# Supporting Informations for "ADEPT–PolyGraphMT: Automated Molecular Simulation and Multi-Task Multi-Fidelity Machine Learning for Polymer Property Generation and Prediction"

Sobin Alosious[1,2], Yuhan Liu[3], Jiaxin Xu[2], Gang Liu[4], Renzheng Zhang[2], Meng Jiang[1,4], Tengfei Luo[1,2,5,6*]

[1]Lucy Family Institute for Data and Society, University of Notre Dame, Notre Dame, IN, 46556, USA.
[2]Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN, 46556, USA.
[3]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN, 46556, USA.
[4]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, 46556, USA.
[5]Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, 46556, USA.
[6]Center for Sustainable Energy at Notre Dame (ND Energy), University of Notre Dame, Notre Dame, IN, 46556, USA.

*Corresponding author(s). E-mail(s): tluo@nd.edu;

## S1 Additional Dataset and Multi-Fidelity Analysis

### S1.1 Property–Fidelity Availability and Dataset Overlap

To clarify the effective multi-fidelity coverage of the dataset, Table S1 summarizes the availability of experimental, MD, DFT, and group-contribution (GC) data for each polymer property. As discussed in the main text, the degree of multi-fidelity overlap

is strongly property-dependent. Substantial overlap between experimental and lower-fidelity computational data is primarily available for properties such as specific heat capacity ($C_p$), glass transition temperature ($T_g$), thermal conductivity ($\kappa$), and density ($\rho$), whereas several mechanical properties are predominantly MD-derived and many electronic or optical descriptors are primarily DFT-derived. Consequently, the effective contribution of fidelity-aware learning varies across property groups depending on the available overlap between fidelity levels.

The overlap analysis shown in Fig. S1 further highlights the heterogeneous nature of the multi-fidelity dataset. For $C_p$, $T_g$, and $\kappa$, the majority of polymers remain unique to either the experimental or lower-fidelity datasets, while only a relatively small subset contains both fidelity levels simultaneously. This indicates that multi-fidelity learning for these properties primarily relies on partially overlapping rather than fully aligned datasets. In contrast, density ($\rho$) exhibits substantially larger overlap between experimental and MD datasets, making it comparatively more suitable for direct cross-fidelity learning. Overall, these results demonstrate that the effective degree of multi-fidelity supervision varies significantly across properties and is strongly constrained by dataset overlap.

### S1.2 Additional Multi-Task Learning Results

Figure S2 summarizes the coefficient of determination ($R^2$) obtained for the different single-task and multi-task learning configurations across the polymer property space. Positive $R^2$ values indicate predictive performance exceeding the mean-prediction baseline, where all polymers are assigned the average property value from the training set. Consistent with the scaled MAE trends discussed in the main text, the results demonstrate that multi-task learning can improve predictive performance for several sparsely sampled properties through shared representation learning, although the degree of improvement remains strongly dependent on the selected task grouping and the intrinsic correlation structure between properties.

### S1.3 PolyGraphMT Architecture and Training Configuration

Table S2 summarizes the principal architectural hyperparameters and training configurations used in the PolyGraphMT framework. The workflow employs a shared graph neural network encoder with task-specific prediction heads for multi-task, multi-fidelity polymer property prediction. Hyperparameters, including the number of message-passing layers, graph embedding dimensions, prediction-head depth, normalization strategy, activation functions, dropout values, and learning rates, were optimized using Bayesian hyperparameter optimization implemented with Optuna. The framework additionally supports fidelity-aware conditioning through learned fidelity embeddings and optional feature-wise linear modulation (FiLM).

## S2 Representative Polymer Structures

Representative repeat-unit structures of selected polymers from the datasets used in this work are shown in Fig. S3. The examples were selected from representative high-

**Table S1** Property–fidelity availability matrix showing the distribution of experimental, MD, DFT, and GC data across the 28 polymer properties used in this work. Exact overlap counts for multi-fidelity properties will be updated in the final revised manuscript.

| Property | Exp. | MD | DFT | GC | Multi-fidelity overlap |
|---|---|---|---|---|---|
| **Thermal** | | | | | |
| $T_m$ | 3671 | – | – | – | Exp. only |
| $T_g$ | 7208 | 152 | – | – | Exp. + MD |
| $\alpha_{\mathrm{T}}$ | – | 799 | – | – | MD only |
| $\kappa$ | 68 | 2259 | – | – | Exp. + MD |
| $C_p$ | 120 | 851 | – | 12133 | Exp. + MD + GC |
| **Mechanical** | | | | | |
| $E$ | – | 1012 | – | – | MD only |
| $G$ | – | 1012 | – | – | MD only |
| $K$ | – | 1017 | – | – | MD only |
| $\nu$ | – | 1012 | – | – | MD only |
| **Transport** | | | | | |
| $\eta$ | – | 704 | – | – | MD only |
| $D$ | – | 700 | – | – | MD only |
| **Gas Permeability** | | | | | |
| $P_{\mathrm{He}}$ | 466 | – | – | – | Exp. only |
| $P_{\mathrm{H_2}}$ | 511 | – | – | – | Exp. only |
| $P_{\mathrm{CO_2}}$ | 756 | – | – | – | Exp. only |
| $P_{\mathrm{N_2}}$ | 798 | – | – | – | Exp. only |
| $P_{\mathrm{O_2}}$ | 807 | – | – | – | Exp. only |
| $P_{\mathrm{CH_4}}$ | 683 | – | – | – | Exp. only |
| **Electronic / Optical** | | | | | |
| $\alpha$ | – | – | 2036 | – | DFT only |
| $E_{\mathrm{HOMO}}$ | – | – | 2916 | – | DFT only |
| $E_{\mathrm{LUMO}}$ | – | – | 2916 | – | DFT only |
| $E_g$ | – | – | 2916 | – | DFT only |
| $\mu$ | – | – | 2916 | – | DFT only |
| $E_{\mathrm{total}}$ | – | – | 2916 | – | DFT only |
| $n$ | – | 744 | – | – | MD/DFT |
| $\varepsilon$ | – | 744 | – | – | MD/DFT |
| $\epsilon_r$ | – | 744 | – | – | MD/DFT |
| **Structural / Physical** | | | | | |
| $R_g$ | – | 2500 | – | – | MD only |
| $\rho$ | 1708 | 1935 | – | – | Exp. + MD |

and low-value regions across multiple experimental and computational datasets to provide chemical context for the property trends discussed in the manuscript.

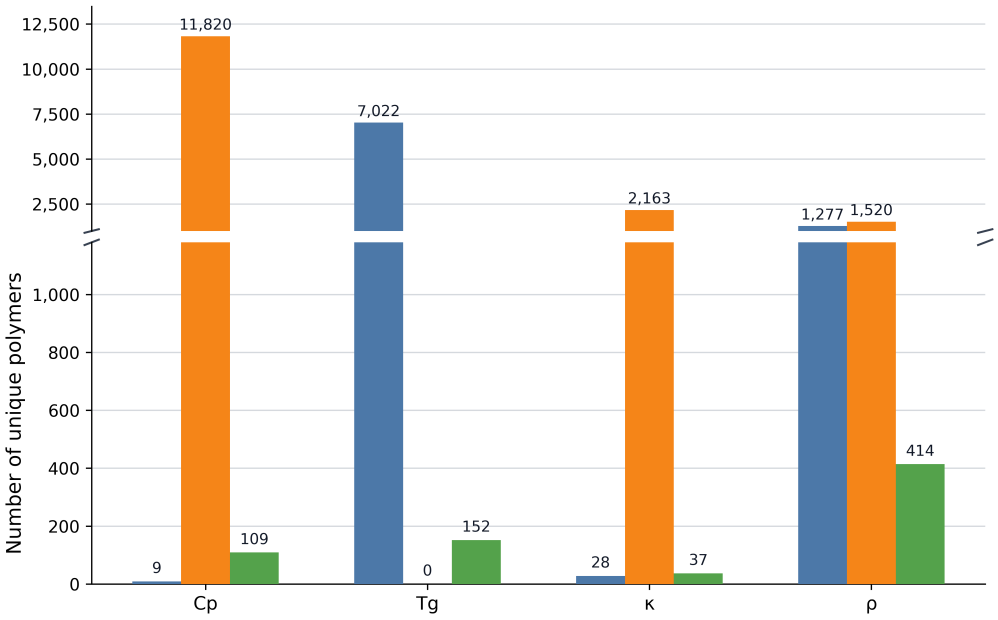


**Fig. S1** Overlap between experimental and lower-fidelity computational datasets for representative multi-fidelity polymer properties. For each property, unique polymer SMILES were extracted from the experimental dataset and the corresponding lower-fidelity dataset, canonicalized with RDKit, and partitioned into three groups: experimental only, lower-fidelity only, and overlap. Comparisons shown are $C_p$ (Experimental vs GC), $T_g$ (Experimental vs MD), thermal conductivity $\kappa$ (Experimental vs MD), and density $\rho$ (Experimental vs MD). A broken y-axis is used to visualize both the large lower-fidelity datasets and the smaller overlap counts in the same panel.

# S3 Post-Equilibration Structural Validation

## S3.1 Validation Strategy

Post-equilibration validation simulations were performed for a representative subset of polymers to assess density stability, volume fluctuations, and chain-statistics behavior. The simulations were initialized from the final equilibrated structures and continued under NPT conditions. Density and volume time series were analyzed to identify possible delayed compaction or void-collapse events, while chain-averaged radius of gyration and mean-squared internal distance scaling were used as structural stability checks. This analysis is intended as a targeted subset validation rather than a complete re-equilibration of the full high-throughput dataset.

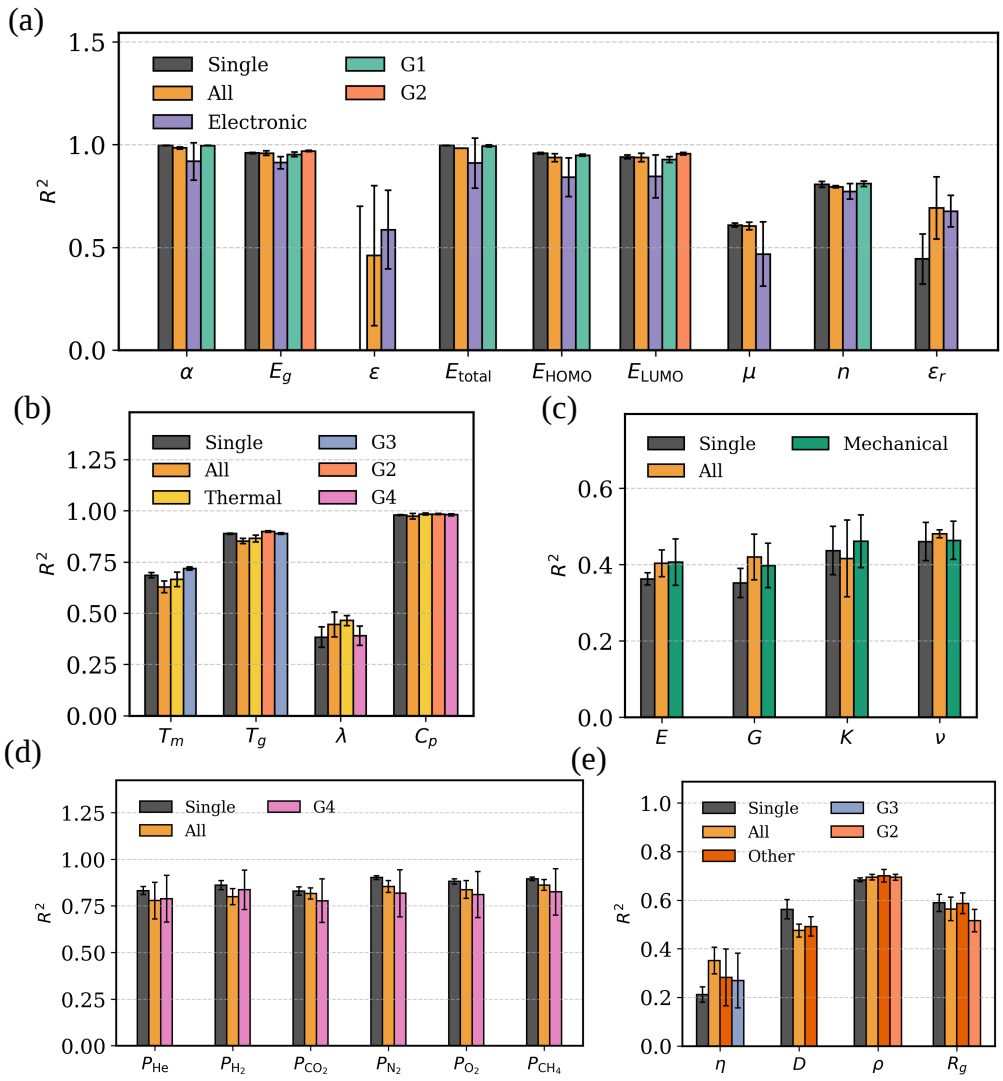


**Fig. S2** Comparison of $R^2$ for different polymer properties obtained from single-task and multi-task learning models. The multi-task models are trained using different sets of target properties, including all properties, property-class-based groups (thermal, mechanical, electronic, permeability, and others), and correlation-based groups (Groups 1–4). Positive $R^2$ values indicate performance exceeding the mean-prediction baseline.

## S3.2 Density and Volume Stability

Density and volume time series for the validation subset are shown in Fig. S4. Most polymers exhibit relatively stable late-time density and volume behavior without large systematic drift or abrupt volume-collapse events over the validation window.

**Table S2** Summary of key architectural hyperparameters used in the PolyGraphMT framework.

| Architecture component | Configuration |
|---|---|
| Node feature dimension | 35 |
| Edge feature dimension | 12 |
| Encoder backbones | GINE, GIN, GCN |
| Backbones explored in HPO | GINE, GIN |
| Message-passing layers | 3–7 |
| Graph embedding dimension | 256–768 |
| Pooling operation | Global mean pooling |
| Residual connections | Enabled |
| Normalization | BatchNorm / LayerNorm |
| Activation functions | ReLU / GELU / SiLU |
| Encoder dropout | 0.0–0.4 |
| Prediction head depth | 1–3 layers |
| Prediction head hidden dimension | 256–768 |
| Task-specific heads | Yes |
| Fidelity embedding | Learned embedding |
| Fidelity conditioning | Concatenation or FiLM |
| Optimizer | AdamW |
| Learning rate | $10^{-4}$–$3 \times 10^{-3}$ |
| Loss function | Masked balanced MSE / Gaussian NLL |

### S3.3 Chain-Statistics Stability Analysis

The corresponding radius-of-gyration trajectories shown in Fig. S5 indicate stable chain-size behavior after equilibration.

Figure S6 presents the mean-squared internal distance scaling obtained from the unwrapped trajectories. The resulting scaling behavior does not show obvious signatures of abrupt chain collapse or unstable late-time structural relaxation within the selected validation subset. Because atom-index separation is used as an approximate contour coordinate, the analysis is intended primarily as a qualitative structural stability check rather than a rigorous polymer-scaling characterization.

### S3.4 Quantitative Validation Metrics

Quantitative stability metrics for the validation subset are summarized in Table S3. Overall, the selected polymers exhibit relatively small density drift, modest late-time volume fluctuations, and stable chain-size behavior over the validation trajectories.

## S4 Additional Multi-Fidelity Benchmarking Results

Additional multi-fidelity benchmarking results for $\rho$, $T_g$, and TC are summarized in Table S4. The results further demonstrate that the effectiveness of multi-fidelity learning depends strongly on the relative balance between experimental-data availability, lower-fidelity dataset coverage, and the degree of correlation between fidelity levels. Properties with sparse experimental datasets and broad lower-fidelity coverage, such as $C_p$ and $\rho$, generally benefited more from heterogeneous fidelity integration, whereas

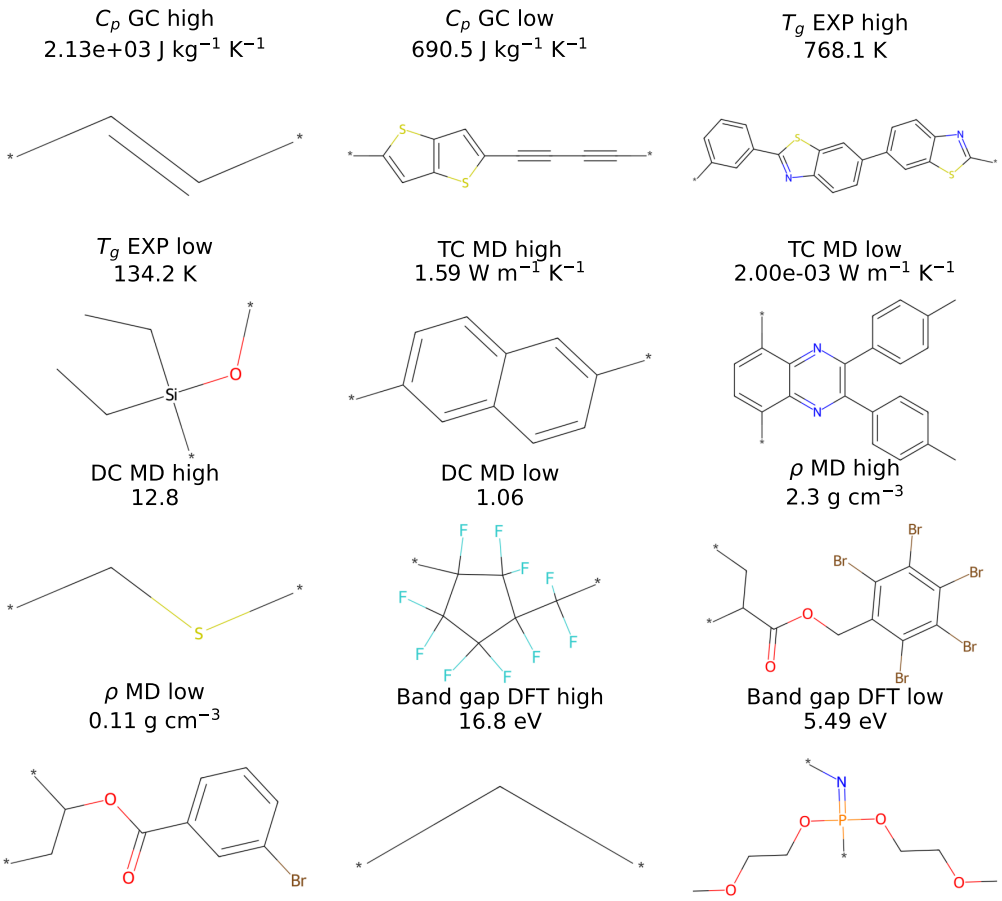


**Fig. S3** Representative repeat-unit structures of polymers selected from the training and validation datasets used in this work. Examples were chosen to span high- and low-value regions across multiple property/source datasets, including heat capacity ($C_p$), glass transition temperature ($T_g$), thermal conductivity (TC), dielectric constant (DC), density ($\rho$), and DFT band gap. Structures are shown using the repeat-unit SMILES representation available in the underlying datasets.

properties already dominated by large experimental datasets, such as $T_g$, showed comparatively limited sensitivity to the fidelity strategy. TC remained challenging across all training strategies because of the extremely limited experimental dataset and the resulting instability of the learned models.

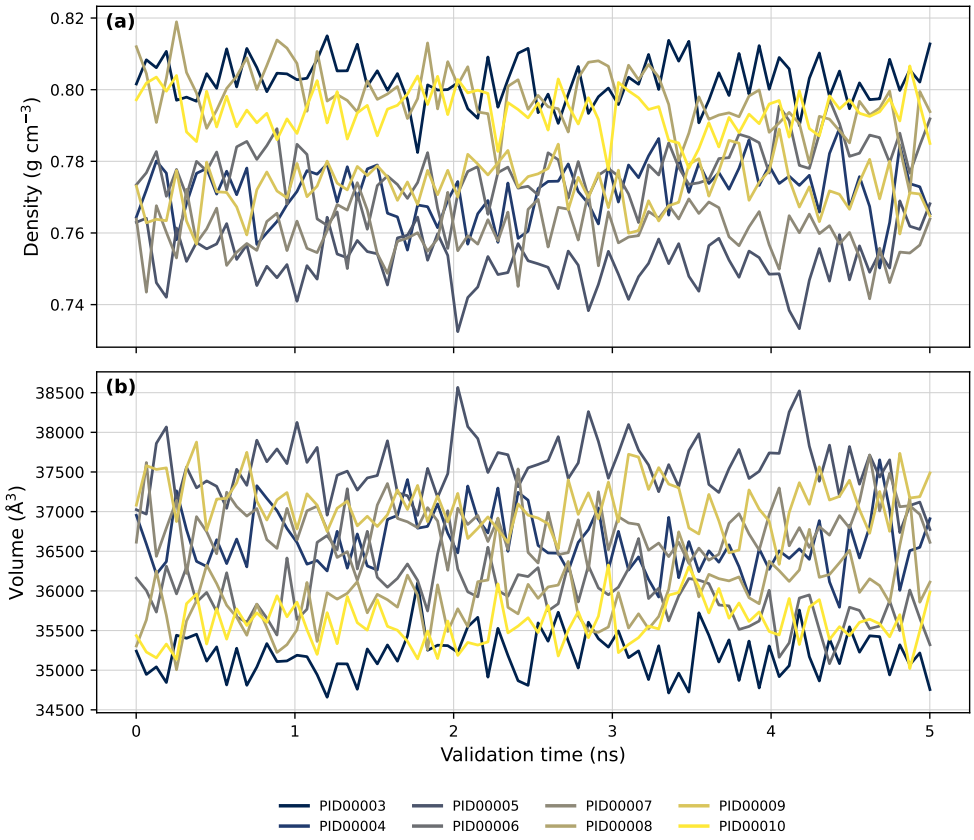


**Fig. S4** Post-equilibration density and volume stability for representative polymers. Time series were obtained from additional NPT validation simulations initialized from the final equilibrated structures.

**Table S3** Summary of post-equilibration validation metrics for the representative polymer subset. Density drift and radius-of-gyration drift were computed from the difference between the initial and final trajectory windows. Volume CV denotes the coefficient of variation of the simulation-cell volume fluctuations over the final half of the validation trajectory.

| PolymerID | Density drift (%) | Volume CV | Max volume jump (%) | $R_g$ drift (%) | Flags |
|---|---|---|---|---|---|
| PID00003 | 0.08 | 0.0076 | 3.82 | -0.15 | none |
| PID00004 | 0.55 | 0.0087 | 4.05 | 0.75 | none |
| PID00005 | -0.28 | 0.0108 | 4.47 | -0.21 | none |
| PID00006 | 1.10 | 0.0102 | 4.04 | -0.21 | none |
| PID00007 | -0.21 | 0.0102 | 4.30 | -0.36 | none |
| PID00008 | -1.59 | 0.0100 | 3.60 | -1.26 | none |
| PID00009 | -0.16 | 0.0092 | 4.33 | 0.22 | none |
| PID00010 | 0.01 | 0.0083 | 3.55 | 0.79 | none |

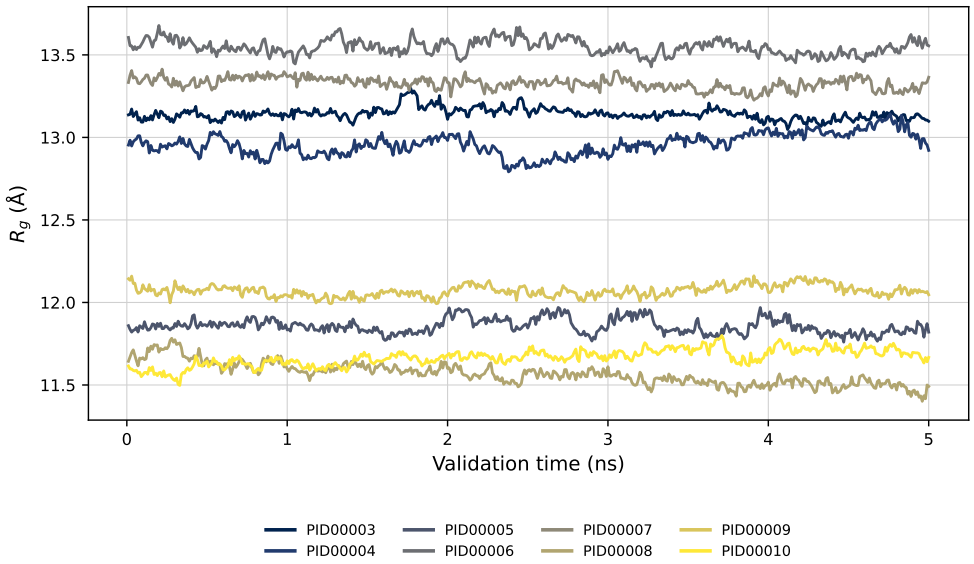


**Fig. S5** Radius-of-gyration stability for representative polymers during post-equilibration validation simulations.

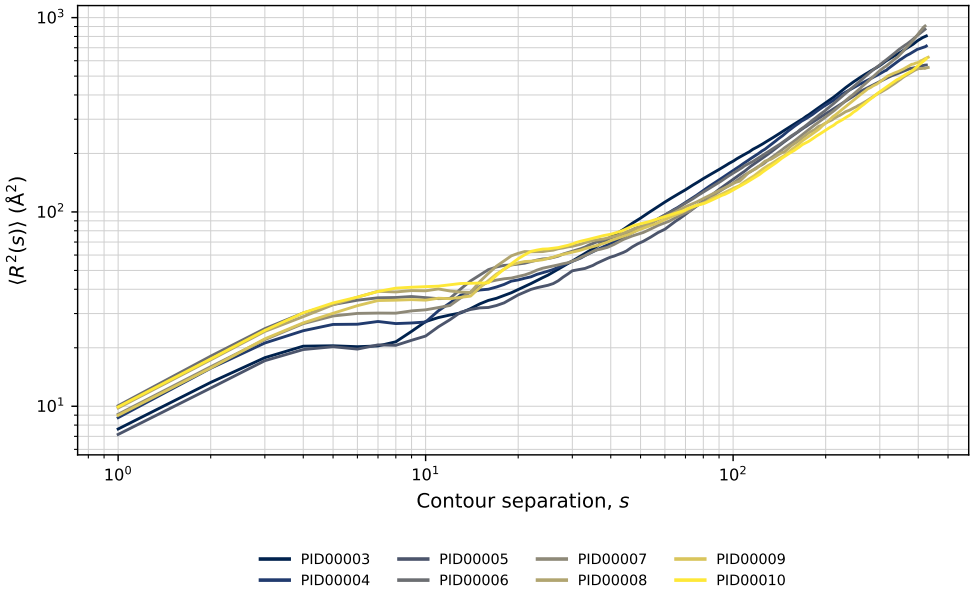


**Fig. S6** Mean-squared internal distance scaling for representative polymer chains computed from unwrapped trajectories. Atom-index separation is used as an approximate contour coordinate.

**Table S4** Summary of multi-fidelity learning performance across representative polymer properties. Reported values correspond to mean ± standard deviation across multiple random seeds evaluated on held-out experimental test sets.

| Property | Exp. | LF data | Exp-only $R^2$ | Best MF $R^2$ | Exp-only MAE | Best MF MAE |
|---|---|---|---|---|---|---|
| $C_p$ | 118 | 12076 (GC) | $0.48 \pm 0.31$ | $0.56 \pm 0.27$ | $186.39 \pm 58.15$ | $171.61 \pm 65.59$ |
| $\rho$ | 1691 | 1935 (MD) | $0.70 \pm 0.17$ | $0.78 \pm 0.06$ | $0.05 \pm 0.01$ | $0.05 \pm 0.01$ |
| $T_g$ | 7174 | 152 (MD) | $0.89 \pm 0.02$ | $0.89 \pm 0.02$ | $23.20 \pm 1.61$ | $23.77 \pm 1.43$ |
| TC | 65 | 2249 (MD) | $-1.54 \pm 2.42$ | $-1.12 \pm 1.28$ | $0.13 \pm 0.08$ | $0.12 \pm 0.06$ |