure that these extracted pockets were suitable for further study:

1. Hydrogen atoms were added or removed as needed to match the expected Lewis structure
2. Ions were adjusted to the expected oxidation states: Li, Na, K, Rb, Cs, Ag to +1, Mg, Ca, Sr, Zn, Cd, Hg to +2, Al, Eu to +3, Fe (if an oxidation state was not specified in the PDB entry) to +2, Cu (if an oxidation state was not specified) to +2, Os (if not specified) to +3, Sn (if not specified) to +4 unless it was part of Sn-C or Sn-S bonding.
3. Formal charges on various functional groups were adjusted
4. Visual inspection of many systems suggest that Epik does not always do a good job with the protonation state of sulfates, phosphates, and ligands bound to metals, especially heme groups. As such, if a fully deprotonated state of these functional groups were not present, then one was added.
5. Disulfide bonds that were disrupted by the extraction procedure were capped with hydrogen
6. Several minor adjustments to bonding and Lewis structure were made to allow the system to work properly with Schrödinger's Desmond MD engine.

Charge and spin states were assigned to these extracted pockets on the basis of the formal charge of the atoms and any metal oxidation states that were determined by this process. After confirming that the charge and spin were physically consistent and restricting the number of atoms to 350 or fewer, this left 416,324 systems.

Since PDB structures are often not suitable for use directly with DFT, owing to, among other things, imprecisely determined atomic bond lengths, and to increase the structural diversity present, the extracted pockets were used as inputs to an MD simulation. Per the recommended procedure from Desmond, the systems were run with Brownian dynamics for 100ps at 10K with all heavy atoms restrained, then Langevin dynamics for 12ps at 10K with all heavy atoms restrained. The restraints were then removed and the system energy was minimized. We then applied restraints to the C $\alpha$  positions of all present residues and 1.02 nanoseconds of Langevin dynamics were simulated. The systems were split randomly into two groups: 85% were run at 300K and 15% were run at 400K. The C $\alpha$  constraints were included to prevent the pocket from disintegrating due to the absence of the rest of the protein structure. We discarded frames where the ligand centroid moved more than 5 Å from the starting centroid position in the final stage, with the idea being that this represented the ligand falling out of the extracted pocket. Frames with an RMSD less than 1 Å from the initial frame were discarded (this occurred in some cases where the combination of the C $\alpha$  constraints and strong ligand-pocket interactions (e.g., with some metals) led to little movement by MD). The remaining frames were evenly subsampled into 10 frames. The first and last, being farthest in time and therefore most likely to be decorrelated, were selected as inputs for DFT. This results in 622,360 DFT inputs from the 300K MD and 122,906 inputs from 400K MD.

## B.2 Fragmented Protein Pockets + Ligand

Having exhausted experimental ligand-protein interaction structures in BioLip2 with 750k datapoints, synthetic ligand-protein structures were required both to increase the sampling of diverse interactions and especially to improve coverage of interactions as they might pertain to docking simulations with novel chemical matter. Receptor pockets obtained in the previous step were identified as "drug-like" pockets (in contrast to e.g., metal-binding pockets) by noting if any binding affinity data (e.g., from PDBbind) was associated with the BioLiP2 entry for the pocket. This procedure identified 32,833 pockets which are likely to be druggable. Drug-like systems were randomly sampled from Zinc20's "clean leads" set (molecular weight between 250 and 350, predicted LogP  $\leq 3.5$ , and 7 or fewer rotatable bonds), ChemBL, and GEOM drugs and docked into these pockets with smina. Up to three protonation/tautomer states were computed for each ligand with Epik before docking. These systems were randomly split 50/50 between 300K and 400K MD using the procedure described above. However, in this case, after MD, instead of the entire pocket being used as input for DFT, only the two receptor fragments that were closest to the ligand are retained (as long as that is less than 5 residues in total). Only the last frame of the MD was used. These fragmented pockets are smaller, which allow a larger number to be run for the same computational cost. This procedure resulted in 2,323,228 inputs from 300K MD and 2,326,639 inputs from 400K MD.

## B.3 Protein + Protein: Core

While there is some degree of protein-protein interaction in the Pocket + Ligand and Fragment Pocket + Ligand data, we sought to increase coverage of protein-protein interaction data. To ensure broad coverage, we focus separately on the cores of proteins and on the interfaces between protein subunits.

We randomly sample proteins in the PDB and, for each structure, randomly sample up to 20 residues whose relative solvent accessible surface areas is 0.2 or less (indicating that the residue was part of the protein core). All residues where any heavy atom of the side-chain was within 4 $\text{\AA}$  (in 80% of cases) or 4.5 $\text{\AA}$  (in 20% of cases) of the heavy atoms in the randomly selected core side-chain was extracted as the residue's protein environment. These residues were treated as for ligand-pockets: adding hydrogens with Protein Preparation Wizard, fixing Lewis structures, capping termini of residues, and/or replacing single-amino acid gaps with glycine. To avoid being too similar to the SPICE dataset of amino acid dimers, we also exclude any system that consists of fewer than three residues. The resulting prepared systems were run with the same MD procedure as described above, 50% at 300K and 50% at 400K. Both the first and last frame of the MD were used for DFT, resulting in 1,432,225 inputs. Additionally, 1,209,514 samples were submitted without running them through the MD procedure first in order to assess whether the DFT machinery could handle the sort of raw geometries typical of PDB structures.

## B.4 Protein + Protein: Interface

Protein-protein interfaces were generated by randomly sampling residues from the DIPS-plus database to identify residues at protein interfaces. Any residue that had a heavy atom of its side-chain within 4 $\text{\AA}$  of the selected residue's side-chain heavy atoms and was on the same protein chain or within 6 $\text{\AA}$  if it was on a different protein chain were extracted as the environment of that interface residue. Preparing, capping, and MD was performed in the same manner as protein-core systems above, giving 929,656 DFT inputs. An additional  $<\mathbf{n}>$  were prepared and capped but instead of running with classical MD, these inputs were run with MD using an MLIP, as described in Section E.

## B.5 DNA/RNA

DNA and especially RNA often are also part of protein structures, and proper coverage of protein-nucleic acid (and nucleic acid-nucleic acid) interactions are critical to ensuring ML models can accurately treat the breadth of biological systems. Fortunately, the BioLiP2 also features extensive coverage of these kinds of interactions. Due to the structural differences of nucleic acid vs. proteins and ligand, we look for specific motifs rather than simple radius cutoffs. Starting from a randomly selected nucleic acid base residue, which we denote as the "core", we classify the types of motifs we will extract as follows. Note that some of these are subsets of each other, but we consider them separate types to obtain a range of sizes of extracted systems.

1. Residues with any atom within 2.5Å of the core and not directly bound to the core. For DNA, this is typically the complementary base pair and any nearby protein residues.
2. Residues with any atom within 2.5Å of the core and not DNA or RNA. This is generally the base itself and any protein nearby.
3. Residues directly attached to the core. This is generally three-some stacks of bases.
4. Residues with any atom within 2.5Å of the core that aren't directly bound to it and not protein. For DNA, this is typically the complementary base.
5. Residues with any atom within 2.5Å of the core and not protein. For DNA, this is typically the complementary base and the residues immediately before and after the core on the same chain.
6. Residues with any atom within 2.5Å of the core or the residue immediately after it in the chain. For DNA, this is typically two consecutive base pairs.

For all nucleic acid ligands in BioLiP2 (which are typically at least 8 nucleic acid bases (note that this implies 16 DNA residues total per chain)), we randomly select three of type 1, two of type 2, one of type 3, one of type 4, and one of either type 5 or type 6 with 50% probability of either. After extracting these residues, proteins are capped as described above and nucleic acids are capped with an O3' OH group. MD was run on these systems as well; in addition to protein Cαs being restrained, the phosphate P and C4' positions of nucleic acids were restrained to keep the nucleic acid backbone intact. For the 400K MD, after discretizing into 10 equally spaced frames, rather than the first and last frame being used, the first two are used. This is because significant melting of nucleic acid structures occurs at this temperature. In our simulations, a large number of base pairs had moved apart to distances greater than 10Å even with the backbones frozen. The first two frames do not feature significant melting yet. For 300K MD, the first two and final frames are taken. This resulted in 770,537 RNA inputs and 570,087 DNA inputs.

Additionally, the Nucleic Acid Knowledge Base (NAKB) was mined for PDB structures of RNA and DNA with unusual structural motifs. Structures had been annotated as being: A-form DNA, Z-form DNA, two-strands forming a parallel helix, a helix composed of base triple steps, a helix composed of base quadruple steps, helix composed of steps with 5 or more hydrogen-bonded bases in plane, or a Holliday junction. The same selection procedure as above was used to sample the various unique environments present in these structures. This yielded 129,029 additional DFT inputs.

## C Metal Complexes

### C.1 Complexes from Architector

Metal complex chemistry spans the periodic table and includes unique features in terms of molecular symmetry and ligand-metal-ligand interactions. To generate complexes with diverse metal centers and ligand chemistry, the open-source Architector package [58] was used. As a brief summary, Architector takes as input a metal center, metal oxidation state, metal spin, metal coordination number, and ligands to surround the metal center. Ligands are specified by both a SMILES string [114] and a list of indices of the atoms of the SMILES string that are bound to the metal center. The combination of this information leads to a full molecular graph encoding a mononuclear molecular complex.

Possible metal oxidation states and coordination numbers were mined from a combination of the mendeleev python package [60] and a previously-published dataset derived from the Cambridge Structural Database (CSD summary) [58, 59]. The mendeleev package contains common oxidation states for each metal center, and the CSD summary dataset contains the frequency of coordination numbers (CNs) observed for each metal center in mononuclear metal complexes. Only CNs with more than 1% frequency in the CSD summary dataset and CNs less than 13 were selected. Since promethium complexes do not exist in the CSD summary dataset due to its radioactivity, relatively short half-life, and abundance, the coordination numbers and oxidation states known for neighboring samarium were applied to promethium. All oxidation states and coordination numbers sampled are found in Table 8. When sampling metal complexes from this space, metal and oxidation state were uniformly sampled followed by a uniform sampling from possible coordination numbers. Metal spin

states were assigned from the mendeleev electron configuration for the given oxidation state, which tends to give the highest reasonable spin for metal at a given oxidation state.

**Table 8** Metals, oxidation states, and coordination numbers sampled for OMol25 metal complexes.

metal	oxidation state	coordination numbers	metal	oxidation state	coordination numbers
Li	1	2,3,4,5,6	Pd	2	4,5
Be	2	4,5	Pd	4	4,5,6
Na	1	2,4,5,6,7,8,9	Ag	1	2,3,4,5,6
Mg	2	4,5,6,7	Cd	2	4,5,6,7,8
Al	3	4,5,6	In	3	4,5,6,7,8
K	1	2,3,4,5,6,7,8,9,10	Sn	2	3,4,5,6,7,8,9
Ca	2	4,5,6,7,8,9	Sn	4	4,5,6,7
Sc	3	4,5,6,7,8	Cs	1	4,5,6,7,8,9,10,11,12
Ti	2	5,6,7,8	Ba	2	4,5,6,7,8,9,10,11,12
Ti	3	4,5,6,7,8	La	3	4,5,6,7,8,9,10,11
Ti	4	4,5,6,7,8	Ce	3	4,5,6,7,8,9,10
V	2	4,5,6,7,8	Ce	4	4,5,6,7,8,9,10
V	3	4,5,6,7	Pr	3	5,6,7,8,9,10
V	4	4,5,6,8	Nd	3	4,5,6,7,8,9,10
V	5	4,5,6	Pm	3	4,5,6,7,8,9,10
Cr	2	4,5,6,7	Sm	3	4,5,6,7,8,9,10
Cr	3	5,6,7	Eu	2	5,6,7,8,9,10
Cr	6	4,5,6,7	Eu	3	6,7,8,9,10
Mn	2	4,5,6,7	Gd	3	5,6,7,8,9,10
Mn	3	4,5,6	Tb	3	4,6,7,8,9,10
Mn	4	4,5,6	Dy	3	4,5,6,7,8,9,10
Mn	6	4,5,6,7	Ho	3	4,5,6,7,8,9,10
Mn	7	4,5,6,7	Er	3	4,5,6,7,8,9,10
Fe	2	4,5,6,7	Tm	3	5,6,7,8,9,10
Fe	3	4,5,6,7	Yb	3	4,5,6,7,8,9,10
Co	2	4,5,6,7	Lu	3	4,5,6,7,8,9,10
Co	3	4,5,6	Hf	4	4,5,6,7,8
Ni	2	4,5,6	Ta	5	4,5,6,7,8
Cu	1	2,3,4,5	W	4	4,5,6,7,8
Cu	2	4,5,6	W	6	4,5,6,7,8
Zn	2	4,5,6	Re	4	4,5,6,7
Ga	3	4,5,6	Re	7	4,5,6
Rb	1	5,6,7,8,9,10,11,12	Os	4	4,5,6
Sr	2	4,5,6,7,8,9,10	Ir	3	4,5,6
Y	3	4,5,6,7,8,9	Ir	4	4,5,6,7
Zr	4	4,5,6,7,8,9	Pt	2	4,5
Nb	5	4,5,6,7,8	Pt	4	4,6
Mo	4	4,5,6,7,8	Au	1	2,3,4
Mo	6	4,5,6	Au	3	4,5
Tc	4	4,6,7	Hg	1	2,3,4,5,6,7,8
Tc	7	4,6,7	Tl	1	2,3,4,5,6,7,8,10
Ru	3	5,6,7	Tl	3	2,4,5,6,7,8
Ru	4	4,5,6	Pb	2	4,5,6,7,8,9,10
Rh	3	4,5,6	Pb	4	4,5,6,7,8
Pd	0	4,5	Bi	3	4,5,6,7,8,9

For ligand sampling, sets of ligands have previously been curated by several groups from the CSD, but these focused on accurately extracting as many different ligands as possible for computational design efforts, rather than for large-scale computation [115, 116]. Further, such datasets contain reported accuracies of only  $\approx 90$ , so

manual curation is regardless required. As the CSD represents a set of experimentally derived structures and it is much easier to synthesize ligands from around known chemistries, these ligand datasets tend to consist of many repeated chemical motifs [117]. To overcome these limitations, here, a set of 723 base ligands was manually curated from the CSD summary dataset based on coordinating atom type diversity and frequency of occurrence in the CSD summary dataset.

The curation process which generated these 723 ligands proceeded as follows. Ligands were selected based on a total of 32 unique SYBYL atom types (*e.g.*, C.2, N.3, N.ar ...) of both metal-coordinated and uncoordinated atoms. For each unique SYBYL atom type the 10 most frequent ligands along with 10 randomized ligands from the CSD summary dataset containing either the metal-bound atom type or non-metal-bound atom type were selected. After removing duplicates, a total of 719 ligands were produced, followed by manual inspection of all ligands resulting in the addition of 3 ligands encoding additional organometallic bonds (cyclopentadienyl, pentamethylcyclopentadienyl, and bulky cyclopentadienyl) and hydride for 723 ligands from which to sample. Basic statistics of the OMol25 sampling ligands and CSD ligands are highlighted in Figure 5. Note the resulting curated ligand set spans a wide range of number of atoms (1-120), metal-coordinating atom types (32), denticities (1-8), and charges (-4 to +1). Additionally, note that metal-coordinating atom type diversity is higher, denticities and number of atoms are lower, and distributions of charges are similar for the OMol25 ligand sampling compared to the unique ligands in the CSD.

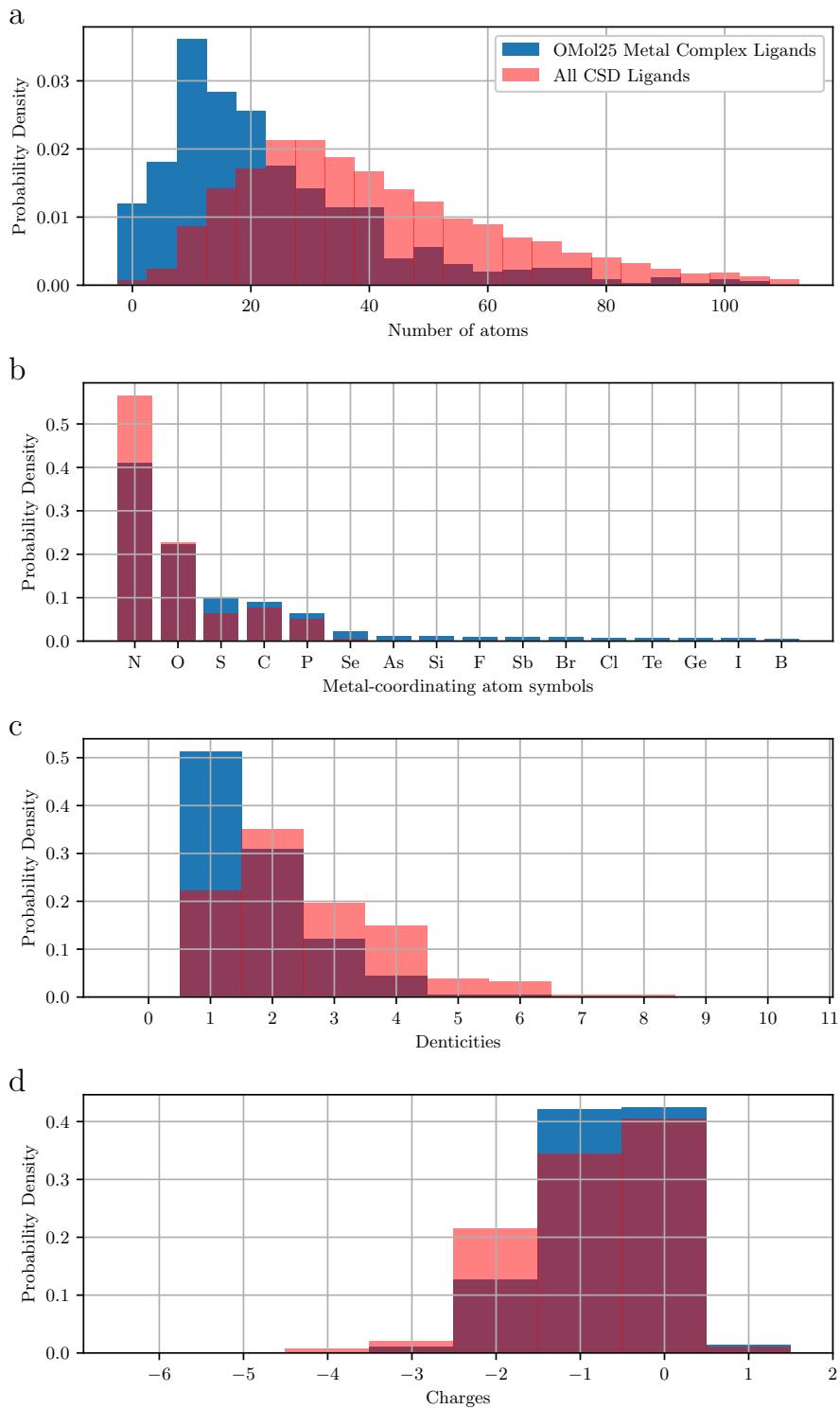
A ligand sampling algorithm was created to build the required Architector inputs. First, a metal, oxidation state, and coordination number were randomly selected. Ligands were randomly and iteratively selected until no remaining coordination sites on the metal remained. With each selection, only ligands that could fit at the remaining coordination sites, did not move the charge of the overall complex outside of the range -2 and +4, and did not make the complex exceed a set number of atoms (both 120 and 250 were used in this dataset) were permitted. From these remaining ligands, a weighted sampling was performed by denticity to ensure the per-site likelihood of all possible binding sites was equal. This was done to avoid under sampling high-denticity ligands. The result of the sampling pipeline is a massively chemically diverse, largely high-spin, set of Architector inputs for metal complexes of a size and charge likely to converge well in DFT calculations.

From here, Architector was run on samples of up to 1M metal complexes. For the most part, Architector was run with default parameters with a few exceptions: to obtain a variety of metal coordination environments and symmetries, up to 10 initial ligand symmetries were screened and up to 3 conformers were requested to be relaxed for every core geometry (*e.g.*, octahedral, trigonal prismatic, hexagonal planar). For initial screening and final relaxation of the generated complexes, GFN2-xTB was used [61]. An additional Architector parameter tuned for this work was allowing a looser graph distance tolerance of 2, which allows generated molecules to have bond distances up to twice as long as the defined molecular graph sum of covalent radii, which was done to allow for structures with more dissociated ligands when sampling at higher coordination numbers. The code to reproduce the sampling workflow and Architector generation are found in the OMol repository, while the curation of metal and ligand datasets can be found in a separate repository [118].

## C.2 Complexes from the Crystallography Open Database (COD)

In order to evaluate the performance of models trained on computationally-generated metal complexes, a test set of experimentally determined metal complexes was obtained. The Crystallography Open Database (COD) contains over 500,000 experimental crystal structures extracted from various journals and publications under a permissive license and a convenient SQL interface. We extracted all crystal structures that were marked as not containing disorder and whose unit cell contents matched the indicated contents and which contained both carbon (as we are interested especially in complexes which contain organic ligands) and hydrogen (as a proxy for checking if the crystal structure has been decorated with hydrogens, as is not always the case, and which would require manual intervention to obtain molecules suitable for DFT). This yields 305,535 structures. These structures were then processed by a workflow to contract molecules into valid structures (*i.e.*, without being split by periodic boundary conditions, crystallographic symmetry, etc.), and assign Lewis structures. This Lewis structure information was used to determine the metal oxidation state and, from there, the spin of the system.

Systems which did not contain exactly one metal center, where the absolute value of the charge on any given molecule was greater than 2, which contained actinides, which had more 250 atoms, or which had colliding



**Figure 5** Histogram cross-sections of ligands sampled for metal complex datasets. a) total number of atoms in the ligands b) metal-coordinating atoms from the ligands c) denticity of the ligands and d) the total charges on the ligands.

atoms (defined as having a bond distance of less than 0.55 times the sum of covalent radii for the two atoms) were discarded. While multi-metallic centers are of considerable interest, substantial manual intervention is required to determine accurate oxidation states in multimetallic complexes. Given the large scale of the OMol25 dataset and the significant effort that would be required, we elected not to include multimetallic complexes at this time. Additionally, systems which contained carbon atoms with more than 4 bonds to non-metals, hydrogen atoms with more than 1 bond to non-metals, and boron atoms forming more than 4 bonds to non-metals were also rejected. Such situations very frequently indicated undocumented disorder in the crystal structure, or, in the case of boron, a cluster structure. Boron clusters are not well described by a 2-electron per bond Lewis structure, and so the automated procedure for assigning charges to these systems often produces nonsensical results for such structural motifs. Lastly, if the assigned metal oxidation states was not one of the common oxidation states for a given metal, the assignment was considered suspect and rejected. We note that rejections based on uncommon oxidation states made up fewer than 10% of the remaining structures at this point. These filters resulted in 70,004 structures with reasonably trustworthy structure and electronic assignment. Second and third row complexes were assigned low-spin configurations and first row complexes were calculated both in their highest and lowest reasonable spin configurations, resulting in 90,359 DFT inputs.

## D Electrolytes

Electrolytes are collections of ions and solvent molecules which feature complex intermolecular interactions, dynamic short-range solvation and in some cases long-range network structures. They are also important for a range of practical applications, from biochemistry, biotechnology, and medicine (the cellular medium is essentially an aqueous electrolyte solution) to battery energy storage, chemical synthesis, and catalysis. In this thrust, we aim to capture solvation structures, the chemical diversity of the small molecules used in electrolyte solutions, and electrolyte reactivity.

### D.1 Bulk Electrolytes

Many electrolyte properties of interest, including transport (*e.g.*, conductivity, transference number, viscosity), (de)solvation, and wetting, are inherently dynamic in nature. These are properties that static calculations of isolated molecules cannot capture. To simulate electrolyte dynamics well, ML models need diverse structures in their training set; for electrolytes specifically, these structures must include clusters of noncovalently bound ions and solvent molecules. To this end, we employ classical molecular dynamics (MD) as a means to sample clusters of solvated ions. The data obtained from MD simulations have been initiated in two generations. Below, we describe each generation of MD-based data:

*First generation.* In the first generation, we aimed to compile a list of electrolyte mixtures that are both relevant for practical applications and chemically diverse. To this end, we used the literature as a reference point and supplemented it with a workflow that generates random mixtures [135–214]. From the literature, we have included systems of technological relevance such as batteries (including Li-, Na-, K-, Mg-, Ca-, Zn-, and Al-ion batteries as well as Pb-acid batteries). We have also included examples from redox flow batteries, catalytic systems (for instance Li-mediated and Ca-mediated electrochemical ammonia synthesis), ionic liquids, and molten salt electrolytes, as well as several common aqueous and non-aqueous solutions. In total, we obtained 223 unique electrolyte systems, encompassing different types of solvents (*e.g.*, protic, aprotic, ionic liquids), concentration (*e.g.*, dilute and high-concentration), electrolyte components (*e.g.*, solvents and salts), and number of components.

Next, we generated random electrolytes in five classes: (1) salts dissolved in protic solvents, (2) salts dissolved in aprotic solvents, (3) ionic liquid mixtures, (4) aqueous solutions, and (5) molten salts. Aqueous solutions, despite being a subset of protic solutions, received special attention due to their importance and the challenges in capturing their dynamic behavior. Each class was chosen with a fixed probability (40% protic, 40% aprotic, 10% ionic liquid, 5% molten salt, 5% aqueous). These main categories are then complemented by fully randomized formulations (110 unique systems), creating a comprehensive exploration of electrolyte chemical space. Every random electrolyte mixture is built from a curated list of electrolyte components based on our initial literature review, where components are labeled to ensure compatibility between cations,

**Table 9** List of cations used for MD simulations of bulk electrolytes and the force field parameters, sorted by valency. Abbreviations: BMPyrr<sup>+</sup> (1-butyl-1-methylpyrrolidinium), BPPyrr<sup>+</sup> (1-butyl-1-propylpyrrolidin-1-ium), EMIM<sup>+</sup> (1-ethyl-3-methylimidazolium), BMIM<sup>+</sup> (1-butyl-3-methylimidazolium), MPA<sup>+</sup> (1-Methoxypalan-2-yiazanium), AMA<sup>+</sup> (acetyl(methyl)azanium), BMEA<sup>+</sup> (bis(2-methoxyethyl)azanium), MPPip<sup>+</sup> (N-methyl-N-propylpiperidinium), PMPyrr<sup>+</sup> (N-propyl-N-methylpyrrolidinium), DEMA<sup>+</sup> (N,N-Diethyl-2-methoxy-N-methylethanaminium), PQ<sup>2+</sup> (paraquat), TBP<sup>+</sup> (tetrabutyl-phosphonium).

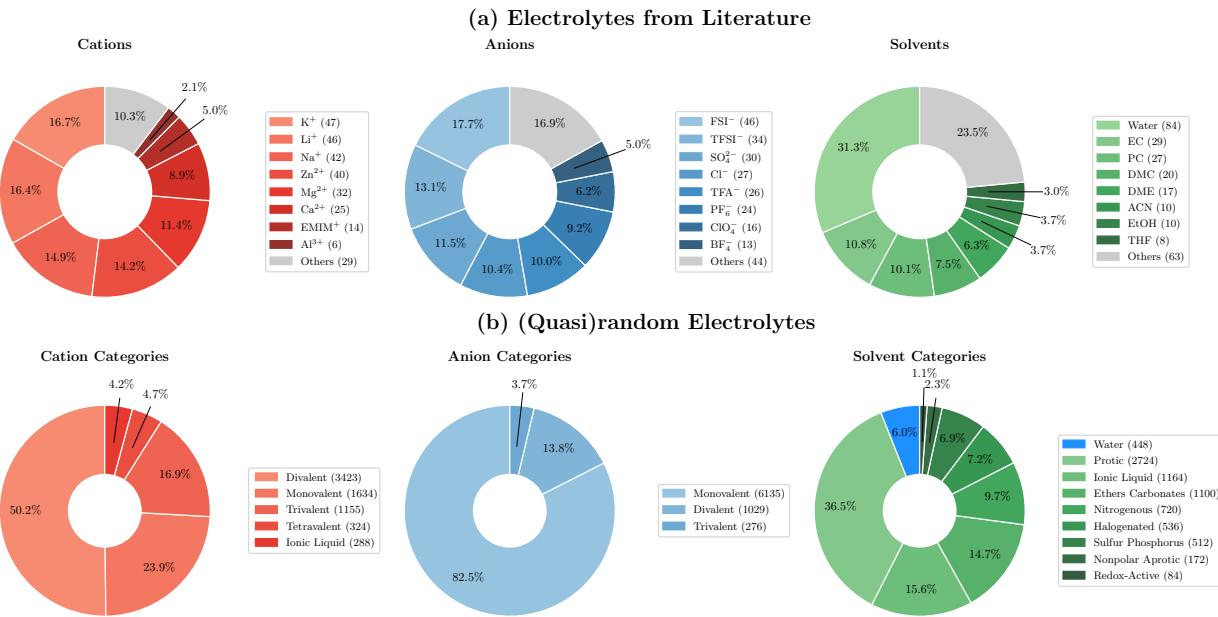
Monovalent Cations			Divalent Cations		
Name	FF Parameters		Name	FF Parameters	
	1st Gen	2nd Gen		1st Gen	2nd Gen
H <sub>3</sub> O <sup>+</sup>	OPLS4 [68]	Jang, et al. (2004) [119]	Be <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Li <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Mg <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Na <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Ca <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
NH <sub>4</sub> <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	V <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
K <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Cr <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Cu <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Mn <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Rb <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Fe <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Ag <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Co <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Cs <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Ni <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Ti <sup>+</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	Cu <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
(VO <sub>2</sub> ) <sup>+</sup>	OPLS4 [68]	Gupta, et al. (2016) [121]	Zn <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
TEMPO <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]	Sr <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
<b>Trivalent Cations</b>					
Name	FF Parameters		Name	FF Parameters	
	1st Gen	2nd Gen		1st Gen	2nd Gen
Al <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]	Pd <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
V <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]	Ag <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Cr <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]	Hg <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Fe <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]	Pb <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
In <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]	PQ <sup>2+</sup>	OPLS4 [68]	OPLS-AA [122–124]
Y <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]	Pt <sup>2+</sup>	OPLS4 [68]	Li, et al. (2020) [126]
Tl <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]			
Ti <sup>3+</sup>	OPLS4 [68]	Li, et al. (2021) [125]	<b>Tetravalent Cations</b>		
Name	FF Parameters		Name	FF Parameters	
	1st Gen	2nd Gen		1st Gen	2nd Gen
EMIM <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]	Ti <sup>4+</sup>	OPLS4 [68]	Li, et al. (2021) [125]
BMIM <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]	Mn <sup>4+</sup>	OPLS4 [68]	Li, et al. (2021) [125]
BMPyrr <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]	Zr <sup>4+</sup>	OPLS4 [68]	Li, et al. (2021) [125]
BPPyrr <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]	<b>Ionic Liquid Cations (continued)</b>		
PMPyrr <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]	Name	FF Parameters	
MPPip <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]		1st Gen	2nd Gen
DEMA <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]	AMA <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]
			MPA <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]
			BMEA <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]
			Choline	OPLS4 [68]	OPLS-AA [122–124]
			TBP <sup>+</sup>	OPLS4 [68]	OPLS-AA [122–124]
			PQ <sup>2+</sup>	OPLS4 [68]	OPLS-AA [122–124]

**Table 10** Anionic species with their associated force field parameters, sorted by valency and grouped by parameter types. Abbreviations: TFSI<sup>-</sup> (bis(trifluoromethylsulfonyl)imide), BOB<sup>-</sup> (bisoxalatoborate), HMDS<sup>-</sup> (hexamethyldisilazide), TEMPO<sup>-</sup> (2,2,6,6-tetramethylpiperidin-1-yl)oxide), FSI<sup>-</sup> (bis(fluorosulfonyl)imide), TFA<sup>-</sup> (triflate), Ac<sup>-</sup> (acetate), MeO<sup>-</sup> (methoxide), EtO<sup>-</sup> (ethoxide), iPrO<sup>-</sup> (isopropoxide), ETFA<sup>-</sup> (ethyl trifluoroacetate), DFP<sup>-</sup> (difluorophosphate), DHB<sup>2-</sup> (2,5-dihydroxy-[1,4]-benzoquinone dianion), EDO<sup>2-</sup> (ethane-1,2-diolate), BOB<sup>-</sup> (bisoxalatoborate).

<b>Monovalent Anions</b>			<b>Divalent Anions</b>		
Name	FF Parameters		Name	FF Parameters	
	1st Gen	2nd Gen		1st Gen	2nd Gen
F <sup>-</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	CO <sub>3</sub> <sup>2-</sup>	OPLS4 [68]	GAFF2 [127, 128]
Cl <sup>-</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	SO <sub>4</sub> <sup>2-</sup>	OPLS4 [68]	GAFF2 [127, 128]
Br <sup>-</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	EDO <sup>2-</sup>	OPLS4 [68]	OPLS-AA [122-124]
I <sup>-</sup>	OPLS4 [68]	Sengupta, et al. (2021) [120]	DHB <sup>2-</sup>	OPLS4 [68]	OPLS-AA [122-124]
OH <sup>-</sup>	OPLS4 [68]	GAFF2 [127, 128]	<b>Trivalent Anions</b>		
BH <sub>4</sub> <sup>-</sup>	OPLS4 [68]	Mamatlukov, et al. [129]	FF Parameters		
AsF <sub>6</sub> <sup>-</sup>	OPLS4 [68]	Ishida, et al. [130]	Name	FF Parameters	
NO <sub>3</sub> <sup>-</sup>	OPLS4 [68]	OPLS-2009IL [131, 132]		1st Gen	2nd Gen
ClO <sub>4</sub> <sup>-</sup>	OPLS4 [68]	OPLS-2009IL [131, 132]	PO <sub>4</sub> <sup>3-</sup>	OPLS4 [68]	GAFF2 [127, 128]
BF <sub>4</sub> <sup>-</sup>	OPLS4 [68]	OPLS-2009IL [131, 132]			
PF <sub>6</sub> <sup>-</sup>	OPLS4 [68]	OPLS-2009IL [131, 132]			
TFSI <sup>-</sup>	OPLS4 [68]	OPLS-2009IL [131, 132]			
TFA <sup>-</sup>	OPLS4 [68]	OPLS-2009IL [131, 132]			
HCO <sub>3</sub> <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
Ac <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
MeO <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
EtO <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
iPrO <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
HMDS <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
TEMPO <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
ETFA <sup>-</sup>	OPLS4 [68]	OPLS-AA [122-124]			
BOB <sup>-</sup>	OPLS4 [68]	Wang, et al. (2014) [133]			
FSI <sup>-</sup>	OPLS4 [68]	GAFF2 [127, 128]			
DFP <sup>-</sup>	OPLS4 [68]	GAFF2 [127, 128]			

**Table 11** Solvent species with their associated force field parameters, organized by solvent type. Abbreviations: ACN (acetonitrile), BuCN (butyronitrile), DMSO (dimethylsulfoxide), DMF (N,N-dimethylformamide), NMA (N-methylacetamide), MPA (1-methoxy-2-propylamine), DHQ (dihydroquinone), TMPOH (2,2,6,6-tetramethylpiperidin-1-ol), BMEA (bis(2-methoxyethyl)amine), DCM (dichloromethane), DBE (1,2-dibromoethane), DCE (1,2-dichloroethane), NM (nitromethane), NB (nitrobenzene), PYR (pyridine), TEA (triethylamine), THF (tetrahydrofuran), DEE (diethyl ether), DME (dimethoxyethane), DTD (ethylene sulfate), EG (ethylene glycol), EC (ethylene carbonate), FEC (fluoroethylene carbonate), PC (propylene carbonate), DMC (dimethyl carbonate), EMC (ethyl methyl carbonate), EtAc (ethyl acetate), ETFA (ethyl trifluoroacetate), GBL ( $\gamma$ -butyrolactone), EMS (ethyl methyl sulfonate), TMP (trimethylphosphate), TEP (triethylphosphate), HMPA (hexamethylphosphoramide), AQ (anthraquinone), BQ (benzoquinone), DHBQ (2,5-dihydroxy-1,4-benzoquinone), TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl), TTE (tris(2,2,2-trifluoroethyl) borate).

<b>Protic Solvents</b>			<b>Ethers, Carbonates, Esters &amp; Ketones</b>		
Name	FF Parameters		Name	FF Parameters	
	1st Gen	2nd Gen		1st Gen	2nd Gen
Water	TIP3P [68]	TIP4P/2005f [134]	THF	OPLS4 [68]	OPLS-AA [122–124]
MeOH	OPLS4 [68]	OPLS-AA [122–124]	DEE	OPLS4 [68]	OPLS-AA [122–124]
EtOH	OPLS4 [68]	OPLS-AA [122–124]	DME	OPLS4 [68]	OPLS-AA [122–124]
iPrOH	OPLS4 [68]	OPLS-AA [122–124]	Diglyme	OPLS4 [68]	OPLS-AA [122–124]
EG	OPLS4 [68]	OPLS-AA [122–124]	EC	OPLS4 [68]	OPLS-AA [122–124]
Urea	OPLS4 [68]	OPLS-AA [122–124]	FEC	OPLS4 [68]	OPLS-AA [122–124]
AcNH <sub>2</sub>	OPLS4 [68]	OPLS-AA [122–124]	PC	OPLS4 [68]	OPLS-AA [122–124]
NMA	OPLS4 [68]	OPLS-AA [122–124]	DMC	OPLS4 [68]	OPLS-AA [122–124]
Pyrrole	OPLS4 [68]	OPLS-AA [122–124]	EMC	OPLS4 [68]	OPLS-AA [122–124]
MPA	OPLS4 [68]	OPLS-AA [122–124]	EtAc	OPLS4 [68]	OPLS-AA [122–124]
DHQ	OPLS4 [68]	OPLS-AA [122–124]	DTD	OPLS4 [68]	OPLS-AA [122–124]
TMPOH	OPLS4 [68]	OPLS-AA [122–124]	ETFA	OPLS4 [68]	OPLS-AA [122–124]
BMEA	OPLS4 [68]	OPLS-AA [122–124]	GBL	OPLS4 [68]	OPLS-AA [122–124]
			Acetone	OPLS4 [68]	OPLS-AA [122–124]
<b>Nonpolar Aprotic Solvents</b>					
Name	FF Parameters		<b>Nitrogen-Containing Solvents</b>		
	1st Gen	2nd Gen	Name	FF Parameters	
Pentane	OPLS4 [68]	OPLS-AA [122–124]	ACN	OPLS4 [68]	OPLS-AA [122–124]
Hexane	OPLS4 [68]	OPLS-AA [122–124]	BuCN	OPLS4 [68]	OPLS-AA [122–124]
Benzene	OPLS4 [68]	OPLS-AA [122–124]	DMF	OPLS4 [68]	OPLS-AA [122–124]
Toluene	OPLS4 [68]	OPLS-AA [122–124]	NM	OPLS4 [68]	OPLS-AA [122–124]
PhF	OPLS4 [68]	OPLS-AA [122–124]	NB	OPLS4 [68]	OPLS-AA [122–124]
CCl <sub>4</sub>	OPLS4 [68]	OPLS-AA [122–124]	PYR	OPLS4 [68]	OPLS-AA [122–124]
			TEA	OPLS4 [68]	OPLS-AA [122–124]
<b>Halogenated Solvents</b>					
Name	FF Parameters		<b>Sulfur &amp; Phosphorus Solvents</b>		
	1st Gen	2nd Gen	Name	FF Parameters	
CHCl <sub>3</sub>	OPLS4 [68]	OPLS-AA [122–124]	DMSO	OPLS4 [68]	OPLS-AA [122–124]
CHBr <sub>3</sub>	OPLS4 [68]	OPLS-AA [122–124]	Sulfolane	OPLS4 [68]	OPLS-AA [122–124]
DCM	OPLS4 [68]	OPLS-AA [122–124]	EMS	OPLS4 [68]	OPLS-AA [122–124]
DBE	OPLS4 [68]	OPLS-AA [122–124]	TMP	OPLS4 [68]	OPLS-AA [122–124]
DCE	OPLS4 [68]	OPLS-AA [122–124]	TEP	OPLS4 [68]	OPLS-AA [122–124]
			HMPA	OPLS4 [68]	OPLS-AA [122–124]
<b>Redox-Active Compounds</b>					
Name	FF Parameters				
	1st Gen	2nd Gen			
AQ	OPLS4 [68]	OPLS-AA [122–124]			
BQ	OPLS4 [68]	OPLS-AA [122–124]			
DHBQ	OPLS4 [68]	OPLS-AA [122–124]			
TEMPO	OPLS4 [68]	OPLS-AA [122–124]			
TTE	OPLS4 [68]	OPLS-AA [122–124]			

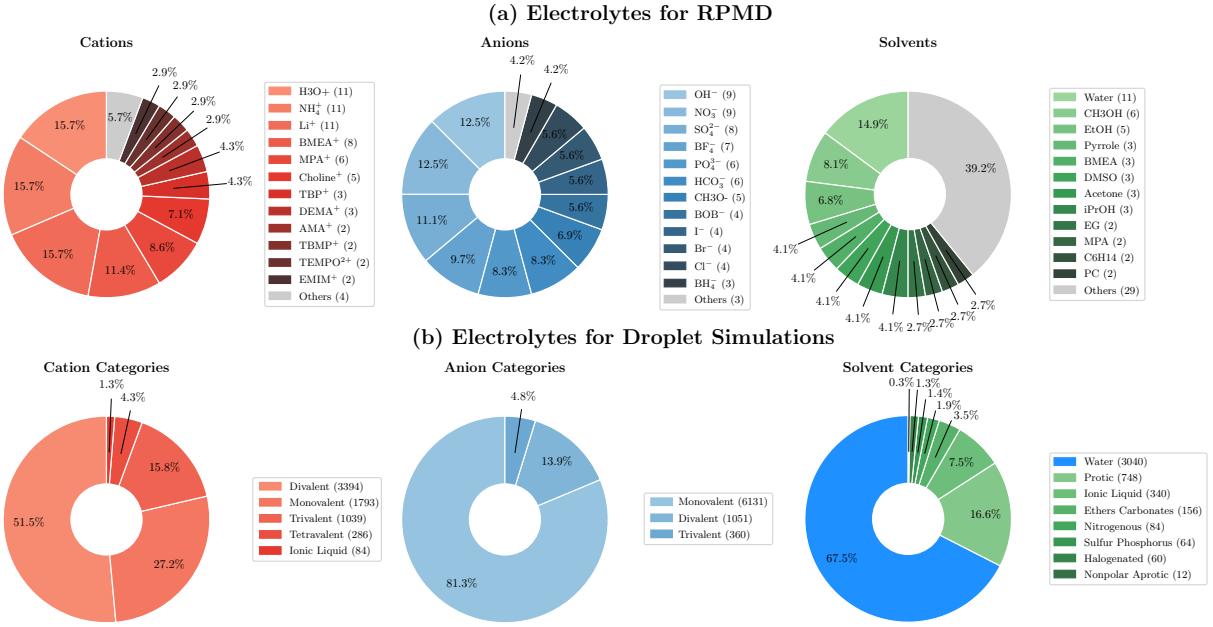


**Figure 6** (a) Statistics of electrolyte systems obtained from the literature. Alkali ions ( $\text{Li}^+$ ,  $\text{K}^+$ , and  $\text{Na}^+$ ) make up the top three cations, while bis(fluorosulfonyl)imide ( $\text{FSI}^-$ ), bis(trifluoromethylsulfonyl)imide ( $\text{TFSI}^-$ ), and sulfate ( $\text{SO}_4^{2-}$ ) make up the top three anions. Water stands out as the most common solvent, with various carbonates like ethylene carbonate (EC) and propylene carbonate (PC) forming the second tier of common solvents. (b) Statistics of (quasi)random electrolyte systems. The final set includes 54 unique cation types, dominated by monovalent ions, 28 anion types, and 79 different solvents with compositions averaging approximately 2 cations, 2 anions, and 2 solvents per unique formulation.

anions, and solvents; see Tables 9 to 11 for the final list. Charge neutrality is maintained by solving charge balance equations, which determines the stoichiometric coefficients for the selected ionic species. For each base composition, we produce four variant simulations, adjusting both temperature and concentration to reasonable high and low values. For temperatures, we ensure that the lowest and highest temperatures are 5% above and below the freezing and boiling/decomposition temperatures, respectively. The concentrations are chosen such that, if the electrolyte system behaves as an ideal solution, the inter-solute separation distances are 2.25 and 1.75 nm, respectively. Note that molten salt systems have no solutes, by definition. For these systems, concentration was not varied, and temperatures were set to 1000 K and 1300 K. Overall, the random electrolyte dataset we produced includes 884 unique compositions and 3,536 simulations in total.

We set up classical MD simulations using Desmond [215] and Schrödinger's disordered system builder [216]. Simulations were performed using the OPLS4 force field [68] and the TIP3P water model with a van der Waals scaling factor of 0.5. Custom charges were generated for  $[\text{AsF}_6]^-$ , TEMPO,  $[\text{VO}]^{2+}$ , and  $[\text{VO}_2]^+$ . After system generation, we prepared a Desmond multisim script (.msj file) that defines a simulation protocol with multiple stages. The protocol begins with an initial relaxation using Brownian dynamics at 10 K, followed by a series of molecular dynamics simulations with gradually increasing time steps. The system then transitions from NVT to NPT ensemble, undergoes a high-temperature annealing phase at 700 K, and concludes with a production run in the NPT ensemble at the specified temperature using a 2 fs time step for 250 ns and snapshots were taken every 2.5 ns.

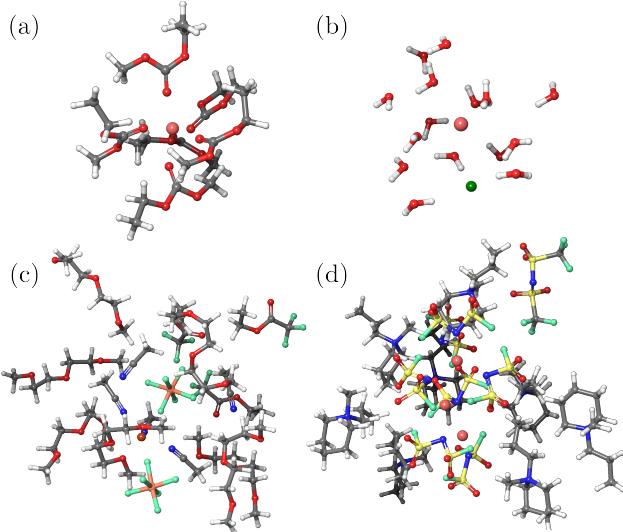
We also prepared additional clusters which contained an OOD anion, cation, or solvent, via the methods described here. The OOD anions were difluorooxalatoborate (DFOB) and hexamethyldisilazide (HMDS), the OOD cations were tetramethylammonium (TMA) and tetraethylphosphonium, and the OOD solvents were toluene and 1,4-dioxane. These clusters contained at least one of the OOD species and any number of ID species. We also prepared a set of clusters that consists only of OOD species. The set containing OOD anions, ID anions, ID cation, and ID solvents was labeled “OOD Anions” and was discussed in the main text Section 2.8.1. Results for the similarly defined sets of cations and solvents, as well as the all OOD set is described in Appendix I.4.



**Figure 7** (a) Statistics of electrolyte systems obtained for RPMD simulations. Overall, the final set contains 16 unique cations, 13 anions, and 38 different solvents. The cation set includes both simple ions like H<sub>3</sub>O<sup>+</sup>, NH<sub>4</sub><sup>+</sup> (ammonium), and Li<sup>+</sup>, as well as more complex organic cations derived from ionic liquids. Anions include simple species like OH<sup>-</sup> and NO<sub>3</sub><sup>-</sup>, as well as boron-containing anions such as BF<sub>4</sub><sup>-</sup>, BH<sub>4</sub><sup>-</sup>, and BOB<sup>-</sup>. The diversity of solvents is particularly broad, including water (H<sub>2</sub>O), alcohols (methanol, ethanol), and organic carbonates. (b) Statistics of random electrolyte systems obtained for spherical droplet simulations. Each electrolyte system allows up to 4 cations, 4 anions, and 4 solvents per system. The resulting dataset for spherical droplet simulations maintains substantial chemical diversity with 50 unique cations, 29 anions, and 60 different solvents.

**Second Generation.** The second generation of electrolyte MD simulations sought to capture specific nuclear quantum (using ring-polymer molecular dynamics, or RPMD) and interfacial effects (using droplet simulations). RPMD approximates quantum nuclei as classical ring polymers to incorporate nuclear quantum effects, which are crucial for accurate descriptions of zero-point energy, tunneling, and hydrogen bonding [217]. RPMD represents each atom as a ring of harmonically bonded beads to capture a wider range of conformations than classical MD [217], making such configurations valuable for our dataset. For RPMD, we selected 50 electrolyte systems containing light atoms (*e.g.*, H, Li, and B) where nuclear quantum effects are expected to be most important. For all systems, we chose relatively low temperatures, ranging from 164.85 to 325.71 K (averaging 247 K or  $-26^{\circ}\text{C}$ ), where quantum effects of the light atoms are expected to be dominant. Meanwhile, spherical droplet simulations provide an opportunity to sample solvation clusters affected by interfacial effects. To this end, we generated 3,734 electrolyte compositions where the compositions are generated in the same way as the first generation with 20% salt in protic solvents, 15% salt in aprotic solvents, and 65% aqueous electrolytes. Note that the second generation focuses heavily on aqueous systems to compensate for their under-representation in the first generation, where we have elected to use rigid water models.

For the second generation, we used OpenMM version 8.1.1 [69], which has the capabilities to perform RPMD and droplet simulations. Force-field parameters were obtained from multiple sources [120, 122–128, 131, 132, 134] and we have set up an additional workflow to translate a curated set of parameters into OpenMM XML force field files that could be readily used and mixed for different electrolyte compositions. Water parameters were derived from the TIP4P/2005f model [134]. Metal cations and some small anions (*e.g.*, I<sup>-</sup>) used Lennard-Jones parameterization benchmarked with the TIP3P-FB water model [120, 125, 126]. Neutral solvents and many ions were parameterized using Optimized Potentials for Liquid Simulations – All-Atom (OPLS-AA) force-field parameters with 1.14\*CM1A atomic partial charges obtained from the LigParGen online server [122–124]. For selected anions, we employed complementary parameters from various sources such as general AMBER force field (GAFF) [127, 128], OPLS-2009IL, which is a variant of OPLS-AA designed for ionic liquids [131, 132],



**Figure 8** Examples of solvation shells extracted from MD simulations. (a)  $\text{Li}^+$  ion in ethylene carbonate (EC) ethyl methyl carbonate (EMC), (b)  $\text{Na}-\text{Cl}$  pair in water, (c)  $\text{Cu}^{2+}$  paired with arsenic hexaflouride in an organic mixture, and (d) Sodium ions in ionic liquid.

and customized parameters from the literature [119, 121, 129, 130, 133]. While mixing parameters from different forcefields is generally discouraged due to potential incompatibilities in energy terms and molecular topologies, our aim is not to simulate accurate dynamics but to sample diverse electrolyte configurations. Chemically implausible or high-energy structures are not a drawback; in fact, they are crucial for training ML models to recognize and avoid unstable regions of the potential energy surface.

For RPMD, we set up an OpenMM simulation using a Langevin integrator with Monte Carlo barostat for NPT ensemble simulations. We first perform energy minimization to relax any high-energy configurations followed by production runs. The simulation operates at a specified temperature and pressure (1.0 bar), with a default timestep of 1 fs, though these parameters were adjusted depending on the system requirements. For light elements, we used 32 beads and employed a ring-polymer contraction [218] to reduce computational overhead, where direct space (short-range) forces use 3 beads and reciprocal space (long-range) forces use just 1 bead, while the rest of intramolecular forces use the full 32 beads. For droplet simulations, we configured a nonperiodic simulation and added a spherical harmonic container force. This container uses a harmonic potential with a force constant of 100 kJ/mol/nm<sup>2</sup> that activates when particles move beyond a defined radius. Both RPMD and droplet simulations were run for 20 ns and snapshots were taken every 0.2 ns.

*Shell Extraction.* As MD simulations of bulk electrolytes often have considerable spatial and temporal redundancy, it is necessary to downselect diverse structures for inclusion in the final dataset. For each electrolyte system, the original MD trajectory was subsampled such that there are 100 equally spaced frames (every 2.5ns for first generation systems and every 0.2 ns for second generation); all of these frames were used as starting points for extraction. Where applicable, shells centered around each solute molecule were then extracted. To make the extraction procedure simple and scalable, we chose system-agnostic initial shell radii of 3Å and 5Å, representing first and second coordination shells. All molecules which had any atom within this distance from the central solute were extracted as part of the shell. Systems larger than a maximum atom count of 350 atoms were discarded. The extracted shells were then sorted into groups of conformers. From each group, a final set of 20 diverse structures were selected according to an approximate maximization of the minimum Root Mean Squared Deviation (RMSD) between structures (a random structure was chosen as seed and the most dissimilar structure by RMSD to all the structures previously taken were added iteratively to the list of accepted structures). We chose this approximate scheme since the Max-Min Diversity problem - the task of selecting a subset of  $m$  elements from a set of  $n$  elements such that the sum of the distances between the chosen elements is maximized - is in general NP hard [219, 220]. To explicitly target solvent-solvent interactions, the extraction procedure was repeated for solvent-centered shells; for this procedure, only shells

that contained no solute were retained. These extracted clusters were then randomly, uniformly sampled for each central solute or solvent species, in the case of first generation systems. In the case of second generation systems, all ion-centered shells were taken, along with a sufficient number of solvent systems to more or less equal the ion-centered clusters. No attempt was made to compare clusters by RMSD between electrolyte system runs with the same species (e.g., between multiple systems with the same solvent for solvent-centered shells). As such, there may be lower diversity than exhaustive application of the Max-Min Diversity algorithm (either exact or approximate) over all electrolyte systems would imply. Samples of the results for the shell extraction can be found in [Figure 8](#).

## D.2 Small Molecule Generation

A library of small, “electrolyte-like” molecules were generated by selecting functional groups from a predefined collection and substituting these functional groups quasirandomly onto “core” structures. This library primarily serves to sample diverse bonded interactions. Note that the term “electrolyte-like” is used because many of the generated molecules are likely not useful for any practical electrolytes and may not even be synthesizable, but the “core” structural motifs reflect common electrolyte structures.

A collection of 77 “core” structures were selected by surveying the literature, in particular focusing on diverse battery chemistries and ionic liquids. Most cores contain only main-group elements, though there are several containing transition metals (*e.g.*, Nb and Ta). We note that 36 of the 77 cores were taken from the Ionic Liquid — Electrochemical Stability Window (IL-ESW) dataset developed by Moraes et al. Cores have as few as one site that can be substituted onto and as many as twelve. Seventy-two core structures were classified as in-distribution (ID), and five were designated to be out-of-distribution (OOD).

Similarly, 240 diverse functional groups containing only main-group elements were collated, including approximately 130 taken from the IL-ESW. Of these 240 functional groups, 228 were designated as ID and twelve were designated as OOD.

The general procedure for generating a small molecule library is as follows: for a given core, choose a random number of randomly selected sites ( $n$ ) to substitute, where  $1 \leq n \leq n_{\max}$  and  $n_{\max}$  is the total number of substitution sites. For each of the  $n$  sites, select a functional group. If the molecule currently has  $m$  atoms, only functional groups smaller than  $N - m$  atoms, where  $N$  is a threshold size (here  $N = 90$  atoms), are allowed. Among the set of allowed functional groups, the selection is weighted using weight =  $1/m_{\text{sub}}$ , where  $m_{\text{sub}}$  is the number of atoms of the substituent functional group. If no functional groups are small enough, then a hydrogen atom is added at the site. Likewise, hydrogen atoms are added to all  $n_{\max} - n$  sites that were not chosen. This procedure is then repeated a set number of times  $p$ , generating a sub-library for the particular core.

Once a sub-library consisting of  $p$  attempt molecules is constructed, it is then filtered. First, duplicate structures are identified by comparing InChI strings and removed. Molecules can also be removed because they contain certain chemical motifs, identified via SMARTS matching using RDKit. Specifically, molecules with uncommon motifs (*e.g.*, P-Br bonds) are removed with a probability  $1 - 0.8^x$ , where  $x$  is the number of uncommon motifs present. Molecules with more than one motif found in energetic materials (*e.g.*, O-O or N-N bonds) are removed, as are all molecules with nitro groups bound to nitrogen (“N-N(=O)O” in SMARTS). For molecules that pass all filters, 3D structures are generated from SMILES strings using the Software for Chemical Interaction Networks Molassembler package. If Molassembler could not generate a valid 3D structure in 25 attempts, for instance because of steric clashes in molecules with multiple bulky groups, the molecule was discarded.

Library generation was conducted in three stages. In the first stage, only ID cores and ID functional groups were used, and 3,000 attempts were made per core. In the second stage, ID cores were used with OOD functional groups with 250 attempts per core; each molecule generated in this stage includes at least one OOD functional group. Finally, in the third stage, OOD cores were used with both ID and OOD functional groups, and 3,000 attempts were again made per core. Molecules generated in the third stage were not required to include OOD functional groups. Molecules generated from the first stage are considered ID, and molecules from the second and third stages are OOD. In total, roughly 149,000 molecules were generated using this procedure, with 249,000 attempts made. The geometry of each molecule generated using this procedure was

optimized for up to five steps.

### D.3 Molecular Clusters

The structures obtained from classical MD (Section D.1) provide valuable information regarding non-covalent and non-bonded interactions but contain a relatively small number of unique species limited by the availability of force-field parameters. To supplement the MD-generated data, quasi-random molecular clusters including more diverse species were constructed.

The Architect package was used to pose initial structures for complexes. Architect can randomly dock a molecule or ion around another molecule and then leverage the semi-empirical quantum chemistry method GFN2-xTB to relax the random complex, providing a reasonable starting point for DFT optimization.

Cluster generation involved two sets of species: a set of central molecules around which complexes were generated, and a set of solvating species placed around the central molecules. In this work, the set of central molecules included small electrolyte-like molecules described in Section D.2 with 50 or fewer atoms, and the set of solvating species, listed in the Appendix, resembles the set of molecules and ions included in MD simulations but includes, for instance, more unique metal ions.

Clusters are generated in three ways: 1) dimers; 2) full solvation-like structures; and 3) random clusters. All clusters were limited to a maximum size of 180 atoms.

A dimer is constructed by placing one random, valid solvating species around the central molecule. If the central molecule is charged, then solvating species of like charge to the central molecule are not considered, to ensure that the clusters do not separate due to electrostatic repulsion.

For full solvation-like structures, a single, neutral solvating species (the solvent) was randomly selected. Then, a target solvation shell size was selected on a Gaussian distribution, with  $\mu = m + 60$  and  $\sigma = 30$ , where  $m$  is the number of atoms in the central molecule. If the randomly selected budget was below  $m + 20$  atoms, then it was shifted to  $m + 20$ . Likewise, if the budget was above the overall maximum size of 180 atoms, then it was set to 180. Then,  $s$  copies of the solvent were placed around the central molecule, where  $s = \lceil (\text{budget} - m) / m_{\text{solvent}} \rceil$  and  $m_{\text{solvent}}$  is the number of atoms in the solvent.

Random clusters began with the determination of a size budget, using the same procedure as the full solvation-like structures. Then, solvating species are added one at a time. In each iteration, the pool of solvating species is reduced to include only those that are small enough to respect the size budget ( $m_{\text{solvating species}} \leq \text{budget} - m$ , where  $m_{\text{solvating species}}$  is the number of atoms in the solvating species), and then one species is randomly selected from this reduced set, weighted by size (weight =  $1/m_{\text{solvating species}}$ ).

For each initial molecule, one dimer, one full solvation-like structure, and three random clusters were constructed using only ID solvating species. In addition, for 10% of initial molecules, either a dimer, a full solvation-like structure, or a random cluster was generated including at least one OOD solvating species; this data was labeled OOD.

### D.4 Water Clusters

We augmented the water-containing frames from above with 150,000 snapshots of 70 water molecules. To capture water structures more accurately, we use the AMOEBA forcefield [70–72], which includes polarization effects absent from the non-polarizable forcefields, e.g., OPLS, used in the MD simulations of bulk electrolytes and flexible water molecules. These snapshots were created by simulating a box of water in OpenMM with Langevin dynamics and a Monte Carlo barostat at 300K and 1 atm for 150 ns, taking frames every 1 ps, following a 100 ps equilibration period. The integration time step was 2 fs and a cubic box 2 nm on a side was used. In each frame, the 70 water molecules closest to the center of the box were saved as a cluster.

### D.5 Noble Gases

Though noble gases rarely form bonds with other elements, they nonetheless do appear variously in chemical applications, often entrained in a larger system. Instead of excluding the noble gases or including only a few examples, we aimed to represent them in environments that are most likely to reveal their intermolecular

interactions, despite their relative lack of reactivity. Hence, a  $40\text{\AA}$  cubic box was filled with five atoms of each of the non-radioactive noble gases (He, Ne, Ar, Kr, Xe) and 5925 TIP4P-D [221] water molecules. This system was equilibrated over 160 ps and then 500 ns of NPT MD with Desmond using the OPLS4 forcefield, taking frames every 250 ps. All molecules within  $4\text{\AA}$  of each gas molecule were extracted as inputs for DFT (10,000 for each element), containing a noble gas atom and several water molecules.

## E Molecular Dynamics with MLIPs

To increase the configurational diversity of our inputs, we incorporated several ML sampling approaches capture reactive interactions that are unattainable by our non-reactive classical MD force fields. For this, we trained three independent EquiformerV2 (EqV2) [222] models using an early snapshot of OMol25 across the different domains: biomolecules, metal complexes, and electrolytes. A pretrained MACE-MP-0 model was also used to provide more diversity.

For metal complexes, we generated input structures following the same pipeline described in C.1. We then ran 1-2 ps of MD with the corresponding metal complex-trained EqV2 model from which 5 samples were randomly selected and used for DFT calculations. The same procedure was also performed with MACE-MP-0. MD was performed at a constant temperature of 300K with a time step of 1 fs. A small sample of data was also generated using a rattled Boltzmann sampling [83] with  $\mu = 0A$ ,  $\sigma = 0.2A$ ,  $N = 1000$ , and  $T = 500K$ . Similarly, five random samples were selected for DFT calculation. For biomolecules, we generated protein interfaces as described in Section B.4, but without running classical MD. The EqV2 model trained on biomolecules was then used to run 2 ps of MD following the same procedure as above. The initial structure was then evaluated with DFT, along with two randomly selected samples. Electrolyte samples were generated in a similar manner; inputs were first generated in the same manner as described in Section D.1 with a new set of random electrolyte boxes. The corresponding EqV2 electrolytes model was used to run 2 ps of MD. The initial structure, along with four randomly selected samples, were then used as inputs for DFT calculations.

While we do not expect these early models to generate accurate MD trajectories, they provide a means to generate reactive configurations. In a similar vein, chemically unreasonable structures generated by these models along the MD trajectory can be particularly useful as training data to help the model learn what to avoid. The above procedure resulted in 1.6M metal complexes, 1.3M electrolytes, and 1.1M biomolecules structures calculated with DFT coming from ML sampling.

## F Reactivity

Reactive geometries were generated via four distinct approaches.

### F.1 RGD1

The RGD1 dataset [79] comprises 126,857 reactant, product, and transition state triplets for neutral, closed-shell species containing C, H, O, and N and up to ten heavy atoms generated via a combination of xTB and DFT simulations. For each reaction, we performed a geodesic interpolation [223] from reactant to transition state and from transition state to product, with each interpolation containing ten configurations. Removing one copy of the duplicate transition state structure, which is present in both interpolations, yielded 19 snapshots per reaction or a total of 2,410,283 snapshots across all reactions.

### F.2 RMechDB and PMechDB

RMechDB [81] contains over 5,300 elementary radical reaction steps, *i.e* each with a single transition state. PMechDB [80] contains over 100,000 elementary polar reaction steps in a similar format. While these reactions are mainly organic, some do contain metals.

Each reaction is provided as a SMIRKS string. We use Schrodinger's Fast3D to generate initial 3D geometries of the reactant and product structures from these. For structures without any metals, we then perform a UFF optimization with an fmax of 0.1 eV/ $\text{\AA}$  and at most 50 relaxation steps. If an RMSD greater than  $100\text{\AA}$

is observed due to UFF optimization, we revert to the unoptimized structure. We then perform a GFN-FF optimization with an  $f_{\max}$  of 0.1 eV/Å and at most 50 relaxation steps on the results from the previous step and for those structures with metals.

Total charge and spin are then detected via Architector utilities. In practice, this only modifies charge and spin for species containing metals for which the charge and spin in RMechDB or PMechDB is suspect. For species with multiple metals, we enforce a low-spin (antiferromagnetic) configuration, which is justified by the fact that the reactions in the MechDBs with multiple metals typically have the metals close together. Species with isolated O<sub>2</sub> molecules have their multiplicity increased by two to account for the fact that O<sub>2</sub> is a triplet in the ground state.

We then apply a custom implementation of the applied force induced reaction (AFIR) procedure [224], built atop Architector and ASE utilities, with MACE-MP-0 [82] as our source of energy and forces to create frames which approximate the reaction pathway. Note that, while the assigned total charge and spin multiplicity have no impact on the energy and force evaluation, they do impact the DFT run on the snapshots that come out of the AFIR procedure.

Our AFIR implementation starts by identifying the broken and formed bonds based on the Architector connectivity graphs of the reactant and product structures. We then apply constraints in ASE that push together pairs of atoms that are supposed to form bonds in the reaction and pull apart atoms in bonds that break with a constant force. We then carry out an optimization of the reactant structure under these constraints using ASE's LBFGSLineSearch, with an  $f_{\max}$  cutoff of 0.15 eV/Å, an initial force constant of 0.1 eV/Å, and at most 50 relaxation steps. If LBFGSLineSearch fails to converge, we switch to BFGSLineSearch instead, which remedies the vast majority of optimization convergence issues. Following optimization convergence, we check the minimum distance between any two atoms in any structure along the optimization trajectory. If a minimum distance less than 0.7 Å is observed, the AFIR procedure is terminated, as this was found to only occur when a pathological region of the PES with this MLIP was encountered. If no problematic interatomic distances are found, then we check if the intended bonds have broken and formed along the optimization trajectory. We use a bond breaking distance cutoff of 1.5 Å and a bond formation distance cutoff of 1.2 Å. We note that we do not check for or take steps to prevent the breakage or formation of bonds besides those intended. If the intended bonds have not all broken or formed, then we increment the constraint force constants by 0.2 eV/Å and run another optimization starting from the end of the preceding optimization trajectory. This iterative procedure repeats until either all intended bonds have broken and formed, a problematic minimum distance is found, or force constants exceed 4 eV/Å (*i.e.*, at the 20th AFIR iteration). This creates one optimization trajectory with various force constants from reactant to product.

We then subsample the optimization trajectories to be simulated with DFT. We are most interested in structures in which bonds are actively breaking/forming, as such structures are unlikely to be captured in other areas of OMol25, and we seek to avoid selecting multiple structures that are too similar. We begin selecting structures with the highest energy structure from along the optimization trajectory.

For each structure in the trajectory, we calculate the average force magnitude (without the AFIR forces) and the change in energy with respect to the previous and following frame. We also compute the minimum RMSD between this frame and all snapshots already selected. We select this frame if:

1.  $\max(\text{abs}(F), \text{deltaE}) > 0.6$  and  $\text{RMSD} > 0.03$  Å
2.  $\max(\text{abs}(F), \text{deltaE}) > 0.4$  and  $\text{RMSD} > 0.04$  Å
3.  $0.6 > \max(\text{abs}(F), \text{deltaE}) > 0.4$  and  $\text{RMSD} > 0.04$  Å
4.  $0.4 > \max(\text{abs}(F), \text{deltaE}) > 0.3$  and  $\text{RMSD} > 0.07$  Å
5.  $0.3 > \max(\text{abs}(F), \text{deltaE}) > 0.18$  and  $\text{RMSD} > 0.1$  Å
6.  $0.18 > \max(\text{abs}(F), \text{deltaE})$  and  $\text{RMSD} > 0.14$  Å

This procedure was empirically developed by testing on many different example reactions and based on the desire to sample approximately ten snapshots per reaction and the observation that the RMSD between adjacent snapshots in the geodesic interpolations described previously almost never went below 0.03 Angstroms.

**Table 12** Reactions removed from previous metal reactivity datasets.

Reactions removed	Reason
ROST17, ROST18, ROST28, ROST58, MOBH18, MOBH20	Duplicative of other reactions when swapping ligands
ROST25, ROST36, ROST43	Too many reactants/products implying not elementary
ROST42	Contained half a product species
ROST55, MOBH21	No bonds breaking or forming
MOR32	Too many active bonds implying not elementary

Running the AFIR and sampling procedure on all reactions yielded 67,302 RMechDB and 1,175,609 PMechDB total snapshots for DFT simulation.

### F.3 Metal Reactivity

Metal complex template reactions were taken from MOR41 [62], ROST61 [63], and MOBH35 [64]. Each reaction was examined, and of the 137 total reactions from these three sources, 14 were removed.

Schrodinger tools were then used to perform atom mapping and, when necessary, assemble reactant and product complexes from component species structures. All endpoint complexes were then manually inspected and validated, during which we realized that we could obtain two additional reactions by manually moving the location of the placed I2 species for MOR36 and MOR23. Thus, in total we ended up with 125 template reactions. We then execute a procedure to perform ligand swaps and metal swaps in order to generate hundreds of thousands of unique reactions from our 125 templates.

We visually and programmatically inspected all reaction templates, assigning possible metal charges in order to facilitate swapping with like-charged metals while simultaneously identifying easily swapped ligands [118]. Briefly, metals were removed and ligand charges were assigned using a modified OpenBabel formal charge assignment scheme accounting for the metal-bound atom valence (routines found in Architector). Visual inspection of the ligand charges resulting in manual overriding of 10-15 ligand charges assigned. With ligand charges assigned, metal oxidation states were selected that were both (1) common oxidation states according to the mendeleev package and (2) kept the total charge of the reaction complex less than +5 and greater than -3 to minimize charge deviations from the intended templates.

To reduce the complexity of swapping ligands, we chose to only perform ligand swaps with ligands curated from the metal complex generation ligand set with matching charge state and denticity. Further, to minimize overlapping atoms and steric clashes with reaction sites, only a single ligand was swapped at a time, only ligands with denticity 2 or 1 were swapped, and no ligands involved in any bond breaking/forming in the template reaction were allowed to be swapped. Additionally, if there are multiple ligands with identical symmetries to a reaction site (*e.g.*, single ligand dissociation from homoleptic octahedral complex, all 4 ligands *cis*- to the dissociating ligand) then all such ligands were treated by a single swappable ligand site.

To perform ligand swaps, we leveraged Architector 3D functionalization capabilities, removing the ligand that is being swapped out, and adding the ligand that is being swapped in, and relaxing with UFF while keeping the rest of the atom indices and positions fixed in both the reactant and product complexes. Spin states were determined for the swapped metal/oxidation states using mendeleev and any additional charge/spin present on the ligands was taken from the pre-tabulated values.

We note that 99 of our reaction templates contained one or more swappable ligands, while 34 did not. Given that those without any swappable ligands yielded only 1436 reactions after metal swapping, we decided to reserve those 1436 reactions for use as a test set. After enumerating all reactions with swappable ligands, we obtained ~1.1M possible reaction templates. Some reaction templates contained vastly more possibilities after swapping ligands and metals, for example, reactions where 4 different monodentate ligands and 3 possible metal oxidation states could be swapped resulted in hundreds of thousands of possible reactions, while reactions with only 1 swappable bidentate ligand and 1 possible metal oxidation state resulted in only ~2K possible reactions. To sample as evenly as possible across reactions given a computational budget of ~5M AFIR snapshots we selected 250k reactions as a target, and selectively down-sampled including all swapped reactions with fewer than 3164 possible swaps, and 3164 randomly sampled swaps from reaction templates with more than 3164 possible swaps. After performing the swaps and removing reactions that

produced overlapping atoms or distorted structures, we obtained 246,785 valid reactions, not including the hold-out test set of 1436 reactions.

All reactions obtained from ligand and metal swapping were subjected to the AFIR procedure described previously. The AFIR procedure can produce different results when run from reactant to product and product to reactant. Thus, when running the AFIR, we randomly swapped reactant and product with a 50% probability to increase overall structural diversity. From our 246,785 total ligand and metal swapped reactions, we obtained 5,145,142 snapshots from AFIR to be simulated with DFT.

#### F.4 Electrolyte Reactivity

Electrolyte reactions were taken from previous work on reaction networks and mechanistic studies of electrolyte decomposition and solid-electrolyte interphase formation [14, 73–76]. Schrodinger tools were used to construct 3D reactant and product complexes when they were not already available. When they were already available, Schrodinger tools were used to construct alternate versions of endpoint complexes. For both original and Schrodinger structures, the RMSD was computed between the reactant and product complexes, and if the RMSD was found to be less than 0.3 Angstroms then the offending pair was discarded. Where both original and Schrodinger endpoints passed this check, if the Schrodinger structures both had a greater than 0.4 Angstrom RMSD compared to the original structures, and the sum of the two RMSDs was greater than 1.4, then we deemed that the Schrodinger and original endpoint pairs were sufficiently different to justify including both in our template set. This procedure was empirically developed via testing and manual inspection of many different example reactions. If these cutoffs were not met, then only the Schrodinger endpoint structures were used. We then performed metal swaps for all reaction templates with at least one metal ion. Li was allowed to be swapped with all eight other metals with common +1 oxidation states according to mendeleev (i.e. Na, K, Cs, Cu, Ag, Rb, Tl, and Hg). Mg was allowed to be swapped with other metals with common +2 or +3 oxidation states according to mendeleev (+2: Ca, Zn, Be, Cu, Ni, Pt, Co, Pd, Ag, Mn, Hg, Cd, Yb, Sn, Pb, Eu, Sm, Cr, Fe, V, Ba, Sr, +3: La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Al, Ga, In, Tl, Bi, Sc, Cr, Fe, Co, Y, Ru, Rh, Ir, Au). The reaction’s charge and/or spin was changed commensurate with the difference in charge/spin between the principal metal and the metal being swapped in. In order to account for the differences in atomic radii between swapped metals, the distance between the metal and each other distinct species in the reaction complex was scaled by the ratio of the new metal’s van der Waals radius and the principal metal’s van der Waals radius, which was found to yield sensible metal-organic distances. A total of 146,962 reactions were obtained following metal swapping.

We attempted to apply our AFIR procedure to these electrolyte reactions, but neither MACE-MP-0 nor even B97-3c had a sufficiently stable/reasonable description of the potential energy surface in the presence of undercoordinated metals/radicals, causing widespread convergence issues and resulting in poor quality snapshots. The geodesic procedure used for RGD1 worked better, but it still encountered some issues, likely owing to the fact that we were interpolating directly from reactant to product rather than from reactant to transition state to product. Thus, we instead employed our recently developed Path Optimization with a Continuous Representation Neural Network aka Popcornn method [77], using the geodesic model potential and loss function but with a continuous path representation during optimization. 17 snapshots were then sampled uniformly from the optimized path, and the middle 15 were selected for simulation with DFT (i.e. endpoints were discarded). Thus, approximately 2.2 million snapshots were generated by running Popcornn on all electrolyte reactions.

We note that, in addition to the four categorizes of reactive geometries just described, other reactive geometries in the training dataset come from recalculated Transition-1x, ANI-1xBB, instances in which bonding changed during optimization, MLIP-MD, Boltzmann rattling, and possibly snapshots from other community datasets in which active learning and/or optimization was employed.

In addition to reactive geometries that ended up in the test set due to composition-based splitting, we also generated two additional sets of reactive snapshots for use in testing which attempt to be explicitly out-of-distribution: Geodesic interpolations of 1782 organic reactions which are neither in Transition1x nor RGD1, and AFIR snapshots of the 1436 reactions obtained from our metal complex templates which had no swappable ligands. These two sets in total contain approximately 67k snapshots for use as an OOD energy and force evaluation set.

## G Community Datasets

SPICE [37, 38], Transition-1x [34], ANI-2x [33], and OrbNet Denali [36] datasets were recalculated at the OMol level of theory without any modifications. Solvated protein fragments [225] structures were randomly split into three groups: 10% had an electron added, 10% had an electron removed, the remaining 80% were unchanged. These systems were then optimized with Sella with up to a maximum of two geometry optimization steps. ANI-1xBB [35] structures were removed if an isolated atom of elements B, C, N, O, Si, P, S, or Se was found, where an atom was considered to be isolated if the nearest other atom was greater than 1.8 times the sum of the two elements' covalent radii. Structures were also filtered if an isolated O<sub>2</sub> or S<sub>2</sub> species were identified based on a bonding graph generated by Architector, due to the ambiguity about the correct spin state for the system. All remaining structures were simulated with the standard OMol25 procedure for open-shell singlets. A random sample of 10% of the structures were additionally simulated as triplets. GEOM [39] structures were taken from ~300k unique molecule families. For each family, ~30% of the conformers were selected for DFT calculations. Of the selected structures, 50% had their atomic positions rattled with displacements sampled from a Gaussian distribution of  $\mu = 0$  and  $\sigma = 0.1\text{\AA}$ . Structures were then optimized to a max per-atom force of 0.05 eV/\text{\AA} using Sella.

## H Evaluation Details

### H.1 Protein-ligand Interactions

Protein-ligand pockets that were added to the BioLiP2 dataset between May 31, 2024 and Mar 13, 2025 were processed by the procedure described in Appendix B.1. Drug-like pockets were identified by scoring the ligands by the QED method as implemented in RDKit [97], retaining those with a value greater than 0.5, and requiring the number of rotatable bonds to be between 3 and 12, inclusive. Ligand-pocket interaction energy and interaction forces are defined as the difference between the energy/forces of the ligand-pocket complex and the isolated ligand and isolated pocket, with atom positions of each isolated component unchanged from the ligand-pocket complex and all calculations performed in vacuum.

### H.2 Ligand Strain

The protein-ligand pockets described above also furnished the bioactive conformations of the systems used to the ligand strain evaluation. The local minimum of the ligand-in-pocket geometry is obtained by taking the union of:

1. CREST, starting from the bioactive conformer
2. RDKit's conformer generation combined with MM94FF optimization
3. Schrödinger's MacroModel

All of these conformers are re-optimized with xTB and de-duplicated if the maximum distance between corresponding atoms was less than 0.25\text{\AA} and the xTB energies are within 1 kcal/mol. The resulting pool of conformers were DFT-optimized with Sella and an fmax of 0.01 eV/\text{\AA}. The 5 lowest energy conformers and a random set of 5 were retained for the evaluation. That is, there are 10 conformers for each bioactive conformation.

### H.3 Conformers

A subset of GEOM molecule families is held out of train for constructing the evaluation set. For each family, optimizations are performed on all conformers with Sella and a max per-atom force of 0.01 eV/\text{\AA}. The 5 lowest energy conformers and a random set of 5 were retained for the evaluation. That is, there are 10 conformers for each molecule family. Linear sum assignment based on RMSD yields a mapping between DFT and MLIP optimized structures which minimizes the sum of the RMSDs across the ensemble, which allows for the scenario of different initial geometries finding the same minimized structure during DFT and MLIP optimization. The average of the RMSDs under the identified mapping is the ensemble RMSD evaluation

metric. In order to construct an additional metric which incorporates the increased importance of the low energy structures which dominate the thermal ensemble, we define the Boltzmann weighted ensemble RMSD as the dot product between the vector of DFT Boltzmann weights and the vector of RMSDs under the identified mapping. Since the Boltzmann weights sum to 1, no further averaging is required. While the resulting metric does not account for the difference in conformer energies between DFT and the MLIP, such differences are directly evaluated via the  $\Delta E$  and  $\Delta E_{reopt}$  evaluation metrics described fully in the main text.

We note that the purpose of the reoptimization component of the conformer evaluation is to have some metrics which are somewhat less tied to the sensitivity of the optimizer. Such sensitivity may cause optimizations from the same far-from-minima conformation to converge to different local minima, even when the potential energy surfaces the two optimizations are conducted on are extremely similar. By additionally starting MLIP optimizations from DFT-minimized structures, we probe how far MLIP minima are from DFT minima in a manner which should reduce the impact of optimizer sensitivity compared to optimizations which start from far-from-minima conformations.

#### H.4 Protonation Energies

Structures for the protonation energy evaluation are taken from Johnston et al. [102], which includes families of organic molecules with variable protonation states. We subject all structures to tightly converged ( $f_{max} = 0.01 \text{ eV}/\text{A}$ ) geometry optimizations with Sella for both DFT and the MLIP being evaluated. For each structure, we can calculate the RMSD between the MLIP and ORCA optimized geometries. We further identify all pairs of structures within a family which differ by exactly one proton and calculate the  $\Delta E$  between the optimized structures separately with DFT and the MLIP, where the MAE between the DFT and MLIP  $\Delta E$  values thus informs upon the MLIP's ability to calculate protonation energy.

#### H.5 Unoptimized IE/EA and Spin Gap

while vertical IE and EA are typically calculated via energy differences between an initially relaxed structure and that same structure with an increased or decreased charge, we are equally interested in the model's ability to predict energy and forces at different charges and energy differences between charge states of the same geometry irrespective of whether or not a geometry is at a PES minima. For our "unoptimized" IE/EA evaluation task, we randomly select 4000 electrolyte structures from our test set with up to two open-shell metals, 1000 metal complexes constructed with Architector from our test set, and 4000 metal complexes from COD. Note that all of these systems are clusters of species, involving both covalent and non-covalent/coordination interactions, where the addition or removal of an electron is more complex than for a single small, isolated, fully connected molecule. For systems that are initially singlets, both charge-increased and charge-decreased snapshots are calculated as doublets. For systems that are initially doublets or a higher multiplicity, charge-increased and charge-decreased snapshots are calculated both with original multiplicity + 1 and original multiplicity - 1 and  $\Delta E$  values are calculated between the original charge/spin and all four modified charge/spins. Our evaluation metrics are thus energy and force MAEs for charge-modified structures and both  $\Delta E$  MAEs and  $\Delta F$  MAEs between principal and charge modified structures.

As with IE/EA, we again opt for vertical spin gaps between static structures without any optimization. We randomly select 1000 electrolyte structures from our test set, 3000 metal complexes constructed with Architector from our test set, and 5000 metal complexes from COD, each constrained to have exactly one open-shell metal and at least two viable spin states. The full ladder of viable spin states are simulated for each complex, and  $\Delta E$  and  $\Delta F$  values are calculated between the highest possible spin and each other spin value.

#### H.6 Distance Scaling: Short-range and Long-range Interactions

For our distance scaling evaluation task, we randomly select 2500 electrolyte structures from our test set without any open-shell metals, 1000 electrolyte structures from our test set with one open-shell metal, 1000 metal complexes constructed with Architector from our test set, 3000 metal complexes from COD, and 2000 PDB fragment structures from our test set. All systems are clusters of at least two non-covalently interacting species. For PDB fragments and both types of electrolyte samples, we scale the magnitude of the vectors between the center of mass of the full structure and the center of mass of each disconnected

component, which modifies the distances between all disconnected components while keeping the intramolecular distances/interactions fixed. We randomly sample two scale factors in the compressive regime between 0.7 and 0.95, and ten scale factors in the short-range expansive regime 1.05 to 1.8. For PDB fragments and closed-shell electrolytes, we additionally sample ten scale factors in the long-range expansive regime 1.8 to 3.0. We do not do any long-range scaling for open-shell electrolyte because the resulting structures would contain a fully isolated open-shell metal, for which we have observed ORCA can struggle to find the lowest energy SCF solution. Given this, for COD and Architector metal complexes we simply scale the distance between the central metal and one coordinating ligand. More specifically, we employ Architect tools and uniformly sample ten points with the metal-ligand distance increased 0.4 Angstroms each up to 4 Angstroms. In the future, we hope to expand this range to include one snapshot in the compressive regime and multiple beyond four Angstroms.

In order to rigorously define which snapshots are in the “short-range” regime versus the “long-range” regime, we construct the neighbor graph for each structure with a 6Å cutoff, and snapshots with multiple disconnected graph components are designated “long-range”. Given an entire scan at variable scaled distances, we then identify the lowest energy short-range structure and calculate  $\Delta E$  and  $\Delta F$  values for each other structure relative to the that structure. If only long-range snapshots exist for a given scan, then we instead reference against the lowest energy long-range snapshot. We consistently reference against a short-range snapshot when one is available to ensure that we are not mixing long-range errors into short-range metrics.

## I Baseline Models and Results

eSEN [30] and GemNet-OC [106] were used as the primary baseline models for this work. eSEN represents the current state-of-the-art equivariant model on many community materials and molecular tasks including Matbench-Discovery [226] and SPICE-MACE-OFF [112]. We train three variants of this model: small (eSEN-sm), medium (eSEN-md), and large (eSEN-lg). GemNet-OC was also trained, representing an accurate invariant model, that was previously state-of-the-art on the OC20 [107] benchmarks. MACE [27] was also trained, but only for the charge-neutral split as significant modifications would have been required to train across varying charges and spins.

### I.1 Training

All models were trained on Nvidia H100 80GB GPU cards, with an AdamW optimizer, a learning rate of 8e-4, and a cosine learning rate scheduler. eSEN-sm was trained with FP32 precision. eSEN-md and eSEN-lg used a multi-stage scheme, where BF16 is used for the entirety of training followed by  $\leq 1$  epoch of finetuning at FP32 at a learning rate of 1e-4. We found this to provide more stable training for larger model sizes. GemNet-OC was also trained with BF16. Batch sizes generally varied to balance model sizes and reasonable training times. All models were trained for 12 epochs on the full dataset and 80 epochs on the 4M. On the neutral split, eSEN and MACE were trained for similar number of optimization steps, but had very different epoch equivalents. MACE incorporated a multi-stage training scheme, beginning with an energy coefficient of 10 but then finetuning with a coefficient of 50 near the end of training.

To account for variable total charge and spin in OMol25 , we incorporate the same architectural change to all models to fairly compare the original baseline models. We first randomly initialize two embeddings based on the total charge and spin of the system, with output dimensions corresponding to the hidden channels of the respective model. The embeddings are concatenated and fed through a single linear layer. This embedding is added to the node embeddings and then subsequently at every layer of message passing. Although we recognize that this may not be an ideal approach for all architectures, we found it to give very positive results and provided a naive baseline for future works.

All energies were referenced using a "heat of formation" (HOF) reference:

$$E_{ref} = E_{DFT} - \sum_i^N [E_{i,DFT} - \Delta H_{f,i}]$$

Where  $E_{DFT}$  is the total system energy coming from ORCA,  $E_i$  is the energy of an isolated atom,  $N$  is the number of atoms in the system, and  $H_{f,i}$  is the heat of formation of atom species  $i$  [60]. Energies are then linearly referenced, following the same procedure in the OC22 dataset [227].

## I.2 Total Energy Test Results

Total energy results on the test splits and the corresponding breakdown results are provided in Tables 13 and 14.

## I.3 Validation Results

Results on the out-of-distribution composition validation set are provided in Table 15. Total energy metrics of the same results are also provided in Table 16.

## I.4 Additional Results

Alongside the main test results, we provide additional out-of-distribution results in Table 17. These include the TorsionNet500 [94] benchmark dataset, curated out-of-distribution cations, solvents, and ions+solvents datasets. For each split, we evaluated the performance in predicting the energy and forces of the structure.

**Table 13** Structure to energy and force results across the different **test** splits. **Total energy** and force mean absolute error metrics are reported across the two different training splits - All and 4M.

Dataset	Model	# of params	Test															
			Comp				M-Lig				PDB-TM				Reactivity			
			Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
All	eSEN-sm-d.	6.3M	105.29	9.09	87.82	17.15	113.86	15.31	123.28	52.21	182.92	23.45	104.33	11.37	119.58	21.43		
	eSEN-sm-cons.	6.3M	93.36	7.39	76.92	14.66	96.71	12.72	103.38	44.25	132.80	20.56	57.37	9.23	93.42	18.14		
	eSEN-md-d.	50.7M	58.46	4.29	50.26	8.79	66.24	9.06	84.86	32.21	110.91	14.30	57.90	5.40	71.44	12.34		
	GemNet-OC	39.1M	34.54	5.85	54.43	12.55	67.54	11.38	94.35	41.98	131.00	18.42	63.07	7.20	74.16	16.23		
	eSEN-lg-d.	690.4M	training in progress															
4M	eSEN-sm-d.	6.3M	149.65	11.86	119.31	21.63	175.33	18.70	160.57	64.56	260.51	30.18	172.54	15.83	172.99	27.13		
	eSEN-sm-cons.	6.3M	140.73	9.89	95.76	18.86	130.12	15.93	125.01	54.17	182.19	27.10	101.24	12.76	129.18	23.12		
	eSEN-md-d.	50.7M	88.68	5.99	73.33	12.02	116.63	12.21	120.70	43.88	162.46	19.01	102.49	8.23	110.72	16.89		
	GemNet-OC-r6	39.1M	89.03	8.85	84.87	17.17	142.13	15.62	135.99	57.06	255.75	26.24	106.55	11.71	135.72	22.78		
	GemNet-OC	39.1M	59.80	8.00	79.71	16.45	124.95	15.71	133.56	55.57	230.45	24.36	89.71	10.99	119.70	21.85		

Energy(meV), Forces(meV/Å)

**Table 14** Out-of-distribution composition **test** results broken across biomolecules, electrolytes, metal complexes, and neutral organics. **Total energy** and force mean absolute error metrics are reported across the two different training splits - All and 4M.

Dataset	Model	Test-Comp									
		Biomolecules				Electrolytes				Metal Complexes	
		Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
All	eSEN-sm-d.	98.94	6.39	123.30	10.41	138.92	31.70	45.01	18.71	training in progress	
	eSEN-sm-cons.	87.17	4.64	110.88	9.09	123.16	27.23	31.58	16.41		
	eSEN-md-d.	51.14	2.62	70.97	5.16	88.21	18.71	22.50	7.66		
	GemNet-OC	24.58	3.95	40.04	6.53	95.22	23.98	28.86	14.52		
	eSEN-lg-d.	training in progress									
4M	eSEN-sm-d.	129.96	8.25	186.73	13.84	185.98	38.77	77.65	27.43	training in progress	
	eSEN-sm-cons.	125.68	6.23	177.72	12.34	148.88	33.31	58.52	25.46		
	eSEN-md-d.	70.29	3.40	114.87	7.54	130.37	25.46	45.76	12.52		
	GemNet-OC-r6	64.78	5.92	116.77	10.29	145.09	31.98	71.79	23.85		
	GemNet-OC	40.93	5.29	74.42	9.16	137.85	31.14	58.79	22.70		

Energy(meV), Forces(meV/Å)

**Table 15** Out-of-distribution composition **validation** results. Alongside the total metrics, results are also broken across biomolecules, electrolytes, metal complexes, and neutral organics. **Energy per atom** and force mean absolute error metrics are reported across the two different training splits - All and 4M.

Val-Comp											
Dataset	Model	Biomolecules		Electrolytes		Metal Complexes		Neutral Organics		Total	
		Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
All	eSEN-sm-d.	0.67	6.30	1.24	9.41	2.53	33.08	1.23	13.84	1.49	9.92
	eSEN-sm-cons.	0.59	4.61	1.01	8.08	2.30	28.86	0.84	11.11	1.27	8.25
	eSEN-md-d.	0.34	2.61	0.69	4.40	1.73	19.99	0.59	5.63	0.84	4.76
	GemNet-OC	0.15	3.88	0.56	5.98	1.83	25.12	0.86	10.38	0.66	6.52
	eSEN-lg-d.	training in progress									
4M	eSEN-sm-d.	0.88	8.12	1.93	12.64	3.37	40.44	2.16	20.17	2.19	13.01
	eSEN-sm-cons.	0.86	6.17	1.61	11.16	2.72	35.33	1.50	16.92	1.89	11.10
	eSEN-md-d.	0.47	3.38	1.18	6.51	2.53	27.31	1.21	9.26	1.32	6.78
	GemNet-OC-r6	0.40	5.84	1.39	9.37	2.74	33.60	1.88	16.55	1.41	9.83
	GemNet-OC	0.25	5.20	1.04	8.42	2.66	32.76	1.64	15.59	1.13	8.98

Energy(meV/atom), Forces(meV/Å)

**Table 16** Out-of-distribution composition **validation** results. Alongside the total metrics, results are also broken across biomolecules, electrolytes, metal complexes, and neutral organics. **Total energy** and force mean absolute error metrics are reported across the two different training splits - All and 4M.

Val-Comp											
Dataset	Model	Biomolecules		Electrolytes		Metal Complexes		Neutral Organics		Total	
		Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
All	eSEN-sm-d.	96.92	6.30	88.20	9.41	145.45	33.08	34.24	13.84	89.34	9.92
	eSEN-sm-cons.	86.94	4.61	73.42	8.08	132.80	28.86	23.10	11.11	76.60	8.25
	eSEN-md-d.	50.94	2.61	48.45	4.40	97.66	19.99	16.63	5.63	49.94	4.76
	GemNet-OC	23.34	3.88	30.72	5.98	103.11	25.12	21.19	10.38	33.85	6.52
	eSEN-lg-d.	training in progress									
4M	eSEN-sm-d.	127.43	8.12	134.78	12.64	192.77	40.44	59.84	20.17	129.77	13.01
	eSEN-sm-cons.	125.74	6.17	117.33	11.16	156.54	35.33	42.07	16.92	114.81	11.10
	eSEN-md-d.	70.13	3.38	78.67	6.51	142.55	27.31	33.37	9.26	77.13	6.78
	GemNet-OC-r6	63.57	5.84	82.08	9.37	152.87	33.60	47.96	16.55	79.67	9.83
	GemNet-OC	39.58	5.20	56.32	8.42	148.49	32.76	40.98	15.59	57.82	8.98

Energy(meV), Forces(meV/Å)

**Table 17** Structure to energy and force results across **additional test** splits. **Energy per atom** and force mean absolute error metrics are reported across the two different training splits - All and 4M.

Dataset	Model	TorsionNet500		OOD Cations		OOD Solvents		OOD Ions+Solvents		
		Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces	
All	eSEN-sm-d.	0.13	2.67	0.67	6.38	0.51	7.20	0.28	3.68	
	eSEN-sm-cons.	0.07	1.88	0.30	4.80	0.30	5.84	0.13	2.53	
	eSEN-md-d.	0.05	0.94	0.32	2.70	0.27	3.56	0.13	1.58	
	GemNet-OC	0.11	2.62	0.22	4.05	0.27	4.82	0.13	2.66	
	eSEN-lg-d.	training in progress								
4M	eSEN-sm-d.	0.23	4.07	1.10	8.99	0.68	9.18	0.42	5.29	
	eSEN-sm-cons.	0.12	2.98	0.54	7.15	0.42	7.77	0.23	3.75	
	eSEN-md-d.	0.10	1.36	0.54	3.84	0.42	5.33	0.22	2.09	
	GemNet-OC-r6	0.18	3.61	0.55	6.13	0.49	7.11	0.30	3.95	
	GemNet-OC	0.18	3.50	0.42	5.52	0.39	6.53	0.20	3.61	

Energy(meV/atom), Forces(meV/Å)

## I.5 Wiggle150

The Wiggle150 evaluation was recently introduced as a high-quality benchmark on strained conformers which uses CCSD(T)/CBS as the reference level of theory [44]. It consists of 50 conformers of three molecules. We evaluated the baseline models trained on OMol25 and compare to results taken from the Wiggle150 paper. As the OMol25 dataset is computed with the  $\omega$ B97M-V functional and with a smaller basis than was used in Wiggle150 (def2-QZVP), it cannot be more accurate than this level of theory.

The errors of the baseline models against the reference theory in OMol25 are sufficiently small that all such models have errors less than or equal to 1 kcal/mol. This level of accuracy against coupled-cluster theory was achieved only by  $\omega$ B97M-V and two double-hybrid functionals in the original Wiggle150 paper.

**Table 18** Comparison of baseline models trained on OMol25 and a sample of methods from the Wiggle150 benchmark [44].<sup>†</sup> GPU-accelerated timings are also provided for baseline models.

Method	RMSE (kcal/mol)	MAE (kcal/mol)	time (s)
DSD-PBEP86	0.90	0.72	2580
DLPNO-MP2	1.13	0.88	15,300
$\omega$ B97M-V	1.18	0.87	1990
$\omega$ B97X-V	2.72	2.41	1930
$\omega$ B97X-D3	3.00	2.49	1870
B3LYP-D3BJ	1.84	1.41	1370
M06-2x	2.42	1.97	1460
PBE-D3BJ	5.68	4.91	244
GFN2-xTB	15.2	14.6	3.21
AIMNet2	3.13	2.39	0.0170
ANI-2X	5.40	3.05	0.008
Orb-V2-D3	11.1	9.25	0.216
MACE-MP0	28.5	26.6	0.272
<b>eSEN-sm-d.</b>	<b>1.31</b>	<b>1.00</b>	0.0612 (0.024 <sup>†</sup> )
<b>eSEN-sm-cons.</b>	<b>1.25</b>	<b>0.91</b>	0.142 (0.063 <sup>†</sup> )
<b>eSEN-md-d.</b>	<b>1.28</b>	<b>0.92</b>	0.218 (0.039 <sup>†</sup> )
<b>GemNet-OC</b>	<b>1.25</b>	<b>0.90</b>	0.192 (0.059 <sup>†</sup> )

<sup>†</sup> GPU-accelerated timings are also provided for baseline models.

## I.6 Hyperparameters

eSEN model and training parameters for the small, medium, and large models are provided in Table 19. GemNet-OC model and training parameters were taken from [the original OC20 configuration](#) and also provided in Table 21. Two configurations of GemNet-OC were trained, the baseline with a cutoff radius of 12 and a cutoff radius of 6, to be consistent with eSEN-sm+md. GemNet-OC required doubling the batch size to train the All set, as compared to its 4M variant. MACE parameters are provided in Table 20.

**Table 19** Model hyperparameters for the three eSEN model configurations trained in this work.

Hyperparameters	eSEN-sm	eSEN-md	eSEN-lg
# sphere channels	128	128	256
lmax	2	4	6
mmax	2	2	2
max neighbors	30	30	30
cutoff radius	6	6	12
# edge channels	128	128	256
distance function	gaussian	gaussian	gaussian
# distance basis	64	128	256
# layers	4	10	16
# hidden channels	128	128	256
normalization type	rms_norm_sh	rms_norm_sh	rms_norm_sh
activation type	gate	gate	gate
ff_type	spectral	spectral	spectral
# gpus	32	64	128
batch size (# atoms)	89600	89600	44800
energy coeff.	10	10	10
force coeff.	5	5	5

**Table 20** Model hyperparameters for training MACE.

MACE	
Hyperparameters	
max_ell	3
correlation	3
max_L	1
num_channels	512
num_interactions	3
num_radial_basis	8
r_max	7
irreps	16x0e
batch size	256
energy coefficient	10 (50 finetuning)
force coefficient	10

**Table 21** Model hyperparameters for training GemNet-OC.

GemNet-OC	
Hyperparameters	
num_spherical	7
num_radial	128
num_blocks	4
emb_size_atom	256
emb_size_edge	512
emb_size_trip_in	64
emb_size_trip_out	64
emb_size_quad_in	32
emb_size_quad_out	32
emb_size_aint_in	64
emb_size_aint_out	64
emb_size_rbf	16
emb_size_cbf	16
emb_size_sbf	32
num_before_skip	2
num_after_skip	2
num_concat	1
num_atom	3
num_output_afteratom	3
cutoff	6,12
cutoff_qint	6,12
cutoff_aeaint	6,12
cutoff_aint	6,12
max_neighbors	30
max_neighbors_qint	8
max_neighbors_aeaint	20
max_neighbors_aint	1000
RBF	gaussian
envelop	polynomial
CBF	spherical_harmonics
SBF	legendre_outer
output_init	HeOrthogonal
activation	silu
quad_interaction	True
atom_edge_interaction	True
edge_atom_interaction	True
atom_interaction	True
num_atom_emb_layers	2
num_global_out_layers	2
# gpus	64
batch size (# systems)	512,1024
energy coefficient	1
force coefficient	10