FlowBack-Adjoint: Energy-Guided Conditional Flow-Matching for Protein Side-Chain Generation

Alex Berlaga
Department of Chemistry
University of Chicago, Chicago, IL 60637, USA
berlaga@uchicago.edu

Michael S. Jones
Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
jones313@llnl.gov

Andrew L. Ferguson
Pritzker School of Molecular Engineering and Department of Chemistry
University of Chicago, Chicago, IL 60637, USA
andrewferguson@uchicago.edu

Abstract

Coarse-grained (CG) molecular models of proteins expand the time and length scales accessible to molecular dynamics simulations, but many scientific applications require recovering accurate all-atom (AA) detail. Recent work introduced FLOWBACK, a deep generative model that reconstructs AA ensembles from protein backbone traces using a flow-matching architecture, achieving state-of-the-art structural fidelity. However, because FLOWBACK is trained only on structural data, it can occasionally generate physically unrealistic or high-energy configurations. We present FLOWBACK-ADJOINT, a lightweight physics-aware enhancement that upgrades a pre-trained FLOWBACK model through a one-time post-training pass. Using adjoint matching, the method steers the generative vector field toward lower-energy regions of configuration space while preserving structural diversity. In benchmark tests against FLOWBACK, FLOWBACK-ADJOINT lowers single-point energies by a median of $\sim$78 kcal/mol.residue, reduces errors in bond lengths by $>$92%, eliminates $>$98% of molecular clashes, and maintains excellent diversity of the AA configurational ensemble. Our results demonstrate that FLOWBACK-ADJOINT offers a practical route to integrating physical realism into protein backmapping and deep generative protein models.

1 Introduction

Recent advances in generative deep learning have made accurate prediction of single native protein backbones routine, but many downstream tasks depend on the *ensemble* of conformations—including variations in side-chain configurations—rather than a single static structure [1, 2, 3, 4, 5, 6]. Molecular dynamics simulations initialized with machine-learned protein backbones can, in principle, recover Boltzmann ensembles, but sufficiently exploring protein dynamics typically requires simulating millisecond-scale trajectories or longer, which is prohibitively expensive for solvated proteins. [7, 8, 9, 10, 11]. Accordingly, several generative methods have recently emerged to efficiently emulate protein ensemble sampling [12, 13, 14, 15]. These strategies can generate realistic backbone ensembles, but often still yield only one most-probable side-chain configuration per backbone. Physics-based and generative pipelines have been developed to produce plausible coordinates,

[16, 17, 18, 19, 20, 21] but these methods typically perform poorly in producing low-energy structure ensembles relative to molecular dynamics trajectories. This limitation hinders downstream tasks such as protein or ligand docking [7, 9], which often require a diverse set of low-energy all-atom structures covering the thermally-relevant configurational space for accurate predictions. Taken together, there have been substantial advancements in protein structure generation, but there remains a need to generate low-energy ensembles of side chain configurations.

In this work, we present a generative modeling tool to fill this gap by transforming backbone-only structures into all-atom ensembles with diverse, physically-plausible, low-energy conformations. Our approach is based on a pre-trained flow-matching model, FLOWBACK, originally developed for coarse-grained backmapping [22]. Coarse-grained (CG) models replace groups of atoms with unified particle "beads", enabling simulations to be conducted at a fraction of the cost of all-atom MD [23, 24, 25, 26]. Backmapping can be conceived as the inverse of coarse-graining, *viz.* the reintroduction of all-atom (AA) detail to a CG scaffold by positioning the remaining atoms so that the resulting structure is physically consistent with the underlying coarse geometry [27, 28].

A number of data-driven protein backmapping approaches have been developed in recent years [22, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42]. FLOWBACK stands out as a deep generative approach based on a conditional flow-matching objective [43] that exhibits state-of-the-art performance in the efficient generation of physically-plausible AA structural ensembles [22]. It is a generic model that works for any protein, scales well to extremely large proteins of lengths far outside the training data, and is available as an open-source Python package [22]. FLOWBACK delivers diverse ensembles of physically-reasonable AA protein structures by placing the backbone and side chain atoms in every residue on top of their parent C_α and learning a flow to correctly guide each atom to a distribution of plausible residue conformations; however, the model possesses no awareness of bonded and non-bonded energetics. Thus, the model does not effectively reproduce the Boltzmann distribution, has no awareness of how to relax conformations in configurational space towards low-energy states, and can produce configurations lying tens of kcal/mol.residue above the local minimum in the potential energy landscape, which is insufficient for the purposes of downstream docking or interaction analysis tasks [7, 9].

To this end, we present FLOWBACK-ADJOINT (Figure 1A), which augments FLOWBACK with an adjoint objective that aligns the learned drift toward low-energy regions from a molecular-mechanics potential. These components are trained end-to-end in the flow-matching framework [43] and can be viewed as an energy-guided variant of adjoint matching [44, 45]. We incorporate information on energy and forces via an inductive bias wherein at the final configuration, $t = 1$, we evaluate a force field energy on the AA coordinates $U(x)$ and back-propagate its gradient $\nabla_x U$ through the time-reversed system. The resulting parameter updates, computed with standard automatic-differentiation tools, nudge the flow toward trajectories that terminate in lower-energy conformations without compromising the structural backmapping and diversity of AA configurations learned by FLOWBACK or risking instabilities or catastrophic forgetting. Empirically, FLOWBACK-ADJOINT improves physicochemical fidelity while preserving diversity: on an out-of-distribution protein benchmark [15], it reduces heavy-atom clash energies by a median of 78 kcal/mol.residue relative to FLOWBACK, and matches or exceeds strong generative baselines such as FLOWBACK and HPACKER on bond and clash error avoidance [15, 20, 22]. Crucially, FLOWBACK-ADJOINT requires as its only input a C_α trace, permitting compatibility with CG protein simulations using a coarse-grained force field [25, 46, 47], incomplete experimental measurements, or structural predictions from computational tools such as ALPHAFLOW [12], BBFLOW [13], or BIOEMU [15] (Figure 1B). We illustrate that FLOWBACK-ADJOINT brings machine-learned biomolecular ensembles into close agreement with Boltzmann-consistent references from molecular dynamics and also yields the most reasonable structures on out-of-distribution proteins.

2 Methods

2.1 FLOWBACK

FLOWBACK reconstructs all-atom protein structures x_1 from initial configurations that place each residue's atoms at its parent C_α position, i.e., $x_0 \sim \mathcal{N}(x_{C_\alpha}, \sigma_p^2 I)$, where σ_p is a prior noise that governs the trade-off between structual accuracy and diversity. FLOWBACK learns a vector field $v_\gamma(x_t, t)$ (with parameters γ learned from the data) under a flow-matching objective [43], with

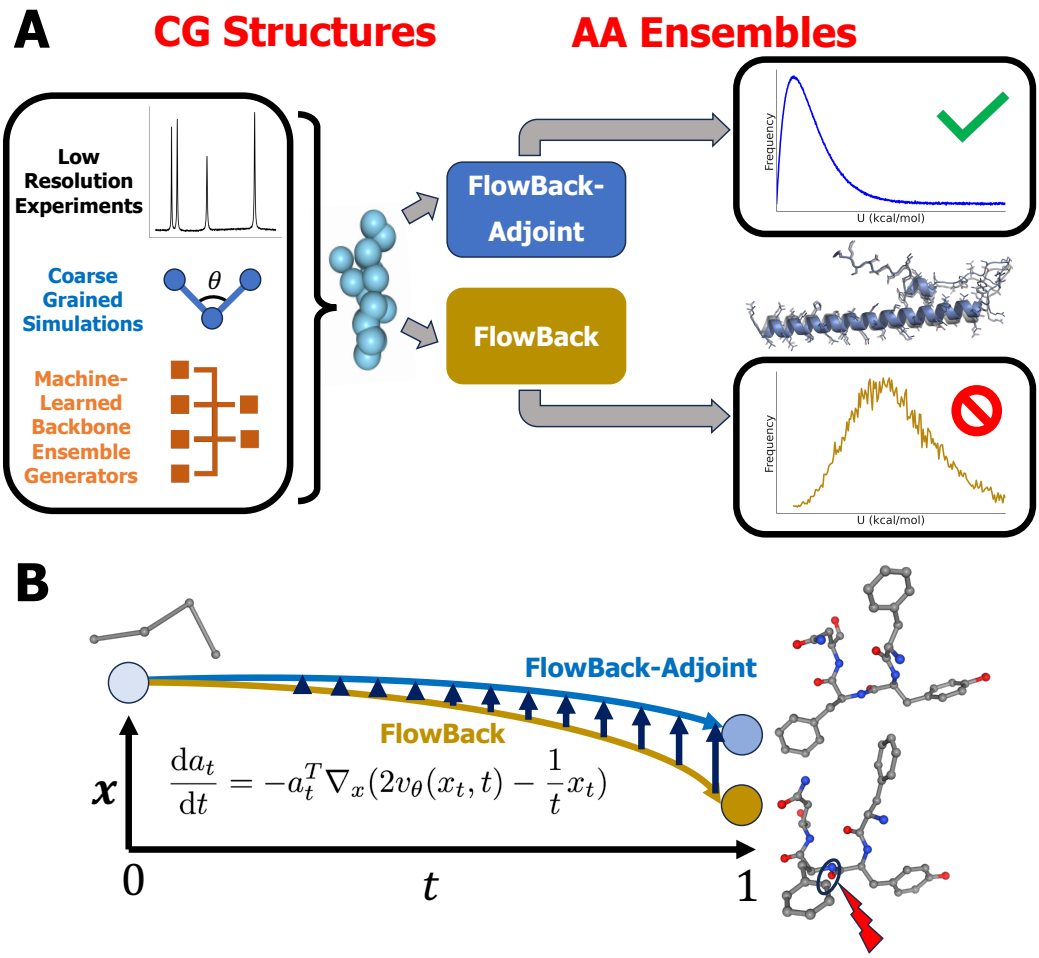

Figure 1: FLOWBACK-ADJOINT improves on the state-of-the art FLOWBACK [22] conditional flow-matching backmapping model. (A) FLOWBACK-ADJOINT accepts C_α traces from low-resolution NMR experiments, CG simulations, generative models, among other sources. Resulting distributions of energies match MD-like distributions more faithfully than state-of-the-art tools like FLOWBACK. (B) FLOWBACK-ADJOINT restores all-atom detail with improved accuracy by using adjoint matching to incorporate inductive biases from a molecular force field, yielding conformationally-diverse and low-energy ensembles of all-atom structures. The incorporation of physics-aware corrections to the flow via adjoint matching reduces the presence of high-energy configurations in the generated ensemble, such as atom-atom clashes indicated by the red lightning bolt.

training target $u_t = x_1 - x_0$ defined between a Gaussian prior centered at the C_α trace and the all–atom endpoint [22]. Conditioning on the C_α geometry enforces backbone fidelity while permitting diverse side–chain ensembles. We use FLOWBACK as the base generator that we guide with energies. (Additional details and hyperparameters are provided in Appendix A.1.1)

2.2 FLOWBACK+LJ/BONDS

As an intermediate step before adjoint matching can be applied, we introduce a coarse–to–fine schedule that progressively activates velocities corresponding to chirality, Lennard-Jones, and bonded corrections so the generator resolves harsh physical errors before learning the fine geometrical changes. These corrections (chirality enforcement, clash removal, bond relaxation) fix residual D-chiralities, atom overlaps, and stretched bonds from FLOWBACK (especially at large σ_p), and the LJ/Bonds step is necessary to smooth the energy gradients so adjoint matching can meaningfully guide the flow. FLOWBACK+LJ/BONDS, which we term v_θ, may also be considered an ablation benchmark for

FLOWBACK-ADJOINT that introduces physics-based corrections on top of which we may discern the additional benefit conveyed by the learned energy-based adjoint matching procedure. Additional details of the Lennard-Jones, bond, and chiral velocities are provided in Appendix A.1.2.

2.3 FLOWBACK-ADJOINT

Adjoint Matching. FLOWBACK-ADJOINT augments inference-time guidance with a training signal that aligns the learned drift to a reward-tilted target distribution via adjoint matching [44]. Given the base distribution generated by FLOWBACK+LJ/BONDS, $p_\theta(x)$, we define a reward-weighted density $p_\phi(x) = p_\theta(x)e^{R(x)}$ and fine-tune the vector field so that trajectories follow updates consistent with this tilt. Operationally, we adjust the flow-matching trajectories toward the tilted distribution as a stochastic optimal control (SOC) problem [48]; such a problem samples flow-matching trajectories, computes the reward signal at the endpoint, and learns how to guide our trajectories toward regions of higher reward. The reward signal is propagated backward along the trajectories' paths, and the drift is updated so future integrations move toward the reward-tilted distribution. Unlike FLOWBACK, which draws initial heavy–atom coordinates from Gaussians centered at their parent C_α, FLOWBACK-ADJOINT samples from a zero–mean prior $x_0 \sim \mathcal{N}(0, \sigma_p^2 I)$ and learns an additive offset relative to the parent C_α. The C_α trace is still provided as input; this reparameterization changes only the internal–coordinate representation and leaves observable metrics unchanged. To satisfy the SOC formulation without bias [44], during the trajectory-sampling portion of training, we replace typical Euler integration with the so-called "memoryless" flow where p_ϕ is faithfully sampled, and that final configurations x_1 are independent of the initially sampled random variable x_0 [44]:

$$\mathrm{d}x_t = 2\, v_\phi(x_t,\, t, C, T)\,\mathrm{d}t \;-\; \frac{1}{t}\, x_t\,\mathrm{d}t \;+\; \sigma_t(t)\,\sigma_p\,\mathrm{d}W_t, \quad \text{with} \quad \sigma_t^2(t) = 2\eta_t,\;\; \eta_t = \frac{1 - t + \Delta t}{t + \Delta t}.$$

Here, C represents the CG trace and T encodes atom-level topologies. While earlier models only accounted for this information implicitly, we now include it explicitly, because we learn the offset from the CG scaffold rather than the absolute coordinates, making this accounting essential. For completeness, we also additionally retrain FLOWBACK and FLOWBACK+LJ/BONDS with these internal coordinates. We then learn the "adjoint state" $a_t \in \mathbb{R}^{3N}$ [44], which represents the vector guiding the trajectory toward high reward and evolves under the "lean adjoint ODE" [44]:

$$\frac{\mathrm{d}a_t}{\mathrm{d}t} = -\,a_t^\top\,\nabla_x\Big(2\, v_\theta(x_t, t, C, T) - \tfrac{1}{t}x_t\Big), \qquad a_1 = -\nabla_x R(x_1),$$

In this work, to reward low-energy structures and steer the flow toward regions of physical accuracy, low clashes and correct bond lengths, we treat $R(x_1) = -\lambda U(x_1)$ with the scalar hyperparameter λ set to 0.01. We integrate this ODE backward from $t = 1$ to $t = 0$ using the same 100 steps as in the forward flow. Mathematically, the SOC problem [49] is optimal at $v_\phi(x_t, t, C, T) = v_\theta(x_t, t, C, T) - \frac{\sigma_t(t)^2}{2}\, a_t$, and we may achieve this objective by minimizing the sum of squared residuals across the flow,

$$\mathcal{L}_{\text{adjoint}} \;=\; \sigma_p^2 \sum_t \left\| \frac{2}{\sigma_t(t)}\big(v_\phi(x_t, t) - v_\theta(x_t, t)\big) \;+\; \sigma_t(t)\, a_t \right\|^2. \tag{1}$$

Implementation. We curate training pairs (C_i, T_i) and reference atom sets x_i by uniformly subsampling single frames from the D. E. Shaw Research (DESRES) all-atom trajectories of 11 fast-folding mini-proteins (See Appendix A.1.3) [50]. We train on nine of these trajectories and hold out two for evaluation (see Section 2.4) Each training iteration samples a DESRES frame index i, feeds the C_α trace and topological informaiton into v_ϕ, draws $x_0 \sim \mathcal{N}(0, \sigma_p^2 I)$, simulates a memoryless trajectory, evaluates $U(x_1)$ and $\nabla_x U(x_1)$, integrates the lean adjoint ODE backward along the sampled path to obtain $\{a_t\}$, and updates v_ϕ by minimizing the adjoint loss over a subset of time points. See detailed hyperparameters in Appendix A.1.3.

At test time we generate ensembles by Euler integration of the fine–tuned vector field. We sample initial heavy–atom offsets from the zero–mean prior $x_0 \sim \mathcal{N}(0, \sigma_p^2 I)$ and integrate from $t = 0$ to $t = 1$ with 100 uniform steps $\Delta t = 0.01$, updating $x_{t+\Delta t} = x_t + v_\phi(x_t, t, C, T)\,\Delta t$. Ensembles are obtained by reseeding x_0.

2.4 Evaluation Metrics and Model Comparisons

We evaluate FLOWBACK-ADJOINT on two test sets, each serving a distinct role.

Test Set 1 (Reference MD Ensembles). This set consists of two held-out DESRES molecular dynamics trajectories: the WW domain and Protein B [50]. As long, high-quality MD simulations, these ensembles provide reliable reference data for computing bond recovery, clash rates, configurational diversity, energy overlap, and stability under a molecular-dynamics integrator

Test Set 2 (Out-of-Distribution Proteins). This set probes generalization to unseen proteins. It includes 19 proteins from the BIOEMU `OOD60` benchmark [15], the 97 kDa LapD c-di-GMP receptor module from *Vibrio cholerae* (PDB `6PWK`), and three single-point mutants of the DNA-binding protein DUX4. Together, these 23 proteins span uncommon folds, long chains, and near-native variants, with no sequence sharing more than 60% identity with training proteins. For each protein we generated 1,000 backbone conformations using BIOEMU and then backmapped side chains with FLOWBACK, FLOWBACK+LJ/BONDS, FLOWBACK-ADJOINT, and HPACKER (BIOEMU's default side chain generator). This set is used to assess whether models can produce low-energy, clash-free AA structures on out-of-sample data.

We assess model performance over these two test sets using five tests designed to probe different aspects of the structure and energy of the backmapped AA configurational ensemble:

- **Tests 1–3 (Bond Score, Clash Score, Diversity Score)** – follow the metrics introduced in Jones et al. [42]. We aim for (100%) high bond recovery, low (0%) clash, and a low diversity score ($\lesssim 0$, see Appendix A.2) indicating a reference configuration bein indistinguishable from the generated ensemble.
- **Test 4 (Energy Divergence)** – quantifies the overlap (K–L divergence) in the CHARMM27 energy distribution of the backmapped configurational ensemble with the ground-truth MD trajectories. Successfully high overlap implies a low (~ 0) K–L divergence.
- **Test 5 (MD Test)** – assesses the stability of MD simulations launched from backmapped AA configurations without any additional energy relaxation. We aim for a high (100%) success rate. For meaningful results, this test is only applied to $\sigma_p = 0.003$, the setting that maximizes accuracy relative to diversity.

Test coverage. Tests 1, 3, and 5 are only applied to Test Set 1 (MD reference ensembles). Tests 2 (Clash Score) and 4 (Energy Divergence) are applied to both Test Set 1 and Test Set 2 (out-of-distribution proteins).

3 Results

3.1 Hold-out DESRES Trajectories: WW domain and Protein B

Our evaluation of FLOWBACK-ADJOINT to the two hold-out trajectories of WW domain and Protein B [50] represents a moderate out-of-sample test of FLOWBACK-ADJOINT in an application to novel mini-proteins not seen during training. In Table 1, we present a comparison of each tested model on all five aforementioned tests for three different choices of the prior noise σ_p (Section 2.4). We recall that σ_p controls the magnitude of the initial noise distribution in the flow-matching path and largely adjusts the trade-off between structural accuracy and diversity of the AA ensemble.

Structure. Relative to FLOWBACK and FLOWBACK+LJ/BONDS, FLOWBACK-ADJOINT preserves the best-in-class bond correctness ($\sim 99.9\%$ at $\sigma_p \in \{0.003, 0.005\}$) while further reducing steric clashes. On both WW and Protein B, unlike any other model, it eliminates the residual clashes at the lowest prior noise ($\sigma_p = 0.003$), as well as for Protein B at $\sigma_p = 0.005$. Under the largest-noise setting ($\sigma_p = 0.010$ nm), clashes are reduced by $\sim 45\%$–60% versus +LJ/Bonds and by $\geq 98\%$ versus FLOWBACK, with bond correctness comparable or modestly higher. Exact per-target values are in Table 1.

Diversity. The prior noise σ_p is the dial that sets ensemble breadth: smaller values compress trajectories and, within a tilted distribution, further concentrate probability mass at energy minima. Hence, at $\sigma_p = 0.003$ nm, FLOWBACK-ADJOINT is modestly less diverse than FLOWBACK and +LJ/Bonds (on the order of 10–20% farther from the target spread). As σ_p increases, this gap nar-

Table 1: Structural, diversity, and energy performance of each model over the two hold-out DESRES trajectories of WW domain and Protein B [50]. The best-performing values in each column are highlighted in **bold**. A KL divergence of ∞ in the energy score implies the distribution of energies over the reconstructed ensemble possesses no overlap with that from the MD trajectory. The optimal diversity score is selected as that which lies closest to zero.

WW domain						
Model	σ_p (nm)	**Bond** (↑)	**Clash** (↓)	**Diversity** (↓)	**Energy** (↓)	**MD** (↑)
FLOWBACK	0.003	99.00%	0.2383%	0.3532	0.1397	66.7%
FLOWBACK+LJ/BONDS	0.003	**99.93%**	0.0077%	0.3627	0.0074	80.9 %
FLOWBACK-ADJOINT	0.003	99.92%	**0.0000%**	0.4010	**0.0018**	**82.6%**
FLOWBACK	0.005	97.28%	1.3107%	0.1368	0.2010	–
FLOWBACK+LJ/BONDS	0.005	99.84%	0.0231%	0.1880	0.0390	–
FLOWBACK-ADJOINT	0.005	99.86%	0.0051%	0.2300	0.0137	–
FLOWBACK	0.010	84.79%	7.6790%	-0.0835	∞	–
FLOWBACK+LJ/BONDS	0.010	98.31%	0.2755%	-0.0624	0.2030	–
FLOWBACK-ADJOINT	0.010	98.63%	0.1166%	**-0.0270**	0.2010	–

Protein B						
Model	σ_p (nm)	**Bond** (↑)	**Clash** (↓)	**Diversity** (↓)	**Energy** (↓)	**MD** (↑)
FLOWBACK	0.003	99.30%	0.2003%	0.3117	0.0859	50.8%
FLOWBACK+LJ/BONDS	0.003	**99.97%**	0.0033%	0.3317	0.0087	63.8%
FLOWBACK-ADJOINT	0.003	**99.97%**	**0.0000%**	0.3730	**0.0017**	**65.4%**
FLOWBACK	0.005	97.64%	0.6517%	0.1185	0.1851	–
FLOWBACK+LJ/BONDS	0.005	99.95%	0.0016%	0.1561	0.0394	–
FLOWBACK-ADJOINT	0.005	99.95%	0.0000%	0.1915	0.0122	–
FLOWBACK	0.010	85.29%	4.3543%	-0.1112	∞	–
FLOWBACK+LJ/BONDS	0.010	99.26%	0.0499%	-0.0711	0.1987	–
FLOWBACK-ADJOINT	0.010	99.37%	0.0278%	**-0.0347**	0.1942	–

rows, and by $\sigma_p = 0.010$ nm, FLOWBACK-ADJOINT is comparably diverse to FLOWBACK on both proteins while retaining much lower bond and clash errors than the base model. Exact per-target values are in Table 1.

Energy. In terms of the energy assessment, we find that FLOWBACK-ADJOINT more faithfully captures the potential energy distribution of the hold-out MD trajectory ensembles. Focusing on the standard noise value of σ_p = 0.003 nm, the KL divergence between the reconstructed and MD potential energy distributions drops by an order of magnitude in going from FLOWBACK to FLOWBACK+LJ/BONDS, and a further approximately 4-5-fold reduction is achieved in moving to FLOWBACK-ADJOINT.

Figure 2A provides a graphical illustration of this improvement by juxtaposing the potential energy distributions generated by each model against those from the MD trajectories. Adding Lennard–Jones and bonded corrections in FLOWBACK+LJ/BONDS shifts the distribution to much lower values and closer to MD. FLOWBACK-ADJOINT further closes the gap, nearly matching the DESRES distribution. As illustrated in Figure A.3, similar trends persist at a noise levels of σ_p = 0.005 nm and 0.01 nm, albeit with overall poorer matching of the reference MD energy distribution.

MD and Forces. We test whether guidance reduces large per–atom gradients that can destabilize short MD integrations (Test 5) and lead to residual physical errors. Relative to the base model, FLOWBACK-ADJOINT yields smaller peak forces and a higher stability rate, and is comparable to or better than FLOWBACK+LJ/BONDS; full definitions, setup, and complete distributions are provided in Appendix A.3 and Figure A.4.

3.2 BIOEMU 23-protein Test Set

We next asked whether the advantages of FLOWBACK-ADJOINT carry over to proteins that lie far outside the training distribution. We inserted our integrator into the BIOEMU pipeline [15] by replacing the default HPACKER side-chain packing module with FLOWBACK-ADJOINT. We conducted this assessment over a 23-protein benchmark comprising the 19 BIOEMU OOD60 targets, the

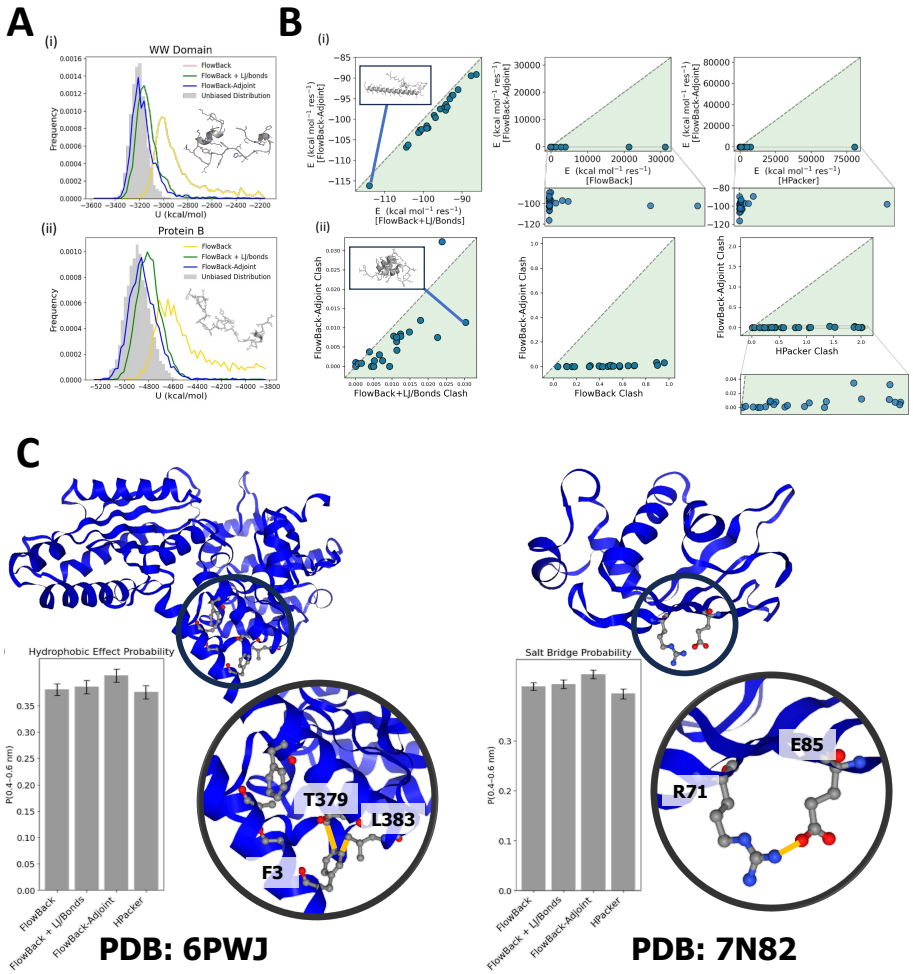

Figure 2: Energy and force performance of each model over the two hold-out DESRES trajectories of WW domain and Protein B [50]. All models employ a noise value of $\sigma_p = 0.003$ nm. (A) Distribution of potential energies over an ensemble of (i) 2230 WW domain configurations or (ii) 2602 Protein B configurations from each model relative to the distribution over the MD simulation trajectory. (B) Performance of FLOWBACK-ADJOINT as a modular replacement for side-chain reconstruction within the BIOEMU backmapping workflow. Comparison of FLOWBACK-ADJOINT (i) potential energies and (ii) clash scores relative to FLOWBACK, FLOWBACK+LJ/BONDS, and HPACKER for each of the 23 test proteins. Backbones are generated with BIOEMU and backmapped to AA resolution by each method. For energies, we report medians and for the clash scores we report means. The green-shaded region of each plot indicates superior FLOWBACK-ADJOINT performance. (C) Assessment of the prevalence of AA backmapped structures containing biologically important features in two test proteins: the heavy-atom contact occupancies between F3 of the S-helix and L383 of the EAL domain in LapD as a key hydrophobic interaction in receptor signaling [51] (left) and the C_ζ-C_δ distance distribution of the R71-E85 salt bridge in the Se0862 cyanobacterial protein [52] (right). The prevalence of each motif over the 1000 AA backmapped structures for each of the four methods is displayed in the accompanying bar charts.

c-di-GMP receptor LapD, and three point mutants of DUX4. We confirmed via a BLAST search that none of these proteins shared more than 60% identity with any entry in the SidechainNet training corpus [53] that was used to train the original FLOWBACK model or any of the nine DESRES proteins [50] used to train FLOWBACK-ADJOINT, thereby guarding against any observed improvements over this test set resulting from data leakage or high sequence similarity with the training set (Appendix A.4). For each protein we used BIOEMU to generate 1,000 backbone-only conforma-

tions. We then generated for each structure full AA backmapped configurations using FLOWBACK, FLOWBACK+LJ/BONDS, FLOWBACK-ADJOINT, and HPACKER, recalling that the three FLOWBACK-based pipelines used only the C_α trace while HPACKER used the full backbone N, C_α, C, and O atoms.

Figure 2B(i) summarizes the performance of FLOWBACK-ADJOINT against the three other approaches in terms of the median potential energy computed over the AA reconstructions from 1000 BIOEMU backbone-only structures. The full energy distributions over the 1000 candidates are presented in Figure A.5. For all 23 test proteins, FLOWBACK-ADJOINT generates AA ensembles with lower energies. Relative to the geometry-centric HPACKER, the improvement is dramatic: median energies drop by roughly two orders of magnitude. Relative to FLOWBACK, the incorporation of Lennard-Jones and bonded interaction inductive biases within FLOWBACK+LJ/BONDS cuts median energies between 6-85%, and FLOWBACK-ADJOINT further reduces them by an additional 1-3 kcal/mol.residue. Overall, FLOWBACK-ADJOINT lowers single-point energies in the OOD60 proteins by a median of $\sim$78 kcal/mol.residue relative to FLOWBACK. Figure 2B (ii) presents the mean clash scores over the 1000 structures for each of the 23 proteins in the test set to provide a complementary measure of steric quality. Consistent with the energy analysis, FLOWBACK-ADJOINT outperforms both FLOWBACK and HPACKER on every protein in the test set. On average, FLOWBACK-ADJOINT reduces clashes by 98.8% compared to the FLOWBACK prior. It also outperforms FLOWBACK+LJ/BONDS on 20/23 proteins. In the remaining three cases, the origin of these marginally better performance to be due to a small number of conformations with poorly-packed side-chains that skew the FLOWBACK-ADJOINT mean clash scores to high values; these rare instances highlight a potential avenue of improvement for FLOWBACK-ADJOINT (Figure A.6).

As a final test, we asked whether FLOWBACK-ADJOINT captures biologically meaningful features that its predecessors may overlook. Figure 2C highlights two examples where the model is challenged to reproduce known favorable interactions [54, 55]. In LapD (left), we track the heavy-atom contact occupancy between F3 of the S-helix and L383 of the EAL domain as a hydrophobic interaction known to be critical for receptor signaling [51]. In the cyanobacterial protein Se0862 (right), we examine the C_ζ–C_δ distance distribution of the R71-E85 salt bridge [52]. In both the hydrophobic contact and the salt bridge, FLOWBACK-ADJOINT reproduces the native interaction with higher probability over the 1000 trials relative to FLOWBACK, FLOWBACK+LJ/BONDS, and HPACKER.

Taken together, these results demonstrate that the adjoint matching refinement not only stabilizes proteins drawn from the training distribution but also generalizes to challenging, unseen sequences, delivering structurally plausible and energetically stable AA configurational ensembles across a wide range of benchmark proteins.

4 Conclusions

We have introduced FLOWBACK-ADJOINT as an energy-aware fine-tuning of the deep generative conditional flow-matching FLOWBACK model for AA backmapping. FLOWBACK-ADJOINT incorporates physics-based corrections to bond lengths and Lennard-Jones interactions, and an energy-based inductive bias that propagates the gradients of a molecular mechanics force field through the model using adjoint matching. The adjoint matching procedure modifies the FLOWBACK flow to promote low-energy configurations in the target ensemble without retraining the model from scratch or inducing catastrophic forgetting. We propose FLOWBACK-ADJOINT as an accurate, efficient physics-aware generative model for AA backmapping, but also as a springboard for future work to address current limitations and broaden its scope.

A first limitation concerns energy: the energy gap to MD stems almost entirely from the chirality–fixing velocity introduced in FLOWBACK+LJ/BONDS (see Appendix A.1.2) applied during sampling. Removing this correction collapses the KL divergence to near zero in held-out proteins (Figure A.7). This shows the vector field already yields near-Boltzmann ensembles, and motivates replacing the *ad hoc* correction with a chirality-aware $SE(3)$-equivariant model.

A second shortcoming is inference speed: incorporating Lennard–Jones and bonded corrections makes FLOWBACK-ADJOINT about 3$\times$ slower than FLOWBACK (1588 s vs 507 s on a set of 83 CASP13 structures). Speed may be recovered by replacing corrective velocities with an intermediate phase of adjoint matching on a smoother, "LJ/Bonds-only" reward.

A third avenue for exploration is the replacement of the classical CHARMM27 force field [56] used throughout this work with a machine learned interatomic potential (MLIP) with broader generalizability beyond proteins. For example, Meta's recently released Universal Model for Atoms (UMA) potential provides energies and forces for arbitrary combinations of 83 elements through a single network [57]. Much like FLOWBACK's extension to DNA [22], this generalization, along with broader training data, will enable us to extend FLOWBACK-ADJOINT to other classes of molecules, including nucleic acids, lipids, peptoids, organic semiconductors, and synthetic polymers.

References

[1] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, and Anna et al. Potapenko. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[2] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

[3] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[4] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024.

[5] Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu, Qi Xie, Bonnie Berger, Jianzhu Ma, and Jian Peng. High-resolution de novo structure prediction from primary sequence. *bioRxiv preprint bioRxiv:10.1101/2022.07.21.500999*, 2022.

[6] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023.

[7] Yong Zhao and Michel F. Sanner. Protein–ligand docking with multiple flexible side chains. *Journal of Computer-Aided Molecular Design*, 22(9):673–679, 2008.

[8] I. Buch, M. J. Harvey, T. Giorgino, D. P. Anderson, and G. De Fabritiis. High-throughput all-atom molecular dynamics simulations using distributed computing. *Journal of Chemical Information and Modeling*, 50(3):397–403, 2010.

[9] Gang Xu, Yilin Wang, Qinghua Wang, and Jianpeng Ma. Studying protein–protein interaction through side-chain modeling method OPUS-Mut. *Briefings in Bioinformatics*, 23(5):bbac330, 2022.

[10] Annett Bachmann, Dirk Wildemann, Florian Praetorius, Gunter Fischer, and Thomas Kiefhaber. Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proceedings of the National Academy of Sciences of the United States of America*, 108(10):3952–3957, 2011.

[11] Karel Berka, Roman A. Laskowski, Pavel Hobza, and Jiří Vondrášek. Energy matrix of structurally important side-chain/side-chain interactions in proteins. *Journal of Chemical Theory and Computation*, 6(7):2191–2203, 2010.

[12] Bowen Jing, Bonnie Berger, and Tommi Jaakkola. AlphaFold meets flow matching for generating protein ensembles. *arXiv preprint arXiv:2402.04845*, 2024.

[13] Nicolas Wolf, Leif Seute, Vsevolod Viliuga, Simon Wagner, Jan Stühmer, and Frauke Gräter. Learning conformational ensembles of proteins based on backbone geometry. *arXiv preprint arXiv:2503.05738*, 2025.

[14] Yan Wang, Lihao Wang, Yuning Shen, Yiqun Wang, Huizhuo Yuan, Yue Wu, and Quanquan Gu. Protein conformation generation via force-guided SE(3) diffusion models. *arXiv preprint arXiv:2403.14088*, 2024.

[15] Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew Y. K. Foong, Victor García Satorras, Osama Abdin, Bastiaan S. Veeling, Iryna Zaporozhets, Yaoyi Chen, Soojung Yang, Adam E. Foster, Arne Schneuing, Jigyasa Nigam, Federico Barbero, Vincent Stimper, Andrew Campbell, Jason Yim, Marten Lienen, Yu Shi, Shuxin Zheng, Hannes Schulz, Usman Munir, Roberto Sordillo, Ryota Tomioka, Cecilia Clementi, and Frank Noé. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, eadv9817:DOI:10.1126/science.adv9817, 2024.

[16] Julia Koehler Leman, Brian D. Weitzner, Steven M. Lewis, Jared Adolf-Bryfogle, Nawsad Alam, Rebecca F. Alford, Melanie Aprahamian, David Baker, Kyle A. Barlow, Patrick Barth, Benjamin Basanta, Brian J. Bender, Kristin Blacklock, Jaume Bonet, Scott E. Boyken, Phil Bradley, Chris Bystroff, Patrick Conway, Seth Cooper, Bruno E. Correia, Brian Coventry, Rhiju Das, RenéM. De Jong, Frank DiMaio, Lorna Dsilva, Roland Dunbrack, Alexander S. Ford, Brandon Frenz, Darwin Y. Fu, Caleb Geniesse, Lukasz Goldschmidt, Ragul Gowthaman, Jeffrey J. Gray, Dominik Gront, Sharon Guffy, Scott Horowitz, Po-Ssu Huang, Thomas Huber, Tim M. Jacobs, Jeliazko R. Jeliazkov, David K. Johnson, Kalli Kappel, John Karanicolas, Hamed Khakzad, Karen R. Khar, Sagar D. Khare, Firas Khatib, Alisa Khramushin, Indigo C. King, Robert Kleffner, Brian Koepnick, Tanja Kortemme, Georg Kuenze, Brian Kuhlman, Daisuke Kuroda, Jason W. Labonte, Jason K. Lai, Gideon Lapidoth, Andrew Leaver-Fay, Steffen Lindert, Thomas Linsky, Nir London, Joseph H. Lubin, Sergey Lyskov, Jack Maguire, Lars Malmström, Enrique Marcos, Orly Marcu, Nicholas A. Marze, Jens Meiler, Rocco Moretti, Vikram Khipple Mulligan, Santrupti Nerli, Christoffer Norn, Shane Ó'Conchúir, Noah Ollikainen, Sergey Ovchinnikov, Michael S. Pacella, Xingjie Pan, Hahnbeom Park, Ryan E. Pavlovicz, Manasi Pethe, Brian G. Pierce, Kala Bharath Pilla, Barak Raveh, P. Douglas Renfrew, Shourya S. Roy Burman, Aliza Rubenstein, Marion F. Sauer, Andreas Scheck, William Schief, Ora Schueler-Furman, Yuval Sedan, Alexander M. Sevy, Nikolaos G. Sgourakis, Lei Shi, Justin B. Siegel, Daniel-Adriano Silva, Shannon Smith, Yifan Song, Amelie Stein, Maria Szegedy, Frank D. Teets, Summer B. Thyme, Ray Yu-Ruei Wang, Andrew Watkins, Lior Zimmerman, and Richard Bonneau. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature Methods*, 17(7):665–680, 2020.

[17] Qiang Wang, Adrian A Canutescu, and Roland L Dunbrack. SCWRL and MolIDE: Computer programs for side-chain conformation prediction and homology modeling. *Nature Protocols*, 3(12):1832–1847, 2008.

[18] Matthew McPartlon and Jinbo Xu. An end-to-end deep learning method for protein side-chain packing and inverse folding. *Proceedings of the National Academy of Sciences of the United States of America*, 120(23):e2216438120, 2023.

[19] Mikita Misiura, Raghav Shroff, Ross Thyer, and Anatoly B. Kolomeisky. DLPacker: Deep learning for prediction of amino acid side chain conformations in proteins. *Proteins: Structure, Function, and Bioinformatics*, 90(6):1278–1290, 2022.

[20] Gian Marco Visani, William Galvin, Michael Neal Pun, and Armita Nourmohammad. H-Packer: Holographic rotationally equivariant convolutional neural network for protein side-chain packing. *arXiv preprint arXiv:2311.09312*, 2023.

[21] Nicholas Z. Randolph and Brian Kuhlman. Invariant point message passing for protein side chain packing. *Proteins: Structure, Function, and Bioinformatics*, 92(10):1220–1233, 2024.

[22] Michael S. Jones, Smayan Khanna, and Andrew L. Ferguson. FlowBack: A generalized flow-matching approach for biomolecular backmapping. *Journal of Chemical Information and Modeling*, 65(2):672–692, 2025.

[23] William George 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. *Journal of Chemical Physics*, 128(24): 244114, 2008.

[24] Cecilia Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Current Opinion in Structural Biology*, 18(1):10–15, 2008.

[25] Paulo C. T. Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart M. H. Bruininks, Tsjerk A. Wassenaar, Peter C. Kroon, Josef Melcr, Vincent Nieto, Valentina Corradi, Hanif M. Khan, Jan J. Domanski, Matti Javanainen, Hector Martinez-Seara, Nathalie Reuter, Robert B. Best, Ilpo Vattulainen, Luca Monticelli, Xavier Periole, D. Peter Tieleman, Alex H de Vries, and Siewert J. Marrink. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nature Methods*, 18:382–388, 2021.

[26] Jaehyeok Jin, Alexander J Pak, Aleksander EP Durumeric, Timothy D Loose, and Gregory A Voth. Bottom-up coarse-graining: Principles and perspectives. *Journal of Chemical Theory and Computation*, 18(10):5759–5791, 2022.

[27] Tsjerk A Wassenaar, Kristyna Pluhackova, Rainer A Böckmann, Siewert J Marrink, and D Peter Tieleman. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation*, 10(2): 676–690, 2014.

[28] Leandro E Lombardi, Marcelo A Martí, and Luciana Capece. CG2AA: Backmapping protein coarse-grained structures. *Bioinformatics*, 32(8):1235–1237, 2016.

[29] Marc Stieffenhofer, Michael Wand, and Tristan Bereau. Adversarial reverse mapping of equilibrated condensed-phase molecular structures. *Machine Learning: Science and Technology*, 1(4):045014, 2020.

[30] Marc Stieffenhofer, Tristan Bereau, and Michael Wand. Adversarial reverse mapping of condensed-phase molecular structures: Chemical transferability. *APL Materials*, 9(3):031107, 2021.

[31] Marc Stieffenhofer, Christoph Scherer, Falk May, Tristan Bereau, and Denis Andrienko. Benchmarking coarse-grained models of organic semiconductors via deep backmapping. *Frontiers in Chemistry*, 10:982757, 2022.

[32] Wei Li, Craig Burkhart, Patrycja Polińska, Vagelis Harmandaris, and Manolis Doxastakis. Backmapping coarse-grained macromolecules: An efficient and versatile machine learning approach. *Journal of Chemical Physics*, 153(4):041101, 2020.

[33] Yaxin An and Sanket A Deshmukh. Machine learning approach for accurate backmapping of coarse-grained models to all-atom models. *Chemical Communications*, 56(65):9312–9315, 2020.

[34] Wujie Wang, Minkai Xu, Chen Cai, Benjamin Kurt Miller, Tess Smidt, Yusu Wang, Jian Tang, and Rafael Gómez-Bombarelli. Generative coarse-graining of molecular conformations. *arXiv preprint arXiv:2201.12176*, 2022.

[35] Kirill Shmilovich, Marc Stieffenhofer, Nicholas E Charron, and Moritz Hoffmann. Temporally coherent backmapping of molecular trajectories from coarse-grained to atomistic resolution. *Journal of Physical Chemistry A*, 126(48):9124–9139, 2022.

[36] Yikai Liu, Ming Chen, and Guang Lin. Backdiff: A diffusion model for generalized transferable protein backmapping. *arXiv preprint arXiv:2310.01768*, 2023.

[37] Jiasheng Li, Zaiqiao Meng, and Shangsong Liang. Towards deep generative backmapping of coarse-grained molecular systems. In *Proceedings of the 2024 2nd Asia Conference on Computer Vision, Image Processing and Pattern Recognition*, pages 1–7, 2024.

[38] Yui Tik Pang, Lixinhao Yang, and James C Gumbart. From simple to complex: Reconstructing all-atom structures from coarse-grained models using cg2all. *Structure*, 32(1):5–7, 2024.

[39] Lim Heo and Michael Feig. One bead per residue can describe all-atom protein structures. *Structure*, 32(1):97–111, 2024.

[40] Daniele Angioletti, Stefano Raniolo, and Vittorio Limongelli. HEroBM: A deep equivariant graph neural network for universal backmapping from coarse-grained to all-atom representations. *arXiv preprint arXiv:2404.16911*, 2024.

[41] Curt Waltmann, Yihang Wang, Chengxi Yang, Siyoung Kim, and Gregory A Voth. MSBack: Multiscale backmapping of highly coarse-grained proteins using constrained diffusion. *Journal of Chemical Theory and Computation*, 2025.

[42] Michael S Jones, Kirill Shmilovich, and Andrew L Ferguson. DiAMoNDBack: Diffusion-denoising autoregressive model for non-deterministic backmapping of Cα protein traces. *Journal of Chemical Theory and Computation*, 19(21):7908–7923, 2023.

[43] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[44] Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. *arXiv preprint arXiv:2409.08861*, 2025.

[45] Jiarui Lu, Xiaoyin Chen, Stephen Zhewen Lu, Aurélie Lozano, Vijil Chenthamarakshan, Payel Das, and Jian Tang. Aligning protein conformation ensemble generation with physical feedback. *arXiv preprint arXiv:2505.24203*, 2025.

[46] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and Garegin A Papoian. AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *Journal of Physical Chemistry B*, 116(29):8494–8503, 2012.

[47] Giulio Tesei and Kresten Lindorff-Larsen. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe*, 2:94, 2023.

[48] Carles Domingo-Enrich, Jiequn Han, Brandon Amos, Joan Bruna, and Ricky T. Q. Chen. Stochastic optimal control matching. *arXiv preprint arXiv:2312.02027*, 2024.

[49] H J Kappen. Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11):P11011, 2005.

[50] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

[51] Giordan Kitts, Krista M. Giglio, David Zamorano-Sánchez, Jin Hwan Park, Loni Townsley, Richard B. Cooley, Benjamin R. Wucher, Karl E. Klose, Carey D. Nadell, Fitnat H. Yildiz, and Holger Sondermann. A conserved regulatory circuit controls large adhesins in vibrio cholerae. *mBio*, 10(6), 2019.

[52] Maria U Johansson, Maarten de Château, Mats Wikström, Sture Forsén, Torbjörn Drakenberg, and Lars Björck. Solution structure of the albumin-binding GA module: A versatile bacterial protein domain. *Journal of Molecular Biology*, 266(5):859–865, 1997.

[53] Jonathan Edward King and David Ryan Koes. SidechainNet: An all-atom protein structure dataset for machine learning. *Proteins: Structure, Function, and Bioinformatics*, 89(11):1489–1496, 2021.

[54] C. Nick Pace, Hailong Fu, Katrina Lee Fryar, John Landua, Saul R Trevino, Bret A Shirley, Marsha Mcnutt Hendricks, Satoshi Iimura, Ketan Gajiwala, J Martin Scholtz, and Gerald R Grimsley. Contribution of hydrophobic interactions to protein stability. *Journal of Molecular Biology*, 408(3):514–528, 2011.

[55] Raphaël Geney, Melinda Layten, Roberto Gomperts, Viktor Hornak, and Carlos Simmerling. Investigation of salt bridge stability in a generalized born solvent model. *Journal of Chemical Theory and Computation*, 2(1):115–127, 2006.

[56] Alexander D. MacKerell Jr., Nilesh Banavali, and Nicolas Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56(4):257–265, 2000.

[57] Brandon M. Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R. Kitchin, Daniel S. Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Ammar Rizvi, Sushree Jagriti Sahoo, Zachary W. Ulissi, and C. Lawrence Zitnick. UMA: A family of universal models for atoms. *arXiv preprint arXiv:2506.23971*, 2025.

[58] Mohammed AlQuraishi. ProteinNet: a standardized data set for machine learning of protein structure. *BMC Bioinformatics*, 20(1):1–10, 2019.

[59] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28 (1):235–242, 2000.

[60] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[61] 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.

[62] Yi Liao and Tess E. Smidt. Equiformer: Equivariant graph attention transformer for 3D atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

[63] Hendrik Antoon Lorentz. Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. *Annalen der Physik*, 248(1):127–136, 1881.

[64] D. Berthelot. Sur le mélange des gaz. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 126:1703–1706, 1898.

[65] J. E. Lennard-Jones. The electronic structure of some diatomic molecules. *Transactions of the Faraday Society*, 25:668–686, 1929.

[66] Michael P Allen and Dominic J Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2nd edition, 2017.

[67] Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: exploring flow and diffusion-based generative models with scalable interpolant transformers. *arXiv preprint arXiv:2401.08740*, 2024.

[68] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.

[69] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yonghai Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13 (7):e1005659, 2017.

[70] Christiam Camacho, George Coulouris, Vahram Avagyan, Nan Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. Blast+: Architecture and applications. *BMC Bioinformatics*, 10(421):1–9, 2009.

A Appendix

A.1 Method Architectures, Derivations, and Implementations

A.1.1 FLOWBACK

Description. FLOWBACK is a deep generative approach based on a conditional flow-matching objective to produce AA structural ensembles of proteins from C_α CG traces [22]. It learns a rotationally and translationally equivariant vector field $v_\gamma(x_t, t)$ whose deterministic integration carries a noisy, CG-conditioned prior distribution $q_0\big(x \mid C_\alpha\big)$ at $t = 0$ to a learned all-atom distribution $q_1(x)$ at $t = 1$. Each generated conformation comes equipped with an exact likelihood and exactly preserves the supplied C_α trace. The prior distribution is constructed by placing every heavy atom in an isotropic Gaussian of variance σ_p^2 around its parent C_α atom. The noise amplitude σ_p is the model's single exposed hyper-parameter that permits users trade off between high structural accuracy (smaller σ_p) and diverse configurational ensembles (larger σ_p). FLOWBACK operates without explicit consideration of hydrogen atoms, leaving these to be placed by downstream tools after the learned flow has positioned all of the heavy atoms.

Training of the learned vector field vector field $v_\gamma(x_t, t)$ proceeds entirely without molecular dynamics or structure relaxation, relying instead on all-atom structural databases to define a supervised learning objective. The model was trained over a corpus of 65,360 all-atom protein structures containing 20–1000 residues that were derived from the SidechainNet [53] extension of ProteinNet [58], which was itself curated from the Protein Data Bank [59]. Hydrogen atoms were stripped to define the AA targets $x_1 \sim q_1(x)$ and the C_α traces extracted to define the corresponding CG conditionings to generate the initial samples x_0 from which we learn the flow. For each training batch, a noisy initial structure $x_0 \sim q_0(x) = \mathcal{N}\big(x_0[M], \sigma_p^2 I\big)$ is drawn, a random time $t \in [0, 1]$ is chosen, a linearly interpolated mean $\mu_t = tx_1 + (1-t)x_0$ is formed, and i.i.d. Gaussian noise of variance σ_{int}^2 added to the mean to obtain $x_t \sim \mathcal{N}\big(\mu_t, \sigma_{int}^2 I\big)$. The coordinates of the C_α atoms are placed under a mask M to ensure that their coordinates in the final backmapped structure exactly match those in the conditioning C_α trace. However, later testing revealed that the C_α mask had little influence on the final all-atom coordinates, so in retraining FLOWBACK as a base model for the development of FLOWBACK-ADJOINT, we eliminated the mask in learning the flow. A six-layer equivariant graph neural network (EGNN) is trained to learn the vector field $v_\gamma(x_t, t)$ by regressing against the reference field $u_t = (x_1 - x_0)$ under an $L1$ loss. Operationally, we "trick" the EGNN into learning the drift $v_\gamma(x_t, t)$ by training it to predict a one-step-ahead configuration $x'_t = EGNN_\gamma(x_t, t)$ and then computing $v_\gamma(x_t, t) = (x'_t - x_t)$.

At inference time, initial structures x_0 are generated by initializing the heavy atoms around their parent C_α atom by drawing from $x_0 \sim \mathcal{N}\big(x_0, \sigma_p^2 I\big)$, where the σ_p used in inference need not be the same σ_p applied in training. The model then uses the learned flow to integrate the ordinary differential equation $\dot{x}_t = v_\gamma(x_t, t)$ for 100 Euler steps.

Amino acids – with the exception of glycine – are chiral, meaning their side chains can exist in two non-equivalent mirror-image forms known as L-form and D-form. Natural proteins employ only the L-form, with 99.99% of the FLOWBACK training data belongs to this stereoisomeric class. However, the FLOWBACK EGNN does not explicitly encode chirality constraints and, as a result, without any

corrections to the learned flow, a small fraction of generated side chains – about 3.7% in the original experiments – emerge in the incorrect D-form. FLOWBACK corrects this by detecting side chains that initially appear as D-isomers both early in the trajectory at $t = 0.2$ and at the end at $t = 1$ to, if necessary, perform a corrective mirror reflection. This operation eliminates all chirality errors while slightly raising the incidence of structural clashes.

Implementation. All calculations were performed on a single NVIDIA L40S GPU, which enabled efficient evaluation of the vector field over large molecular graphs using tensor cores. Training was conducted with a batch size of one protein, and the Gaussian noise added to the interpolated structure was fixed at $\sigma_{int} = 0.005$ nm. The prior noise applied to coarse-grained inputs was $\sigma_p = 0.003$ nm. The EGNN was trained to learn a vector field $v(x, t, M)$ by regressing against the reference field $u_t = (x_1 - x_0)$ under an $\ell 1$ loss. Operationally, we "trick" the EGNN into learning the drift $v_\gamma(x_t, t, M)$ by training it to predict a one-step-ahead configuration $x'_t = EGNN_\gamma(x_t, t, M)$ and then computing $v_\gamma(x_t, t, M) = (x'_t - x_t)$. Training is conducted for 15 epochs using the Adam optimizer [60] with a learning rate of 0.001. The model architecture consists of a six-layer $E(3)$-equivariant graph neural network [61, 62], with message passing between each node and its 15 nearest neighbors. The latent-space dimension of was 32. Both node and message embeddings were passed with `SiLU` nonlinearity. Vector field updates were clamped to a maximum magnitude of 2 nm/timestep.

A.1.2 FLOWBACK+LJ/BONDS

Description. FLOWBACK-ADJOINT is a sophistication of FLOWBACK that employs adjoint matching to fine-tune the learned flow by steering it towards low-energy configurations. As an intermediate stepping stone to this model, we introduce FLOWBACK+LJ/BONDS as a physically motivated augmentation to FLOWBACK that adds three auxiliary time-gated velocity fields to the learned drift v_γ without changing the network parameters, numerical integrator, or model architecture. *First*, in place of FLOWBACK's original discrete side-chain flipping protocol, we inject a continuous, differentiable velocity field v_{chiral} for $t \geq 0.25$ that smoothly steers any D-form residue toward the correct L-form configuration by pushing it along the axis perpendicular to the reflection plane, thereby enforcing native stereochemistry without breaking gradient flow. *Second*, a repulsive Lennard-Jones (LJ) velocity is introduced for $t \geq 0.85$, wherein heavy-atom pairs separated by more than two covalent bonds and residing within a distance $d < 0.42$ nm are subjected to a Lennard-Jones interaction under the CHARMM27 force field [56]. *Third*, for ($t \geq 0.95$), a harmonic bond length interaction is added around the equilibrium bond length mandated by the CHARMM27 force field [56] in order to draw each covalent bond towards its low-energy equilibrium configuration. We denote the FLOWBACK+LJ/BONDS drift – the FLOWBACK drift v_γ plus these three auxiliary corrections – by v_θ. This is the baseline flow that will be modified by energy-based adjustment in FLOWBACK-ADJOINT. Additional algorithmic details, including velocity clamping, scaling, and force constants, are provided in Appendix A.2.

These three corrections – chirality enforcement, steric clash removal, and bond relaxation – address the residual steric overlaps, stretched bonds, and occasional D-side chains that FLOWBACK is prone to produce, particularly for large choices of σ_p, and represent an inductive bias that alleviates the degree of additional steering that must be done by the molecular mechanics energy gradients propagated by adjoint matching in FLOWBACK-ADJOINT.

Implementation. To transform coarse side-chain coordinates into clash-free, force-field-compatible structures we employ a lightweight Euler integrator. Coordinates are advanced along a schedule $t_0 = 0 \to t_N = 1$ with uniform step $\Delta t = 1/N$,

$$x_{k+1} = x_k + v_\theta(x_k, t_k)\Delta t,$$

where the drift,

$$v_\theta = v_\gamma(x_k, t_k) + \mathbf{1}_{t_k \geq 0.25}\, v_{\text{chiral}} + \mathbf{1}_{t_k \geq 0.85}\, \frac{t_k^\alpha}{10}\, v_{\text{LJ}} + \mathbf{1}_{t_k \geq 0.95}\, t_k^\alpha\, v_{\text{bond}},$$

combines augments the learned, unmasked EGNN velocity field $v_\gamma(x_k, t_k)$ from FLOWBACK, with three auxiliary time-gated velocity fields that adjust the chirality, Lennard-Jones (LJ) interactions, and bond lengths. $\mathbf{1}_{t \geq \tau}$ denotes an indicator function that is zero for $t < \tau$ and unity for $t \geq \tau$. The scaling of the Lennard–Jones (L-J) velocities by $\frac{t_k^\alpha}{10}$ and the bonded velocities by t_k^α is intended as a

heuristic annealing schedule: at early integration steps, the scale factors are tiny, so both non-bonded and bonded forces are almost zero. As the trajectory advances and $t_k \to 1$, the factors grow rapidly with scaling t_k^α, progressively switching on the full force constants when the atoms become close to their physically realistic geometry. Empirically, we find an exponent of $\alpha = 20$ together with the $\frac{1}{10}$ pre-factor on the LJ term to perform well in practice.

Chirality. In place of FLOWBACK's original discrete side-chain flipping protocol, we inject a continuous, differentiable velocity field v_{chiral} for $t \geq 0.25$ that smoothly steers any D-form residue toward the correct L-form configuration thereby enforcing native stereochemistry without breaking gradient flow. For a residue with atoms $(\mathrm{N}, \mathrm{C}_\alpha, \mathrm{C}, \mathrm{C}_\beta)$ we monitor the volume,

$$V = \big(N - C_\alpha\big) \times \big(C - C_\alpha\big) \;\cdot\; \big(C_\beta - C_\alpha\big),$$

where a positive volume implies a desirable L-chirality, whereas a negative volume implies an undesirable D-chirality. For each Euler step we compute a target increment for each amino acid residue, ΔV,

$$\Delta V = \begin{cases} \varepsilon - V, & V < 0 \quad \texttt{(FLIP)}, \\ V_{\text{ref}}(t) - V, & 0 \leq V < V_{\text{ref}}(t) \quad \texttt{(PUSH)}, \\ 0, & \text{otherwise}, \end{cases}$$

where FLIP is a velocity that reflects the side chain across the symmetry plane, PUSH is a velocity that guides the residue from an aphysical position on or near the reflection plane to its expected positive-volume positions relative to the reflection plane, $V_{\text{ref}}(t) = 2{\times}10^{-3} t^3$ nm^3 acts as a cutoff set to about 80–90% of, yet always below, the volume expected at integration time t, and $\varepsilon = 2{\times}10^{-4}$ nm^3 prevents instabilities in normalizing the velocity vectors. This increment is then converted into a velocity as follows. First, to determine the magnitude of the velocity on every atom, we compute,

$$k_{\text{side}} = \texttt{clip}\left(\frac{\Delta V}{A\,\Delta t},\, 0,\, k_{\max} \right),$$

where $A = \big\|(N - C_\alpha) \times (C - C_\alpha)\big\|$ is the area of the parallelogram spanned by these three backbone atoms – and hence, k_{side} increases in proportion to the magnitude of the chirality error component normal to the reflection plane – and the clipping operation delimits the outputted value to the range $[0, k_{\max}]$, where $k_{\max}$ = 1.2 nm/timestep. We then compute the direction of the chriality velocity by projecting the normal vector of the chirality reflection plane onto the plane normal to the C_α–C_β bond to keep that bond length stable,

$$\hat{n}_\perp = \hat{n} - (\hat{n} \cdot u_\beta)\, u_\beta, \quad \text{where} \quad u_\beta = \frac{C_\beta - C_\alpha}{\|C_\beta - C_\alpha\|} \quad \text{and} \quad \hat{n} = ((N - C_\alpha) \times (C - C_\alpha))/A.$$

However, not all atoms are applied a uniform velocity: we note that the backbone N, C_α, and C are not affected by this operation, as they define the chirality reflection plane. All remaining atoms in the residue are subject to velocities,

$$v_{\text{chiral}} = \begin{cases} +k_{\text{side}}\, \hat{n}_\perp & \text{for side-chain atoms} \\ -\dfrac{1}{2} k_{\text{side}}\, \hat{n}_\perp & \text{for the backbone oxygen}, \end{cases}$$

with the backbone O atom velocity scaled by 50% to avoid backbone distortions and instability due to overcorrection.

Lennard-Jones. A Lennard-Jones (LJ) velocity term v_{LJ} is introduced for $t \geq 0.85$, wherein close-range heavy-atom pairs are subjected to a Lennard-Jones interaction under the CHARMM27 molecular mechanics force field [56]. Specifically, for each protein heavy-atom pair (i, j) topologically separated by more than two covalent bonds and possessing an inter-atomic distance $r_{ij} = \|x_i - x_j\| < 0.42$ nm, we compute the Lennard-Jones force,

$$F_{ij} = 24 \frac{\epsilon_{ij}}{r_{ij}^2} \left[2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] (x_i - x_j),$$

where $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ for chemically dissimilar atom types are computed as prescribed by the Lorentz-Berthelot mixing rules [63, 64, 65, 66].

The Lennard-Jones force on atom i is converted into a corresponding velocity correction for atom i by summing over all pairwise interactions to resolve the net force F_i experienced by atom i, dividing by the mass of atom i to compute the net acceleration $a_i = F_i/m_i$, and then multiplying by a single time step,

$$v_{\mathrm{LJ}_i} = \frac{\Delta t}{m_i} \sum_{j \neq i} F_{ij}.$$

The additional velocity introduced by this auxiliary force is capped at a magnitude of v_{LJ} = 5 nm/timestep to preserve stability of the Euler integrator.

Bonds. For ($t \geq 0.95$), a harmonic bond length interaction is added around the equilibrium bond length mandated by the CHARMM27 molecular mechanics force field MacKerell Jr. et al. [56]. For each bonded pair (i, j) with ideal bond length b_0 and spring constant k_b we evaluate the bonded force,

$$F_{ij}^{\text{bond}} = k_b\big(r_{ij} - b_0 t\big)(x_i - x_j),$$

and compute the corresponding velocity update as,

$$v_{\text{bond}_i} = -\frac{\Delta t}{m_i} F_{ij}^{\text{bond}}, \qquad v_{\text{bond}_j} = +\frac{\Delta t}{m_j} F_{ij}^{\text{bond}}.$$

The additional velocity introduced by this auxiliary force is capped at a magnitude of v_{bond} = 1 nm/timestep to preserve stability of the Euler integrator.

A.1.3 FLOWBACK-ADJOINT

Derivation of the FLOWBACK-ADJOINT method. Unlike FLOWBACK, which samples its initial coordinates from a Gaussian distribution of heavy atoms centered on each parent C_α, FLOWBACK-ADJOINT generates initial heavy atom positions from a zero-centered normal prior $x_0 \sim q_0(x) = \mathcal{N}\big(0, \sigma_p^2 I\big)$ and learns an additive offset relative to the parent C_α. The C_α coordinates are still passed as an input, and the predicted atom positions are placed relative to those coordinates before the EGNN featurization as well as for all subsequent geometric and energetic evaluations. The distribution of generated structures q_1 is identical to one obtained with the C_α-centered prior since the choice of base distribution changes only an internal coordinate representation and has no impact on any observable metrics, ensuring that all comparisons to the previous FLOWBACK model remain fully valid. The re-parameterization is required because the adjoint matching framework of Domingo-Enrich et al. [44] employed in this work assumes a zero-mean Gaussian base distribution in order to derive an unbiased estimator between the original flow and a reward-tilted target. Adopting $\mathcal{N}(0, \sigma_p^2 I)$ priors preserves the theoretical consistency of the adjoint correction while still allowing the model to condition explicitly on the supplied backbone. For completeness we also retrain the baseline FLOWBACK and FLOWBACK+LJ/BONDS model with the same $\mathcal{N}(0, \sigma_p^2 I)$ prior.

To train FLOWBACK-ADJOINT, we provide the C_α positions C, protein topologies (i.e., covalent connectivity of the constituent atoms) T, atomic positions x, and interpolation time t, to an EGNN that learns the trainable weights and biases ϕ of the learned conditional flow $v_\phi(x_t, t, C, T)$. The learned flow v_ϕ, represents a fine-tuned adjustment of the FLOWBACK+LJ/BONDS flow v_θ via adjoint matching. Adjoint matching is a post-training procedure that refines an existing diffusion or flow-matching generator so its outputs better satisfy a user-defined score or reward [44] by deriving a gradient correction that nudges the model's velocity field to raise the final reward under a specified objective function. In the present work, the adjoint matching procedure modifies a flow that samples AA configurations x from a base target distribution $p_\theta(x)$ by steering the flow to instead sample from a new reweighted distribution $p_\phi(x) = p_\theta(x)e^{R(x)}$, where $R(x)$ is the reward function.

Tilting a generator from its base distribution $p_\theta(x)$ to the new, reward-weighted density can be cast as a SOC problem wherein one seeks a time-dependent control $u_t(x)$ that minimizes the expected path cost $\mathbb{E}\Big[\int_0^1 \frac{1}{2}\|u_t\|^2 \,\mathrm{d}t - R(x_1)\Big]$ such that the $t = 1$ distribution follows the reward-weighted density $p_\phi(x) = p_\theta(x)e^{R(x)}$ [48]. To solve this SOC problem, we follow the approach presented in Domingo-Enrich et al. [44]. A key requirement for this approach is that the underlying diffusion must be memoryless, meaning that the variance of its driving noise satisfies $\sigma_t^2(t) = 2\eta_t$ with $\eta_t = (1 - t)/t$. Because $\eta_t \to \infty$ as $t \to 0$ and $\eta_t \to 0$ as $t \to 1$, the process randomizes the initial draw and gradually "de-randomizes" toward the end of the flow as the target distribution

increases in importance.In practice, the divergence at $t \to 0$ is avoided by adding a small offset in both the numerator and denominator, giving $\eta_t = (1 - t + \Delta t)/(t + \Delta t)$. The memoryless schedule guarantees that the controlled dynamics reproduce p_ϕ without bias. Violations of the memoryless requirement can leave a residual imprint of the starting noise and corrupt the tilt [44]. Adjoint Matching provides a practical solution to this SOC by computing, via an adjoint-state recursion, the minimal-variance estimator of the optimal control and realizing it as a small additive correction to the velocity field. Applied to the memoryless diffusion above, the method drives the generator toward the reward-weighted density $p_\phi(x) = p_\theta(x)e^{R(x)}$.

To generate a heavy-atom backbone from a fixed C_α trace within conditional flow-matching framework, we model the unknown coordinates as a time-indexed random path $\big(x_t\big)_{t\in(0,1]}$ that evolves from an easy-to-sample prior at $t = 0$ to the target distribution at $t = 1$. In conventional flow-matching, samples are advanced by the deterministic ODE $\mathrm{d}x_t = v_\theta(x_t, t)\,\mathrm{d}t$, whose associated continuity equation $\partial_t p_\theta(x_t) + \nabla\cdot\big(p_\theta(x_t)\, v_\theta(x_t)\big) = 0$ guarantees probability conservation [67]. Any dynamics that satisfy this continuity equation are admissible, and to satisfy the SOC formulation, we adopt the memoryless stochastic variant,

$$\mathrm{d}x_t = 2\, v_\phi(C,\, T,\, x_t,\, t)\,\mathrm{d}t \;-\; \frac{1}{t}\, x_t\,\mathrm{d}t \;+\; \sigma_t(t)\,\sigma_p\,\mathrm{d}W_t, \tag{2}$$

where W_t is a standard Brownian motion and $\sigma_t^2(t) = 2\eta_t$ with $\eta_t = (1 - t + \Delta t)/(t + \Delta t)$ [44].

We then compute the adjoint costate vector $a_t \in \mathbb{R}^{3N}$ that measures how an infinitesimal perturbation of the system state x_t would change the expected terminal reward $\mathbb{E}[R(x_1)]$ as measured by the partial derivative $a_t = \partial\mathbb{E}[R(x_1)]/\partial x_t$. The adjoint therefore carries backward in time the sensitivity information needed to turn local updates of the base velocity field v_θ into an update velocity field v_ϕ that produces a global improvement of the reward objective. Because the underlying diffusion uses the memoryless noise schedule $\sigma_t^2(t) = 2(1 - t)/t$, it has been shown that the adjoint obeys a simpler ordinary differential equation known as the "lean adjoint ODE" [44],

$$\frac{\mathrm{d}a_t}{\mathrm{d}t} \;=\; -\,a_t^\top\,\nabla_x\Big(2\, v_\theta(x_t, t, C, T) - \tfrac{1}{t}x_t\Big), \qquad a_{\,1} = -\nabla_x R(x_{\,1}), \tag{3}$$

which contains only the deterministic drift and a terminal condition given by the negative reward gradient. Operationally within FLOWBACK-ADJOINT, we evaluate the required Jacobian-vector products using automatic differentiation and integrate this equation backwards over the same 100 explicit Euler steps used in the forward simulation.

Having numerically solved the lean adjoint ODE for a_t, we now steer the flow v_θ toward the reward-weighted density $p_\phi(x) = p_\theta(x)e^{R(x)}$ under a time-dependent control force $u_t(x)$, which, under the dynamics of Eqn. 2, becomes $u_t(x) = -\big(\sigma_t(t)^2/2\big)\, a_t$ [49]. The reward-adjusted velocity therefore becomes,

$$v_\phi(x_t, t, C, T) \;=\; v_\theta(x_t, t, C, T) \;-\; \frac{\sigma_t(t)^2}{2}\, a_t.$$

Re-arranging gives a residual that should vanish under perfect control,

$$\frac{2}{\sigma_t(t)}\big(v_\phi - v_\theta\big) \;+\; \sigma_t(t)\, a_t \;=\; 0,$$

which informs the adjoint matching loss–scaled by a factor σ_p^2 for consistency with the flow-matching integration,

$$\mathcal{L}_{\text{adjoint}} \;=\; \sigma_p^2 \sum_t \left\| \frac{2}{\sigma_t(t)}\big(v_\phi(x_t, t, C, T) - v_\theta(x_t, t, C, T)\big) \;+\; \sigma_t(t)\, a_t \right\|^2, \tag{4}$$

defined as the sum of squared residuals over all time steps of the flow. Operationally, we minimize $\mathcal{L}_{\text{adjoint}}$ over the trainable parameters ϕ of the EGNN encoding the flow v_ϕ using the Adam optimizer [60].

Until now, we have left the form of the reward function unspecified. In principle, any differentiable function of x in Eqn. 3 is permissible. In this work, we choose $R(x) = -\lambda U(x)$, where λ is a hyperparameter controlling the degree to which the reward function modulates the base distribution and $U(x)$ is the energy assigned to an AA configuration x by the CHARMM27 molecular mechanics

force field [56]. This choice of reward function can be conceived as promoting a reweighted distribution favoring low-energy configurations by penalizing those with high values of $U(x)$. In empirical tests, we have found that setting λ = 0.01 provides a good balance between training stability and efficient modulation of the backmapped distribution towards low-energy AA configurations.

FLOWBACK-ADJOINT Training. Adjoint matching was performed over all-atom trajectories of nine fast-folding mini-proteins in water conducted by D.E. Shaw Research (DESRES) [50]:

1. BBA - 28-residue $\beta\beta\alpha$ miniprotein (PDB ID: `1FME`)
2. BBL - 47 residue ultrafast protein folder (PDB ID: `2WXC`)
3. λ-repressor fragment - 80-residue switch protein fragment (PDB ID: `1LMB`)
4. NTL9 - 39-residue N-terminal domain of ribosomal protein L9 (K12M mutation, PDB ID: `2HBA`)
5. Chignolin - 10-residue synthetic mini-protein (PDB ID: `1UAO`)
6. Trp-cage (TC5b) - 20-residue α/π miniprotein (PDB ID: `2JOF`)
7. Protein G - 56-residue fast-folding GB1 variant (PDB ID: `1IGD`)
8. UVF - 52-residue homeodomain variant (PDB ID: `2P6J`)
9. α3D - 73-residue de novo three-helix bundle (PDB ID: `2A3D`).

For each trajectory frame, training occurred according to Algorithm 1. we generated one output sample using an Euler integrator with 100 integration steps through the reference trajectory. Weights were loaded from the pre-trained FLOWBACK model. The coarse-grained prior noise was fixed at 0.003 nm. Energies are computed in the gas phase: each structure is converted to a GROMACS topology with `pdb2gmx` using CHARMM27 [56, 68], then the topology is loaded into OpenMM [69]. To ensure consistency with the generator, we exclude hydrogens and termini from the energy model while redistributing their partial charges to the bonded heavy atoms; bonded and nonbonded terms are evaluated only over the atoms produced by the model. The adjoint matching objective was optimized using the Adam optimizer with a learning rate of 10^{-5}. Gradient accumulation was used across 16 mini-batches and energy gradients were clipped at a a maximum gradient norm of 20 nm/timestep. In the selection of $\mathcal{T}$ during the computation of the adjoint matching loss, using a similar protocol to Domingo-Enrich et al. [44], we included all timesteps $t > 0.8$, as well as 10 values of $t \leq 0.8$ selected form a uniform distribution. The regularization parameter for the adjoint loss was set to $\lambda = 0.01$, and the CHARMM27 molecular mechanics force field MacKerell Jr. et al. [56] used to evaluate energies and forces. We scaled GPU memory usage to 24 GB to fit within hardware constraints by controlling the accumulating batch size during the calculation of $\mathcal{L}_{\text{adjoint}}$.

Inference. The coordinates x are successively advanced by Euler steps $x \leftarrow x + v_\phi \Delta t$. The final frame x_1 resulting from the Euler integration is anchored via a fixed offset relative to the conditioned parent C_α coordinates. By virtue of the physics-based auxiliary corrections within v_θ due to FLOWBACK+LJ/BONDS and the energy-based corrections incorporated by adjoint matching due to FLOWBACK-ADJOINT, the terminal backmapped AA configurations are anticipated to be stereochemically correct (i.e., L-form), possess equilibrium bond lengths, contain very few clashes, and be significantly lower in energy relative to the FLOWBACK output. Pseudo-code detailing the inference procedure is provided in Algorithm 2.

A.2 Evaluation Metrics

• **Test 1 – Bond Score (↑).** We gauge physical plausibility of the covalent bond network by computing the fraction of bonds whose lengths are within 10% of the amino-acid reference. A 100% bond quality score is ideal.

• **Test 2 – Clash Score (↓).** We quantify steric clashes as the share of residues that sit within 1.2 Å of any other residue. A clash score of 0% is ideal.

• **Test 3 – Diversity Score (↓).** This metric evaluates how well the all-atom (AA) reference structure is represented within the ensemble of AA structures generated from a single C_α trace, and is defined as $DIV = 1 - RMSD_{\text{gen}}/RMSD_{\text{ref}}$, where $RMSD_{\text{gen}}$ is the mean pairwise heavy atom

Algorithm 1 FLOWBACK-ADJOINT Training Loop

input Base model v_θ, fine-tuned model v_ϕ, C_α coordinates $\{C\}$, protein topologies $\{T\}$, prior noise σ_p, number of Euler steps N

for each training step **do**

Sample random training configuration $i \sim \mathcal{U}\{0, |C|\}$

Sample initial coordinates of heavy atoms from prior distribution $x_0 \sim \mathcal{N}(0, \sigma_p^2 I)$

Compute the memoryless noise schedule regularized to avoid the divergence at $t \to 0$

$$\sigma_t^2(t) = \frac{2(1 - t + \Delta t)}{t + \Delta t}$$

Simulate the memoryless trajectory $\{x_t\}_{t=0}^{1}$ in time-steps of $\Delta t = 1/N$ from $t = 0$ to $t = 1$ using Euler's method

$$x_{t+\Delta t} = \left[2v_\phi(C_i, T_i, x_t, t) - \tfrac{1}{t} x_t\right]\Delta t + \sigma_t(t)\sigma_p(t)\sqrt{\Delta t}\,\epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I).$$

Compute negative energy gradient (i.e., atomic forces) from molecular mechanics force field $a_1 = -\nabla_x U(x_1)$

Solve backward in time using Euler's method to obtain $\{a_t\}_{t=0}^{1-\Delta t}$

$$a_{t-\Delta t} = a_t + (\Delta t) a_t^T \nabla_x (2v_\theta(x_t, t) - \frac{1}{t} x_t)$$

Select a subset of discrete time points $\mathcal{T} \subseteq \{0, \Delta t, 2\Delta t, \ldots, 1\}$

Compute loss

$$\mathcal{L}_{\text{adjoint}} = \sigma_p^2 \sum_{t \in \mathcal{T}} \left| \frac{2}{\sigma_t(t)} \left(v_\phi(C_i, T_i, x_t, t) - v_\theta(C_i, T_i, x_t, t)\right) + \sigma_t(t)\, a_t \right|^2$$

Take gradient step with respect to parameters of v_ϕ

end for

Algorithm 2 FLOWBACK-ADJOINT Inference Loop

input Fine-tuned model v_ϕ, C_α coordinates $\{C\}$, protein topology $\{T\}$, prior noise σ_p, number of Euler steps N

Sample initial coordinates of heavy atoms from prior distribution $x_0 \sim \mathcal{N}(0, \sigma_p^2 I)$

Set $\Delta t = 1/N$

Set $x_t = x_0$

while $t < 1.0$ **do**

$x_t \leftarrow x_t + v_\phi(x_t, t, C, T)\Delta t$

$t \leftarrow t + \Delta t$

end while

Return trajectory $\{x_0, x_{\Delta t}, x_{2\Delta t} \ldots x_1\}$

root mean squared deviation (RMSD) between an ensemble of G backmapped AA configurations generated from the same C_α trace, and $RMSD_{\text{ref}}$ is the mean heavy atom RMSD of these G configurations from the ground-truth reference configuration [42]. Since CG to AA backmapping is inherently one-to-many, a desirable backmapping algorithm should yield a physically plausible, diverse ensemble that includes the reference conformation among its possibilities. A score of $DIV = 1$ indicates a fully deterministic method (i.e., all generated structures are identical to one another), whereas values of $DIV = 0$ indicates both high diversity of the generated ensemble and good consistency with the reference (i.e., the reference structure lies within the distribution of generated configurations). In general, the generated configurations are anticipated to exhibit a tighter distribution around their own mean than around the reference configuration (i.e., $RMSD_{\text{gen}} < RMSD_{\text{ref}}$ such that DIV generally lies on the $[0, 1]$ interval. However, this inequality is not strict and we do observe minor violations such that $DIV < 0$, indicative of tighter dispersion around the reference than around the generated ensemble mean. For evaluation purposes, we treat small negative values of $-0.1 < DIV < 0$ as operationally equivalent to an ideal score of $DIV = 0$.

• **Test 4 – Energy Divergence (↓).** This test measures how closely the potential energy distribution of the generated ensemble matches that of the reference ensemble. Let p_k^{gen} and p_k^{ref} denote the discretized probability distributions of observing a configuration with a CHARM27 energy in bin k under, respectively, the generated backmapped AA ensemble and the reference data harvested from the DESRES MD trajectory. We quantify the distance of the reference distribution from the generated distribution using the Kullback-Leibler (KL) divergence,

$$KL(p_{\text{ref}} \parallel p_{\text{gen}}) = \sum_{k=1}^{B} p_k^{\text{ref}} \ln \frac{p_k^{\text{ref}}}{p_k^{\text{gen}}},$$

where B is the number of bins. The KL divergence, also known as the relative entropy, is non-negative statistical distance that attains its ideal minimum at $KL = 0$ indicating that the two distributions are identical.

To further probe the influence of the adjoint matching upon the energy distribution, we sought to determine whether the energies could be driven to even lower values by fitting customized single-protein FLOWBACK-ADJOINT models for each of the WW domain and for Protein B and comparing their predictions to the general-purpose model trained over the nine DESRES training trajectories. Training was conducted for the same number of optimization steps used for the nine-trajectory model and the trained models then used to generate in-sample AA configurations for WW domain and for Protein B. The energy distributions for the two single-protein models show no statistically significant reduction in energies relative to the multi-protein model at an α = 0.05 significance threshold under a Kolmogorov-Smirnov test (Figure A.8; p = 0.65 (WW domain) and p = 0.36 (Protein B)). This observation suggests that the multi-protein FLOWBACK-ADJOINT model has achieved a excellent out-of-sample performance in generating low-energy AA configurations that cannot be significantly improved by dedicated training on the specific protein of interest.

• **Test 5 – MD Test (↑).** This test empirically assesses the fraction of backmapped AA configurations that can be used to initialize a stable MD run in OPENMM [69] without additional energy minimization. Commencing from N_{start} backmapped AA configurations, we count N_{fin}, the number of configurations that produce a stable 20 ps gas phase MD trajectory. A success fraction $\Phi = N_{\text{fin}}/N_{\text{start}}$ of 100% is ideal. Details of the simulation protocol are provided in Appendix A.3.

As illustrated in the top panels of Figure A.5, conformations generated by FLOWBACK-ADJOINT exhibit smaller instantaneous force magnitudes during the validation MD run (median maximum atom-wise force = 2820 kcal/nm for WW domain and 3297 kcal/nm for Protein B) relative to FLOWBACK (3729 kcal/nm and 5588 kcal/nm) and smaller or comparable to FLOWBACK+LJ/BONDS (3217 kcal/nm and 3262 kcal/nm). Moreover, a higher proportion of MD simulation trajectories initialized with backmapped AA configurations without any additional energy minimization remain stable over a 20 ps simulation trajectory using FLOWBACK-ADJOINT (82.6% for WW domain and 65.4% for Protein B) relative to FLOWBACK (66.7% and 50.8%) or FLOWBACK+LJ/BONDS (80.9% and 63.8%).

Nevertheless, a substantial gap remains relative to the MD simulation data, which possesses a median maximum atom-wise force = 1644 kcal/nm for WW domain and 2001 kcal/nm for Protein B, and corresponding stable simulation success rates of 99.1% and 99.5%. This residual discrepancy appears not to be a failing of the adjoint matching procedure, but rather a consequence of limitations

in hydrogen atom placement on top of the AA backmapped structures produced by FLOWBACK-ADJOINT, which, like the original FLOWBACK model, lack explicit hydrogen atoms. Analysis of our trajectories reveal that the large magnitude atom-wise forces tend to be associated with clashing or bonded hydrogen interactions, suggesting that the instabilities may be introduced by suboptimal H atom placement using the `gmx pdb2gmx` function [68], and is also the reason the MD runs initialized with MD trajectory frames do not reach a 100% success rate. As illustrated in the bottom panels of Figure A.5, if we neglect these forces from our analysis, the median maximum atom-wise forces drop substantially for all four cases – FLOWBACK: 1557 kcal/nm for WW domain and 1946 kcal/nm for Protein B, FLOWBACK+LJ/BONDS: 1188 kcal/nm and 1264 kcal/nm, and FLOWBACK-ADJOINT: 1159 kcal/nm and 1214 kcal/nm – with both FLOWBACK-ADJOINT and FLOWBACK+LJ/BONDS achieving a value commensurate with the values of 1212 kcal/nm and 1271 kcal/nm computed for the MD data. After deleting the hydrogen atoms and termini, reassigning their partial charges to the adjacent heavy atoms, and rerunning the simulation with these merged groups treated as united atoms that experience exactly the same bonded forces as their original forms, all trajectories initiated from the MD snapshots remained stable to attain a 100 % success rate. This ideal behavior is closely approached by all three models – 96.1% for WW domain and 97.0% for Protein B under FLOWBACK, 99.4% and 99.9% under FLOWBACK+LJ/BONDS, and 99.6% and 99.9% under FLOWBACK-ADJOINT. Improved algorithms for hydrogen atom placement are therefore anticipated to essentially completely close the force and stability gap of FLOWBACK-ADJOINT.

A.3 Simulation Parameters for Molecular Dynamics Stability Test

All-atom MD simulations of the backmapped configurations are initialized by generating from each heavy-atom structure a CHARMM27 topology file using `gmx pdb2gmx` to place hydrogens and employing pH 7 protonation states. The resulting files are converted into an OpenMM simulation input file [69]. The cubic unit cell is padded by 2 nm in each dimension. Electrostatics are evaluated using Particle Mesh Ewald (PME) with a 1.0 nm real-space cutoff. Dynamics are propagated in the canonical (NVT) ensemble at $T = 300$ K using a Langevin integrator with collision frequency $\gamma = 1$ ps^{-1} and time-step 1 fs. All covalent bonds involving hydrogens are constrained and center-of-mass motion is removed every 1 ps. A 100-step pre-run is used to detect immediate pathologies in the initialization (e.g., NaN positions or energies). Upon passing this pre-check, the simulation is continued for 20,000 time steps (20 ps). A trajectory is deemed stable when (i) the simulation finishes without exception, (ii) no NaNs appear in energies or coordinates, and (iii) the largest Cartesian force on any heavy atom during the pre-check remains finite. These simulations are very inexpensive and typically complete in $<$10 s on a single NVIDIA V100 GPU.

To assess the influence of hydrogen-atom placement on the stability of the MD simulation, we also conducted calculations in which the hydrogen atoms and termini were deleted, their partial charges to their neighboring atoms, and simulations conducted treating these groups as united atoms that experience exactly the same bonded forces as their original forms. In this case, the absence of hydrogen atoms permitted us to employ a larger 2 fs time step and conduct 20,000 time steps (40 ps) simulations. All other simulation parameters remained the same.

A.4 BLAST

We confirmed via a BLAST search that none of these proteins shared more than 60% identity with any entry in the SidechainNet training corpus [53] that was used to train the original FLOWBACK model or any of the nine DESRES proteins [50] used to train FLOWBACK-ADJOINT, thereby guarding against any observed improvements over this test set resulting from data leakage or high sequence similarity with sequences in the training set.

1. **Database construction.** All SidechainNet chains were exported to a single FASTA file with the SidechainNet Python API (v1.0.1) and converted into a local protein database using BLAST+ (v2.15.0) [70]. The same process was applied to the 9 DESRES training proteins.

2. **Query preparation.** The primary sequence of each `OOD60` test chain was written to an individual FASTA file, preserving its PDB chain identifier.

3. **BLAST search.** Each query was aligned against the SidechainNet database with default composition-based statistics, soft segmentation, an *E*-value cutoff of 10^{-3}, and retrieval of up to 5000 target sequences.
4. **Filtering criteria.** Alignments were retained only if they satisfied all of the following: (i) alignment length greater than $70\,\%$ of the query chain, (ii) percentage identity $\geq 30\,\%$, and (iii) *E*-value $\leq 10^{-3}$. For each query we recorded the maximum percentage identity observed across all remaining hits.
5. **Sequence identity threshold.** A protein was included in our `OOD60` test set only if its best SidechainNet hit exhibited $< 60\,\%$ sequence identity.

A.5 Supplementary Figures

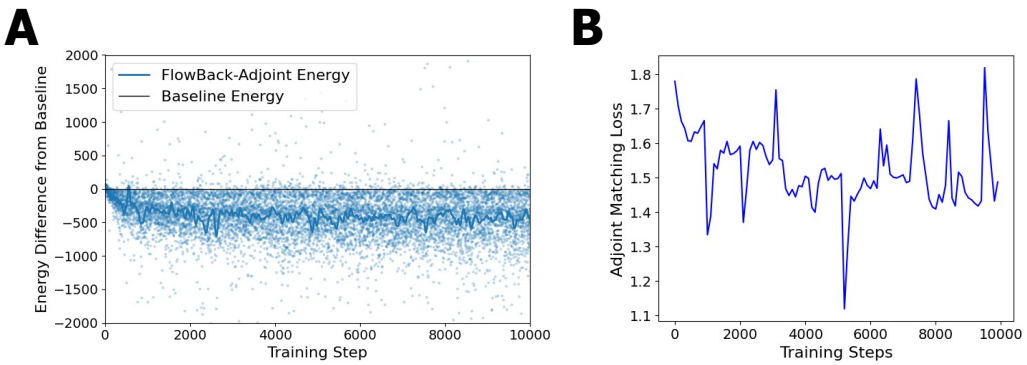

Figure A.1: Fine-tuning of FLOWBACK-ADJOINT. (A) Plot of the FLOWBACK-ADJOINT CHARMM27 potential energy relative to the FLOWBACK+LJ/BONDS baseline. The difference is plotted as blue points and the dark blue line represents a smoothed average using a Gaussian kernel. (B) The average adjoint matching loss recorded at 100 training step intervals. Several checkpoints are taken for validation against the WW Domain and Protein B. The checkpoint at step 7000 is used as the final model.

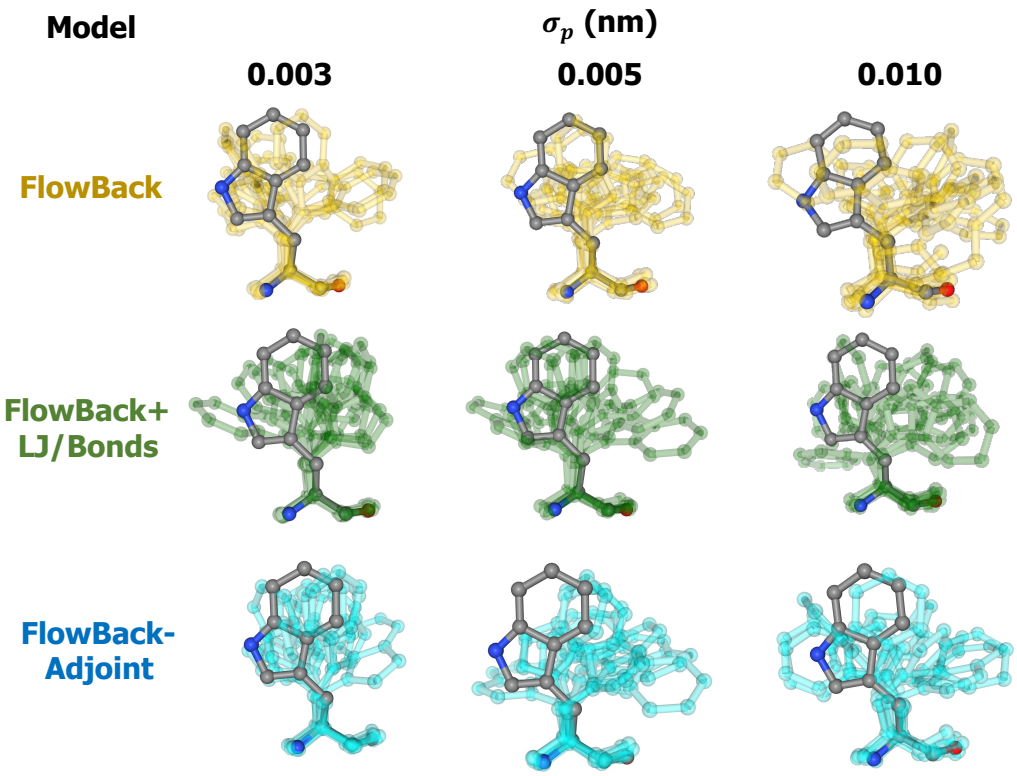

Figure A.2: Visualization of ensembles of 10 AA reconstructions produced by FLOWBACK, FLOWBACK+LJ/BONDS, and FLOWBACK-ADJOINT at choices of the prior noise σ_p = 0.003 nm, 0.005 nm, and 0.010 nm around the Trp8 residue of Protein B. Higher values of σ_p produce more diverse structural ensembles, but whereas this results in a significant degradation of the structural accuracy of the configurations produced by FLOWBACK and FLOWBACK+LJ/BONDS – including distortions in the aromatic tryptophan rings and other unrealistic elements of the chemical structure – FLOWBACK-ADJOINT breaks this trade-off and can achieve high diversity ensembles that maintain high structural accuracy in terms of bond and clash scores even at high noise values (cf. Table 1).

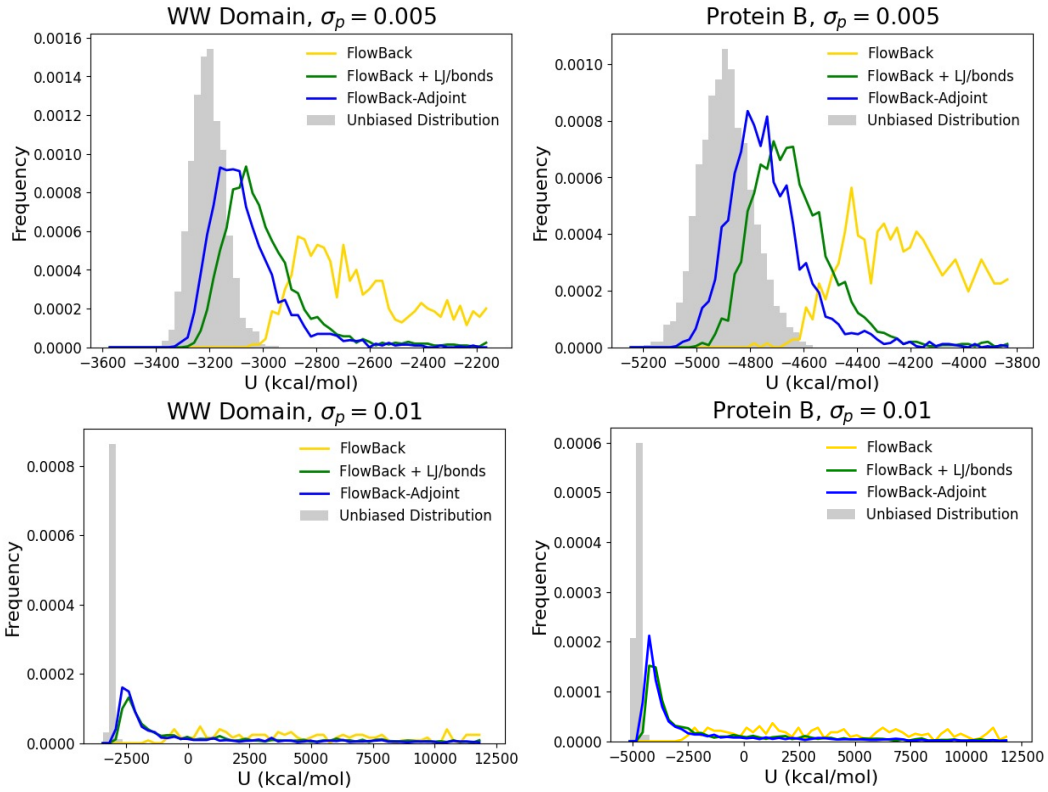

Figure A.3: Distribution of potential energies over an ensemble of 2230 (WW domain) or 2602 (Protein B) backmapped AA configurations from FLOWBACK (yellow), FLOWBACK+LJ/BONDS (green), and FLOWBACK-ADJOINT (blue) at noise values of $\sigma_p = 0.003$ nm (top row) and 0.010 nm (bottom row) for the two hold-out DESRES trajectories of WW domain (left column) and Protein B (right column) [50].

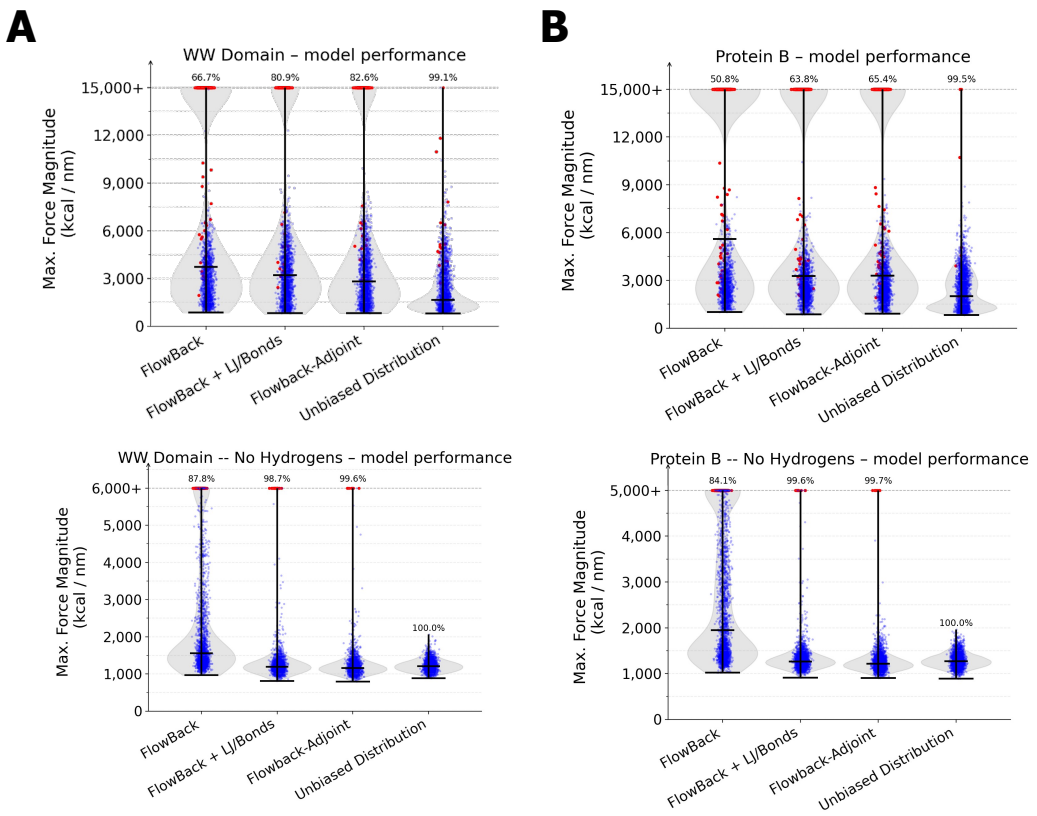

Figure A.4: Swarm and violin plots of the distribution of the maximum force magnitudes experienced by any atom in over the course of a short MD simulation initialized from a backmapped AA configuration using FLOWBACK, FLOWBACK+LJ/BONDS, FLOWBACK-ADJOINT, or harvested from the MD simulation trajectory. All models employ a noise level of σ_p = 0.003 nm. In the top plots, all atoms are included in force and energy calculations. In the bottom plots, hydrogen atoms and termini were deleted, their partial charges to their neighboring atoms, and simulations conducted treating these groups as united atoms that experience the same bonded forces as their original forms. Statistics for each model are aggregated over 2230 (WW domain) or 2602 (Protein B) initial AA configurations and the maximum force identified on any atom in first frame of the molecular-dynamics simulation (i.e. the model-generated structure). For clarity of exposition, large forces in excess of 15,000 kcal/nm (top) or 6,000 kcal/nm (bottom) are collapsed together at the top of the plot. Test runs in which the MD simulation remained stable over the course of the run (i.e. the simulation doesn't crash due to energy, position, or velocity values exceeding floating point limits) have their corresponding points in the swarm colored blue and those which destabilized are colored red. The lowest and median maximum atom-wise force in each swarm plot are indicated by horizontal lines and the overall success rate of stable MD simulation runs is noted as a percentage at the top of each violin.

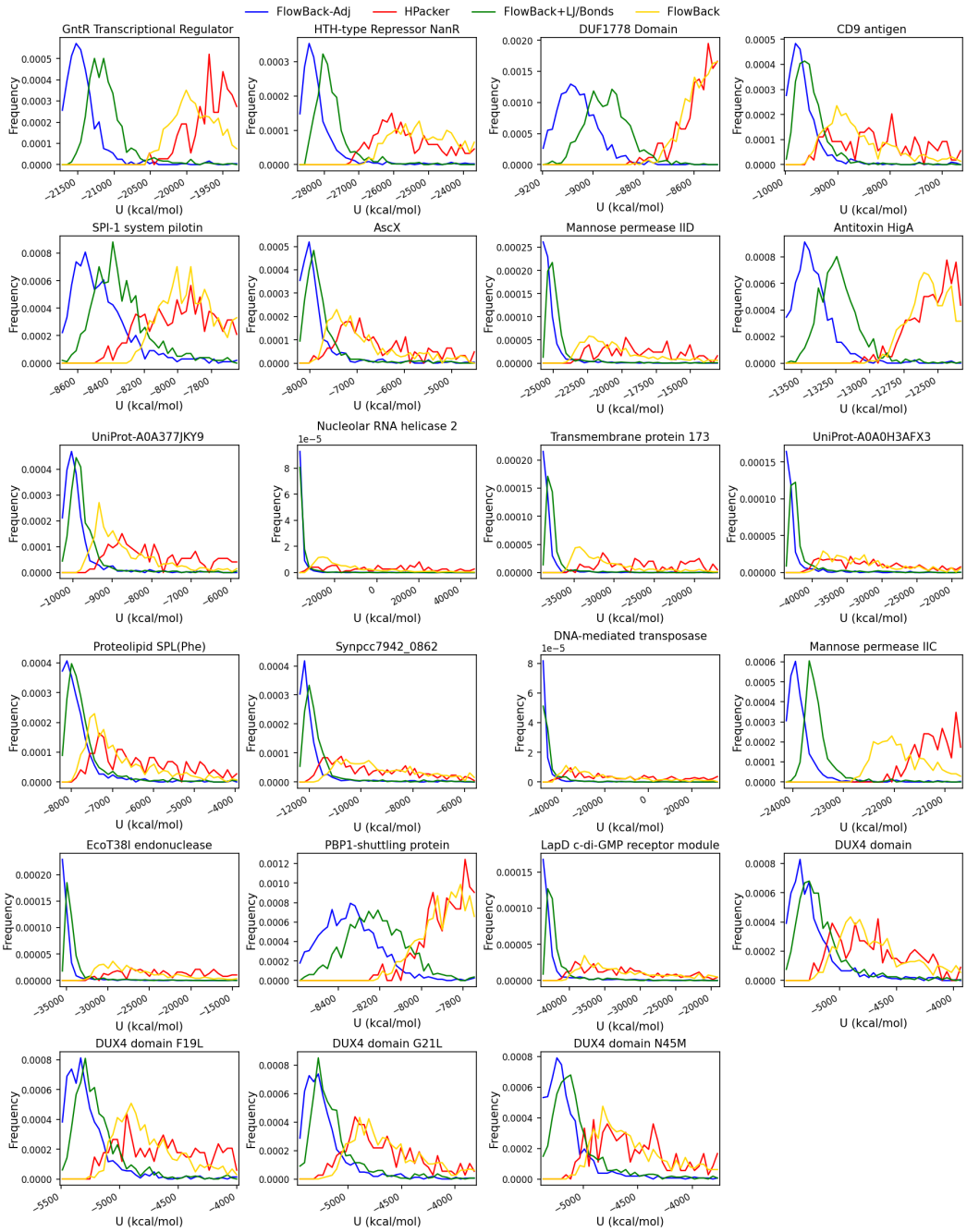

Figure A.5: Potential energy distributions for the AA configurations generated by FLOWBACK-ADJOINT, FLOWBACK+LJ/BONDS, FLOWBACK, and HPACKER from 1000 backbone-only BIOEMU structures generated for the 23 test proteins. In all cases, FLOWBACK-ADJOINT produces lower-energy (i.e., more stable) distributions.

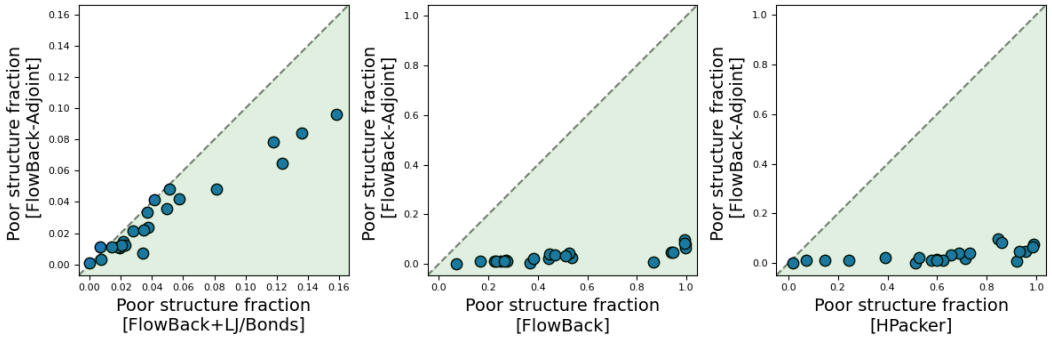

Figure A.6: Fractions of "poor structures" in each of the proteins in our out-of-distribution test. A "poor structure" is defined as one with an energy over 10,000 kcal/mol higher than the median FLOWBACK-ADOINT energy for a given protein, indicating a serious distortion in one or more of the backmapped residues. FLOWBACK-ADJOINT produced fewer poor structures than any other model for any protein, except for 3/23 proteins for which FLOWBACK+LJ/BONDS produced an equal number or fewer poor structures than FLOWBACK-ADJOINT. Both FLOWBACK and HPACKER have several proteins for which more than half of the structures are deemed poor.

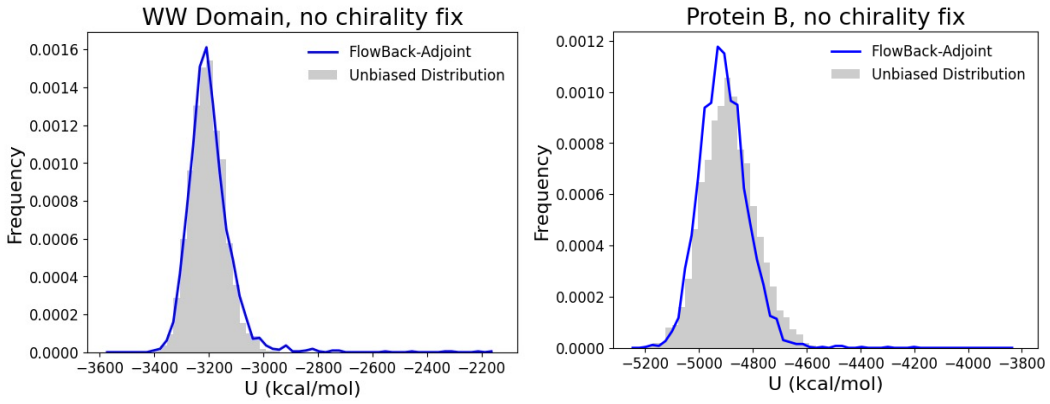

Figure A.7: Distributions of potential energies over an ensemble of 2230 AA configurations of the WW domain and 2602 configurations of Protein B generated FLOWBACK-ADJOINT at a noise level of $\sigma_p = 0.003$ nm in the absence of the chirality correction. The blue distribution corresponds to that generated by the general FLOWBACK-ADJOINT model trained over nine DESRES protein trajectories, which did not include WW domain or Protein B, such that the predictions therefore represent an out-of-sample test. The gray distribution represents the energy distribution over the DESRES MD simulation trajectories of WW domain and Protein B. Eliminating the chirality correction results in 1-3% of residues with the incorrect D-form chirality, but the energy distributions show far better agreement than when this correction is active with KL divergences falling from $0.0018 \rightarrow 0.0004$ for WW domain and $0.0017 \rightarrow 0.0004$ for Protein B, indicating almost perfect matching.

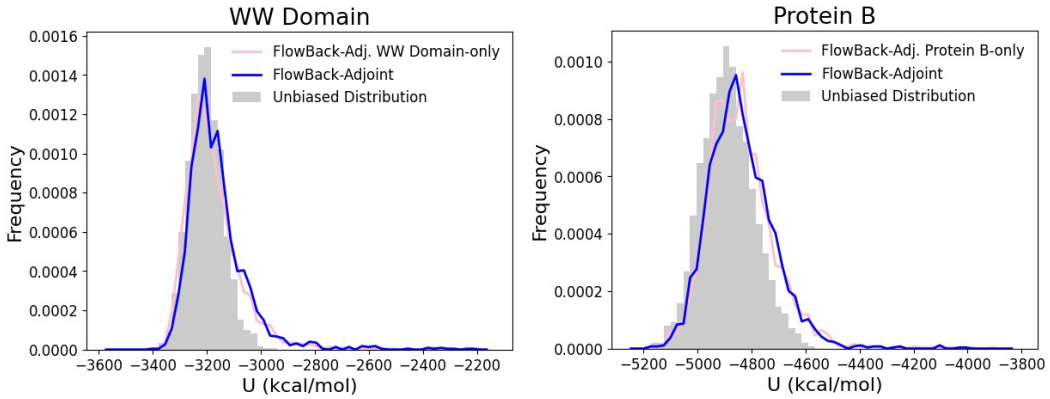

Figure A.8: Distributions of potential energies over an ensemble of 2230 AA configurations of the WW domain and 2602 configurations of Protein B generated by FLOWBACK-ADJOINT at a noise level of σ_p = 0.003 nm. The blue distribution corresponds to that generated by the general FLOWBACK-ADJOINT model trained over nine DESRES protein trajectories, which did not include WW domain or Protein B, such that the predictions therefore represent an out-of-sample test. The pink distribution corresponds to that generated by the FLOWBACK-ADJOINT model trained on either WW domain or Protein B and then used to generate in-sample predictions. The gray distribution represents the energy distribution over the DESRES MD simulation trajectories of WW domain and Protein B.