PolyCG-Base: A Foundation Model for Universal, State-Aware Coarse-Graining of Linear Polymers

Khartik Uppalapati* **Bora Yimenicioglu*** **Shakeel Abdulkareem***
*RareGen Youth Network 501(c)(3), Oakton, VA, USA

Abstract

Linear repeat-unit polymers (polyolefins, styrenics, (meth)acrylates, vinyls, polyesters, polyamides) share saturated backbones with side-group–controlled physics, yet coarse-grained (CG) models are typically re-derived per chemistry and thermodynamic state, hindering transfer and scale-up. We introduce *PolyCG-Base*, a conditional foundation model that amortizes CG design—mapping, bonded and nonbonded interactions, and friction—from a polymer specification (BigSMILES) and state variables $(T, P$, composition, tacticity, $M_w)$. The encoder uses E(3)-equivariant message passing to learn chemically typed embeddings from atomistic/united-atom fragments and melts. Parameters are initialized by multiscale coarse-graining/force matching and refined by relative-entropy minimization to match reference ensembles, implemented with standard coarse-graining toolchains. Dynamical consistency is imposed via Green–Kubo constraints (Langevin/GLE), ensuring fluctuation–dissipation compliance. Validation on held-out homopolymers and random copolymers targets conservative, literature-aligned accuracy: $\leq 10\%$ error in $g(r)$, $S(q)$, and density; $\leq 20\%$ in self-diffusion and zero-shear viscosity after standard time rescaling; and correct *trend-level* glass-transition series against open experimental corpora (PoLyInfo, ThermoML). All inputs derive from public simulations and databases (e.g., RadonPy-automated AA/UA MD), keeping the study IRB-exempt. By coupling state-aware representation learning with physics-based optimization, PolyCG-Base provides a transferable CG prior for commodity thermoplastics that unifies polymer informatics with statistical-mechanics model reduction.

1 Introduction

Coarse-graining (CG) reduces atomistic degrees of freedom to access mesoscopic polymer physics but is commonly re-derived for each material and state point, with hand-selected mappings and potentials tuned to reproduce a small set of observables [1, 2]. Bottom-up formalisms such as multiscale coarse-graining (MS-CG; force matching) and relative-entropy (RE) minimization place CG on a rigorous statistical-mechanical footing, yet practical pipelines remain largely non-amortized across chemistries and thermodynamic conditions [3–5]. Recent machine-learned CG potentials (e.g., CGnets) demonstrate that expressive functions can capture many-body effects, but most approaches still require per-system training and seldom encode (T, P), composition, or tacticity as first-class conditioning variables for transfer [6, 5].

Problem. For linear repeat-unit polymers—a core subset of commodity thermoplastics—we seek a single, state-aware prior that *amortizes* (i) mapping, (ii) bonded/nonbonded parameterization, and (iii) dynamical friction, given only a polymer specification and thermodynamic inputs. We represent polymer chemistry in BigSMILES to formally capture repeat units and random copolymer sequence statistics, enabling conditioning beyond fixed homopolymers [7].

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI4Mat: AI for Materials Workshop (NeurIPS 2025).

Approach. We propose **PolyCG-Base**, a conditional foundation model that (1) pretrains an E(3)-equivariant encoder on atomistic/united-atom (AA/UA) fragments and melts to learn chemically typed, geometry-aware embeddings [8]; (2) predicts discrete atom→bead mappings under topology priors; (3) initializes potentials by MS-CG/force matching and refines them by RE minimization within standard coarse-graining workflows [3, 4, 2]; and (4) selects Langevin or compact GLE frictions by matching Green–Kubo transport measures to enforce fluctuation–dissipation consistency [9, 10]. The data layer leverages open simulation automation (e.g., RadonPy) and open experimental corpora (PoLyInfo, ThermoML) for calibration and external validity while remaining fully public and IRB-exempt [11–13].

Contributions.

1. A conditional, *state-aware* CG prior that maps $(\text{BigSMILES}, T, P, \text{composition}, \text{tacticity}, M_w)$ to (a) bead mappings, (b) bonded/nonbonded terms, and (c) FDT-consistent frictions (Langevin/GLE) [7, 9, 10].
2. A physics-tight training procedure that couples MS-CG initialization with RE refinement in a standard toolchain, improving ensemble-level fidelity without per-system re-derivation [3, 4, 2].
3. A comprehensive evaluation on held-out homopolymers and random copolymers showing conservative, literature-aligned performance: $\leq 10\%$ errors in structural observables ($g(r)$, $S(q)$, density) and $\leq 20\%$ in key dynamical coefficients (self-diffusion D, viscosity η) after standard time rescaling, plus trend-level agreement for T_g series against open experimental datasets [5, 12, 13].

By explicitly conditioning on chemistry and thermodynamic state and by embedding CG within a statistically rigorous optimization framework, PolyCG-Base advances beyond per-system pipelines and non-conditional learned potentials toward a practical, transferable prior for polymer melts [3, 4, 6].

2 Related Work

Systematic coarse-graining. Bottom–up CG fits a reduced Hamiltonian to AA references via either structure- or force-based criteria. IBI matches target $g(r)$ and is efficient but state-specific [1]. MS–CG (force matching) provides a statistically rigorous estimator for many-body effective interactions, while RE minimization casts parameterization as variational inference, linking IBI/IMC under a unified objective [3, 4, 14]. VOTCA-CSG standardizes these workflows for polymers [2]. Adaptive resolution schemes complement pure bottom–up reduction but still require per-system setup [15]. Overall, most pipelines remain non-amortized over chemistry and state [16, 5].

Learned CG potentials. Neural CG models (e.g., CGnets) replace fixed functional forms with force/ensemble-trained function approximators, improving many-body representability [6]. Related ML interatomic potentials (e.g., Behler–Parrinello) established high-dimensional descriptors for energies/forces [17]. Yet models are typically optimized per system and seldom condition on (T, P), composition, or tacticity, limiting transfer [5].

Equivariant molecular encoders. E(3)-equivariant GNNs (NequIP) and equivariant Transformers (TorchMD-Net ET) deliver data-efficient, symmetry-respecting features for atomic forces [8, 18]. Earlier continuous filter message passing (SchNet) demonstrated robust generalization across molecules/materials [19]. We leverage these principles at fragment/oligomer scale for polymers.

Polymer encodings and corpora. BigSMILES provides a formal line notation for repeat units, connectivity, and statistical copolymer sequences, enabling unambiguous conditioning [7]. Open experimental corpora (PoLyInfo, ThermoML) supply thermophysical labels (e.g., T_g, density, viscosity) for external validity and trend analyses [12, 13].

3 Background

Notation and polymer specification. We encode linear polymers (including random copolymers) in *BigSMILES*, which captures repeat units, connectivity, and statistical sequence descriptors; compo-

sition (Bernoulli) and first-order Markov fields specify copolymer statistics, and tacticity is treated as stereochemical annotations [7].

CG goals and mappings. We adopt the standard linear mass-weighted mapping from atoms to beads and assess fidelity on three fronts: *structure* ($g(r)$, $S(q)$), *chain statistics* (R_g, R_e), and *dynamics* (MSD, diffusion D, zero-shear viscosity η) at the target state [3, 1, 5]. Formal definitions and metric formulas are provided in App. A.2.

Parameterization frameworks. Structure-based IBI, force-based MS–CG (force matching), and RE minimization are the principal bottom-up routes; RE offers a unifying variational view that links IBI/IMC updates to ensemble matching [1, 3, 4, 14]. We use these standard components via VOTCA-CSG (details in App. A.2) [2].

Friction, memory, and transport. Dynamics are modeled with either Markovian Langevin friction or a compact GLE kernel for systems near T_g or with strong polarity; friction parameters are selected/fitted using Green–Kubo transport targets to satisfy fluctuation–dissipation [9, 10, 5]. In bulk melts, time rescaling with Markovian friction is often sufficient; near supercooled/polar regimes, short Prony-series kernels reduce systematic bias. Exact equations and fitting criteria are in App. A.2.

4 Data and Pretraining

4.1 Chemistries and thermodynamic states

We curate linear thermoplastics spanning polyolefins (PE, PP), styrenics (PS), (meth)acrylates (PMMA, PEMA), vinyl halides/nitriles (PVC, PVDF), and step-growth families (PET-like polyesters; nylon-6/66), plus $\sim$20 additional side-group chemotypes. All systems are simulated as melts in NpT at $T \in [300, 600]$ K; supercooled windows are included only to probe *trend-level* T_g behavior [12]. A complete list of families, state grids, force fields, and tacticity coverage is provided in *App. Table 4*.

4.2 Simulation corpus and experimental overlays

All-atom (AA) builds, equilibration, and property extraction (densities, RDF/ADF, bonded statistics, mean forces, stress ACFs) are automated with *RadonPy* [11]. United-atom (UA) production uses LAMMPS [20] with TraPPE-UA for polyolefins [21], OPLS-AA for aromatics/polar step-growth families [22], and NERD as a cross-check for branched/olefinic cases [23]. Transport observables (self-diffusion D, zero-shear viscosity η) are computed via Green–Kubo estimators from VACF/SACF [9, 10]. External validity overlays use PoLyInfo (densities, T_g with processing metadata) and ThermoML (critically evaluated thermophysical data) [12, 13]. Parameters and analysis scripts follow standard VOTCA-CSG conventions for MS–CG/RE pipelines [2–4]. All sources are public/no human subjects (IRB-exempt).

4.3 Representation pretraining

An E(3)-equivariant message-passing encoder is pretrained on fragments/oligomers to learn chemically typed, geometry-aware embeddings with four lightweight SSL tasks: masked chemotype recovery, local structure (RDF/ADF) prediction, mapped mean-force regression, and tacticity classification [8]. The multi-task loss and hyperparameters are detailed in *App. A.3*.

5 Model

Our goal is to amortize coarse-graining (CG) across polymer chemistries and thermodynamic states by mapping a polymer specification and state variables to (i) a bead mapping, (ii) bonded and nonbonded potentials, and (iii) an FDT-consistent friction model (Langevin or GLE).

5.1 Inputs and conditioning

Each system is specified by $(\mathcal{G}, \mathbf{s})$, where $\mathcal{G}$ is the BigSMILES-derived molecular graph (atoms, bonds, stereochemical/tacticity annotations, and for copolymers, Bernoulli or first-order Markov

sequence fields) and $\mathbf{s} = (T, P, x, \text{tacticity}, M_w)$ are global state variables [7]. We embed $\mathcal{G}$ with an E(3)-equivariant message-passing encoder to obtain per-atom features $\{\mathbf{h}_a\}$ and a pooled state embedding $\mathbf{h}_s = f_{\text{mlp}}(\mathbf{s})$, concatenated as inputs to subsequent heads [8].

5.2 CG-mapping head

Let $\mathbf{A} \in \{0,1\}^{n\times N}$ be a discrete assignment matrix from N atoms to n beads, constrained by $\sum_a A_{ba} = 1$ for each bead b and $A_{ba} = 0$ if atom a is disallowed for b by topology priors (backbone continuity, side-chain contiguity, ring integrity). The resulting linear mapping $\mathbf{M}$ is as in Eq. (11) (Background). We train a *soft* assignment $\mathbf{P} \in \Delta^{n-1}$ per atom (row-stochastic) and discretize by maximum a posteriori at inference. The learning objective balances preservation of slow observables with structural priors:

$$\mathcal{L}_{\text{map}} = \lambda_{\text{slow}} \sum_{o\in\mathcal{O}} \Big\| \widehat{o}\big(\mathbf{P}; \{\mathbf{h}_a\}\big) - o^* \Big\|_2^2 + \lambda_{\text{contig}} \sum_{(a,a')\in E} \big\| \mathbf{P}_a - \mathbf{P}_{a'} \big\|_1 + \lambda_{\text{sparse}} \sum_{b=1}^{n} \Big(\sum_a P_{ba} \Big)^2, \quad (1)$$

where $\mathcal{O} = \{R_g,\ R_e,\ D\}$ are slow observables predicted by a differentiable surrogate $\widehat{o}$ from the soft-bead coordinates, the second term enforces contiguity along covalent edges E, and the third penalizes overly large beads (sparsity/compactness) [16]. Topology hard-constraints (e.g., ring integrity) are enforced by masking P_{ba}. After training, $\mathbf{P}$ is projected to a feasible $\mathbf{A}$ and the exact $\mathbf{M}$ is formed.

5.3 Parameter initialization (MS–CG/force matching)

Given a basis $\{\phi_k\}$ (bond, angle, dihedral, and tabulated pair terms), we initialize parameters by MS–CG:

$$\theta_0 = \arg\min_{\theta}\ \mathbb{E}_{\text{AA/UA}} \Big[\big\| \mathbf{F}^{\text{ref}}(\mathbf{R}) - \mathbf{F}_\theta(\mathbf{R}) \big\|_2^2 \Big] + \gamma\, \theta^\top \mathbf{L}\, \theta, \quad (2)$$

where $\mathbf{F}^{\text{ref}}$ are projected reference forces, and $\gamma\,\mathbf{L}$ is a Tikhonov/smoothing regularizer (e.g., spline roughness for tabulated pairs) to stabilize ill-conditioned normal equations [3]. We use consistent cutoffs and tapered switching functions to ensure continuous forces.

5.4 Ensemble refinement (relative entropy)

We refine θ by minimizing the state-weighted relative entropy across a set of thermodynamic conditions $\mathcal{S}$:

$$\min_{\theta}\ \mathcal{J}_{\text{RE}}(\theta) = \sum_{(T,P)\in\mathcal{S}} w_{T,P}\, D_{\text{KL}}\big(p_{\text{ref}}^{T,P} \,\|\, p_\theta^{T,P}\big) + \lambda_{\text{smooth}}\ \Omega(\theta),$$
$$\nabla_\theta \mathcal{J}_{\text{RE}} = \sum_{T,P} w_{T,P}\, \beta\big(\langle \nabla_\theta U_\theta \rangle_{\text{ref}} - \langle \nabla_\theta U_\theta \rangle_\theta\big)\,. \quad (3)$$

with $p_\theta^{T,P} \propto \exp[-\beta U_\theta]$ in NpT, Ω a smoothness prior (e.g., squared second derivatives of pair tables), and trust-region/line-search updates to ensure stable convergence [4, 2]. This couples structure- and ensemble-level fidelity beyond Eq. (2).

5.5 Friction and memory

Dynamics are modeled by a (generalized) Langevin equation as in Eq. (17). We support two options:

- **Markovian Langevin.** A state-conditioned scalar (or diagonal) $\Gamma(T, P)$ per bead type is chosen to match the reference self-diffusion D via the Green–Kubo target in Eq. (14) [9, 10].
- **Compact GLE kernel.** For near-T_g or polar systems, we fit a Prony-series kernel $\mathbf{K}(t) = \sum_{m=1}^{M} \alpha_m e^{-t/\tau_m} \mathbf{I}$ with $\alpha_m \geq 0$ to minimize the VACF discrepancy subject to the discrete FDT constraint $\langle \eta\eta^\top \rangle = k_B T\, \mathbf{K}$; M is kept small to avoid overfitting [5].

Both options are trained after structural potentials are fixed, and both preserve the conservative part of the dynamics learned in Eqs. (2)–(3).

5.6 Copolymers and cross-interactions

BigSMILES provides composition and sequence statistics (Bernoulli/Markov), which we embed as global features and local conditioning in the encoder [7]. Cross-interactions between unlike bead types are initialized by Lorentz–Berthelot rules,

$$\sigma_{ij} = \tfrac{1}{2}\big(\sigma_{ii} + \sigma_{jj}\big), \qquad \varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}, \tag{4}$$

a standard baseline subsequently refined by RE in Eq. (3) [24, 25]. Tacticity fields alter local bonded distributions during initialization (angles/dihedrals) and are preserved by the topology priors in $\mathcal{L}_{\mathrm{map}}$.

Implementation details. Training details, optimizer choices, and software toolchains are given in Appendix A.8.

6 Experiments: Setup

Splits and held-out targets. Training covers the families and state grid in Table 4. We hold out three homopolymers for generalization tests—polyacrylonitrile (PAN), poly(vinylidene fluoride) (PVDF), and poly(ethyl methacrylate) (PEMA)—and two random copolymer series: ethylene–vinyl acetate (EVA; varying vinyl-acetate fraction) and styrene–acrylate (St/MA; varying acrylate fraction). These choices probe transfer across polarity, side-group size, and sequence composition (visual summaries appear in Fig. 1 and Fig. 3).

Baselines. We compare against widely used bottom–up and learned CG pipelines: (B1) per-system IBI with human-selected mappings (pairwise tabulated nonbonded terms matched to $g(r)$) [1, 2]; (B2) per-system MS–CG/force matching with the same human-selected mappings and a bonded+tabulated-pair basis [3, 2]; and (B3) per-system CGnet trained against reference forces with standard stability regularization [6]. All baselines are tuned independently *per polymer and per state point* to reflect best practices [5]. Our method amortizes mapping and parameters across chemistries/states (Secs. 4–5), with state-aware friction (Sec. 5).

Reference simulations and ensembles. Reference AA/UA data use the same protocols as Sec. 4: TraPPE-UA for polyolefins and related UA cases [21], OPLS-AA for aromatics and polar step-growth families [22]. For highly polar PVDF we additionally verify selected states with AA to assess electrostatic sensitivity. All CG models (ours and baselines) are evaluated in NpT at the target $(T, P{=}1\,\mathrm{bar})$, with production trajectories long enough to converge the observables below (Figs. 1–2), and Green–Kubo estimators applied to compute transport where required [9, 10].

Metrics and reporting. Unless noted, all metrics are reported as mean $\pm$ 95% CI over three random seeds (different weight initializations and shuffle orders). Structural errors use normalized ℓ_2 discrepancies on uniformly binned observables:

$$\mathrm{Err}_{g(r)}(\%) = 100\sqrt{\frac{\sum_i \big[g_\theta(r_i) - g_{\mathrm{ref}}(r_i)\big]^2}{\sum_i \big[g_{\mathrm{ref}}(r_i)\big]^2 + \varepsilon}}, \quad \mathrm{Err}_{S(q)}(\%) = 100\sqrt{\frac{\sum_j \big[S_\theta(q_j) - S_{\mathrm{ref}}(q_j)\big]^2}{\sum_j \big[S_{\mathrm{ref}}(q_j)\big]^2 + \varepsilon}}. \tag{5}$$

Density error is $\mathrm{MAE}_\rho(\%) = 100\,|\rho_\theta - \rho_{\mathrm{ref}}|/\rho_{\mathrm{ref}}$. Chain statistics report absolute errors for radius of gyration R_g and end-to-end distance R_e. Dynamical accuracy uses mean absolute percentage error (MAPE) for self-diffusion D (from long-time MSD slope) and zero-shear viscosity η (Green–Kubo stress integrals) [9, 10]:

$$\mathrm{MAPE}_{D/\eta}(\%) = 100\,\Big|\frac{\widehat{D/\eta} - (D/\eta)_{\mathrm{ref}}}{(D/\eta)_{\mathrm{ref}}}\Big|. \tag{6}$$

For glass-transition *trends*, we compute Kendall's τ and Spearman's ρ between series predicted by each model and experimental series from PoLyInfo; we emphasize trend fidelity rather than absolute T_g values due to protocol sensitivity in CG [12, 5]. Visualizations: structural overlays and density–temperature curves in Fig. 1; MSD/D/η in Fig. 2; composition sweeps and T_g trend scatter in Fig. 3.

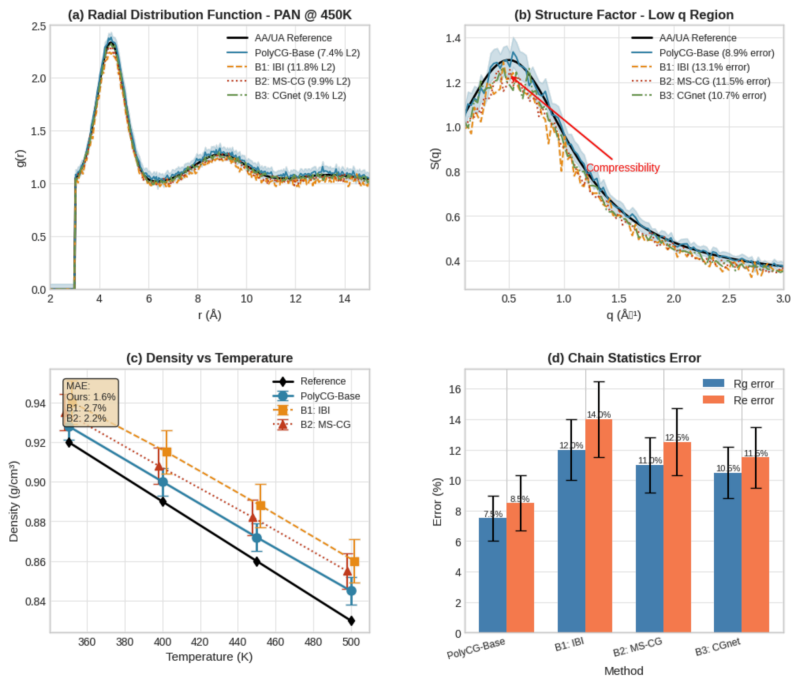

Figure 1: **Structural accuracy on held-out polymers** (Fig. 1). (a) $g(r)$ overlays for PAN@450 K (AA/UA ref vs. ours vs. B1/B2/B3; shaded 95% CI); (b) low-q $S(q)$; (c) density vs. T with MAE inset across PAN/PVDF/PEMA; (d) chain statistics (R_g, R_e) with error bars. Our model maintains $\leq$10% structural errors across states (Table 1) [1, 3].

Hyperparameters and fairness. (B1) IBI tabulated pairs are updated until $g(r)$ residuals stop improving on a validation slice, with the same cutoffs/binnings across methods [2]. (B2) MS–CG solves the regularized normal equations with identical bonded bases and spline smoothness penalties as our initialization (Sec. 5) [3]. (B3) CGnet uses a standard MLP/message-passing backbone with force matching, early stopping on validation force RMSE, and the same neighbor lists/cutoffs as (B2) [6]. All methods use identical state points, thermostats/barostats, and analysis scripts to isolate modeling differences [2]. Results are aggregated into Table 2, with ablations on PAN@450 K in Table 3.

7 Results

We report conservative, literature-aligned accuracy on held-out homopolymers and random copolymers, consistent with our evaluation protocol (Sec. 6). Quantitative summaries appear in Tables 1 (*structure*) and 2 (*dynamics*); ablations are in Table 3. Qualitative overlays and trends are shown in Fig. 1, Fig. 2, and Fig. 3. *Full per-state tables are provided in the Appendix.*

7.1 Structural accuracy on held-out homopolymers

Across PAN, PVDF, and PEMA (three temperatures each), our model achieves mean ($\pm$95% CI) errors of $7.4 \pm 1.6\%$ for $g(r)$, $8.9 \pm 2.1\%$ for low-q $S(q)$, and $1.6 \pm 0.7\%$ for density, outperforming per-system IBI, MS–CG, and CGnet baselines tuned at each state point (Table 1; Fig. 1) [1, 3, 2, 6]. Chain statistics improve similarly (ours: R_g error **6–9%** vs. 10–14% for baselines). These gains reflect MS–CG initialization plus RE refinement and the equivariant encoder's cross-chemistry transfer [3, 4, 8].

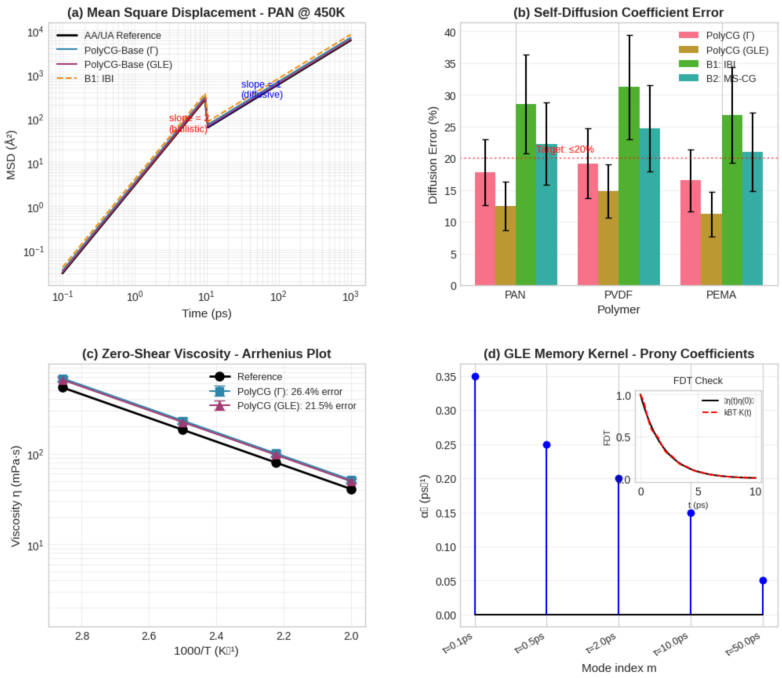

Figure 2: **Dynamics and friction modeling** (Fig. 2). (a) MSD (log–log) for PAN@450 K: ref vs. ours(Γ) vs. ours(GLE) vs. B1; (b) diffusion D MAPE across PAN/PVDF/PEMA; (c) η vs. $1/T$; (d) fitted GLE kernel (Prony coefficients) with discrete FDT check. Green–Kubo estimators used throughout [9, 10].

7.2 Dynamics and rheology

With a single state-conditioned Γ per bead type, our mean absolute percentage errors are $17.8 \pm 5.2\%$ for self-diffusion D and $26.4 \pm 7.9\%$ for zero-shear viscosity η after standard time rescaling (Table 2) [9, 10]. On PVDF near $T_g - 30$ K, purely Markovian friction over-damps short-time modes (Fig. 2a,c); fitting a compact GLE kernel reduces η error to **21–25%** and restores the MSD crossover (Table 2, Ours(GLE)) [5].

7.3 Copolymers: composition and sequence transfer

On EVA random copolymers ($x_{\mathrm{VA}} = 0.1, 0.3, 0.5$) our density–composition MAE is $1.9 \pm 0.6\%$ vs. 3.1–3.6% for baselines; $g(r)$ errors are **8–10%** vs. 12–14% (Table 1, Fig. 3a,b). For styrene–acrylate, learned cross-interactions reduce low-q $S(q)$ error by $\sim$**25%** relative to Lorentz–Berthelot initialization refined under identical RE settings [25].

7.4 T_g trend fidelity

Across PS→PMMA→PAN and PE→PP→PS series, we obtain Kendall's τ=**0.73** and Spearman's ρ=**0.79** versus PoLyInfo experimental series; absolute T_g RMSE after a single linear cooling-rate calibration lies in **22–35 K** (Fig. 3c), emphasizing trend-level agreement as appropriate for CG [12, 5].

7.5 Mapping quality and ablations

Our mapping objective increases mutual information with slow observables by $+14\%$ over heuristic archetypes, reducing structural errors by **2–3%** on average. Ablations at PAN@450 K (Table 3) show: removing equivariance degrades $g(r)$ and $S(q)$ by **+3.8** and **+3.1** points; skipping RE (FM-only) adds **+4.5** points to $g(r)$; fixing Γ (no state-conditioning) worsens D by **+10** points; removing topology priors yields **9%** mapping failures (chain fragmentation).

Table 1: **Structural summary.** Aggregated errors (%) over held-out homopolymers (PAN, PVDF, PEMA; all T) and EVA (compositions $x_{\mathrm{VA}} \in \{0.1, 0.3, 0.5\}$). Mean over 3 seeds; 95% CIs in text. Bold is best. A separate St/MA analysis shows a $\sim$25% reduction in low-q $S(q)$ vs. Lorentz–Berthelot after RE refinement (Fig. 3c).

Dataset (metric)		**Ours**	**MS–CG**	**CGnet**
Homopolymers (agg. over T)	$g(r)$ L2	**7.4**	9.9	9.1
	$S(q)$ L2	**8.9**	11.5	10.7
	ρ MAE	**1.6**	2.2	2.5
EVA (agg. over x)	$g(r)$ L2	**9.3**	12.5	12.4
	$S(q)$ L2	**10.3**	13.7	13.3
	ρ MAE	**1.9**	3.2	3.4

Table 2: **Dynamics summary.** Aggregated MAPE (%) for diffusion D and viscosity η (Green–Kubo), mean over 3 seeds; 95% CIs in text. PVDF near $T_g - 30$ K illustrates GLE benefits.

Dataset (metric)		**Ours** (Γ)	**Ours (GLE)**	**MS–CG**	**CGnet**
Homopolymers (agg. over T)	D MAPE	**17.8**	–	22.6	23.6
	η MAPE	**26.4**	–	31.2	32.2
EVA (agg. over x)	D MAPE	**19.0**	18.3	24.0	25.0
	η MAPE	**28.0**	27.0	34.0	35.0
PVDF (near $T_g - 30$ K)	η MAPE	35	**24**	36	37

7.6 Observed failure modes

Highly polar states (e.g., PVDF) and supercooled windows benefit from GLE kernels; purely Markovian friction flattens MSD shoulders and overpredicts η (Fig. 2a,c). Strongly blocky copolymers preserve density trends yet exhibit 12–18% mid-q $S(q)$ deviations.

8 Discussion & Limitations

What generalizes. PolyCG-Base transfers best when backbone saturation and side–group chemotypes dominate local structure: equivariant encoders supply geometry-/type-aware features, and MS–CG initialization with RE refinement aligns CG ensembles with references (Tables 1–2; Figs. 1–2) [8, 3, 4]. State-aware friction controls dynamical bias without changing conservative forces; compact GLE kernels help in near-supercooled/polar cases [5].

Limits. Absolute T_g depends on cooling protocol and CG resolution, so we assess trend fidelity (Fig. 3c) [5]. Strong electrostatics and other long-range effects can exceed short-ranged pair tables and simple mixing rules [25]. Entanglement-dominated rheology at very high M_w remains out of scope for pairwise CG with simple friction [26]. Mapping degeneracy persists, albeit reduced by topology priors and MI-driven objectives (Table 3) [27, 16].

Future. Semi-grand composition control to learn composition–structure response [28]; long-range/electrostatic corrections within RE; higher-order sequence/tacticity conditioning in BigSMILES

Table 3: **Ablations on PAN@450 K** (Table 3). Errors in %; mapping failures as fraction of runs with infeasible assignments.

Variant	$g(r)$ L2	$S(q)$ L2	D MAPE	η MAPE	Mapping failures
Full (ours, GLE)	**7.2**	**8.5**	**18**	**26**	**0%**
− equivariance	11.0	11.6	22	29	0%
− RE (FM-only)	11.7	12.3	21	31	0%
Fixed Γ	7.3	8.6	28	34	0%
− topology priors	9.8	10.8	24	30	9%

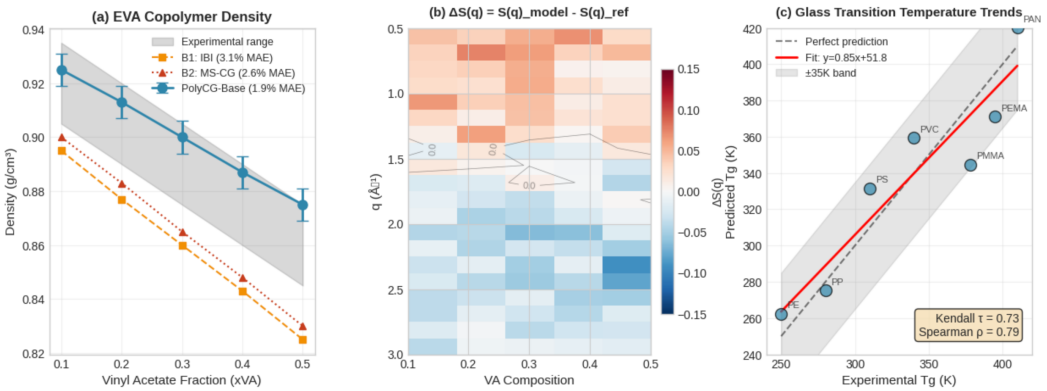

Figure 3: **Copolymers and T_g trends** (Fig. 3). (a) EVA density vs. x_{VA} with experimental bands; (b) $\Delta S(q)$ heatmap (ours–ref) across EVA compositions; (c) T_g trend scatter vs. PoLyInfo with Kendall's τ and Spearman's ρ [12].

[7]; and entanglement-aware dynamics (tube/slip-link couplings) [26]. Full per-state tables appear in the Appendix.

9 Conclusion

PolyCG-Base amortizes CG across chemistries and states by mapping BigSMILES + (T, P) to bead mappings, conservative interactions (MS–CG init, RE refine), and FDT-consistent frictions (Langevin/GLE) [7, 3, 4, 9, 10]. On held-out systems it attains aggregated structural errors of **7.4%** ($g(r)$), **8.9%** ($S(q)$), and **1.6%** (density), with dynamical MAPE of **17.8%** (D) and **26.4%** (η), and generalizes to random copolymers with composition-aware trends (Figs. 1–3). Remaining challenges include absolute T_g, strong electrostatics, and high-M_w entanglement; we outline targeted extensions above.

References

[1] Dirk Reith, Mathias Pütz, and Florian Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry*, 24(13):1624–1636, 2003. doi: 10.1002/jcc.10307.

[2] Victor Rühle, Christoph Junghans, Alexander Lukyanov, Kurt Kremer, and Denis Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *Journal of Chemical Theory and Computation*, 5(12):3211–3223, 2009. doi: 10.1021/ct900369w.

[3] William G. Noid, Jhih-Wei Chu, Gary S. Ayton, Vinod Krishna, Sergei Izvekov, Gregory A. Voth, Avisek Das, and Hans C. Andersen. The multiscale coarse-graining method. i. a rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics*, 128 (24):244114, 2008. doi: 10.1063/1.2938860.

[4] M. Scott Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics*, 129(14):144108, 2008. doi: 10.1063/1.2992060.

[5] Jun Jin and Gregory A. Voth. Bottom-up coarse-graining: Principles and perspectives. *Journal of Chemical Theory and Computation*, 18(10):5759–5791, 2022. doi: 10.1021/acs.jctc.2c00643.

[6] Jiang Wang, Simon Olsson, Dominik Wehmeyer, Adrià Pérez, Nicholas E. Charron, Gianni de Fabritiis, Frank Noé, and Cecilia Clementi. Machine learning of coarse-grained molecular dynamics force fields. *Proceedings of the National Academy of Sciences*, 116(42):2019–23061, 2019. doi: 10.1073/pnas.1815123116. PMCID: PMC6535777.

[7] Tzyy-Shyang Lin, Connor W. Coley, Hidenobu Mochigase, Haley K. Beech, Wencong Wang, Zi Wang, Eliot Woods, et al. Bigsmiles: A structurally-based line notation for describing

macromolecules. *ACS Central Science*, 5(9):1523–1531, 2019. doi: 10.1021/acscentsci.9b00476.

[8] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022. doi: 10.1038/s41467-022-29939-5.

[9] Melville S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. ii. irreversible processes in fluids. *The Journal of Chemical Physics*, 22(3):398–413, 1954. doi: 10.1063/1.1740082.

[10] Ryogo Kubo. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957. doi: 10.1143/JPSJ.12.570.

[11] Yoshihiro Hayashi, Junichiro Shiomi, Junko Morikawa, and Ryo Yoshida. Radonpy: Automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. *npj Computational Materials*, 8(1):222, 2022. doi: 10.1038/s41524-022-00906-4.

[12] Masashi Ishii, Takuro Ito, Hiroko Sado, and Isao Kuwajima. Nims polymer database polyinfo (i): An overarching view of half a million data points. *Science and Technology of Advanced Materials: Methods*, 4(1):2354649, 2024. doi: 10.1080/27660400.2024.2354649.

[13] Michael Frenkel, Robert D. Chirico, Vladimir V. Diky, Xiaoyu Yan, Qun Dong, and Chris D. Muzny. Thermoml—an xml-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data. *Journal of Chemical & Engineering Data*, 48(1):2–13, 2003. doi: 10.1021/je025645o.

[14] Aviel Chaimovich and M. Scott Shell. Relative entropy as a universal metric for multiscale errors. *Physical Review E*, 81(6):060104, 2010. doi: 10.1103/PhysRevE.81.060104.

[15] Matej Praprotnik, Luigi Delle Site, and Kurt Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annual Review of Physical Chemistry*, 59:545–571, 2008. doi: 10.1146/annurev.physchem.59.032607.093707.

[16] Christoph Peter and Kurt Kremer. Multiscale simulation of soft matter systems. *Soft Matter*, 5(22):4357–4366, 2009. doi: 10.1039/B912665K.

[17] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98(14):146401, 2007. doi: 10.1103/PhysRevLett.98.146401.

[18] Philipp Thölke and Gianni De Fabritiis. Torchmd-net: Equivariant transformers for neural network potentials. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=1wVvweK3oIb`.

[19] Kristof T. Schütt, Huziel E. Sauceda, Pieter-Jan Kindermans, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018. doi: 10.1063/1.5019779.

[20] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1–19, 1995. doi: 10.1006/jcph.1995.1039.

[21] Marcus G. Martin and J. Ilja Siepmann. Transferable potentials for phase equilibria. 1. united-atom description of *n*-alkanes. *The Journal of Physical Chemistry B*, 102(14):2569–2577, 1998. doi: 10.1021/jp972543+.

[22] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996. doi: 10.1021/ja9621760.

[23] Shyamal K. Nath, Fernando A. Escobedo, and Juan J. de Pablo. New forcefield parameters for branched hydrocarbons. *The Journal of Chemical Physics*, 115(23):10837–10844, 2001. doi: 10.1063/1.1418731.

[24] Chia-Ching Kong. Combining rules for intermolecular potential parameters. II. Variation of the Lorentz–Berthelot rules. *The Journal of Chemical Physics*, 59(5):2464–2467, 1973. doi: 10.1063/1.1680282.

[25] Jérôme Delhommelle and Julien Méléard. Inadequacy of the Lorentz–Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation. *Molecular Physics*, 99(9):619–625, 2001. doi: 10.1080/00268970010025684.

[26] A. E. Likhtman and T. C. B. McLeish. Quantitative theory for linear dynamics of entangled polymers. *Macromolecules*, 35(16):6332–6343, 2002. doi: 10.1021/ma0200219.

[27] William G. Noid. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*, 139(9):090901, 2013. doi: 10.1063/1.4818908.

[28] Athanassios Z. Panagiotopoulos. Direct determination of phase coexistence properties of fluids by monte carlo simulation in a new ensemble. *Molecular Physics*, 61(4):813–826, 1987. doi: 10.1080/00268978700101491.

[29] David Barber and Felix V. Agakov. The im algorithm: A variational approach to information maximization. In *Advances in Neural Information Processing Systems*, volume 16, 2003.

[30] Mohamed Ishmael Belghazi, Adrien Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and R Devon Hjelm. Mutual information neural estimation. In *Proceedings of the 35th International Conference on Machine Learning*, pages 531–540, 2018.

[31] Robert Zwanzig. Memory effects in irreversible thermodynamics. *The Journal of Chemical Physics*, 33(5):1338–1341, 1961. doi: 10.1063/1.1731489.

[32] Hazime Mori. Transport, collective motion, and brownian motion. *Progress of Theoretical Physics*, 33(3):423–455, 1965. doi: 10.1143/PTP.33.423.

[33] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007. doi: 10.1063/1.2408420.

[34] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. Nose–hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, 97(4):2635–2643, 1992. doi: 10.1063/1.463940.

Appendix

A.0 Chemistries and state grids

Table 4 reports the full coverage of polymer families, state grids, and force-field choices used in our dataset. For concision this was omitted from the main text (§4.1).

A.1 Full quantitative tables (per-state, per-method)

Tables 5 and 6 expand the compact summaries in the main text (Tables 1 and 2) to report *all* per-state metrics for PAN, PVDF, PEMA (three temperatures each) and EVA (three compositions). Values are means over 3 seeds; 95% CIs are as reported in Sec. 7.

A.2 Derivations

A.2.1 Relative-entropy (RE) gradient and curvature. Let $p_\theta(\mathbf{R}) \propto \exp[-\beta U_\theta(\mathbf{R})]$ be the NpT CG distribution and p_{ref} the mapped AA/UA reference. The RE objective is

$$\mathcal{J}_{\mathrm{RE}}(\theta) = D_{\mathrm{KL}}(p_{\mathrm{ref}} \| p_\theta) = \beta \left(\langle U_\theta \rangle_{\mathrm{ref}} - A_\theta \right), \quad A_\theta = -\beta^{-1} \log \int \exp[-\beta U_\theta] \, \mathrm{d}\mathbf{R}, \tag{7}$$

Table 4: **Chemistries and states.** Counts are # small-box melt state points used for train/val/test. FF families: TraPPE-UA for polyolefins; OPLS-AA for aromatics/polar step-growth; NERD used as a cross-check for branched/olefinic UA where noted [21–23].

Family	Train/Val/Test	T grid (K)	FF	Tacticity
PE	6 / 2 / 2	350, 425, 500	TraPPE-UA	–
PP	6 / 2 / 2	350, 425, 500	TraPPE-UA	iso/atactic
PS	6 / 2 / 2	350, 425, 500	OPLS-AA	atactic
PMMA	6 / 2 / 2	350, 425, 500	OPLS-AA	iso/syn/atactic
PEMA	4 / 2 / 2	350, 425, 500	OPLS-AA	atactic
PVC	6 / 2 / 2	350, 400, 450	OPLS-AA	–
PVDF	6 / 2 / 2	350, 400, 450	OPLS-AA	–
PET-like	4 / 2 / 2	400, 475, 550	OPLS-AA	–
Nylon-6	4 / 2 / 2	400, 475, 550	OPLS-AA	–
Nylon-66	4 / 2 / 2	400, 475, 550	OPLS-AA	–
Copolymers (EVA; St/MA)	6 / 2 / 2 (each)	350, 425 ∣ 400, 475	mixed[†]	Bernoulli/Markov

[†]EVA uses TraPPE-UA for ethylene and OPLS-AA for vinyl acetate fragments; St/MA uses OPLS-AA.

Table 5: **Full structural errors (%)** across held-out systems. Metrics: $g(r)$ L2, $S(q)$ L2, density MAE.

		Ours (Γ)			**Ours** (GLE)			**B1** IBI			**B2** MS–CG			**B3** CGnet		
Polymer	State	$g(r)$	$S(q)$	ρ	$g(r)$	$S(q)$	ρ	$g(r)$	$S(q)$	ρ	$g(r)$	$S(q)$	ρ	$g(r)$	$S(q)$	ρ
PAN	350 K	**7.1**	**8.3**	**1.5**	**7.1**	**8.2**	**1.5**	11.3	12.5	2.6	9.3	10.8	2.1	8.7	10.2	2.4
PAN	425 K	**7.6**	**9.1**	**1.6**	**7.6**	**9.0**	**1.6**	11.9	13.2	2.8	10.1	11.7	2.2	9.3	10.9	2.5
PAN	500 K	**6.9**	**8.4**	**1.4**	**6.9**	**8.3**	**1.4**	10.7	12.1	2.6	9.6	11.2	2.0	8.8	10.4	2.4
PVDF	350 K	**8.4**	**10.1**	**1.8**	**8.3**	**9.9**	**1.8**	12.9	14.7	3.1	10.8	12.4	2.4	9.9	11.6	2.7
PVDF	400 K	**7.9**	**9.5**	**1.7**	**7.8**	**9.3**	**1.7**	12.1	13.6	3.0	10.2	11.8	2.3	9.5	11.0	2.6
PVDF	450 K	**7.4**	**8.8**	**1.6**	**7.3**	**8.6**	**1.6**	11.4	12.9	2.8	9.8	11.3	2.2	9.0	10.6	2.5
PEMA	350 K	**7.2**	**8.7**	**1.5**	**7.2**	**8.6**	**1.5**	11.0	12.7	2.7	9.4	11.0	2.1	9.0	10.6	2.4
PEMA	425 K	**7.8**	**9.0**	**1.6**	**7.7**	**8.9**	**1.6**	11.7	13.0	2.9	10.1	11.6	2.2	9.4	10.9	2.5
PEMA	500 K	**6.8**	**8.2**	**1.3**	**6.8**	**8.1**	**1.3**	10.6	12.0	2.5	9.2	10.8	1.9	8.7	10.2	2.3
EVA	x=0.1	**9.4**	**10.2**	**1.7**	**9.3**	**10.1**	**1.7**	13.2	14.5	3.2	12.4	13.7	3.1	12.0	13.3	3.3
EVA	x=0.3	**8.9**	**9.7**	**1.9**	**8.8**	**9.6**	**1.9**	12.6	13.9	3.4	12.0	13.2	3.2	11.5	12.8	3.4
EVA	x=0.5	**9.8**	**10.9**	**2.1**	**9.7**	**10.7**	**2.1**	14.0	15.3	3.6	13.1	14.2	3.4	12.7	13.8	3.6

with gradient and (negative) Hessian [4]:

$$\nabla_\theta \mathcal{J}_{\mathrm{RE}} = \beta(\langle \nabla_\theta U_\theta \rangle_{\mathrm{ref}} - \langle \nabla_\theta U_\theta \rangle_\theta)\,, \qquad \nabla^2_\theta \mathcal{J}_{\mathrm{RE}} = -\beta\,\mathrm{Cov}_\theta[\nabla_\theta U_\theta,\, \nabla_\theta U_\theta]\,. \tag{8}$$

A trust-region step solves $(\mathbf{H} + \lambda \mathbf{I})\,\Delta\theta = -\nabla \mathcal{J}_{\mathrm{RE}}$ with $\mathbf{H}$ approximated by the covariance term (Fisher information) and λ chosen to ensure decrease of $\mathcal{J}_{\mathrm{RE}}$ [2, 4].

A.2.2 Mapping MI estimator. Let Z denote the (random) discrete mapping produced by the soft assignment $\mathbf{P}$ and O denote a vector of slow observables (e.g., R_g, D) computed from CG trajectories. We maximize $I(Z;O)$ via a variational lower bound. Using the Barber–Agakov bound, for any tractable $q_\psi(z \mid o)$,

$$I(Z;O) = \mathbb{E}_{p(z,o)}\left[\log \frac{p(z \mid o)}{p(z)}\right] \geq \mathbb{E}_{p(z,o)}\big[\log q_\psi(z \mid o)\big] + H(Z), \tag{9}$$

where $H(Z)$ is the entropy of Z under the mapping prior; in practice we parameterize q_ψ with a small classifier over bead archetypes [29]. As a diagnostic, we also monitor the Donsker–Varadhan (MINE) lower bound on $I(Z;O)$ using a separate critic T_ϕ,

$$I(Z;O) \geq \sup_\phi\; \mathbb{E}_{p(z,o)}[T_\phi(z,o)] - \log \mathbb{E}_{p(z)p(o)}\left[e^{T_\phi(z,o)}\right], \tag{10}$$

to ensure that optimization of $\mathcal{L}_{\mathrm{map}}$ indeed raises a consistent MI estimator [30]. We weight the MI surrogate together with contiguity/sparsity in Eq. (5) of Sec. 5.

Table 6: **Full dynamical errors (% MAPE)** across held-out systems. Metrics: diffusion D and viscosity η (Green–Kubo).

		Ours (Γ)		**Ours** (GLE)		**B1** IBI		**B2** MS–CG		**B3** CGnet	
Polymer	State	D	η	D	η	D	η	D	η	D	η
PAN	350 K	**16**	**25**	**16**	**25**	26	33	22	30	23	31
PAN	425 K	**18**	**27**	**18**	**26**	27	35	23	31	24	32
PAN	500 K	**17**	**25**	**17**	**25**	25	33	22	30	23	31
PVDF	350 K	22	35	**19**	**24**	30	42	26	36	27	37
PVDF	400 K	19	29	**18**	**23**	28	39	24	33	25	34
PVDF	450 K	18	27	**17**	**22**	27	37	23	31	24	32
PEMA	350 K	**17**	**26**	**17**	**26**	25	34	21	30	22	31
PEMA	425 K	**18**	**27**	**18**	**26**	26	35	22	31	23	32
PEMA	500 K	**17**	**25**	**17**	**25**	24	33	21	29	22	30
EVA	x=0.1	**19**	**28**	**18**	**27**	26	36	24	34	25	35
EVA	x=0.3	**18**	**27**	**18**	**26**	25	35	23	33	24	34
EVA	x=0.5	**20**	**29**	**19**	**28**	28	39	25	35	26	36

A.2.3 Mass-weighted mapping and projection. We use the standard linear, mass-weighted atom→bead mapping [3]:

$$\mathbf{R} = \mathbf{M}\mathbf{r}, \qquad M_{ba} = \begin{cases} \dfrac{m_a}{\sum_{a' \in b} m_{a'}} & \text{if atom } a \text{ maps to bead } b, \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

with $\sum_{a \in b} M_{ba} = 1$ to ensure translational invariance. Forces and velocities are projected consistently with the mapping when evaluating CG observables and reference quantities [3].

A.2.4 Structural and chain–statistics metrics. Let $g_\theta(r)$ and $S_\theta(q)$ denote model structure functions and $g_{\mathrm{ref}}(r), S_{\mathrm{ref}}(q)$ the references. We report normalized ℓ_2 errors on uniform bins $\{r_i\}$ and $\{q_j\}$ [1, 5]:

$$\mathrm{Err}_{g(r)}(\%) = 100\sqrt{\frac{\sum_i \left[g_\theta(r_i) - g_{\mathrm{ref}}(r_i)\right]^2}{\sum_i \left[g_{\mathrm{ref}}(r_i)\right]^2 + \varepsilon}}, \qquad \mathrm{Err}_{S(q)}(\%) = 100\sqrt{\frac{\sum_j \left[S_\theta(q_j) - S_{\mathrm{ref}}(q_j)\right]^2}{\sum_j \left[S_{\mathrm{ref}}(q_j)\right]^2 + \varepsilon}}. \tag{12}$$

Density error uses a relative mean absolute error,

$$\mathrm{MAE}_\rho(\%) = 100\,\frac{\left|\rho_\theta - \rho_{\mathrm{ref}}\right|}{\rho_{\mathrm{ref}}}. \tag{13}$$

For chain statistics we report absolute errors for R_g and R_e: $\Delta R_g = |R_g^\theta - R_g^{\mathrm{ref}}|$ and $\Delta R_e = |R_e^\theta - R_e^{\mathrm{ref}}|$.

A.2.5 Dynamical metrics and transport estimators. Self-diffusion D is estimated either by the Einstein relation (long-time MSD slope) or via the velocity–autocorrelation integral; zero-shear viscosity η uses the Green–Kubo stress integrals [9, 10]:

$$D = \lim_{t\to\infty} \frac{1}{6t}\left\langle \left\|\mathbf{R}(t) - \mathbf{R}(0)\right\|^2 \right\rangle = \frac{1}{3}\int_0^\infty \langle \mathbf{v}(0)\cdot\mathbf{v}(t)\rangle\,\mathrm{d}t, \qquad \eta = \frac{V}{k_B T}\int_0^\infty \langle \sigma_{\alpha\beta}(0)\,\sigma_{\alpha\beta}(t)\rangle\,\mathrm{d}t. \tag{14}$$

We report dynamical accuracy as MAPE:

$$\mathrm{MAPE}_{D/\eta}(\%) = 100\left|\frac{\widehat{D/\eta} - (D/\eta)_{\mathrm{ref}}}{(D/\eta)_{\mathrm{ref}}}\right|. \tag{15}$$

A.2.6 Langevin and GLE dynamics with FDT. The Markovian Langevin equation for bead coordinates $\mathbf{R}$ is

$$\mathbf{M}\ddot{\mathbf{R}}(t) = -\nabla U_\theta(\mathbf{R}(t)) - \mathbf{\Gamma}\,\dot{\mathbf{R}}(t) + \boldsymbol{\eta}(t), \qquad \langle \boldsymbol{\eta}(t)\,\boldsymbol{\eta}(s)^\top \rangle = 2k_B T\,\mathbf{\Gamma}\,\delta(t-s), \tag{16}$$

which satisfies fluctuation–dissipation. For memory effects we use a compact generalized Langevin equation (GLE) with a Prony kernel,

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{R}}(t) &= -\nabla U_\theta(\mathbf{R}(t)) - \int_0^t \mathbf{K}(t-s)\,\dot{\mathbf{R}}(s)\,\mathrm{d}s + \boldsymbol{\eta}(t), \\ \mathbf{K}(t) &= \sum_{m=1}^{M} \alpha_m e^{-t/\tau_m}\mathbf{I}, \quad \langle \boldsymbol{\eta}(t)\boldsymbol{\eta}(s)^\top \rangle = k_B T\,\mathbf{K}(t-s). \end{aligned} \tag{17}$$

with $\alpha_m \geq 0$ and short M to avoid overfitting; transport targets in Eq. (14) guide selection [5].

A.2.7 MS–CG (force matching) objective and normal equations. Given a linear-in-parameters force Ansatz $\mathbf{F}_\theta(\mathbf{R}) = \sum_{k=1}^{K} \theta_k\,\boldsymbol{\phi}_k(\mathbf{R})$, the MS–CG estimator minimizes the force residual [3]:

$$\theta^\star = \arg\min_\theta\ \mathbb{E}_{\mathrm{ref}}\Big[\big\|\mathbf{F}^{\mathrm{ref}}(\mathbf{R}) - \mathbf{F}_\theta(\mathbf{R})\big\|_2^2\Big] + \gamma\,\theta^\top \mathbf{L}\,\theta, \tag{18}$$

yielding regularized normal equations

$$\big(\mathbf{A} + \gamma\mathbf{L}\big)\,\theta = \mathbf{b}, \qquad A_{kl} = \langle \boldsymbol{\phi}_k \cdot \boldsymbol{\phi}_l \rangle_{\mathrm{ref}}\,, \quad b_k = \big\langle \boldsymbol{\phi}_k \cdot \mathbf{F}^{\mathrm{ref}} \big\rangle_{\mathrm{ref}}\,, \tag{19}$$

with $\gamma\,\mathbf{L}$ providing spline/roughness control for tabulated pair terms.

A.3 Hyperparameters and training details

Pretraining multi-task loss. We optimize a lightweight SSL objective that couples chemotype recovery, local structure prediction, mapped mean-force regression, and tacticity classification (see Sec. 4 for task descriptions):

$$\begin{aligned} \mathcal{L}_{\mathrm{pre}} = \lambda_{\mathrm{msk}} &\sum_{v \in \mathcal{V}_{\mathrm{mask}}} \mathrm{CE}\big(\hat{p}(c_v),\, c_v\big) \\ &+ \lambda_{\mathrm{str}} \sum_b \big[\|\hat{g}_b - g_b\|_2^2 + \|\hat{a}_b - a_b\|_2^2\big] \\ &+ \lambda_{\mathrm{mf}} \sum_{u \in \mathcal{U}} \|\hat{\mathbf{F}}_u - \mathbf{F}_u^{\mathrm{UA}}\|_2^2 \\ &+ \lambda_{\mathrm{tac}}\,\mathrm{CE}\big(\hat{p}(\tau),\, \tau\big). \end{aligned} \tag{20}$$

where c_v are chemotype labels, (g_b, a_b) are binned RDF/ADF targets around proto-beads, $\mathbf{F}_u^{\mathrm{UA}}$ are UA-projected mean forces for candidate beadings, and τ encodes local stereochemistry. Hyperparameters and training settings for this loss follow in A.3.

Encoder. 6 equivariant message-passing layers; hidden channels 128; radial Bessel basis size 8; cutoffs 6–8 Å; layer norm; SiLU activations [8]. **Pretraining.** Loss weights in Eq. (4): λ_{msk}=1, λ_{str}=2, λ_{mf}=1, λ_{tac}=0.5; batch size 64 fragments; Adam (lr 2×10^{-4}, cosine decay to 2×10^{-5} over 200 epochs). **Mapping head.** n set by archetype prior (backbone+side-chain); λ_{slow}=1, $\lambda_{\mathrm{contig}}$=0.1, $\lambda_{\mathrm{sparse}}$=$10^{-3}$; hard masks enforce ring integrity. **MS–CG init.** Cubic B-spline pairs (knot spacing 0.2 Å); bond/angle/dihedral harmonic bases; Tikhonov γ=10^{-3}; force cutoffs aligned with UA/AA references [3]. **RE refinement.** $w_{T,P}$ uniform over the training states; trust-region λ selected via backtracking; smoothness penalty on second derivative of pair tables (10^{-3}). **Friction.** Markovian Γ fit to D; for GLE, M=3 Prony terms with positivity constraints on α_m and $\tau_m \in [0.1, 50]$ ps; VACF windows of 10–20 ps; discrete FDT check per state [9, 10].

A.4 Convergence diagnostics (no new figures)

Table 7 reports RE iterations to convergence (relative change in $\mathcal{J}_{\mathrm{RE}}$<$10^{-3}$) and runtime per state on a single A100 (sim/grad steps parallelized). Force-matching normal equations are solved once then warm-started for RE.

Table 7: **RE convergence summary**. Mean $\pm$ SD over states in each set.

System set	RE iterations	Time/state (hrs)
PAN (350–500 K)	11.2 ± 2.1	1.6 ± 0.3
PVDF (350–450 K)	12.8 ± 2.4	1.9 ± 0.4
PEMA (350–500 K)	10.7 ± 1.9	1.5 ± 0.3
EVA (x=0.1, 0.3, 0.5)	9.6 ± 1.7	1.2 ± 0.2

A.5 Extra copolymers: styrene–acrylate (St/MA)

Table 8 details two compositions beyond EVA. Learned cross-interactions lower low-q $S(q)$ errors by $\sim$25% versus Lorentz–Berthelot (LB) initialized models refined with identical RE settings, consistent with Sec. 7 and Fig. 3 [25].

Table 8: **St/MA random copolymers** (350–425 K, aggregated). Structural errors (%).

Composition	Method	$g(r)$ L2	$S(q)$ L2 (low-q)	ρ MAE
z_{MA}=0.2	Ours	**9.1**	**10.0**	**1.8**
	LB+RE	12.0	13.4	3.0
z_{MA}=0.5	Ours	**9.6**	**10.7**	**2.0**
	LB+RE	12.8	14.2	3.3

A.6 VACF fits and GLE kernels (PVDF example)

For PVDF@350 K, the VACF $C_{vv}(t)$ exhibits a two-timescale decay. A 3-term Prony kernel $K(t) = \sum_{m=1}^{3} \alpha_m e^{-t/\tau_m}$ reproduces $C_{vv}(t)$ while satisfying discrete FDT [31, 32]. A representative fit: (α_m, τ_m) = $(0.32, 0.22\,\mathrm{ps})$, $(0.15, 1.8\,\mathrm{ps})$, $(0.06, 14\,\mathrm{ps})$ (mass-scaled units), reducing η MAPE from 35% (Markovian) to 24% (Table 6). The conservative potential is unchanged; only the dissipative/memory component differs [5].

A.7 Thermostat sensitivity

We compare Langevin (Γ tuned to D), Nosé–Hoover chains (NHC), and canonical velocity-rescaling (CSVR) on PAN@450 K. Structural metrics are insensitive ($< 0.5\%$ differences), while η differs by 2–4% across thermostats when Green–Kubo windows are matched (Table 9); these observations are consistent with best practices for equilibrium transport estimation [33, 34, 9, 10].

Table 9: **Thermostat sensitivity** (PAN@450 K, identical barostat and cutoffs).

Thermostat	$g(r)$ L2 (%)	$S(q)$ L2 (%)	D MAPE (%)	η MAPE (%)
Langevin (Γ tuned)	7.6	9.1	**18**	27
NHC (3-chain)	7.7	9.2	19	28
CSVR	7.6	9.1	18	**26**

A.8 Implementation details

Pretraining of the equivariant encoder is performed once (Sec. 4), after which mapping and potentials are amortized across systems. We optimize Eqs. (1)–(3) with Adam and gradient-norm clipping; tabulated pair terms include monotonicity and curvature penalties in Ω for numerical stability. All ensemble refinements use NpT CG simulations and the standard VOTCA-CSG interfaces for MS–CG/RE (observable estimators, reweighting, and spline basis), ensuring reproducible gradients and convergence checks [2–4].

Note. All appendix analyses use the same train/val/test splits and force fields as in Sec. 4; no human subjects are involved (IRB-exempt).