HyFF: Atomic Foundation Models on the Speed–Accuracy Frontier

Hyra Team

Abstract

Atomic foundation models for molecular simulation, trained on large density-functional-theory (DFT) datasets, now describe organic molecules, biomolecules, electrolytes and metal complexes close to DFT accuracy, but their cost per molecular-dynamics (MD) step limits the time and length scales they reach. We present HyFF, a family of five atomic foundation models for molecular simulation, trained on OMol25 and covering 83 elements as well as charged and open-shell systems. The five models, Presto, Vivace, Moderato, Andante and Largo, share one equivariant architecture, span 2.3 million to 1.2 billion parameters, and take 0.71 to 123 ms per MD step for 2,586 atoms on one NVIDIA H20. The architecture is built on three principles: few edges, light edges, and model capacity placed on atoms rather than on atom pairs. We compare HyFF with 16 published models, including DPA4, DPA4C, UMA, MACE-OMol, OrbMol and UBio-MolFM, on the full OMol25 validation set (2.76 million structures), with every model timed over long MD runs in its fastest officially supported configuration. On this benchmark, all five HyFF models lie on the speed–accuracy frontier for both forces and energies. For 15 of the 16 baselines, a HyFF model is at least as accurate on forces and as fast or up to 15 times faster. HyFF Andante reaches a force error comparable to that of UMA-S-1.2 (7.65 versus 7.55 meV/Å) at 29 times lower cost, and HyFF Largo has the lowest energy error of all models scored (0.364 meV/atom), 9.7 times faster than DPA4-Pro and 7.0 times faster than UMA-S-1.2. In applications, HyFF runs 100 ns of a 58,560-atom aquaporin membrane system, keeps protein–ligand hydrogen-bond networks that a classical force field loses, and follows bond rearrangement inside a 46,282-atom enzyme.

1 Introduction

Atomic foundation models for molecular simulation are interatomic potentials trained once on large quantum-chemistry datasets that span organic molecules, biomolecules, electrolytes and metal complexes, and then applied to such systems without refitting (Levine et al., 2025; Wood et al., 2025; Li et al., 2026a). Their accuracy now approaches that of the underlying DFT calculations, but the most accurate of them need hundreds of milliseconds to more than a second per MD step for a few thousand atoms, so that a nanosecond of a solvated protein takes weeks. Compact models, at the other end, run at the speed of empirical potentials but give up most of that accuracy (Li et al., 2026b). Between these extremes, a user who needs a given accuracy often pays far more per step than that accuracy requires.

HyFF is designed to reach each level of accuracy at a low cost per step; on our benchmark, each of its five sizes is faster than every tested model of equal or lower error. The five sizes, named after musical tempo markings from the fastest to the most deliberate (Presto, Vivace, Moderato, Andante, Largo), share one architecture and one inference engine, and each is tuned for MD throughput on a single GPU. The speed comes from where the network spends its arithmetic. In a message-passing potential, most of the work is done on edges, the neighbour pairs within a cutoff, and an atom in condensed phase has about 17 neighbours within 3.5 Å but more than 200 within 8 Å. HyFF therefore keeps edges *few*, by giving each interaction block its own cutoff and reserving long cutoffs for the last blocks, and *light*, by passing narrow messages with tabulated radial filters. It places its capacity on *atoms*, where each atom is updated by a network specific to its element: most parameters sit in 83 element networks, of which an atom evaluates only one. Model capacity can therefore grow much faster than the cost per atom.

Our main results are the following.

- **Every HyFF model is on the frontier.** On the full OMol25 validation set and a common 2,586-atom MD benchmark, all five HyFF models are Pareto-optimal in force error among the 21 models scored

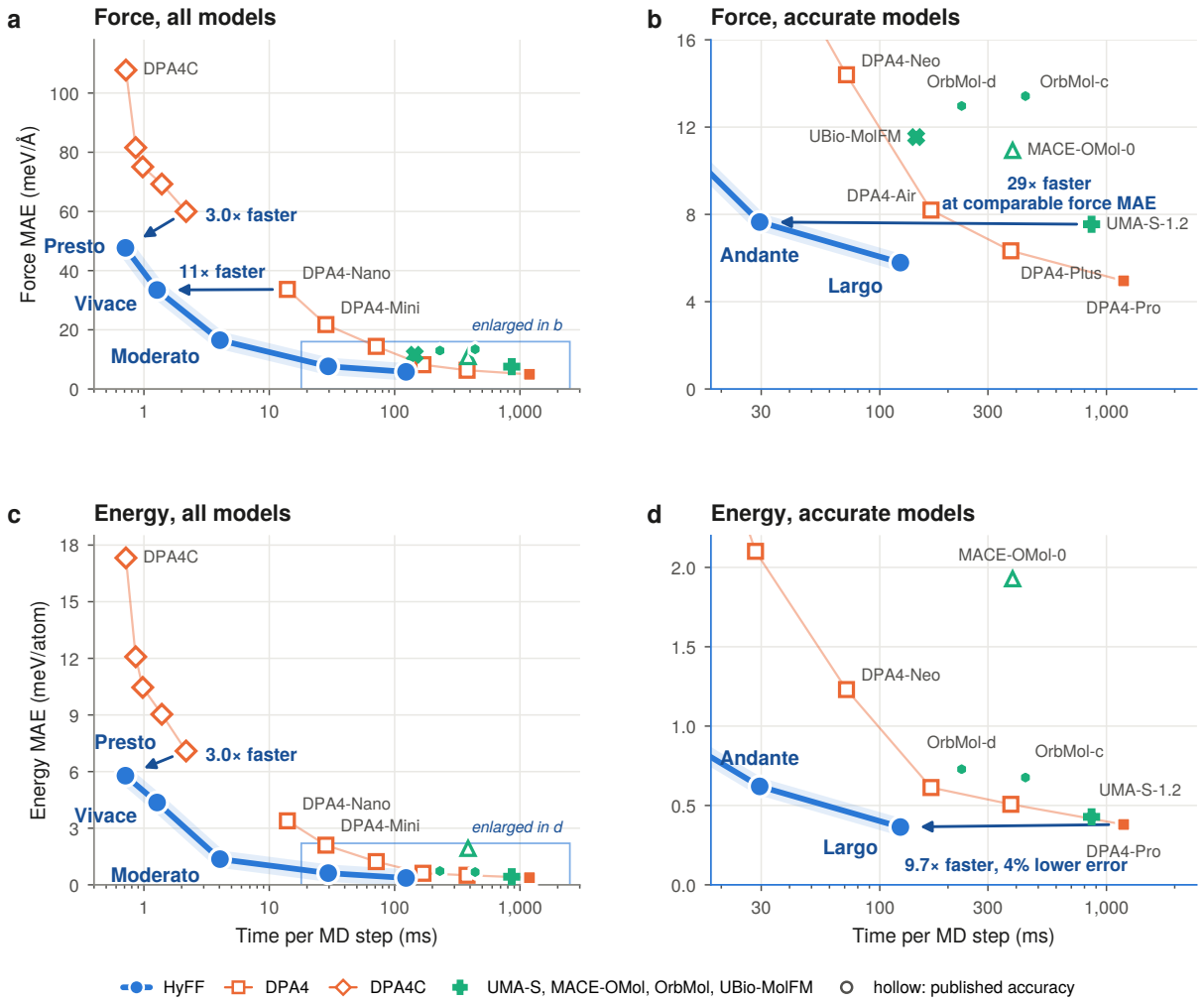


Figure 1: HyFF on the speed–accuracy frontier of OMol25. Force MAE (a, b) and energy MAE per atom (c, d) on the full validation set (2,762,021 structures) against the time of one MD step for 2,586 atoms on one NVIDIA H20. Panels b and d enlarge the accurate region outlined in a and c. Arrows connect a baseline with the fastest HyFF model of equal or lower error, or of comparable force error for UMA-S-1.2 in b. Filled markers are our accuracy measurements; hollow markers use published accuracy. All times are our measurements over long MD runs, each in the fastest officially supported configuration of the model (Section 3.1); DPA4 and DPA4C run in LAMMPS with the KOKKOS package. UBio-MolFM has no energy score.

and in energy error among the 20 models with an energy score (Figure 1). The family spans 0.71 to 123 ms per step, or 121 to 0.70 simulated nanoseconds per day.

- **Matched accuracy at a fraction of the cost.** For 15 of the 16 baselines a HyFF model is at least as accurate on forces and as fast or up to 15 times faster, and for every baseline with an energy score a HyFF model is at least as accurate on energies and as fast or up to 95 times faster (Table 2). Even the fastest DPA4C model, a compact single-layer model without multi-hop message passing, runs no faster than HyFF Presto at 2.3 times its force error.
- **Foundation-model accuracy at the top.** HyFF Largo has the lowest energy error of all models scored, 0.364 meV/atom, and a force error of 5.79 meV/Å, 23% below that of UMA-S-1.2.
- **Stable, large-scale MD.** In NVE simulations Presto, Vivace and Moderato drift by less than 0.035 meV per atom and nanosecond, and HyFF has run unbiased 100 ns simulations of protein systems with up to 58,560 atoms.

2 The HyFF family

All five models share one architecture (Figure 2). Atoms carry Cartesian tensor features of rank at most two. Two to five interaction blocks each pass equivariant messages between neighbours within a block-specific cutoff, optionally mix the messages in the frame of each short bond, and update every atom with the network of its element. The energy is read out from all blocks and the forces are its exact gradient. The sizes differ in width, depth and cutoffs and in which optional components they use (Table 1).

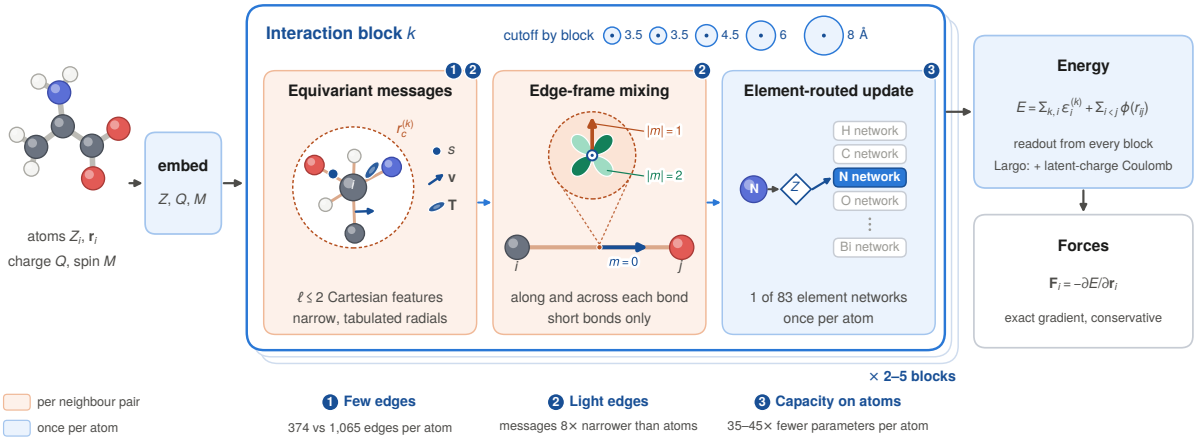


Figure 2: Architecture and design principles. Atoms, total charge and spin are embedded and passed through two to five interaction blocks whose cutoffs grow with depth. In each block, equivariant messages carry scalar (s), vector ($\mathbf{v}$) and rank-2 tensor ($\mathbf{T}$) features from every neighbour j within the block cutoff to atom i; in HyFF Andante and HyFF Largo, edge-frame mixing then splits the features of each short bond into components along the bond ($m = 0$) and across it ($|m| = 1, 2$) and mixes them; finally each atom is updated by the network of its own element. Warm stages do work for every neighbour pair, blue stages once per atom. The energy is read out from every block and the forces are its exact gradient. Below the block, the three design principles with their numbers for HyFF Andante on the 2,586-atom benchmark system; the numbered badges mark the stages that implement them.

2.1 Design principles

The cost of a message-passing potential is dominated by work on edges, because every operation applied to a neighbour pair is repeated for each of the tens to hundreds of neighbours of every atom, whereas work on atoms is done once. HyFF follows three rules that move work from edges to atoms (Figure 2).

Few edges. The number of neighbours grows with the cube of the cutoff: in the benchmark system an atom has on average 16.8 neighbours within 3.5 Å, 37.6 within 4.5 Å, 89.3 within 6 Å and 213 within 8 Å. Each block therefore uses its own cutoff, short in the first blocks, where features are refined from local bonding, and long only in the last blocks, which collect the environment that the earlier blocks have already summarised. The receptive field still grows to the sum of the cutoffs, 25.5 Å for Andante and Largo, but one force call visits 374 edges per atom instead of the 1,065 required by 8 Å in all five blocks. HyFF Presto visits 54 edges per atom and HyFF Vivace 71.

Light edges. The operations that remain on edges are kept narrow. Messages carry 8 to 84 channels, four to eight times fewer than the atom features they update, and their radial filters are smooth functions of distance alone that the HyFF inference engine evaluates from precomputed tables. The most expensive edge operation, edge-frame mixing in Andante and Largo, is applied only to bonds shorter than 3.5 Å, which visits 67 edges per atom across the four blocks that use it instead of the 161 that the block cutoffs would give.

Capacity on atoms. The capacity freed on edges is spent on atom updates, which are evaluated once per atom. HyFF uses a separate update network for each of the 83 elements, so that the update for a carbon atom never has to share its parameters with those of a metal. These networks hold 89–98% of all parameters, but each atom evaluates only one of them, and one atom uses 35 to 45 times fewer parameters than the model holds. HyFF Largo holds 1.2 billion parameters and uses 34 million per atom.

Together these rules let the family grow from 2.3 million to 1.2 billion parameters while the cost per step grows much more slowly: from Presto to Largo the parameter count rises 520-fold, the time per step 173-fold, and the force error falls 8.2-fold.

2.2 Features and inputs

Atom i carries H scalar channels s_i, T vector channels $\mathbf{v}_i$ and T symmetric traceless rank-2 tensor channels $\mathbf{T}_i$, the irreducible Cartesian representations of degree $\ell = 0$, 1 and 2. Working with Cartesian vectors and tensors, as in PaiNN (Schütt et al., 2021) and TensorNet (Simeon & De Fabritiis, 2023), avoids spherical-harmonic expansions and Clebsch–Gordan products while retaining the angular resolution that $\ell = 2$ equivariant models such as NequIP and MACE have shown to be important (Batzner et al.,

2022; Batatia et al., 2022). Because the network uses only polar vectors, it is equivariant under rotations and reflections. The initial scalar features are an embedding of the element plus a linear embedding of the total charge and the spin excess per atom, Q/N and $(M-1)/N$, so that one model covers neutral, charged and open-shell systems in the manner of UMA and SpookyNet (Wood et al., 2025; Unke et al., 2021).

2.3 Equivariant message passing

Each interaction block k uses the neighbours within its own cutoff $r_c^{(k)}$ (Table 1). Pair distances are expanded in 32 Gaussian radial functions (Schütt et al., 2017) multiplied by a cosine envelope, and each block applies a smooth switch $S(x) = x^3(10 - 15x + 6x^2)$, $x = (r_c^{(k)} - r)/w$, over the last $w = 1$ Å so that energies and forces vary continuously when neighbours cross a cutoff.

A message from atom j to atom i combines the sender features, projected to the narrow edge channels, with the bond direction $\hat{\mathbf{u}}_{ij}$ and its traceless square $\mathcal{Q}(\hat{\mathbf{u}}_{ij})$. The first block, whose input is scalar only, creates vectors and tensors through $s_j\hat{\mathbf{u}}_{ij}$ and $s_j\mathcal{Q}(\hat{\mathbf{u}}_{ij})$. Later blocks use seven paths that keep the degree of the sender feature or change it by contraction or product with the bond direction: s_j; $\mathbf{v}_j$, $s_j\hat{\mathbf{u}}_{ij}$ and $\mathbf{T}_j\hat{\mathbf{u}}_{ij}$ for vectors; and $\mathbf{T}_j$, $s_j\mathcal{Q}(\hat{\mathbf{u}}_{ij})$ and the symmetric traceless part of $\mathbf{v}_j \otimes \hat{\mathbf{u}}_{ij}$ for tensors. The scalar contractions $\mathbf{v}_j \cdot \hat{\mathbf{u}}_{ij}$ and $\hat{\mathbf{u}}_{ij}^\top \mathbf{T}_j \hat{\mathbf{u}}_{ij}$ feed back into the scalar message. The weight of each path is the product of a learned function of the sender's scalar features and a learned radial function shared across elements, so the message is a continuous filter in the sense of SchNet (Schütt et al., 2017) extended to $\ell \leq 2$. Messages are summed over neighbours and projected back to the full atom width.

2.4 Element-routed atom updates

After aggregation, each atom forms two learned channel mixtures of its vector and tensor features and computes their rotation invariants: the norms $|\mathbf{v}|$ and $|\mathbf{T}|$ and, from smoothly normalised copies, the three higher invariants $\mathrm{tr}(\mathbf{T}^3)$, $\mathbf{v}^\top\mathbf{T}\mathbf{v}$ and $|\mathbf{T}\mathbf{v}|^2$. These invariants and the layer-normalised scalars enter an update network that returns a scalar update and gates for the vector and tensor channels, in the gated style of PaiNN and NequIP (Schütt et al., 2021; Batzner et al., 2022).

The update network is element-specific. HyFF holds one copy for each of the 83 elements from H to Bi and evaluates, for every atom, the copy that matches its element. This carries the element-specific networks of Behler–Parrinello potentials, ANI and Deep Potential (Behler & Parrinello, 2007; Smith et al., 2017; Zhang et al., 2018) into a message-passing model. It resembles a mixture of experts with fixed routing; unlike the mixture of linear experts in UMA (Wood et al., 2025), which is selected per system from global properties, the routing here is per atom and needs no gating network. In HyFF Andante and HyFF Largo, the two norm invariants are passed through asinh and every output channel of the update network carries a learnable gate, which keeps the residual streams bounded on rare high-force structures.

2.5 Edge-frame mixing

HyFF Andante and HyFF Largo add edge-frame mixing (EFM) to their first four blocks. EFM follows the SO(2) edge convolutions introduced by eSCN (Passaro & Zitnick, 2023) and developed further in the EMFA SO(2) convolution of DPA4 (Li et al., 2026a). In a frame aligned with a bond, equivariant features separate into components of definite azimuthal order m, and these can be mixed by ordinary dense layers. HyFF computes the decomposition directly in Cartesian form with the projector $\mathbf{P} = \mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^\top$:

$$m = 0: \quad s, \quad \hat{\mathbf{u}}\cdot\mathbf{v}, \quad \sqrt{\tfrac{3}{2}}\,\hat{\mathbf{u}}^\top\mathbf{T}\hat{\mathbf{u}}; \qquad |m| = 1: \quad \mathbf{P}\mathbf{v}, \quad \sqrt{2}\,\mathbf{P}\mathbf{T}\hat{\mathbf{u}}; \qquad |m| = 2: \quad \text{the in-plane traceless part of } \mathbf{T}.$$

No local axes are constructed, so the operation is smooth for every bond orientation. The $m = 0$ components are mixed by a dense layer with a SiLU nonlinearity and an additive radial term, and they gate the $|m| = 1$ and $|m| = 2$ components, which are mixed channel by channel. HyFF omits the in-plane rotation that SO(2) convolutions apply between $+m$ and $-m$ components, because it acts as a cross product with the bond direction and breaks reflection symmetry. The mixed components are recombined into scalar, vector and tensor messages and added before the atom update. EFM is applied to edges shorter than 3.5 Å in all four blocks, so its cost scales with the number of close contacts rather than with the block cutoff; HyFF Andante uses 64, 64, 64 and 32 EFM channels, and HyFF Largo 168, 168, 168 and 84.

Table 1: The HyFF family. Active parameters are those used for one atom, since only one element network is evaluated per atom. Edge channels are the width of the messages. Time is the mean time of a complete 1 fs NVE MD step with bond constraints for a 2,586-atom solvated system on one NVIDIA H20, over a long run with the HyFF inference engine. Errors are over the full OMol25 validation set; energy MAE per atom is the total absolute energy error divided by the total number of atoms.

Model	Parameters	Active per atom	Width $H/T/$edge	Blocks	Cutoffs (Å)	EFM	ms per step	ns per day	Force MAE (meV/Å)	Energy MAE (meV/atom)
HyFF Presto	2.31 M	0.05 M	32/16/8	2	3.5, 4.5	–	0.71	121	47.7	5.78
HyFF Vivace	6.66 M	0.16 M	64/32/8	3	3.5, 3.5, 4.5	–	1.27	68	33.5	4.36
HyFF Moderato	34.8 M	0.83 M	128/64/16	4	3.5, 3.5, 6, 8	–	4.04	21	16.5	1.36
HyFF Andante	175 M	5.0 M	256/128/32	5	3.5, 3.5, 4.5, 6, 8	✓	29.5	2.9	7.65	0.620
HyFF Largo	1,202 M	34 M	672/336/84	5	3.5, 3.5, 4.5, 6, 8	✓	123	0.70	5.79	0.364

HyFF Largo also carries electronic embeddings and latent charges. All models are trained on OMol25 (Levine et al., 2025).

2.6 Energy, charge and spin

The energy is the sum over atoms of a readout of the layer-normalised scalar features of every block, plus a per-element offset and a learned pair term $\sum_{i<j} \phi_{Z_i Z_j}(r_{ij})$ for every element pair. Forces are its exact negative gradient, $\mathbf{F}_i = -\partial E/\partial \mathbf{r}_i$, so all models are conservative.

HyFF Largo adds two components for charged and open-shell systems. After every block, an attention over the atoms of each structure distributes the total charge Q and the spin excess $M-1$ to individual atoms, and the per-atom shares are embedded back into the scalar features, as in the electronic embeddings of SpookyNet (Unke et al., 2021). In addition, HyFF Largo predicts a latent atomic charge for each atom, bounds it smoothly, and shifts all charges of a structure uniformly so that they sum exactly to Q, as in PhysNet (Unke & Meuwly, 2019). The charges interact through a Gaussian-screened Coulomb energy

$$E_{\text{elec}} = \lambda\, k_e \sum_{i<j} q_i q_j \, \frac{\operatorname{erf}\!\left(r_{ij}/\sqrt{2}\sigma\right)}{r_{ij}},$$

with a learned width σ and a learned scale λ that starts at zero. The charges are not fitted to reference charges or dipoles; they are latent variables determined by the energies and forces alone, as in latent Ewald summation (Cheng, 2025) and the charge-equilibration electrostatics of fourth-generation neural network potentials (Ko et al., 2021). In the released HyFF Largo the learned scale remains small, and information beyond the block cutoffs reaches each atom mainly through the 25.5 Å receptive field and the global charge and spin embeddings.

2.7 Implementation for molecular dynamics

The released models run on the HyFF inference engine, a dedicated engine for inference and MD. Message passing uses fused kernels that reduce over neighbour lists without storing per-edge tensors and have hand-written backward passes for the forces. Radial filters are read from quintic interpolation tables. The element networks run as grouped matrix multiplications. Matrix products are computed at FP32 accuracy, on tensor cores with three-pass TF32 arithmetic where this is faster, and energies are accumulated in FP64. Each force evaluation is captured as a CUDA graph, and for Presto, Vivace and Moderato the complete MD loop, including velocity-Verlet integration with SHAKE/RATTLE constraints (Ryckaert et al., 1977; Andersen, 1983) and GPU neighbour-list rebuilds, runs on the device for up to 1,024 steps without returning to the host. HyFF Largo can distribute the atoms of a large system over several GPUs of one node, exchanging atom features between blocks.

3 The speed–accuracy frontier

3.1 Protocol

Accuracy. All models are scored on the complete OMol25 validation set (Val-Comp; 2,762,021 structures and 283,298,012 atoms). Force MAE pools absolute Cartesian force-component errors over all atoms. Energy MAE per atom is the sum of absolute total-energy errors divided by the total number of atoms, without any fitted offset. Because every model is scored on the same structures, this quantity equals the total-energy MAE per structure divided by 102.57, the mean number of atoms per structure, and is exact for published total-energy MAEs as well. For HyFF, UMA-S-1.2 (Wood et al., 2025), UBio-MolFM-V1.5 (Huang et al., 2026a;b), the direct and conservative OrbMol models (Rhodes et al., 2025; Orbital Materials, 2025) and DPA4-Pro (Li et al., 2026a) we evaluated all 2,762,021 structures ourselves. For the other

DPA4 models (Li et al., 2026a), the DPA4C models (Li et al., 2026b) and MACE-OMol-0 (ACEsuit, 2025; Batatia et al., 2025; 2022) we use the published Val-Comp errors (Li et al., 2026a;b; Levine et al., 2025). OrbMol was evaluated with the official calculator of `orb-models` 0.7.0 (checkpoints `orb-v3-direct-omol` and `orb-v3-conservative-omol`) in batches of up to 32 structures at highest FP32 matrix precision; on 128 validation structures, evaluating each structure separately lowers the mean energy error by 1.2% and 1.6% and leaves the force error unchanged.

Speed. All models simulate the same 2,586-atom solvated chorismate system in NVE at 1 fs on one exclusive NVIDIA H20, and every time includes neighbour-list work, energy and force evaluation, and integration. Each model runs in the fastest configuration that its developers officially support, and all baselines were run with their developers' released code, unmodified. HyFF runs on the HyFF inference engine with bond constraints and a 1.5 Å neighbour-list skin; its time is the wall time of a complete run divided by its length, 3,000 steps for Presto, Vivace and Moderato, 1,000 for Andante and 500 for Largo (median of three runs), after a separate 100-step warm-up run, and includes all neighbour-list rebuilds. DPA4 and DPA4C run in LAMMPS with the KOKKOS package and the device-resident `deepmd/kk` pair style, the configuration that their model cards prescribe for MD on GPUs, using compressed archives for DPA4C; their time is the LAMMPS loop time over 500 steps after 100 warm-up steps (DPA4-Air and DPA4-Plus 100 after 20, DPA4-Pro 30 after 10) with a 2 Å skin and without constraints, which slightly favours them; their forces in LAMMPS agree with the official Python calculator to 7×10^{-6} eV/Å. The other baselines run through their official Python calculators and spend almost all of a step in the model; their time is the median of 30 steps after 10 warm-up steps. On 128 size-stratified validation structures, the reduced-precision mode of the HyFF inference engine changes the force MAE by less than 10^{-4} meV/Å relative to strict FP32 evaluation.

3.2 Results

Figure 1 and Table 2 place the five HyFF models among 16 baselines. The five HyFF models form the whole energy frontier and, with DPA4-Pro at the most accurate and slowest end, the force frontier.

- **Fast regime.** The fastest baselines are the DPA4C models, compact single-layer models without multi-hop message passing (Li et al., 2026b). HyFF Presto is as fast as the fastest of them, DPA4C-Nano (0.71 versus 0.72 ms per step), at 2.3 times lower force error and 3.0 times lower energy error, and it is faster and more accurate than the other four; DPA4C-Plus, the most accurate, takes 3.0 times longer per step with 26% higher force error. HyFF Vivace matches the force error of DPA4-Nano at 11 times lower cost, and HyFF Moderato is 7.0 times faster than DPA4-Mini with 24% lower force error and 35% lower energy error.
- **Intermediate regime.** HyFF Andante reaches a force error comparable to that of UMA-S-1.2 (mean 7.65 versus 7.55 meV/Å, identical medians of 3.58 meV/Å) at 29 times lower cost. It is 5.7 times faster than DPA4-Air, 13 times faster than MACE-OMol-0 and 15 times faster than OrbMol (conservative), each at lower force error.
- **Accurate regime.** HyFF Largo has the lowest energy error of all models scored, 0.364 meV/atom, compared with 0.380 for DPA4-Pro, 0.428 for UMA-S-1.2 and 0.507 for DPA4-Plus. It is more accurate than UMA-S-1.2 on every metric (force MAE 23% lower, energy MAE 15% lower) at 7.0 times lower cost, and more accurate than DPA4-Plus on forces and energies at 3.1 times lower cost. DPA4-Pro reaches a lower force error (4.96 versus 5.79 meV/Å) at 9.7 times the cost of HyFF Largo.

For the three models with complete per-structure predictions, HyFF Largo, DPA4-Pro and UMA-S-1.2, we also compared errors on identical structures with 2,000 paired bootstrap resamples. The energy advantage of HyFF Largo over DPA4-Pro is significant (4.2% lower mean error, 95% interval 3.3–5.2%). Against UMA-S-1.2, HyFF Largo is more accurate on force and energy, in mean and median, with all intervals excluding zero, and has lower force error on 94.9% of individual structures.

Table 2: All models on the full OMol25 validation set, ordered by MD time. Force MAE in meV/Å; energy MAE per atom in meV/atom. All times are our measurements; accuracy is ours except for the DPA4 models other than DPA4-Pro, the DPA4C models and MACE-OMol-0, whose published values are used. Models marked • are on the force frontier, ○ on the energy frontier. HyFF times are long-run means on the HyFF inference engine with bond constraints.

Model	ms per step	Force MAE	Energy MAE
HyFF Presto•○	0.71	47.7	5.78
DPA4C-Nano	0.72‡	107.8	17.32
DPA4C-Mini	0.86‡	81.6	12.08
DPA4C-Neo	0.98‡	75.1	10.46
HyFF Vivace•○	1.27	33.5	4.36
DPA4C-Air	1.39‡	69.3	9.03
DPA4C-Plus	2.17‡	60.0	7.09
HyFF Moderato•○	4.04	16.5	1.36
DPA4-Nano	13.9‡	33.7	3.39
DPA4-Mini	28.3‡	21.8	2.10
HyFF Andante•○	29.5	7.65	0.620
DPA4-Neo	71.3‡	14.4	1.23
HyFF Largo•○	123	5.79	0.364
UBio-MolFM-V1.5	145	11.5	–
DPA4-Air	168‡	8.20	0.613
OrbMol (direct)	230	13.0	0.728
DPA4-Plus	379‡	6.33	0.507
MACE-OMol-0	386	10.9	1.93
OrbMol (conservative)	439	13.4	0.676
UMA-S-1.2	860	7.55	0.428
DPA4-Pro•	1,190‡	4.96	0.380

‡LAMMPS with the KOKKOS package (`deepmd/kk`), without constraints. HyFF Presto and DPA4C-Nano take 0.712 and 0.718 ms.

4 Molecular dynamics at scale

4.1 Cost across system sizes

HyFF runs MD entirely on the GPU, including neighbour-list construction, and for Presto, Vivace and Moderato the integration loop itself is device-resident. At 33,426 atoms, HyFF Presto takes 3.5 ms per step, or 25 simulated ns per day, while UMA-S-1.2 takes 10.4 s. HyFF Largo simulates 103,440 atoms in 6.3 s per step on one GPU; distributing the atoms over the GPUs of one node reduces this to 2.2 s on four GPUs and 1.17 s on eight, a 5.4-fold speed-up. These times are measured as in Section 3.1, as the wall time of a long NVE run divided by its length.

4.2 Energy conservation

Table 3 summarises NVE simulations at 1 fs with bond constraints. Each run starts from the same structure with a different velocity seed, and the drift is the slope of a linear fit to the total energy per atom. All drifts are below 0.035 meV atom^{-1} ns^{-1} at both system sizes.

Table 3: Fitted total-energy drift in NVE simulations at 1 fs.

Model	Atoms	Runs	Drift (meV atom^{-1} ns^{-1})
HyFF Presto	19,366	10×100 ps	-0.023 to $+0.032$
HyFF Vivace	19,366	3×100 ps	-0.014 to $+0.020$
HyFF Moderato	2,182	10×100 ps	$\lvert d\rvert \leq 0.029$
HyFF Moderato	19,366	3×100 ps	$\lvert d\rvert \leq 0.011$

5 Applications

The simulations below show what the fast and intermediate models make practical. Solvated systems used constrained bonds to hydrogen, rigid water and 1 fs steps, with a stochastic velocity-rescaling thermostat (Bussi et al., 2007) at 300 K for NVT runs.

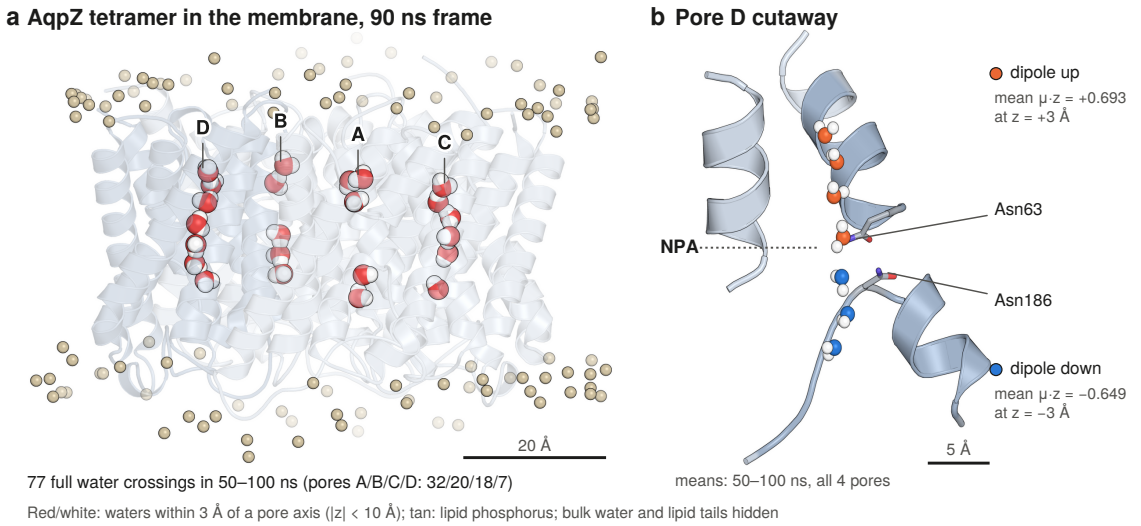


Figure 3: AqpZ water channel in a POPE membrane, 100 ns of unbiased 300 K dynamics with HyFF Presto (58,560 atoms). Water molecules in the pores form a single file whose dipoles reverse at the NPA motif; 77 molecules crossed a pore completely between 50 and 100 ns.

5.1 Aquaporin water channel in a membrane

The *E. coli* aquaporin AqpZ tetramer (PDB 1RC2) in a POPE bilayer with 9,000 water molecules (58,560 atoms) was simulated for 100 ns with HyFF Presto without any bias (Figure 3). The backbone stayed within 1.81 Å of the crystal structure, and 95.2% of crystal α-helical residues kept their geometry. All four pores held 6.7–7.4 water molecules arranged predominantly in single file, and the water dipoles reversed orientation at the conserved NPA motif, the hallmark of aquaporin water transport. Between 50 and 100 ns, 77 water molecules crossed a pore completely, corresponding to a diffusive water permeability of $(0.58 \pm 0.08) \times 10^{-14}\,\mathrm{cm^3\,s^{-1}}$.

5.2 Protein–ligand binding against a classical force field

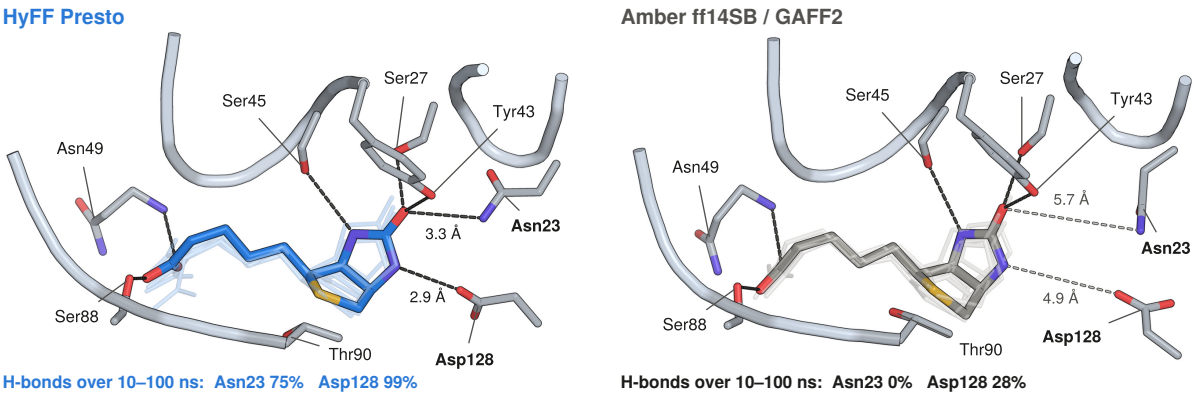
The streptavidin–biotin complex (PDB 1STP, 19,366 atoms) was simulated for 100 ns with HyFF Presto and with a classical force field (Amber ff14SB protein parameters (Maier et al., 2015), GAFF2 (Wang et al., 2004) for biotin and TIP3P water (Jorgensen et al., 1983)) from the same starting structure. With HyFF Presto, biotin stayed closer to its crystal pose (0.64 versus 0.88 Å), and the hydrogen bonds from Asn23 and Asp128 were present 75% and 99% of the time, compared with 0% and 28% with the classical force field. In a snapshot at 95 ns, both contacts are at hydrogen-bond distance with HyFF Presto (3.25 and 2.91 Å) and broken in the classical simulation (5.73 and 4.93 Å; Figure 4).

5.3 Chemical reactions

HyFF uses no fixed bond topology, so the same models describe bond breaking and formation.

Claisen rearrangement inside an enzyme. In *B. subtilis* chorismate mutase with substrate and explicit water (46,282 atoms), HyFF Moderato followed the conversion of chorismate to prephenate (Figure 5a). The initial state was generated by unbiased shooting: a structure on a steered reaction path in the enzyme was given random 300 K velocities, and its backward-in-time branch relaxed into the reactant basin. Started from that reactant endpoint with the velocities reversed, a single unbiased 1 ps NVE trajectory broke the C–O bond (1.50 to 3.33 Å) and formed the C–C bond (4.19 to 1.56 Å), reaching the product geometry after 0.57 ps. The product then remained stable for 1 ns of unbiased NVT simulation.

Proton transfer and cycloaddition. In 550 ps of unbiased 300 K dynamics of malonaldehyde, the shared proton moved between the two oxygen atoms 23 times while the carbon–oxygen framework stayed intact (Figure 5b). For a Diels–Alder reaction between butadiene and ethylene in 986 explicit water molecules (2,974 atoms), the reactants were steered together for 2 ps and released. The ring closed 0.9 ps after release and remained closed for the remaining 19 ps of unbiased NVE, whereas an unsteered control from identical coordinates and velocities separated without reacting (Figure 5c).



Pocket at 95.01 ns in both simulations, Cα of the pocket aligned. Dark dashes: N/O–N/O < 3.5 Å; light dashes: contact lost; faint biotin: other frames, 90–100 ns.

Figure 4: Streptavidin–biotin binding pocket at the same time point (95.01 ns) of 100 ns simulations with HyFF Presto (left) and a classical force field (right), started from the same structure. Biotin's closest contacts to Asn23 and Asp128 are 3.25 and 2.91 Å with HyFF Presto and 5.73 and 4.93 Å with the classical force field; over 10–100 ns these hydrogen bonds were present 75% and 99% of the time with HyFF Presto, compared with 0% and 28%.

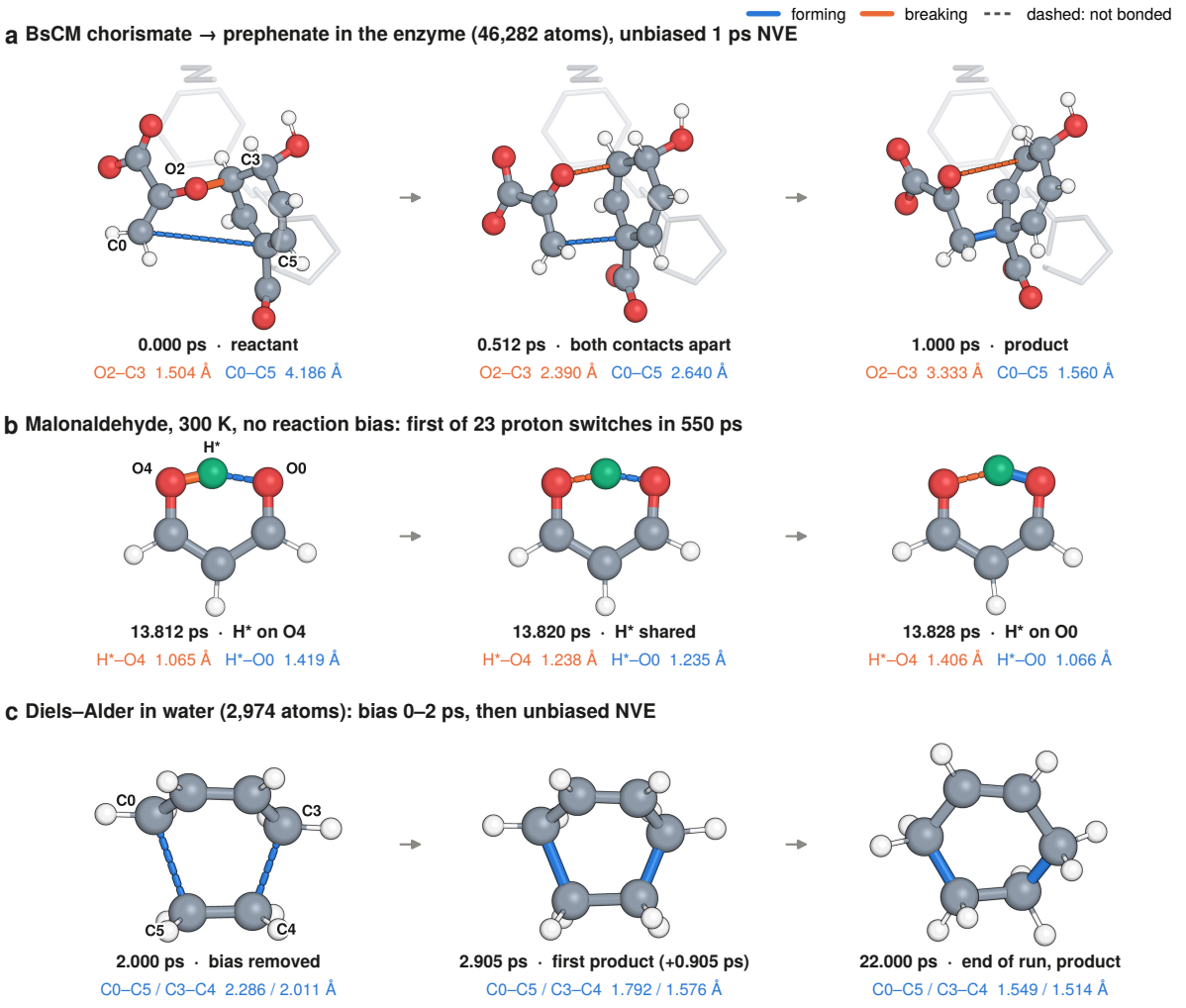


Figure 5: Reactions with HyFF Moderato. (a) Chorismate to prephenate inside *B. subtilis* chorismate mutase (46,282 atoms), unbiased 1 ps NVE trajectory. (b) Proton transfer in malonaldehyde at 300 K. (c) Diels–Alder cycloaddition in water after 2 ps of steering.

6 Availability

The five models are distributed with the `hyff` Python package, which provides GPU inference, batched inference over many structures, and molecular dynamics with SHAKE/RATTLE constraints (Ryckaert et al., 1977; Andersen, 1983), NVE and stochastic velocity-rescaling integration, including multi-GPU MD for HyFF Largo.

References

ACEsuit. MACE-OMol-0 model release. Model release, 2025. URL https://github.com/ACEsuit/mace-foundations.

Hans C. Andersen. Rattle: a "velocity" version of the Shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52:24–34, 1983. doi: 10.1016/0021-9991(83)90014-1.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436, 2022.

Ilyes Batatia, Philipp Benner, Yuan Chiang, et al. A foundation model for atomistic materials chemistry. *The Journal of Chemical Physics*, 163:184110, 2025. doi: 10.1063/5.0297006.

Simon Batzner, Albert Musaelian, Lixin Sun, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13:2453, 2022. doi: 10.1038/s41467-022-29939-5.

Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98:146401, 2007. doi: 10.1103/physrevlett.98.146401.

Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126:014101, 2007. doi: 10.1063/1.2408420.

Bingqing Cheng. Latent Ewald summation for machine learning of long-range interactions. *npj Computational Materials*, 11:80, 2025. doi: 10.1038/s41524-025-01577-7.

Lin Huang, Arthur Jiang, XiaoLi Liu, et al. UBio-MolFM: a universal molecular foundation model for bio-systems. arXiv:2602.17709, 2026a. URL https://arxiv.org/abs/2602.17709.

Lin Huang, Frank Peng, JiaJun Cheng, et al. UBio-MolFM: enabling biomolecular dynamics at DFT accuracy and 10^5 atoms with one untuned potential. arXiv:2608.18623, 2026b. URL https://arxiv.org/abs/2608.18623.

William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79:926–935, 1983. doi: 10.1063/1.445869.

Tsz Wai Ko, Jonas A. Finkler, Stefan Goedecker, and Jörg Behler. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nature Communications*, 12:398, 2021. doi: 10.1038/s41467-020-20427-2.

Daniel S. Levine, Muhammed Shuaibi, Evan Walter Clark Spotte-Smith, et al. The Open Molecules 2025 (OMol25) dataset, evaluations, and models. arXiv:2505.08762, 2025. URL https://arxiv.org/abs/2505.08762.

Tiancheng Li, Wentao Li, Anyang Peng, et al. DPA4: pushing the accuracy-cost frontier of interatomic potentials with EMFA SO(2) convolution. arXiv:2606.02419, 2026a. URL https://arxiv.org/abs/2606.02419.

Tiancheng Li, Jianming Xue, Linfeng Zhang, Duo Zhang, and Han Wang. Universal machine-learning molecular dynamics at the speed of empirical potentials. arXiv:2608.19041, 2026b. URL https://arxiv.org/abs/2608.19041.

James A. Maier, Carmenza Martinez, Koushik Kasavajhala, et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11:3696–3713, 2015. doi: 10.1021/acs.jctc.5b00255.

Orbital Materials. OrbMol: models trained on OMol25. Model release, 2025. URL https://github.com/orbital-materials/orb-models.

Saro Passaro and C. Lawrence Zitnick. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 27420–27438, 2023. URL https://proceedings.mlr.press/v202/passaro23a.html.

Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, et al. Orb-v3: atomistic simulation at scale. arXiv:2504.06231, 2025. URL https://arxiv.org/abs/2504.06231.

Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C. Berendsen. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, 1977. doi: 10.1016/0021-9991(77)90098-5.

Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, et al. SchNet: a continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, volume 30, pp. 992–1002, 2017.

Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 9377–9388, 2021. URL https://proceedings.mlr.press/v139/schutt21a.html.

Guillem Simeon and Gianni De Fabritiis. TensorNet: Cartesian tensor representations for efficient learning of molecular potentials. In *Advances in Neural Information Processing Systems*, volume 36, pp. 37334–37353, 2023.

J. S. Smith, O. Isayev, and A. E. Roitberg. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science*, 8:3192–3203, 2017. doi: 10.1039/c6sc05720a.

Oliver T. Unke and Markus Meuwly. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of Chemical Theory and Computation*, 15:3678–3693, 2019. doi: 10.1021/acs.jctc.9b00181.

Oliver T. Unke, Stefan Chmiela, Michael Gastegger, et al. SpookyNet: learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 12:7273, 2021. doi: 10.1038/s41467-021-27504-0.

Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25:1157–1174, 2004. doi: 10.1002/jcc.20035.

Brandon M. Wood, Misko Dzamba, Xiang Fu, et al. UMA: a family of universal models for atoms. arXiv:2506.23971, 2025. URL https://arxiv.org/abs/2506.23971.

Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and Weinan E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical Review Letters*, 120:143001, 2018. doi: 10.1103/physrevlett.120.143001.