Augmenting Control over Exploration Space in Molecular Dynamics Simulators to Streamline De Novo Analysis through Generative Control Policies

Paloma Gonzalez-Rojas [1] Andrew Emmel [2] Luis Martinez [3] Neil Malur [3] Gregory Rutledge [4]

Abstract

This study introduces the P5 model - a foundational method that utilizes reinforcement learning (RL) to augment control, effectiveness, and scalability in molecular dynamics simulations (MD). Our novel strategy optimizes the sampling of target polymer chain conformations, marking an efficiency improvement of over 37.1%. The RL-induced control policies function as an inductive bias, modulating Brownian forces to steer the system towards the preferred state, thereby expanding the exploration of the configuration space beyond what traditional MD allows. This broadened exploration generates a more varied set of conformations and targets specific properties, a feature pivotal for progress in polymer development, drug discovery, and material design. Our technique offers significant advantages when investigating new systems with limited prior knowledge, opening up new methodologies for tackling complex simulation problems with generative techniques.

1. Introduction

Molecular dynamics has emerged as an invaluable tool for examining molecular systems, albeit with certain limitations. These methods often prove computationally intensive, particularly when handling complex systems such as macromolecules, for instance, polymer chains with more than 10^4 atoms. In addition, uncertainties may arise due to this increasing system complexity, inaccuracies in complex def-

[1]Department of Chemical Engineering, MIT Climate and Sustainability Consortium, Massachusetts Institute of Technology, Cambridge, United States [2]Department of Biological Engineering, Massachusetts Institute of Technology [3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology [4]Department of Chemical Engineering, Massachusetts Institute of Technology. Correspondence to: Paloma Gonzalez-Rojas <palomagr@mit.edu>.
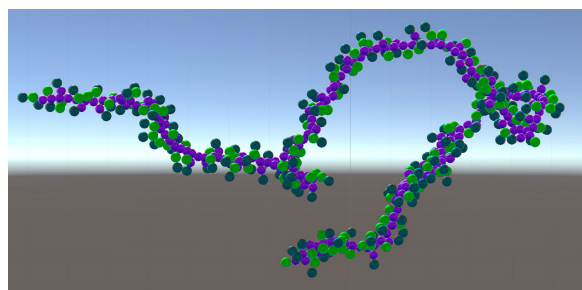
Figure 1. Simulated polymer chain of Cellulose Acetate with P5.

initions such as force fields, or limited knowledge about de Novo molecular systems. Machine learning has shown promise in addressing these challenges with data-driven methods like property predictions using discriminative techniques (Karuth et al., 2021; Tao et al., 2021), and machine learning force fields (Doerr et al., 2021; Kleinschmidt et al., 2022; Unke et al., 2021). Computational chemistry has seen machine learning techniques gradually incorporate 3D (Ganea et al., 2021) and graph representations (Antoniuk et al., 2022; Park et al., 2022; You et al., 2018; Zuo et al., 2021). Recent work has included physics analysis to inform generative molecular machine learning (Liu et al., 2023). Despite these advances, deriving molecular properties that demand expectations over an ensemble of states or time-dependent behavior of atoms and molecules remains relatively unexplored.

This paper presents P5, an RL-based control policy system in a three-dimensional physics engine simulator to enhance molecular simulation processes. P5 stands for Predicting Polymer Properties and Processability with Physics-Informed Reinforcement Learning. By utilizing reinforcement learning with a molecular dynamics module, including Brownian Dynamics, we extend the scope of application for control policies in molecular dynamics simulations. Our approach demonstrated increased control over uncertainty and addressed previous limitations in modeling and manipulating the polymer chain's structural dynamics.

Our work builds upon existing molecular dynamics sampling machine learning algorithms. While Holdijk et al. (2022) demonstrates how a physics-inspired approach for

transition sampling may provide direct control over molecular systems, P5's RL approach offers flexibility, adaptability, and environmental feedback. Klein et al. (2023) introduce 'TimeWarp,' which utilizes the Metropolis-Hastings algorithm for probabilistic sampling, while P5 actively targets specific states. Rose et al. (2021) deploy RL to sample rare events in their molecular dynamics research, and P5 seeks an array of targets. Our approach focuses on increasing control and efficiency over macromolecule dynamics. Furthermore, it builds upon RL extensive applications in robotics through training on 3D physics engines (Muratore et al., 2022; Brockman et al., 2016; Todorov et al., 2012; Juliani et al., 2020), which are widely used for model-based control learning. Reinforcement learning algorithms, notably A3C (Mnih et al., 2016) and PPO (Schulman et al., 2017), form the foundation of these platforms, driving control systems for kinetic simulators.

The P5 framework, as a physics-informed reinforcement learning model, contrasts with previous machine learning approaches like Generative Adversarial Networks (GAN), Variational Autoencoders (VAE), Adversarial Autoencoders (AAE), Long Short-Term Memory Networks (LSTM), Transformer networks, and the Generative Network Complex (GNC) in its methodology (Martinelli, 2022). While these models primarily focus on generating novel molecular structures based on statistical learning from large datasets, they don't inherently model the physical behaviors of molecular systems. The framework P5, conversely, utilizes a simulation-based approach to optimize molecular dynamics simulations, thereby modeling and predicting the behavior of known and novel molecular systems. In the context of de novo drug discovery, while the other models contribute by generating new candidate molecules, the current state of the P5 model can control and optimize molecular conformations, a vital aspect affecting a molecule's properties and activities, thus opening up new avenues for designing novel drug candidates.

While RL is not inherently generative, it can be used within a generative modeling framework to learn policies that guide data generation. In the case of P5, RL is employed to learn control policies that influence the forces and behaviors of polymer chains in molecular dynamics simulations. These learned policies allow P5 to generate trajectories and explore the conformational space of the polymer, or any molecule in general, in a goal-directed manner. This can be particularly useful in de novo molecule design, where the goal is to create novel molecules with desired properties.

The primary rationale for considering the P5 model as a promising tool for de novo molecular discovery is its proven adaptability to novel, unencountered states, described in the Results section. The reinforcement learning (RL) mechanism in P5, underpinned by a rich state space encapsulating geometric, dynamic, and environmental components, is designed to dynamically manipulate Brownian forces acting on the constituent beads of the polymer chain. This adjustment directly responds to alterations in the initial states or changes in the radii of gyration, enabling the model to traverse a vast conformational space not strictly constrained by its training corpus. By appropriately augmenting the reward function, observational parameters, and the action space, the model's capabilities can be extended to cater to the de novo design of molecular structures.

Through simulating a molecule of Cellulose Acetate (renewable biopolymer chain) with P5, we demonstrate how RL-based control policies offer improved exploration of the configuration space, adaptation to specific objectives, and optimization of computational resources in over a 37.1% improvement of polymer chain conformation sampling targeting a known radius of gyration range. This approach boosts the efficiency of the simulation process and fosters the development of RL-based control policies, which enhance the precision and efficiency of the simulations. Furthermore, P5 can be extended to various molecular systems, underscoring its transferability to de Novo molecules, applicability to cross-scale problems, and generating new conformations. This development with inexpensive computation and data makes P5 a promising tool for accelerating polymer interdisciplinary research and development. Furthermore, this simulation technology advancement can accelerate discoveries in various scientific and engineering domains.

2. Methods

2.1. Simulation Architecture

We present P5, a generative simulator developed to predict polymer properties and processing using physics-informed reinforcement learning to advance molecular dynamics simulations. This system was built on Unity 3D, a physics game engine that allows for the simulation of Newtonian physics of motion and 3D geometric representation of chemical structures. We incorporated the ML-agents module within the Unity environment, including standard reinforcement learning algorithms, Proximal Policy Optimization (PPO) (Schulman et al., 2017), and other similar techniques. We selected PPO as our reinforcement learning algorithm due to its flexibility and balanced exploration-exploitation trade-off. We 'froze' the components of the MD, 3D-physics, RL, and the molecular system, thus concentrating our efforts on designing control policies for molecular dynamics[1].

Our simulation framework incorporated the extensively-validated Martini Force Field (Bereau & Kremer, 2015) as shown in Figure 2 for representing organic molecules explained in Appendix 3. The geometric modeling of P5 is

---

[1]See Appendix 1, including error detection.

performed in Unity, with hinge joints employed to emulate bonds between polymer chain beads, facilitating precise rotational dynamics and an accurate depiction of the structural intricacies of the polymer chain. We utilized SMILE and Gromacs readers for monomer topology, enabling precise bead position retrieval in Cartesian space using RDKit (Landrum, 2010). The nonbonded interactions were modeled through Lennard-Jones and Brownian forces as an implicit solvent. Collectively, these methodologies contribute to a high-fidelity representation of polymeric molecular systems.
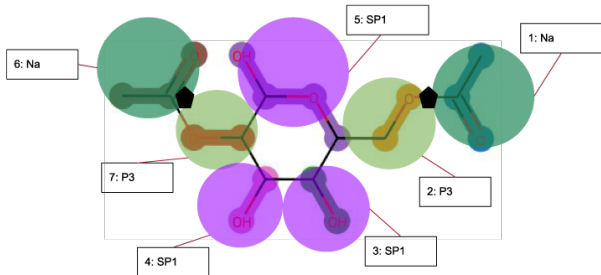

Figure 2. Martinization of Cellulose Acetate Molecule).

In the P5 model, the operational timescale is denoted as a dimensionless timestep, attributable to its implementation in the physics engine. This dimensionless timestep can be converted to physical time units such as femtoseconds using characteristic values of the relevant properties (e.g., bond lengths, bond angles, atomic masses) to derive the appropriate scaling factor. The selection of the exact scaling factor is contingent on the specific system under study and forms a crucial part of the model calibration process [2].

The development and training phase of the P5 model was conducted on a standard laptop, demonstrating its compatibility with commonly accessible computing resources. However, the model can be executed on cloud computing platforms for increased performance by 10x and scalability.

2.2. Control Policy Development

The control policy is optimized by a neural network (NN) agent that manipulates Brownian forces acting on the backbone beads of the polymer chain (PC). This process propels the beads to swiftly alter their positions, enabling a transition from an initial conformation to a target ensemble defined by a pre-specified radius of gyration (RG) range.

States S_t: The state space, denoted by S, combines several elements: the Geometry component, $S_{geometry}$, dihedral angles $dihedral_s$, and bond angles $bond_s$ are calculated based on the position vectors. The Dynamics component, $S_{dynamics}$, captures the position r_s, velocity v_s, and angular velocity w_s of the rigid body of the beads. The Environmen-

---

[2]See Appendix 2.

tal component, $S_{environment}$, reflects the positions of other neighboring beads' rigid bodies, encapsulated in the set $others_s$, within a specific observation radius around the current rigid body, to provide environmental feedback. The totality of the state space S is thus a combination of these three components: $S = S_{geometry} \cup S_{dynamics} \cup S_{environment}$.

In the state space S_t, there will be a series of states corresponding to target conformations with the desired radius of gyration. These states capture the specific geometric and dynamic properties of the polymer chain. The radius of gyration (RG) is calculated in equation (1):

$$\text{RG} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (r_i - r_{cm})^2} \qquad (1)$$

Where N is the number of beads in the chain, r_i represents the position of each bead, and r_{cm} is the position of the center of mass of the chain, quantifying the spatial extent of the polymer chain. The RG is an intrinsic characteristic of a polymer, a known property for widely studied polymers, and obtained from the literature for P5. Its implementation established a clear target state for developing a proof of concept. In further P5 development, the RG can be estimated from the simulation data to increase generalization. In addition, the objective function is flexible enough to accommodate a broad array of target objectives.

Actions A_t: Our system's neural network (NN) is the agent that controls the Brownian dynamics by adding a random "kick" or disturbance to the particles at each timestep, simulating the effect of random collisions with a fluid medium, and rotations to the backbone beads of each monomer in the chain, influencing the system's dynamics. The force applied to the bead is denoted as $\mathbf{F}_{learned}$ and is calculated as follows: $\mathbf{F}_{learned} = \begin{bmatrix} x_a & y_a & z_a \end{bmatrix} \cdot (\text{coef} \cdot \alpha_a)$. Here, x_a, y_a, and z_a represent the directional components of the force, while coef and α_a are factors to control force magnitude. The rotation angle action is denoted as $\theta_{rotation}$ and is calculated as $\theta_{rotation} = \text{angle}_a \cdot \theta_{max}$. Here, angle_a represents the rotation angle action provided by the NN agent, and θ_{max} denotes the maximum rotation angle defined as a parameter of the system. The rotation angle $\theta_{rotation}$ influences the dihedral angles of the beads and ultimately affects the overall conformation of the polymer chain.

Rewards R_t: The reward function consists of three components that are functions of the radius of gyration (RG) and its deviation from the target range $targetRG_{min}$ and $targetRG_{max}$. The distance reward $R_{dist\ reward}$ penalizes the agent in proportion to the squared distance between the current radius of gyration and the target range. The radius of gyration reward $R_{RG\ reward}$, this component incentivizes the agent to maintain the radius of gyration within the target range. The third reward is the shaping reward

$R_{\text{RG shaping reward}}$, which guides the agent's exploration and encourages stability. It rewards the agent for maintaining a radius of gyration near the middle of the target range and penalizes deviations.

With this reward structure, the RL agent explores conformations where the radius of gyration falls within the target range $targetRG_{min} = 0$ [nm], $targetRG_{max} = 20$ [nm]. Consequently, conformations within the target range were sampled more frequently, leading to a higher concentration of conformations $\rho_\pi(\tau)$ around the target radius of gyration, resulting in desired trajectories. The probability of the trajectory τ of the polymer chain can be described based on the expression:

$$\rho_\pi(\tau) = p(s_0) \prod_{t=0}^{T-1} \pi(f_{t\text{ learned}}|s_t)p(s_{t+1}|s_t, f_{t\text{ learned}}) \quad (2)$$

In expression (2), $p(s_0)$ represents the initial state s_0 probability, which captures the likelihood of the polymer chain starting in a particular conformation. This probability depends on the initial conditions and can influence the exploration of different regions of the conformational space. The term $\pi(f_{t\text{ learned}}|s_t)$ represents the policy π that selects the action $f_{t\text{ learned}}$ given the current state s_t. The policy π is responsible for shaping the trajectory of the polymer chain by guiding the selection of actions that maximize expected rewards.

Transition T_t: The transition probability T_t defined by $p(s_{t+1}|s_t, f_{t\text{ learned}})$ which captures the likelihood of transitioning from the current state s_t to the next state $st + 1$ given the reward r_t and the applied force $f_{t\text{ learned}}$. This probability reflects how the dynamics of the polymer chain, influenced by the applied forces and the rewards obtained, lead to transitions between different conformations. The probability $\rho_\pi(\tau)$ represents the observability of the trajectory τ of the polymer chain under the policy π. It captures the likelihood of traversing a sequence of states as determined by the policy π optimized by the P5 model.

3. Results

Our study focused on Cellulose Acetate, a renewable and biodegradable biopolymer. The cellulose acetate molecule typically has between 65 to 350 monomers and is represented by a coarse-grained model following the Martini Force Field with 75 monomers of 458 [g/mol]. The P5 model was trained for approximately 35 million episodes with 20,000 simulation steps each. The trained policy was then implemented in real-time simulation, enabling continuous adjustment to ensure the polymer chain remained within the target RG range of 0 to 20 nanometers throughout the runtime. During the training phase, the cumulative reward versus the environment optimized the policy quickly, as

shown in Figure 3. Intermittent dips in the cumulative reward were attributed to the potential energy excess induced by initially compressed conformations and high potential energy spikes. Despite these dips, the NN converged with sustained high values after six hours of training on a conventional laptop. The P5 model demonstrates the effectiveness of machine learning in efficiently controlling the conformational behavior of the Cellulose Acetate polymer.
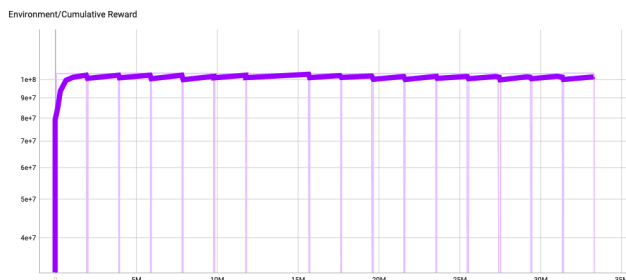

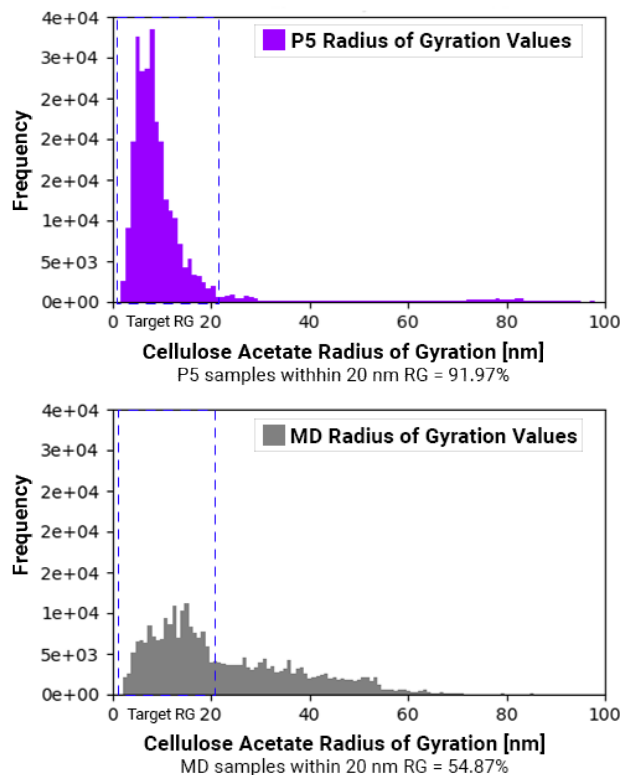*Figure 3.* Learning curve used to monitor training progress.

Figure 4. Comparing Conformations Control: P5 Model vs. Molecular Dynamics Simulations in Cellulose Acetate.

Our results demonstrate the effectiveness of the P5 model in efficiently controlling the conformational behavior of a simulated molecule of Cellulose Acetate. With optimized policies, P5 successfully guides the polymer chain to stay

within the target range of the RG, resulting in a 37.1% increase in sampling targeted states, as depicted in Figure 4. P5 effectively corrected conformation spikes due to high initial potential energy. Notably, P5 achieves these target conformations in half the wall clock time than MD.

The histogram trajectories shown in Figure 5, generated by P5, demonstrate the robustness and effectiveness of the model in correcting and guiding the polymer chain's conformation toward the targeted RG values. By manually changing the initial state and starting from different RG values, the policy network within P5 dynamically adjusts the forces acting on the beads to ensure the trajectory aligns with the desired RG range. This showcases the model's ability to adapt to unseen states and highlights its computational viability and transferability to other polymers and potentially de Novo molecules.
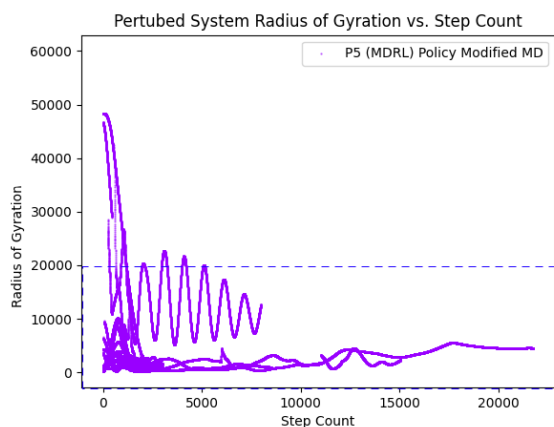

Figure 5. Trajectory Corrections of Polymer Chain Conformations by P5 Model from Unseen States (manually perturbed initial states explained in Appendix 4 and 5).

4. Contributions and Future work

In this work, we introduced the P5 model, a powerful tool that leverages machine learning control policies to optimize molecular dynamics simulations and expand their generative capabilities. The P5 simulator's unique integration of enhanced optimization, computational efficiency, and intuitive usability offers a novel approach for property evaluation, morphology analysis, and mechanical property prediction in complex molecular systems. Regarding efficiency and cost-effectiveness, P5 significantly outperforms traditional MD simulations by close to 40%, marking a pivotal step towards the computational design of molecular systems using generative machine learning techniques.

In acknowledging the current limitations of our empirical evaluation performed on cellulose acetate, we envision promising future directions. While the presented results of

the P5 model derive from a single molecule application, it is essential to clarify that these findings merely represent an initial demonstration of the methodology rather than an exhaustive exploration of its capabilities. The underlying framework of the P5 model is designed with flexibility and extensibility, capable of assimilating novel physics, diverse target systems, and various training regimes. This robust design positions P5 as a potent tool capable of performing de novo analysis and expanding its applicability to various molecules. As we progress, the crucial avenue for research lies in undertaking further empirical evaluations across a wider range of molecular structures. This research path will bolster our claims regarding the broad applicability of the P5 model, ushering in a new phase of enhanced performance in molecular dynamics simulations.

We plan to extend the P5 model to a broader range of polymer structures, benchmark against several other MD software, and expand the model to full-atom simulations, providing a more comprehensive understanding of the model's efficacy and scalability. Additionally, we aim to explore innovative control algorithms and integrate other physics machine learning models, analytical approaches, and time-coarse optimization to enhance the P5 model's performance further. These research paths are promising approaches that combine physics-based machine learning and molecular dynamics simulations.

References

Antoniuk, E. R., Li, P., Kailkhura, B., and Hiszpanski, A. M. Representing Polymers as Periodic Graphs with Learned Descriptors for Accurate Polymer Property Predictions, May 2022. URL http://arxiv.org/abs/2205.13757. arXiv:2205.13757 [cond-mat].

Bereau, T. and Kremer, K. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules. *Journal of Chemical Theory and Computation*, 11(6):2783–2791, June 2015. ISSN 1549-9618. doi: 10.1021/acs.jctc.5b00056. URL https://doi.org/10.1021/acs.jctc.5b00056. Publisher: American Chemical Society.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W. OpenAI Gym, June 2016. URL http://arxiv.org/abs/1606.01540. arXiv:1606.01540 [cs].

Doerr, S., Majewski, M., Pérez, A., Krämer, A., Clementi, C., Noe, F., Giorgino, T., and De Fabritiis, G. TorchMD: A Deep Learning Framework for Molecular Simulations. *Journal of Chemical Theory and Computation*, 17(4):2355–2363, April 2021. ISSN 1549-9618. doi: 10.1021/acs.jctc.0c01343. URL https://doi.org/

10.1021/acs.jctc.0c01343. Publisher: American Chemical Society.

Ganea, O.-E., Pattanaik, L., Coley, C. W., Barzilay, R., Jensen, K. F., Green, W. H., and Jaakkola, T. S. GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles, June 2021. URL http://arxiv.org/abs/2106.07802. Number: arXiv:2106.07802 arXiv:2106.07802 [physics].

Holdijk, L., Du, Y., Hooft, F., Jaini, P., Ensing, B., and Welling, M. Path Integral Stochastic Optimal Control for Sampling Transition Paths, June 2022. URL http://arxiv.org/abs/2207.02149. arXiv:2207.02149 [physics, q-bio].

Juliani, A., Berges, V.-P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., and Lange, D. Unity: A General Platform for Intelligent Agents, May 2020. URL http://arxiv.org/abs/1809.02627. arXiv:1809.02627 [cs, stat].

Karuth, A., Alesadi, A., Xia, W., and Rasulev, B. Predicting glass transition of amorphous polymers by application of cheminformatics and molecular dynamics simulations. *Polymer*, 218:123495, March 2021. ISSN 0032-3861. doi: 10.1016/j.polymer.2021.123495. URL https://www.sciencedirect.com/science/article/pii/S003238612100118X.

Klein, L., Foong, A. Y. K., Fjelde, T. E., Mlodozeniec, B., Brockschmidt, M., Nowozin, S., Noé, F., and Tomioka, R. Timewarp: Transferable Acceleration of Molecular Dynamics by Learning Time-Coarsened Dynamics, February 2023. URL http://arxiv.org/abs/2302.01170. arXiv:2302.01170 [cond-mat, physics:physics, stat].

Kleinschmidt, A. T., Chen, A. X., Pascal, T. A., and Lipomi, D. J. Computational Modeling of Molecular Mechanics for the Experimentally Inclined. *Chemistry of Materials*, 34(17):7620–7634, September 2022. ISSN 0897-4756. doi: 10.1021/acs.chemmater.2c00292. URL https://doi.org/10.1021/acs.chemmater.2c00292. Publisher: American Chemical Society.

Landrum, G. RDKit, 2010. URL https://www.rdkit.org/.

Liu, Z., Luo, D., Xu, Y., Jaakkola, T., and Tegmark, M. GenPhys: From Physical Processes to Generative Models, April 2023. URL http://arxiv.org/abs/2304.02637. arXiv:2304.02637 [physics, physics:quant-ph].

Martinelli, D. D. Generative machine learning for de novo drug discovery: A systematic review. *Computers in Biology and Medicine*, 145, June 2022. ISSN 00104825. doi:

10.1016/j.compbiomed.2022.105403. URL https://www.proquest.com/docview/2663091743/abstract/F3732178F6134807PQ/1. Place: Oxford, United Kingdom Publisher: Elsevier Limited.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., Silver, D., and Kavukcuoglu, K. Asynchronous Methods for Deep Reinforcement Learning, June 2016. URL http://arxiv.org/abs/1602.01783. arXiv:1602.01783 [cs].

Muratore, F., Ramos, F., Turk, G., Yu, W., Gienger, M., and Peters, J. Robot Learning From Randomized Simulations: A Review. *Frontiers in Robotics and AI*, 9, 2022. ISSN 2296-9144. URL https://www.frontiersin.org/articles/10.3389/frobt.2022.799893.

Park, J., Shim, Y., Lee, F., Rammohan, A., Goyal, S., Shim, M., Jeong, C., and Kim, D. S. Prediction and Interpretation of Polymer Properties Using the Graph Convolutional Network. *ACS Polymers Au*, January 2022. doi: 10.1021/acspolymersau.1c00050. URL https://doi.org/10.1021/acspolymersau.1c00050. Publisher: American Chemical Society.

Rose, D. C., Mair, J. F., and Garrahan, J. P. A reinforcement learning approach to rare trajectory sampling. *New Journal of Physics*, 23(1):013013, January 2021. ISSN 1367-2630. doi: 10.1088/1367-2630/abd7bd. URL https://dx.doi.org/10.1088/1367-2630/abd7bd. Publisher: IOP Publishing.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal Policy Optimization Algorithms, August 2017. URL http://arxiv.org/abs/1707.06347. arXiv:1707.06347 [cs].

Tao, L., Varshney, V., and Li, Y. Benchmarking Machine Learning Models for Polymer Informatics: An Example of Glass Transition Temperature. *Journal of Chemical Information and Modeling*, 61(11):5395–5413, November 2021. ISSN 1549-9596. doi: 10.1021/acs.jcim.1c01031. URL https://doi.org/10.1021/acs.jcim.1c01031. Publisher: American Chemical Society.

Todorov, E., Erez, T., and Tassa, Y. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, October 2012. doi: 10.1109/IROS.2012.6386109. ISSN: 2153-0866.

Unke, O. T., Chmiela, S., Sauceda, H. E., Gastegger, M., Poltavsky, I., Schütt, K. T., Tkatchenko, A., and Müller, K.-R. Machine Learning Force Fields. *Chemical Reviews*, 121(16):10142–10186, August

2021. ISSN 0009-2665. doi: 10.1021/acs.chemrev. 0c01111. URL https://doi.org/10.1021/acs. chemrev.0c01111. Publisher: American Chemical Society.

You, J., Liu, B., Ying, Z., Pande, V., and Leskovec, J. Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings. neurips.cc/paper/2018/hash/ d60678e8f2ba9c540798ebbde31177e8-Abstract. html.

Zuo, Y., Qin, M., Chen, C., Ye, W., Li, X., Luo, J., and Ong, S. P. Accelerating Materials Discovery with Bayesian Optimization and Graph Deep Learning, April 2021. URL http://arxiv.org/abs/2104. 10242. arXiv:2104.10242 [cond-mat].

A. Appendix

A.1. Simulation Architecture

The development of the P5 model was a complex endeavor that required integrating a molecular dynamics (MD) module within Unity3D, a robust physics engine, and the inclusion of reinforcement learning (RL). This project's complexity was heightened due to the need for accurate representation and simulation of atomic and molecular interactions, including Brownian random kicks that eventually evolved into a learned force within the system. Unity3D facilitated the intricate modeling and real-time visual inspection of simulations, which is critical for verifying model accuracy. Meanwhile, the integration of RL demanded the creation of a high-performing reward function and a balance between exploration and exploitation trade-offs. Amid these challenges, we opted for Proximal Policy Optimization (PPO) for its robustness, computational efficiency, and relative simplicity, aiding more straightforward integration with the MD module. Furthermore, despite the lack of extensive precedents, our pioneering effort in developing control policies for molecular dynamics contributed to establishing a robust system, setting the stage for future refinement and optimization.

Quantifying the effects and efficiency of the reinforcement learning-driven Brownian "kicks" in P5 molecular dynamics (MD) simulations is crucial. Unity 3D visualization plays a significant role in this process by providing an additional layer of validation, contributing to the robustness of the simulations. The raw data generated from MD simulations relevant for P5 include particle coordinates, velocity, and acceleration data, energy data (kinetic and potential energies), bond lengths, angles, dihedrals, and diffusion coefficients. This information helps track particles' motion and energy changes influenced by the Brownian "kicks" and identify structural changes or bond breakages that could result from these rapid perturbations.

Despite the comprehensiveness of this raw data, it might overlook certain rapid phenomena, such as bond breakages induced by Brownian "kicks." These events occur so swiftly that their detection might be challenging when solely relying on raw data analysis. Unity 3D visualization becomes invaluable in this error detection, providing an interactive, visually rich platform for identifying and rectifying inconsistencies or errors. By making potential issues visible that the raw data may fail to detect, Unity 3D visualization ensures the simulations' accuracy and validity, proving indispensable in the analytical toolkit.

A.2. Timescale Factor Calculation

To convert the dimensionless timescale to physical units, specifically milliseconds, we use the fundamental relationship $\tau = \frac{r}{v}$, where τ is the characteristic time, r is the characteristic length, and v is the characteristic velocity.

Characteristic Length (r): In the P5 model, the units of length are in Angstroms (Å), which equals 10^{-10} m.

Characteristic Mass (m): The model uses mass units based on the Van der Waals mass. For a polymer composed of monomers with seven beads, each with a total molecular weight of 458 g/mol, the characteristic mass of each bead (m_{bead}) is calculated by dividing this molecular weight by 7 (to account for the seven beads), and by Avogadro's number $N_A = 6.022 \times 10^{23}$ to convert to atomic mass units.

Characteristic Velocity (v): The characteristic velocity comes from the kinetic theory of gases and the equipartition theorem, which states that the mean kinetic energy of a molecule due to its motion through space is given by $\frac{1}{2}mv^2 = \frac{3}{2}kT$, where m is mass, v is velocity, k is Boltzmann's constant, and T is temperature. Here, we can calculate the characteristic velocity v_{char} when T is set to a typical room temperature (say, 298 K) and m is set to m_{bead}.

Characteristic Time (τ): The characteristic time can then be found by substituting r and v_{char} into the equation $\tau = \frac{r}{v_{\text{char}}}$.

The timestep factor to convert from dimensionless to femtoseconds is 209.7915273799608.

A.3. Martnini Model and Hyperparameters

This model results in seven types of beads, each categorized as "Na" (non-polar, hydrogen-acceptor), "P3" (polar, polarity level 3), or "SP1" (small, in ring, polar, polarity level 1). These beads carry assigned masses according to the Martini Model. Van der Waals' diameters are 5.2 Angstroms for regular beads and 4.7 Angstroms for smaller beads in a ring (designated with an 'S' prefix). Additionally, we incorporate an implicit solvent model in the form of Brownian dynamics to provide a more representative simulation environment, accounting for the effects of the solvent on the behavior of our molecular system.
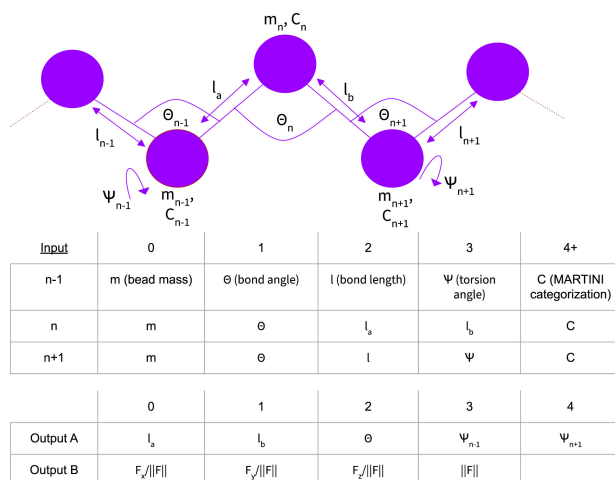
Input	0	1	2	3	4+
n-1	m (bead mass)	Θ (bond angle)	l (bond length)	Ψ (torsion angle)	C (MARTINI categorization)
n	m	Θ	l_a	l_b	C
n+1	m	Θ	l	Ψ	C

	0	1	2	3	4
Output A	l_a	l_b	Θ	Ψ_{n-1}	Ψ_{n+1}
Output B	$F_x/\|F\|$	$F_y/\|F\|$	$F_z/\|F\|$	$\|F\|$	

Figure 6. Parameters of the polymer chain agent.

A.4. Generative P5

We conducted a series of experiments to demonstrate the generative capabilities of the P5 model. We manually perturbed the initial state of 10 episodes, each consisting of 20,000 steps, while keeping the P5-MD system in standby mode. By changing the positions of the beads in the initial state, we introduced variations in the starting configurations. Remarkably, the P5 model effectively corrected the trajectory in each episode, guiding the polymer chain toward the desired gyration (RG) range radius. This showcases the model's ability to generate new conformations within a target space by leveraging its learned control policies. These generated conformations serve as valuable data points that can be further analyzed to gain insights into specific functional groups' structural properties, energy landscape, and behavior. Moreover, these conformations can be utilized in various downstream applications, such as predicting molecular properties, exploring structure-activity relationships, or conducting virtual screening for drug discovery. The P5 model's generative capabilities open up exciting opportunities for researchers to expand their understanding of molecular behavior, explore previously uncharted regions of chemical space, and drive innovation in diverse scientific domains.

Further, P5 autonomously generates data through modified molecular dynamics simulations, obviating the reliance on substantial experimental chemical datasets typically required in similar approaches. P5 finds conformations that may not be accessible to pure molecular dynamics.

A.5. compares gyration radius (RG) trajectories from P5 and traditional molecular dynamics (MD) simulations that provide critical insights into the conformational latent space. RG trajectories from pure MD showcase the molecule's inherent behavior. In contrast, P5 trajectories, driven by reinforcement learning-guided Brownian "kicks," demonstrate an enhanced ability to explore broader or alternative regions within this latent space. This comparison reveals the influence and effectiveness of the learned "kicks" in navigating the latent space, serving as a pivotal tool in assessing P5's generative capabilities and directing future model improvements.
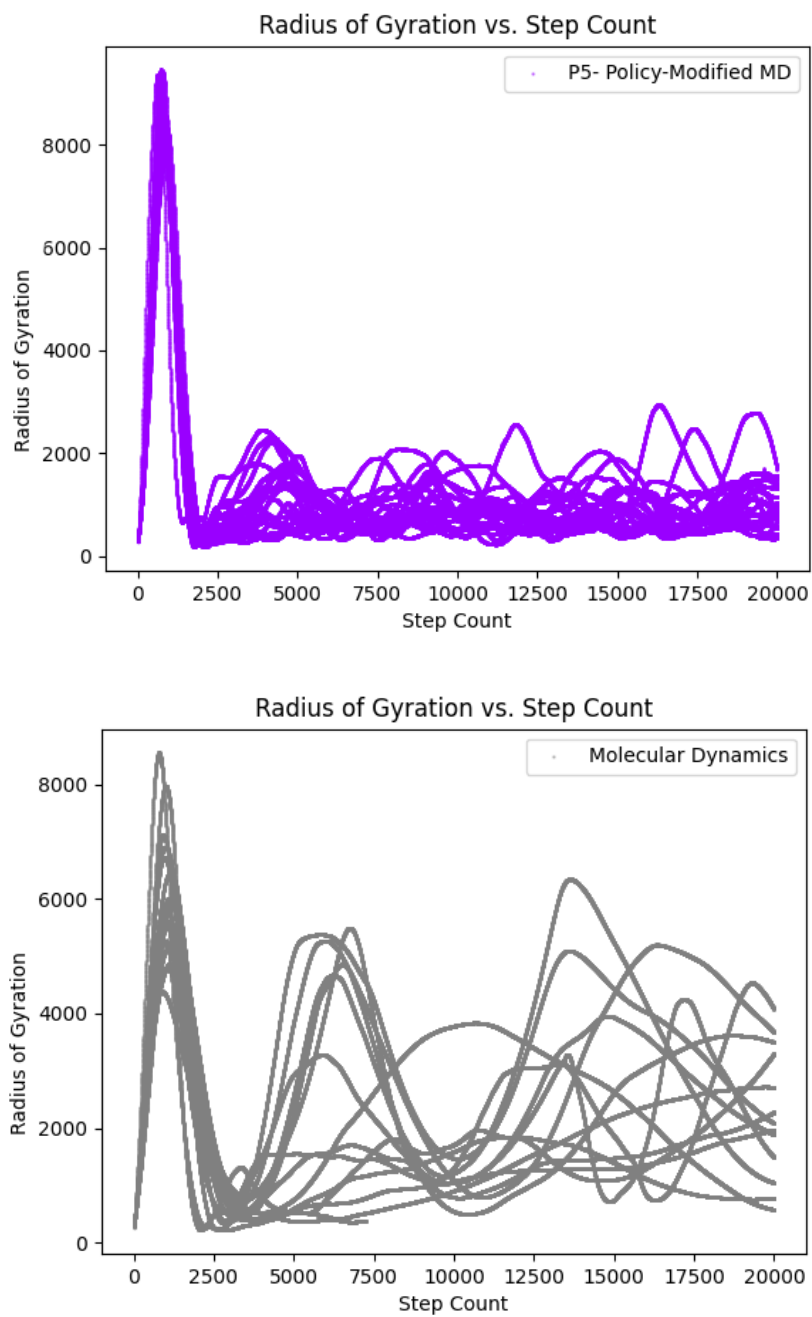
A.5. Trajectories Comparison

Figure 7. P5 & MD trajectories for the Radius of Gyration Values.