A131 · Understanding and Correcting Representation-Specific Failure Modes in SE(3) Flow-Matching Protein Backbone Generation

Presenter: Michael Widener — Northeastern University / Boston College

Authors: Michael 'Xander' Widener, Shantanu Jain, Predrag Radivojac

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Generative models for protein backbones typically commit to one geometric representation: rigid SE(3) frames per residue via globally consistent poses or internal torsion angles reconstructed via NeRF. The former provides globally well-posed structure but lacks explicit chemical bond constraints, while the latter yields chemically valid bonds by construction but suffers from recursive error propagation.

We quantify this tradeoff directly on a sub-quadratic, O(L·K)-attention SE(3) Riemannian flow-matching backbone generator. FATCAT's hinge-tolerant structural alignment shows that torsion-generated chains need on average 1.91 rigid-body 'twists' to align well against 0.61 for frame-generated chains. A controlled noise-injection sweep shows dihedral error compounds superlinearly along the chain (growth exponent ≈1.0 – 1.3) while bond-length error does not (≈0.65 – 0.7) – a mechanistic account of why each representation fails the way it does.

We then tested five mechanisms for combining strengths of the two representations. Two of five toy-scale positives (a model-capacity increase; an anchor-constrained hinge-inpainting scheme) vanished completely at larger scales. This underscores that the inherent ease with which torsion-based models encode secondary structure at small scales often fails to translate to larger, data-rich architectures. The strongest surviving mechanism, a differentiable Lie-Trotter operator-splitting scheme that alternates one frame-space and one torsion-space integration step and is fine-tuned end-to-end via short bridge-supervised windows, is the first to outperform both single-representation baselines outright (TM-score 0.41 vs. 0.34/0.34) with no exposed tradeoff. Production-scale verification of this result is in progress.