SUPRATITO: TRANSFERABLE GENERATIVE MOLECULAR DYNAMICS FOR SUPRAMOLECULAR SYSTEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Peptide sequence governs the structures and dynamics that emerge during supramolecular assembly, but predicting them requires resolving slow collective processes among many interacting molecules. Molecular dynamics (MD) provides microscopic insight, yet long assembly times and the vast peptide sequence space make systematic exploration costly. We introduce SupraTITO, a transferable generative molecular dynamics (GenMD) framework for supramolecular dynamics, demonstrated for peptide self-assembly. SupraTITO learns transferable implicit transfer operators (TITO) conditioned on peptide sequence, molecular topology, and periodic geometry, allowing configurations to be propagated over intervals much longer than an MD integration step. On a comprehensive dipeptide benchmark, SupraTITO generalizes to held-out sequences, reproduces sequence-dependent structures and dynamics, and maintains molecular integrity over long rollouts. Compared with direct ensemble prediction trained on the same trajectory data, it produces more accurate assembly structures while also recovering their temporal evolution. SupraTITO further generalizes across peptide concentrations, including dilute conditions absent from training. These results extend transferable GenMD to collective dynamics in periodic supramolecular systems and related processes beyond peptide assembly.

1 INTRODUCTION

Predicting how amino-acid sequence determines molecular structure and function is a central problem in molecular science. Deep learning has transformed static structure prediction (Jumper et al., 2021; Baek et al., 2021; Abramson et al., 2024), but many biological and materials processes depend on conformational changes and collective interactions among multiple molecules (Lehn, 2002; Whitesides & Grzybowski, 2002). Peptide assembly is a prominent example of this broader supramolecular prediction problem (Yang et al., 2026): peptide sequences can encode fibers, nanotubes, sheets, and other morphologies (Reches & Gazit, 2003), while concentration, solvent, temperature, and other physicochemical conditions can redirect the assembly process (Kaygisiz et al., 2025). Molecular dynamics (MD) provides microscopic access to these processes, but assembly involves collective rearrangements that are slow relative to the MD integration step. Even with coarse-grained force fields such as MARTINI (Marrink et al., 2007), exploring many sequences and conditions requires substantial simulation effort (Frederix et al., 2015; Shmilovich et al., 2020).

Generative molecular dynamics (GenMD) addresses this by learning probabilistic surrogates of molecular simulations (Olsson, 2026). Whereas equilibrium generative models sample stationary molecular ensembles (Noé et al., 2019; Zheng et al., 2024; Lewis et al., 2025; Chen et al., 2026), dynamical methods learn molecular evolution over finite physical intervals (Schreiner et al., 2023; Klein et al., 2023; Vlachas et al., 2022; Costa et al., 2024). In particular, implicit transfer operator (ITO) learning parameterizes the finite-lag transition density $p_\tau(\mathbf{x}_{t+\tau} \mid \mathbf{x}_t)$ with a conditional generative model (Schreiner et al., 2023). Repeated sampling generates stochastic trajectories without resolving every intermediate MD step, while shared representations can enable transfer to systems excluded from training (Diez et al., 2026; Antoniadis et al., 2026). Existing dynamical GenMD methods, however, have primarily addressed individual molecules. Supramolecular dynamics additionally requires coordinating many interacting molecules while respecting molecular topology and periodic boundary conditions throughout long recursive rollouts.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

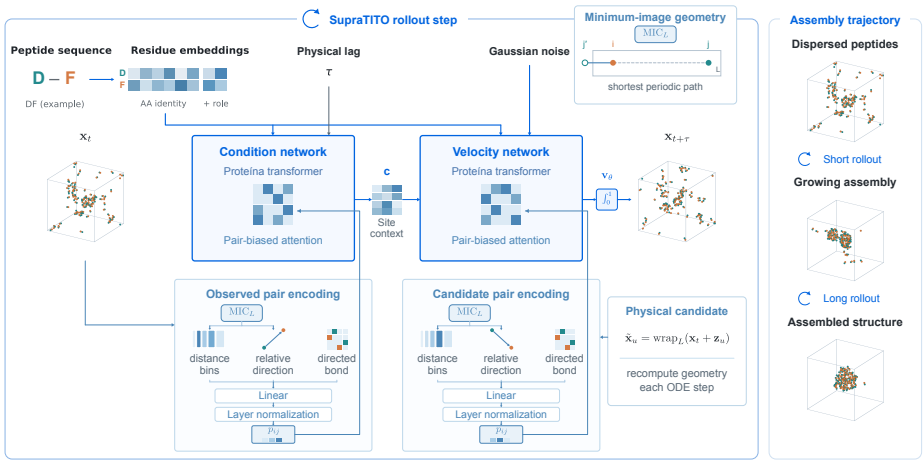


Figure 1: **SupraTITO architecture and generation of peptide assembly trajectories.** Sequence and topology features condition a pair-aware velocity network that generates periodic molecular displacements. Pair geometry is evaluated from the observed configuration and updated throughout flow integration. Repeated finite-lag transitions generate a stochastic assembly trajectory.

We introduce SupraTITO, a GenMD framework for **Supra**molecular dynamics with **T**ransferable **I**mplicit **T**ransfer **O**perators (figure 1). SupraTITO learns finite-lag transition densities conditioned on peptide sequence, molecular topology, and periodic geometry. On a comprehensive dipeptide benchmark, it generalizes to held-out sequences, reproduces sequence-dependent structures and dynamics, and preserves molecular integrity over long rollouts. Compared with direct ensemble prediction trained on the same trajectories, SupraTITO predicts assembly structures more accurately while additionally recovering their temporal evolution. Its learned dynamics also transfer across peptide concentrations, including lower concentrations where larger simulation volumes and longer molecular encounter times can make direct MD particularly costly. Although demonstrated for peptide assembly, the formulation applies to other periodic supramolecular systems.

Our main contributions are:

- **A GenMD framework for supramolecular dynamics.** SupraTITO extends GenMD beyond single-molecule conformational dynamics to collective processes. The formulation can, in principle, support related problems such as coassembly and other supramolecular processes, including micelles, self-assembled monolayers, phospholipid bilayers, viral capsids, and DNA origami.
- **Implicit transfer operators under periodic boundary conditions.** We extend ITO learning to periodic molecular systems using molecule-consistent minimum-image displacement targets, wrapped trajectory generation, and relative periodic geometry.
- **Evaluation of transfer across sequences and environments.** We systematically assess generalization across molecular sequences and physical conditions using structural and dynamical observables, intramolecular fidelity, direct ensemble prediction and training-sequence coverage experiments.

2 Background and Preliminaries

2.1 Molecular Dynamics and Transfer Operators

Molecular dynamics (MD) integrates equations of motion under a potential energy model and specified simulation conditions (Frenkel et al., 1997). Let $\mathbf{x}_t = (\mathbf{x}_{t,1}, \ldots, \mathbf{x}_{t,N}) \in \Omega \subseteq \mathbb{R}^{3N}$ denote the configuration of N retained particles at physical time t.

For time-homogeneous dynamics, the transition density $p_\tau(\mathbf{x}' \mid \mathbf{x})$ describes the distribution of configurations reached after a physical lag τ, conditioned on the current configuration $\mathbf{x}$. It induces

the transfer operator $\mathcal{P}_\tau$:

$$\rho_{t+\tau}(\mathbf{x}') = (\mathcal{P}_\tau \rho_t)(\mathbf{x}') = \int_\Omega p_\tau(\mathbf{x}' \mid \mathbf{x})\, \rho_t(\mathbf{x})\, d\mathbf{x}. \tag{1}$$

Here, ρ_t and $\rho_{t+\tau}$ denote the distributions of configurations at times t and $t+\tau$, respectively. Whereas the transition density describes evolution from a given configuration, the transfer operator propagates an ensemble (Prinz et al., 2011; Nuske et al., 2014).

For observables $f, g : \Omega \to \mathbb{R}$, structural properties are characterized by averages $\langle f \rangle_t = \mathbb{E}_{\mathbf{x}\sim\rho_t}[f(\mathbf{x})]$, while two-time correlations $C_{f,g}(t,\tau) = \mathbb{E}[f(\mathbf{x}_t)g(\mathbf{x}_{t+\tau})]$ describe temporal dependence. The transition density therefore determines both the evolution of configurational statistics and the temporal correlations of molecular observables.

2.2 Conditional Flow Matching

Let p_0 be a tractable reference distribution on $\mathbb{R}^d$ and p_1 a target distribution. A continuous normalizing flow transports samples from p_0 to p_1 through the ordinary differential equation (Chen et al., 2018)

$$\frac{d\mathbf{z}_u}{du} = \mathbf{v}_\theta(\mathbf{z}_u, u), \qquad \mathbf{z}_0 \sim p_0, \qquad u \in [0,1]. \tag{2}$$

Here, u denotes generative flow time, distinct from physical time t and lag τ.

Conditional flow matching trains the vector field without integrating this ODE (Lipman et al., 2022; Liu et al., 2022; Albergo & Vanden-Eijnden, 2023; Tong et al., 2023). For independent samples $\mathbf{z}_0 \sim p_0$, $\mathbf{z}_1 \sim p_1$, and $u \sim \mathcal{U}(0,1)$, the linear path

$$\mathbf{z}_u = (1-u)\mathbf{z}_0 + u\mathbf{z}_1 \tag{3}$$

has conditional velocity $\mathbf{z}_1 - \mathbf{z}_0$. The training objective is

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{\mathbf{z}_0,\mathbf{z}_1,u}\left[\|\mathbf{v}_\theta(\mathbf{z}_u, u) - (\mathbf{z}_1 - \mathbf{z}_0)\|_2^2\right]. \tag{4}$$

At inference, samples are generated by integrating equation (2) from $u = 0$ to $u = 1$.

2.3 Implicit Transfer Operator Learning

Implicit transfer operator (ITO) learning approximates finite-lag molecular transition densities with conditional generative models (Schreiner et al., 2023):

$$p_\theta(\mathbf{x}' \mid \mathbf{x}, \tau) \approx p_\tau(\mathbf{x}' \mid \mathbf{x}). \tag{5}$$

The operator is represented through a sampler of future configurations rather than an explicit transition matrix or basis expansion (Prinz et al., 2011; Nuske et al., 2014; Husic & Pande, 2018).

Training uses MD tuples $(\mathbf{x}_t, \mathbf{x}_{t+\tau}, \tau)$, with the current configuration and lag providing the conditions and the future configuration providing the target. Sampling different lags allows one parameterization to represent a family of finite-time transition densities. Starting from an initial configuration, recursive sampling generates a trajectory:

$$\widehat{\mathbf{x}}_{(k+1)\tau} \sim p_\theta(\cdot \mid \widehat{\mathbf{x}}_{k\tau}, \tau), \qquad k = 0, 1, \ldots. \tag{6}$$

Each sample advances physical time by τ. This recursion treats the retained state as Markovian at the selected lag, although position-only and coarse-grained representations can retain memory through unresolved degrees of freedom (Mori, 1965; Klippenstein et al., 2021). Consequently, consistency between a direct long-lag transition and a composition of shorter transitions is not guaranteed and should be evaluated empirically (Christians et al., 2026).

3 Related Work

Dynamical GenMD methods generate molecular transitions or trajectory segments (Costa et al., 2024; Jing et al., 2024), while TITO and PLaTITO extend implicit transfer operator learning across molecular systems (Diez et al., 2026; Antoniadis et al., 2026). SupraTITO extends this approach

to collective dynamics in periodic many-molecule systems, accounting for molecular connectivity and changing intermolecular environments. Machine learning for peptide assembly has primarily focused on predicting assembly propensity and designing sequences (Batra et al., 2022; Shmilovich et al., 2020; Njirjak et al., 2024; Costa & Zavadlav, 2026). SupraTITO complements these approaches by generating assembly trajectories, providing both structural and dynamical information. An extended discussion is provided in appendix A.

4 SupraTITO: Generative Dynamics for Supramolecular Systems

Peptide assembly couples fixed molecular attributes, including sequence and covalent topology, with evolving intermolecular organization. SupraTITO models the distribution of molecular displacements over finite physical intervals using a transformer architecture that integrates molecular features with periodic geometry.

4.1 Implicit Transfer Operators under Periodic Boundary Conditions

Consider a periodic system containing M peptide molecules, with solvent degrees of freedom excluded from the retained representation. Molecule m has amino-acid sequence

$$s^{(m)} = \left(a_1^{(m)}, \ldots, a_{R_m}^{(m)}\right) \in \mathcal{A}^{R_m}, \tag{7}$$

where $\mathcal{A}$ is the amino-acid alphabet, $a_r^{(m)}$ is the identity of residue r, and R_m is the sequence length. The retained representation contains N sites in total, with wrapped coordinates $\mathbf{x}_t \in \mathbb{R}^{N\times 3}$, molecular graph $\mathcal{G}$, and orthorhombic box lengths $\mathbf{L} = (L_x, L_y, L_z)$. Peptide self-assembly is the homogeneous case in which all molecules share the same sequence.

SupraTITO generates molecular displacements rather than absolute future coordinates. Applying the minimum-image convention independently to bonded sites can select different periodic images, causing them to separate during flow interpolation. We therefore decompose each molecular displacement into a common translation and an internal change.

For the two-site per molecule representation used here, molecule m contains sites $2m-1$ and $2m$. Let $\Delta\mathbf{g}_{t,\tau}^{(m)}$ denote the minimum-image displacement of its geometric center and $\Delta\mathbf{b}_{t,\tau}^{(m)}$ the change in its periodic bond vector. The site displacements are

$$\mathbf{d}_{t,\tau,2m-1} = \Delta\mathbf{g}_{t,\tau}^{(m)} - \frac{1}{2}\Delta\mathbf{b}_{t,\tau}^{(m)}, \qquad \mathbf{d}_{t,\tau,2m} = \Delta\mathbf{g}_{t,\tau}^{(m)} + \frac{1}{2}\Delta\mathbf{b}_{t,\tau}^{(m)}. \tag{8}$$

Collecting these displacements gives $\mathbf{d}_{t,\tau} \in \mathbb{R}^{N\times 3}$. This construction exactly reconstructs the wrapped target while assigning bonded sites a common molecular translation (appendix D.1).

Let $\mathcal{S} = \{s^{(1)}, \ldots, s^{(M)}\}$ denote the molecular sequences. SupraTITO approximates

$$p_\theta(\mathbf{d}_{t,\tau} \mid \mathbf{x}_t, \mathcal{S}, \mathcal{G}, \tau, \mathbf{L}) \simeq p_\tau(\mathbf{d}_{t,\tau} \mid \mathbf{x}_t, \mathcal{S}, \mathcal{G}, \mathbf{L}). \tag{9}$$

A generated displacement is mapped back to periodic configuration space as

$$\widehat{\mathbf{x}}_{t+\tau} = \mathrm{wrap}_{\mathbf{L}}\left(\mathbf{x}_t + \widehat{\mathbf{d}}_{t,\tau}\right). \tag{10}$$

One parameterization represents finite-lag transfer operators across sequences and lags.

4.2 Molecular Representation and Neural Architecture

SupraTITO uses separate condition and velocity networks adapted from the Proteína transformer architecture (Geffner et al., 2025; Antoniadis et al., 2026). We extend this architecture to periodic many-molecule systems by introducing directed molecular topology, minimum-image pair geometry, and relative coordinate features. The condition network encodes the observed configuration, while the velocity network generates its finite-lag displacement.

Node features. Each retained site i combines learned embeddings of its amino-acid type a_i and molecular role r_i: $\mathbf{f}_i^{\mathrm{node}} = \mathbf{e}_{\mathrm{aa}}(a_i) \,\|\, \mathbf{e}_{\mathrm{role}}(r_i)$. Roles distinguish terminal and internal residues, including N-terminal and C-terminal sites in the dipeptide representation. Peptides are represented through their constituent site types, roles, and connectivity rather than a categorical sequence label.

Pair features. Each ordered pair (i, j) has a minimum-image vector, distance, and direction:

$$\mathbf{r}_{ij}(\mathbf{x}) = \mathrm{MIC}_{\mathbf{L}}(\mathbf{x}_i - \mathbf{x}_j), \quad d_{ij}(\mathbf{x}) = \|\mathbf{r}_{ij}(\mathbf{x})\|_2, \quad \mathrm{dir}_{ij}(\mathbf{x}) = \begin{cases} \mathbf{r}_{ij}(\mathbf{x})/d_{ij}(\mathbf{x}), & d_{ij}(\mathbf{x}) > 0, \\ \mathbf{0}, & d_{ij}(\mathbf{x}) = 0. \end{cases} \tag{11}$$

A directed bond category b_{ij} distinguishes nonbonded pairs and the two directions along an intramolecular bond. The complete pair feature is

$$\mathbf{f}_{ij}^{\mathrm{pair}}(\mathbf{x}) = \mathrm{onehot}(\mathrm{bin}(d_{ij}(\mathbf{x}))) \,\|\, \mathrm{dir}_{ij}(\mathbf{x}) \,\|\, \mathrm{onehot}(b_{ij}). \tag{12}$$

Nonuniform distance categories provide greater resolution for bonded and short-range intermolecular geometry while retaining coverage of more distant pairs. A learned projection followed by layer normalization (LN) gives $\mathbf{p}_{ij}(\mathbf{x}) = \mathrm{LN}(W_{\mathrm{pair}}\mathbf{f}_{ij}^{\mathrm{pair}}(\mathbf{x}))$. The complete pair representation is $\mathbf{p}(\mathbf{x}) = (\mathbf{p}_{ij}(\mathbf{x}))_{i,j=1}^{N}$.

Condition and velocity networks. The condition network combines a sinusoidal lag embedding with molecular features as $\boldsymbol{\eta}_i^{\mathrm{cond}} = W_{\mathrm{cond}}[\mathbf{e}_\tau(\tau) \,\|\, \mathbf{f}_i^{\mathrm{node}}]$. It jointly processes all site inputs and pair representations to produce one context vector per site:

$$\mathbf{c}_{t,\tau} = (\mathbf{c}_{t,\tau,i})_{i=1}^{N} = f_{\mathrm{cond}}\big((\boldsymbol{\eta}_i^{\mathrm{cond}})_{i=1}^{N}, \mathbf{p}(\mathbf{x}_t)\big). \tag{13}$$

The observed structure enters the condition network through translation-invariant periodic pair geometry, without absolute coordinates.

The velocity network receives the intermediate displacement $\mathbf{z}_u$, the site contexts $\mathbf{c}_{t,\tau}$, the molecular features, and the flow-time embedding $\mathbf{e}_u(u)$. The physical lag therefore enters the velocity network through $\mathbf{c}_{t,\tau}$, whereas u is provided directly. The intermediate configuration is

$$\mathbf{x}_u^{\mathrm{flow}} = \mathrm{wrap}_{\mathbf{L}}(\mathbf{x}_t + \mathbf{z}_u). \tag{14}$$

The velocity network computes pair geometry from $\mathbf{x}_u^{\mathrm{flow}}$, allowing it to observe the bond lengths, contacts, and periodic neighborhoods implied by the current displacement.

In each transformer layer, pair features affect attention weights and information passed between sites:

$$A_{ij}^{(h)} = \frac{\langle \mathbf{q}_i^{(h)}, \mathbf{k}_j^{(h)} \rangle}{\sqrt{d_h}} + g_b^{(h)}(\mathbf{p}_{ij}), \tag{15}$$

$$\alpha_{ij}^{(h)} = \mathrm{softmax}_j(A_{ij}^{(h)}), \tag{16}$$

$$\mathbf{o}_i^{(h)} = \sum_j \alpha_{ij}^{(h)}[\mathbf{v}_j^{(h)} + g_v^{(h)}(\mathbf{p}_{ij})]. \tag{17}$$

The pair bias $g_b^{(h)}$ controls which sites exchange information, while $g_v^{(h)}$ communicates their geometry and topology directly. After repeated transformer updates, the velocity network returns one three-dimensional generative flow velocity per site. These outputs describe motion along the generative path and are not physical MD velocities.

4.3 Training and Trajectory Generation

For each time-lagged MD pair, we instantiate the linear path in equation (3) with $\mathbf{z}_1 = \mathbf{d}_{t,\tau}$ and add a small Gaussian perturbation. Sampling $\mathbf{z}_0, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{3N})$ and $u \sim \mathcal{U}(0, 1)$ gives

$$\mathbf{z}_u = (1 - u)\mathbf{z}_0 + u\mathbf{d}_{t,\tau} + \sigma\boldsymbol{\epsilon}. \tag{18}$$

The velocity network is trained using

$$\mathcal{L}_{\mathrm{SupraTITO}} = \mathbb{E}\left[\frac{1}{3\sum_i w_i}\sum_{i=1}^{N} w_i \,\|\mathbf{v}_{\theta,i}(\mathbf{z}_u, u; \mathbf{c}_{t,\tau}) - (\mathbf{d}_{t,\tau,i} - \mathbf{z}_{0,i})\|_2^2\right], \tag{19}$$

where w_i masks padded sites. No assembly labels, force targets, or bond-length losses are used.

During inference, the condition network is evaluated once for the current configuration. The resulting site contexts remain fixed while the flow ODE is integrated from Gaussian noise to a sampled displacement. The resulting configuration conditions the next finite-lag transition, and fresh Gaussian noise is drawn at each step. Repeating this generates a stochastic trajectory.

Symmetry. The architecture is permutation equivariant under a consistent reindexing of sites and pair features. Minimum-image geometry is independent of the periodic image used to represent each site, while relative geometry in the condition network and displacement coordinates in the velocity network make the architecture invariant to global translation. Common wrapped translations of training pairs are retained as numerical augmentation. Because Cartesian directions are used directly, the architecture is not exactly rotation equivariant.

5 RESULTS

Dataset. We evaluate SupraTITO on all $20^2 = 400$ ordered dipeptides, with one $1.25\,\mu$s MD trajectory per sequence generated using the Martini 2.1 force field (Marrink et al., 2007; He et al., 2026). Each simulation contains 100 peptides at a concentration of approximately $100\,\mathrm{mM}$, with configurations saved every $2.5\,\mathrm{ps}$. Retaining one backbone bead per residue gives 200 sites per configuration. The split contains 320 training and 80 held-out sequences, keeps reverse-sequence pairs together, and balances amino-acid coverage and late-time MD assembly behavior. Training lags are sampled uniformly from 1 to 1000 saved frames, and no assembly labels are used. Simulation and split details are provided in appendices C.1 and C.2.

Metrics. Solvent-accessible surface area (SASA) measures solvent exposure. Contact count and largest-cluster fraction quantify local association and collective aggregation, while cluster radius of gyration (R_g), nematic order, and relative shape anisotropy describe aggregate size, orientational order, and shape. Bond-length distributions assess molecular integrity. Dynamical comparisons use SASA and cluster-fraction trajectories, contact persistence, and cluster-fraction autocorrelations. Inference uses a lag of 1000 saved frames. Generated observables are averaged over five independent rollouts from the same initial configuration and compared with the corresponding MD trajectory. Metric definitions, rollout settings, and computational costs are provided in appendices D.3, D.4 and E.

5.1 SUPRATITO TRANSFERS ASSEMBLY DYNAMICS ACROSS PEPTIDE SEQUENCES

SupraTITO reproduces the strong sequence dependence of peptide assembly. Generated late-window averages closely match MD for all six observables, across dispersed systems, compact aggregates, and intermediate structures (figure 2). The largest differences and rollout variability occur for intermediate systems, where association and dissociation continue and trajectories may occupy different partially assembled structures. Late-window averaging reduces these fluctuations while retaining differences among sequences. Final-frame measurements vary more widely, particularly for orientational order and aggregate shape, but preserve the same trends (figure 7).

These results suggest that SupraTITO does not need to learn each peptide sequence as an independent system. Training exposes residue types to different neighboring residues and intermolecular environments, allowing the learned operator to reuse these interaction patterns in held-out sequences.

SupraTITO also preserves the internal structure of individual peptides throughout generated trajectories. Its bond-length distributions closely match MD (appendix F.4), and intramolecular connections remain intact in the final configurations (appendix F.3), retaining molecular fidelity over long rollouts.

Beyond these structural comparisons, representative trajectories show assembly dynamics. SupraTITO reproduces the distinct cluster-growth behavior and fluctuations of dispersed (AT), intermediate (LQ), and assembled (YP) systems (figure 3a–c). All held-out trajectories are shown in figures 8 and 9, with representative MD and SupraTITO snapshots in figures 12 to 14.

The processed trajectories contain wrapped coordinates without molecular winding information, preventing an unambiguous estimate of long-time diffusion. We therefore characterize relative assembly dynamics through contact memory. For each sequence and trajectory, we remove the chance-contact baseline set by its mean contact occupancy. Integrating the corrected persistence over the first $25\,\mathrm{ns}$ measures retention of pair identity. SupraTITO reproduces its sequence-dependent variation, with modest underestimation for some of the longest-lived contacts (figure 3d). Complete persistence curves and windowed autocorrelations are provided in figures 10 and 11.

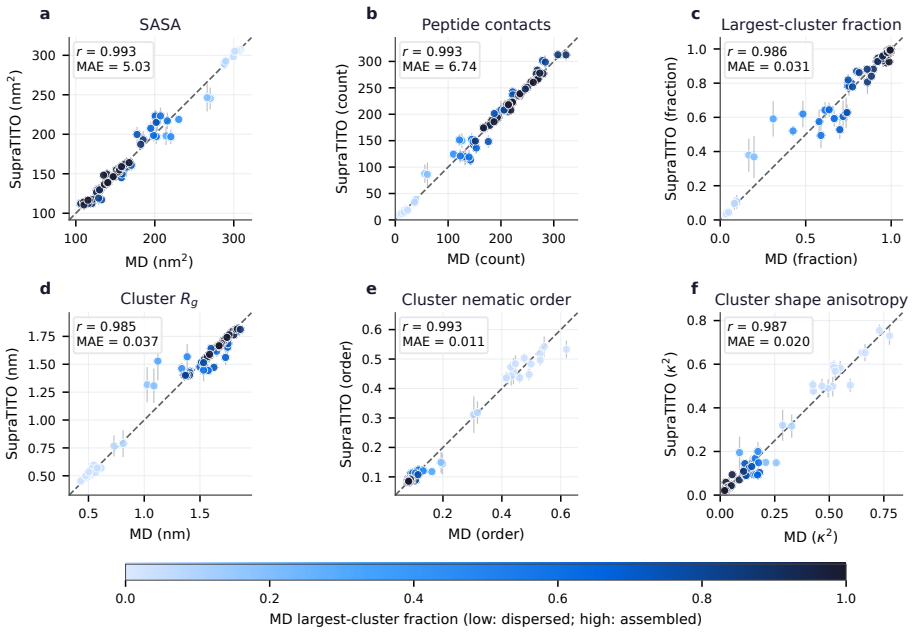


Figure 2: **Transfer to held-out dipeptide sequences.** Each point is one sequence. Markers and error bars show the mean and one standard deviation across five rollouts; MD and generated observables are averaged over the final 125 ns. Color shows the MD largest-cluster fraction; dashed lines indicate agreement. Annotations report Pearson correlation (r) and mean absolute error (MAE).

Dependence on training-sequence coverage. We further train SupraTITO on nested subsets of 40, 80, and 160 ordered sequences and compare them with the model trained on the complete partition. All models are evaluated on the same held-out sequences. Errors decrease and correlations improve across all six observables as coverage grows (figure 4), even after every amino-acid type is represented. Transfer therefore depends not only on amino-acid coverage, but also on exposure to diverse residue combinations and intermolecular environments. The subset construction is described in appendix C.3.

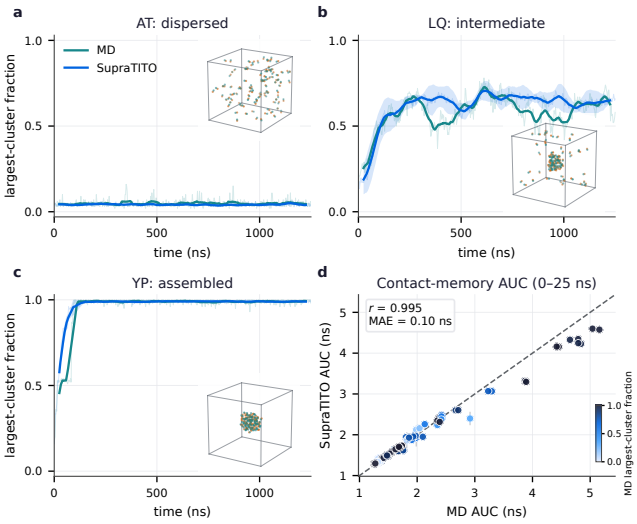


Figure 3: **Dynamical fidelity on held-out sequences. a–c**, largest-cluster fraction for representative dispersed, intermediate, and assembled sequences. Green and blue show MD and SupraTITO; solid curves are moving averages and blue bands show one rollout standard deviation. Insets show final MD configurations. **d**, corrected contact-persistence area over 0–25 ns, colored by the MD late-window largest-cluster fraction.

5.2 Finite-lag learning outperforms direct ensemble prediction

We compare SupraTITO with a sequence-conditioned direct ensemble prediction (DEP) baseline trained on the same sequences and MD trajectories. DEP uses the same residue features, directed molecular topology, and periodic pair geometry, but directly learns $p_\phi(\mathbf{x}_{\text{late}} \mid s, \mathcal{G}, \mathbf{L})$, generating independent

assembly configurations rather than a trajectory. Training targets are sampled from the final 125 ns of each MD trajectory, where assembly structures are comparatively stable. This provides a structurally varied target ensemble rather than only one endpoint per sequence. Implementation and evaluation details are provided in appendix G.

SupraTITO follows the sequence-dependent MD SASA more closely than DEP, which shows larger deviations for several sequences (figure 5a). SupraTITO also achieves lower errors and higher correlations for all six structural observables, together with smaller Wasserstein and Kolmogorov–Smirnov distances for SASA (figure 5b). Full-data DEP performs only slightly better than SupraTITO trained on half as many sequences, whereas full-data SupraTITO clearly outperforms DEP (figure 4).

These results indicate that finite-lag learning uses the available trajectory data more effectively. DEP must infer the complete late assembly distribution from sequence and molecular geometry alone. SupraTITO instead learns configuration changes from many overlapping pairs along each trajectory, with the current configuration providing information about the assembly state. Similar benefits have been reported for protein conformational dynamics (Antoniadis et al., 2026). SupraTITO therefore produces more accurate late structures while also providing the temporal evolution absent from direct ensemble prediction.

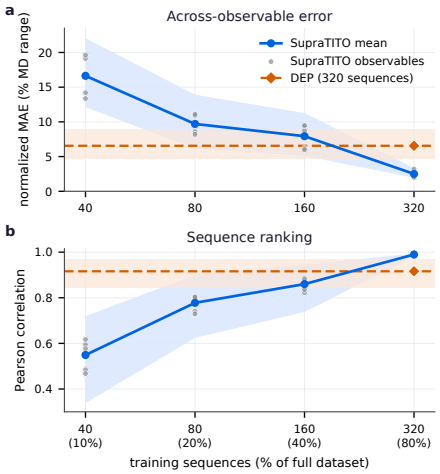


Figure 4: **Transfer improves with training-sequence coverage. a**, normalized MAE; **b**, Pearson correlation on the same held-out set. Gray markers show individual observables; blue shows their mean with 95% bootstrap confidence bands. Orange shows full-data DEP.

5.3 SupraTITO transfers across peptide concentrations

We next test whether the learned dynamics respond to physical conditions outside the training distribution. Peptide concentration is varied by changing the periodic box volume while keeping peptide number fixed. SupraTITO receives the resulting changes in intermolecular separation and periodic geometry rather than a separate concentration variable.

The benchmark contains 22 systems from six representative sequences spanning distinct concentration-dependent behavior. AG and KR remain largely dispersed, FF and FH form stable aggregates, and IK exhibits a marked response at low concentration. QI changes its assembly behavior across the tested range and provides a more challenging case because its held-out sequence is evaluated at changed concentrations. Simulation conditions are provided in appendix C.4.

SupraTITO follows the MD concentration response, including at the most dilute conditions (figure 6). It preserves the contrasting behavior of dispersed and strongly assembling peptides and captures changes in largest-cluster fraction with concentration. The response of QI further shows that concentration transfer can be combined with transfer to a held-out sequence.

The corresponding SASA trajectories show that this agreement extends beyond the late assembly state (appendix H.1). SupraTITO reproduces the progression of molecular exposure and its sequence-dependent late values. These results suggest that the learned operator responds to changing intermolecular environments rather than assigning a fixed assembly outcome to each sequence.

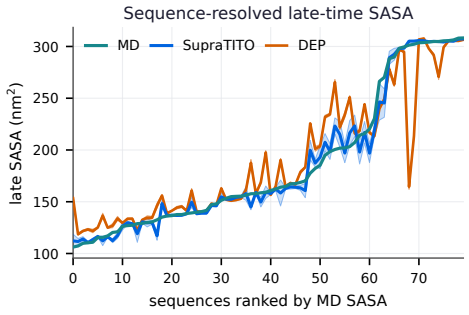
We additionally vary peptide number beyond the single system size used for training (appendix H.2). SupraTITO accepts these configurations and reproduces some MD trends, but agreement is less consistent than in the concentration study. Architectural support for variable site counts therefore does not itself establish size transfer; training across multiple sizes may be needed to learn the associated collective effects.

6 CONCLUSION

We introduced SupraTITO, a transferable GenMD framework for collective dynamics in periodic molecular systems. On the dipeptide benchmark, SupraTITO reproduces structural and dynamical observables for held-out sequences while preserving molecular integrity. Transfer improves with training-sequence coverage, and finite-lag learning produces more accurate assembly structures than direct ensemble prediction while also providing their temporal evolution. Generalization across concentrations shows the potential of learned transfer operators for exploring supramolecular dynamics beyond the training conditions.

Limitations and future work. Training covers homogeneous dipeptides at one concentration and system size; transfer to longer peptides, coassembly, and larger systems remains to be established. Broader training conditions and independent MD trajectories would help assess these settings and the variation among assembly pathways. Accuracy is also limited by the residue-level representation and reference Martini force field, motivating atomistic reconstruction and higher-fidelity reference data. Riemannian flow matching and more scalable attention mechanisms offer further architectural directions. These extensions are discussed in appendix B.

a



b

	Error/distance ↓		r ↑	
Observable/metric	SupraTITO	DEP	SupraTITO	DEP
SASA	**5.733(429)**	15.167(125)	**.990(1)**	.924(1)
Contacts	**7.720(606)**	18.802(231)	**.991(1)**	.921(1)
Cluster fraction	**.0343(27)**	.0779(10)	**.982(1)**	.917(2)
Cluster R_g	**.0405(36)**	.1032(28)	**.981(2)**	.897(5)
Nematic order	**.0132(11)**	.0311(10)	**.989(2)**	.903(5)
Shape anisotropy	**.0231(3)**	.0404(6)	**.983(2)**	.929(3)
SASA W_1	**6.03(32)**	15.55(15)	–	–
SASA KS	**.340(14)**	.507(5)	–	–

Figure 5: **Finite-lag learning versus direct ensemble prediction. a**, late-window SASA for MD, SupraTITO, and DEP; bands show one standard deviation across five generations. **b**, structural MAE and Pearson r, plus SASA Wasserstein (W_1) and Kolmogorov–Smirnov (KS) distances. Parentheses give generation standard deviations in the final digits.

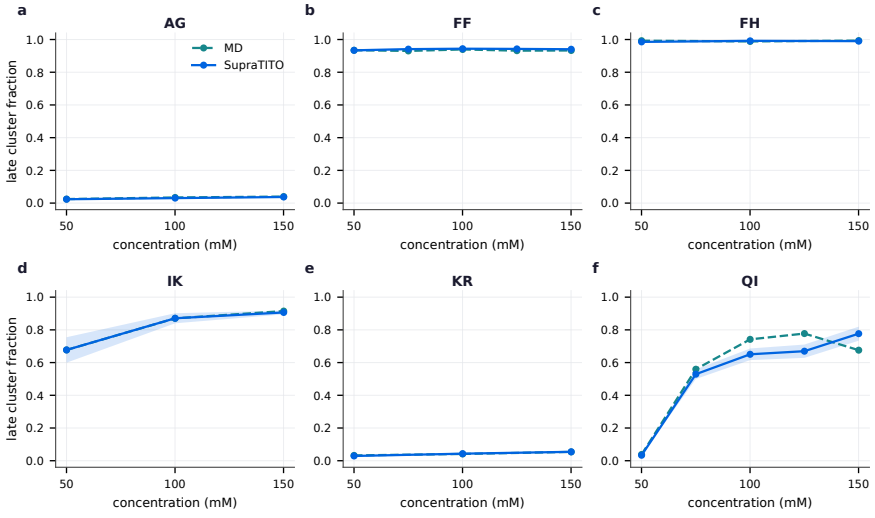


Figure 6: **Transfer beyond the training concentration.** Late largest-cluster fraction for six representative sequences at fixed peptide number. Dashed green curves show MD; blue curves and bands show the SupraTITO mean and one rollout standard deviation. Complete SASA trajectories are provided in appendix H.1.

AI USE STATEMENT

Generative AI tools were used to aid and polish writing, retrieve and identify related literature, support research ideation and execution, and draft sections of the manuscript. Research assistance included discussing methodological and experimental choices, implementing and debugging code, preparing figures, and interpreting results. The authors retain responsibility for the final text, references, methodology, code, and scientific claims.

ETHICS STATEMENT

This work develops a generative molecular dynamics framework for supramolecular systems, demonstrated through peptide assembly. Predictions are limited by the reference force field and training conditions and should not be treated as experimentally validated molecular behavior. Applications to molecular design require independent validation and appropriate assessment of safety and potential misuse.

REPRODUCIBILITY STATEMENT

The simulation protocol, retained molecular representation, and evaluation splits are described in appendices C.1 to C.4. Appendices D.1 to D.3 specify the periodic displacement construction, architecture, training hyperparameters, and trajectory-generation procedure. Metric definitions and the direct ensemble prediction baseline are detailed in appendices E and G. The data and code supporting this study will be made publicly available upon acceptance.

REFERENCES

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. doi: 10.1016/j.softx.2015.06.001.

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023.

Panagiotis Antoniadis, Beatrice Pavesi, Simon Olsson, and Ole Winther. Protein language model embeddings improve generalization of implicit transfer operators. *arXiv preprint arXiv:2602.11216*, 2026.

Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

Rohit Batra, Troy D Loeffler, Henry Chan, Srilok Srinivasan, Honggang Cui, Ivan V Korendovych, Vikas Nanda, Liam C Palmer, Lee A Solomon, H Christopher Fry, et al. Machine learning overcomes human bias in the discovery of self-assembling peptides. *Nature chemistry*, 14(12):1427–1435, 2022.

Ricky TQ Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, volume 2024, pp. 47922–47945, 2024.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Weilong Chen, Bojun Zhao, Jan Eckwert, and Julija Zavadlav. Coarse-grained boltzmann generators. *arXiv preprint arXiv:2602.10637*, 2026.

Luc F Christians, Anna Wojnar, and Alexander J Pak. Probabilistic forecasting for coarse-grained molecular dynamics. *Journal of Chemical Theory and Computation*, 22(9):4231, 2026.

Allan Dos Santos Costa, Ilan Mitnikov, Franco Pellegrini, Ameya Daigavane, Mario Geiger, Zhonglin Cao, Karsten Kreis, Tess Smidt, Emine Kucukbenli, and Joseph Jacobson. Equijump: Protein dynamics simulation via so (3)-equivariant stochastic interpolants. *arXiv preprint arXiv:2410.09667*, 2024.

Nuno Costa and Julija Zavadlav. Morphology-aware peptide discovery via masked conditional generative modeling. *ACS nano*, 20(17):13323, 2026.

Juan Viguera Diez, Mathias Schreiner, and Simon Olsson. Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics. *Science advances*, 12(15):eaed2333, 2026.

J Thorben Frank, Stefan Chmiela, Klaus-Robert Müller, and Oliver T Unke. Machine learning global atomic representations with euclidean fast attention. *Nature Machine Intelligence*, 8(3):388–402, 2026.

Pim WJM Frederix, Gary G Scott, Yousef M Abul-Haija, Daniela Kalafatovic, Charalampos G Pappas, Nadeem Javid, Neil T Hunt, Rein V Ulijn, and Tell Tuttle. Exploring the sequence space for (tri-) peptide self-assembly to design and discover new hydrogels. *Nature chemistry*, 7(1): 30–37, 2015.

Daan Frenkel, Berend Smit, and Mark A Ratner. Understanding molecular simulation: from algorithms to applications, 1997.

Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Proteina: Scaling flow-based protein structure generative models. In *International Conference on Learning Representations*, 2025.

Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-proteina: Atomistic protein generation via partially latent flow matching. In *International Conference on Learning Representations*, volume 2026, pp. 67350–67395, 2026.

Z. He, C. Bai, R. Ma, S. Zhang, X. Fan, J. Li, S. Guo, C. Luo, J. Jin, and F. Ye. Comprehensive evaluation and optimization of martini simulations for dipeptide self-assembly. *The Journal of Physical Chemistry B*, 130:3358–3368, 2026. doi: 10.1021/acs.jpcb.5c08159.

Brooke E Husic and Vijay S Pande. Markov state models: From an art to a science. *Journal of the American Chemical Society*, 140(7):2386–2396, 2018.

Bowen Jing, Hannes Stärk, Tommi Jaakkola, and Bonnie Berger. Generative modeling of molecular dynamics trajectories. *Advances in Neural Information Processing Systems*, 37:40534–40564, 2024.

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

Kübra Kaygisiz, Deborah Sementa, Vignesh Athiyarath, Xi Chen, and Rein V Ulijn. Context dependence in assembly code for supramolecular peptide materials and systems. *Nature Reviews Materials*, 10(6):449–472, 2025.

Leon Klein, Andrew Foong, Tor Fjelde, Bruno Mlodozeniec, Marc Brockschmidt, Sebastian Nowozin, Frank Noé, and Ryota Tomioka. Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics. *Advances in Neural Information Processing Systems*, 36:52863–52883, 2023.

Viktor Klippenstein, Madhusmita Tripathy, Gerhard Jung, Friederike Schmid, and Nico FA Van Der Vegt. Introducing memory in coarse-grained molecular simulations. *The journal of physical chemistry. B*, 125(19):4931, 2021.

Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-based long-range message passing for molecular graphs. In *International Conference on Machine Learning*, pp. 17544–17563. PMLR, 2023.

Jean-Marie Lehn. Toward self-organization and complex matter. *Science*, 295(5564):2400–2403, 2002.

Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew YK Foong, Victor García Satorras, Osama Abdin, Bastiaan S Veeling, Iryna Zaporozhets, et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761): eadv9817, 2025.

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

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The martini force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B*, 111(27):7812–7824, 2007.

Hazime Mori. Transport, collective motion, and brownian motion. *Progress of theoretical physics*, 33(3):423–455, 1965.

Marko Njirjak, Lucija Žužić, Marko Babić, Patrizia Janković, Erik Otović, Daniela Kalafatovic, and Goran Mauša. Reshaping the discovery of self-assembling peptides with generative ai guided by hybrid deep learning. *Nature machine intelligence*, 6(12):1487–1500, 2024.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

Feliks Nuske, Bettina G Keller, Guillermo Pérez-Hernández, Antonia SJS Mey, and Frank Noé. Variational approach to molecular kinetics. *Journal of chemical theory and computation*, 10(4): 1739–1752, 2014.

Simon Olsson. Generative molecular dynamics. *Current Opinion in Structural Biology*, 96:103213, 2026.

Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of chemical physics*, 134(17), 2011.

Meital Reches and Ehud Gazit. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, 300(5619):625–627, 2003.

Mathias Schreiner, Ole Winther, and Simon Olsson. Implicit transfer operator learning: Multiple time-resolution models for molecular dynamics. *Advances in Neural Information Processing Systems*, 36:36449–36462, 2023.

Kirill Shmilovich, Rachael A Mansbach, Hythem Sidky, Olivia E Dunne, Sayak Subhra Panda, John D Tovar, and Andrew L Ferguson. Discovery of self-assembling π-conjugated peptides by active learning-directed coarse-grained molecular simulation. *The Journal of Physical Chemistry B*, 124(19):3873–3891, 2020.

Hythem Sidky, Wei Chen, and Andrew L Ferguson. Molecular latent space simulators. *Chemical Science*, 11(35):9459–9467, 2020.

Henry D Smith, Nathaniel L Diamant, and Brian L Trippe. Calibrating generative models to distributional constraints. *arXiv preprint arXiv:2510.10020*, 2025.

Fabian L Thiemann, Thiago Reschützegger, Massimiliano Esposito, Tseden Taddese, Juan D Olarte-Plata, and Fausto Martelli. Force-free molecular dynamics through autoregressive equivariant networks. *Nature Machine Intelligence*, pp. 1–13, 2026.

Matthew Z. Tien, Dariya K. Sydykova, Andrew G. Meyer, and Claus O. Wilke. Peptidebuilder: A simple python library to generate model peptides. *PeerJ*, 1:e80, 2013. doi: 10.7717/peerj.80.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

Pantelis R Vlachas, Julija Zavadlav, Matej Praprotnik, and Petros Koumoutsakos. Accelerated simulations of molecular systems through learning of effective dynamics. *Journal of Chemical Theory and Computation*, 18(1):538–549, 2022.

Tsjerk A Wassenaar, Kristyna Pluhackova, Rainer A Bockmann, 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.

Marius Wehner and Frank Würthner. Supramolecular polymerization through kinetic pathway control and living chain growth. *Nature Reviews Chemistry*, 4(1):38–53, 2020.

George M Whitesides and Bartosz Grzybowski. Self-assembly at all scales. *Science*, 295(5564): 2418–2421, 2002.

Wei Yang, Shunzhi Wang, Gyu Rie Lee, Jason Z Zhang, Alexis Courbet, David Juergens, Xinru Wang, Thomas Schlichthaerle, Mohamad Abedi, Robert Ragotte, et al. The past, present and future of de novo protein design. *Nature*, 652(8112):1139–1152, 2026.

Shuxin Zheng, Jiyan He, Chang Liu, Yu Shi, Ziheng Lu, Weitao Feng, Fusong Ju, Jiaxi Wang, Jianwei Zhu, Yaosen Min, et al. Predicting equilibrium distributions for molecular systems with deep learning. *Nature Machine Intelligence*, 6(5):558–567, 2024.

A Extended Related Work

Generative molecular dynamics. Generative molecular dynamics learns molecular distributions to support equilibrium sampling or the generation of dynamical trajectories (Olsson, 2026). Alongside the ITO framework, several approaches learn molecular evolution over finite physical intervals. Molecular latent space simulators learn slow collective variables, propagate dynamics in latent space, and reconstruct continuous atomistic trajectories (Sidky et al., 2020). Timewarp learns transferable normalizing flow proposals for molecular moves and combines them with Markov chain Monte Carlo sampling (Klein et al., 2023). EquiJump uses equivariant stochastic interpolants to generate transitions between protein conformations (Costa et al., 2024), whereas MDGen learns a joint distribution over trajectory segments for forward generation and other conditional trajectory tasks (Jing et al., 2024). Autoregressive approaches also propagate molecular states at a learned coarse temporal resolution (Vlachas et al., 2022; Thiemann et al., 2026). Within the ITO framework, TITO extends finite-lag generation across molecular systems, with PLaTITO incorporating protein language model embeddings to improve generalization to unseen proteins (Diez et al., 2026; Antoniadis et al., 2026). SupraTITO extends this line of work from primarily single molecule conformational dynamics to peptide assembly in periodic many-molecule systems.

Machine learning for peptide assembly. Machine learning for peptide assembly has focused on predicting assembly propensity and navigating peptide sequence space. Supervised models relate sequence and molecular descriptors to assembly propensity (Batra et al., 2022), while active learning uses these predictions to guide further simulations and reduce the number of candidates that require evaluation (Shmilovich et al., 2020). Generative design methods take this further by proposing new self-assembling sequences (Njirjak et al., 2024), with recent work guiding generation toward desired aggregate morphologies through isolated peptide descriptors (Costa & Zavadlav, 2026). These approaches help identify promising peptide sequences, but primarily predict assembly outcomes rather than how assemblies form and evolve. SupraTITO complements this work by generating sequence-conditioned assembly trajectories, enabling the evaluation of both structural and dynamical properties for held-out peptides.

B Limitations and Future Work

The present study considers homogeneous dipeptide systems, with training performed at a single concentration and system size. It therefore does not yet establish transfer to longer peptides, co-assembly, or arbitrary system sizes. A natural next step is to train across peptide lengths, molecular compositions, concentrations, and box sizes, and to evaluate these forms of generalization independently.

The reference data contain one MD trajectory for each condition. Peptide assembly can follow multiple pathways and produce distinct metastable or polymorphic morphologies, particularly in larger systems (Wehner & Würthner, 2020). Future benchmarks should therefore include independent initializations and compare distributions of assembly times, morphologies, and physical observables.

SupraTITO currently operates on a residue-level representation and reproduces the dynamics generated by the Martini 2.1 force field. Its accuracy is consequently limited by both the coarse-grained mapping and the reference simulator. Atomistic descriptions are valuable when specific interactions, such as hydrogen bonding and side-chain packing, determine assembly structure and function. Future work could introduce atomistic detail directly through architectures such as La-Proteina (Geffner et al., 2026), or reconstruct it using backmapping (Wassenaar et al., 2014). Backmapped configurations could then be relaxed and propagated with a classical all-atom MD engine, combining SupraTITO's efficient generation of long assembly trajectories with a higher-fidelity description of local molecular interactions. Training on more accurate reference simulations and validating the resulting dynamics against experimental structural and kinetic observables would further improve their physical reliability (Smith et al., 2025).

At the model level, flow matching directly on the periodic torus (Chen & Lipman, 2024) could provide a more natural alternative to the current Euclidean interpolation of periodic displacements. Global self-attention also scales quadratically with the number of retained sites, which may become restrictive for longer peptides and larger supramolecular systems. A scalable extension could combine local geometric message passing with an efficient periodic long-range mechanism to preserve collective interactions without evaluating every molecular pair explicitly (Kosmala et al., 2023; Frank et al., 2026).

C Molecular Dynamics Data and Evaluation Sets

C.1 Dipeptide MD Dataset

The reference dataset contains one independently initialized CG-MD trajectory for each of the 400 ordered dipeptides formed from the canonical amino acids. Peptides are constructed with PeptideBuilder (Tien et al., 2013) and mapped to the Martini 2.1 protein force field with neutral termini and a β-sheet backbone assignment, following the screening parametrization of He et al. (He et al., 2026). Each simulation contains 100 peptides of one sequence in an initially cubic $11.840\,\mathrm{nm}$ periodic box, corresponding to approximately $100\,\mathrm{mM}$. The systems are solvated with standard Martini water, and Na^+ or Cl^- counterions neutralize the side-chain charge. No background salt is included.

Simulations are performed using GROMACS 2024.5 (Abraham et al., 2015). Following energy minimization, each system is equilibrated for $10\,\mathrm{ns}$ in the NPT ensemble at $303\,\mathrm{K}$ and $1\,\mathrm{bar}$ using a $25\,\mathrm{fs}$ integration step, stochastic velocity-rescaling temperature coupling, and Berendsen isotropic pressure coupling. Production continues for $1.25\,\mu\mathrm{s}$ under the same integration step and thermodynamic conditions. Electrostatics use reaction-field treatment with a $1.1\,\mathrm{nm}$ cutoff for nonbonded interactions. Coordinates are stored every $2.5\,\mathrm{ps}$, yielding 500,001 production configurations per sequence. Peptide molecules are subsequently made whole across periodic boundaries.

The stored trajectories contain all Martini peptide beads. SupraTITO retains the backbone bead of each residue, giving two sites per peptide and $N = 200$ sites for each 100-peptide system.

C.2 Training and Held-Out Sequences

The 400 ordered sequences are divided into 320 training sequences and 80 held-out sequences. For every heterodipeptide, the two reverse orderings (XY) and (YX) are assigned to the same parti-

tion, preventing a held-out composition from occurring in reverse order during training. All 20 homodipeptides remain in the training set. The 40 held-out unordered pairs balance amino-acid coverage and the distributions of late-time MD SASA and relative SASA change. Both orderings of the pairs in table 1 form the evaluation set; all remaining ordered sequences form the training set.

Table 1: Held-out unordered dipeptide pairs. Both ordered versions of every listed pair are included in the evaluation set.

Held-out pair groups			
AC	AE	AI	AT
CH	CR	CT	DG
DH	DN	DY	EQ
ER	ES	FI	FL
FV	FW	GK	GS
GT	HK	HN	IQ
IY	KM	KV	LQ
LV	LY	MP	MT
MW	NQ	NV	PR
PS	PY	RW	SW

C.3 Sequence-Coverage Subsets

To measure how performance depends on sequence coverage, we construct three nested training sets containing 40, 80, and 160 ordered sequences from the full 320-sequence training partition. Reverse orderings remain grouped at every coverage. The deterministic selection balances amino-acid frequencies and the training-set distributions of late-time MD SASA and relative SASA change. The 40-sequence set is contained in the 80-sequence set, which is contained in the 160-sequence set. All four models use the same architecture, optimization schedule, rollout protocol, and 80 held-out evaluation sequences.

C.4 Condition and System-Size Evaluation Sets

We construct a further 28-system panel to evaluate transfer beyond the concentration and system size used for training. It contains 22 concentration conditions at fixed peptide count and six evaluations at new system sizes. Table 2 summarizes the complete panel. Here, `C` denotes concentration changes at fixed peptide count, `S` denotes changes in system size at the baseline concentration, and `R` denotes concentration changes obtained by varying peptide number at fixed box size.

Table 2: Concentration and finite-size transfer panel. M denotes the number of peptides and L_0 the initial cubic box length. The `C100` conditions are the corresponding baseline trajectories reused from the exhaustive corpus.

Peptides	Condition IDs	Comparison	Conditions
FF, QI	`C050`, `C075`, `C100`, `C125`, `C150`	Concentration	$M = 100$, concentrations of 50, 75, 100, 125, and 150 mM obtained with $L_0 = 14.917$, 13.032, 11.840, 10.991, and 10.343 nm, respectively.
FF, QI	`S050`, `C100`, `S200`	Finite size	$M = 50$, 100, and 200 at 100 mM, with $L_0 = 9.397$, 11.840, and 14.917 nm, respectively.
FF	`R050`, `R150`	Fixed-box route	$M = 50$ and 150 in the fixed $L_0 = 11.840$ nm box, corresponding to 50 and 150 mM.
FH, IK, AG, KR	`C050`, `C100`, `C150`	Concentration	$M = 100$ at 50, 100, and 150 mM, with $L_0 = 14.917$, 11.840, and 10.343 nm, respectively.

All additional systems follow the same simulation protocol as the exhaustive corpus. Only peptide count, initial box size, and random insertion seed differ.

D Implementation, Training, and Rollout Details

D.1 Molecule-Consistent Periodic Displacements

The flow-matching target is the molecule-consistent periodic displacement in equation (8). The two sites of a peptide share one minimum-image center displacement, while the change in their periodic bond vector describes internal motion. Molecule consistency therefore changes the target construction without introducing additional model parameters.

For an orthorhombic box with lengths $\mathbf{L} = (L_x, L_y, L_z)$, the minimum-image operation is defined component-wise as

$$[\mathrm{MIC}_{\mathbf{L}}(\boldsymbol{\delta})]_k = \delta_k - L_k \,\mathrm{round}\left(\frac{\delta_k}{L_k}\right), \qquad k \in \{x, y, z\}. \tag{20}$$

For the two-site per molecule representation used in this work, the sites of molecule m are indexed by $2m-1$ and $2m$. The periodic bond vector and geometric center are

$$\mathbf{b}_t^{(m)} = \mathrm{MIC}_{\mathbf{L}}\left(\mathbf{x}_{t,2m} - \mathbf{x}_{t,2m-1}\right), \qquad \mathbf{g}_t^{(m)} = \mathbf{x}_{t,2m-1} + \frac{1}{2}\mathbf{b}_t^{(m)}. \tag{21}$$

The molecular translation and internal change are

$$\Delta\mathbf{g}_{t,\tau}^{(m)} = \mathrm{MIC}_{\mathbf{L}}\left(\mathbf{g}_{t+\tau}^{(m)} - \mathbf{g}_t^{(m)}\right), \qquad \Delta\mathbf{b}_{t,\tau}^{(m)} = \mathbf{b}_{t+\tau}^{(m)} - \mathbf{b}_t^{(m)}. \tag{22}$$

Substitution into equation (8) gives the site displacements. Adding them to the observed coordinates and wrapping into the box recovers the target configuration, with a common periodic translation for both sites of each molecule.

D.2 Architecture and Hyperparameters

The reported SupraTITO model uses the translation-invariant, topology-aware, and molecule-consistent formulation in section 4.2 and appendix D.1. Its condition and velocity networks are specified in table 3. The amino-acid and chain-role embeddings have widths 24 and 8, the physical-lag embedding has width 32, and the flow-time embedding has width 128. The complete model contains approximately 3 million parameters.

The flow-time embedding modulates the velocity-network pair representation through adaptive layer normalization. Pair representations are constructed once from the input geometry and topology during each network evaluation and remain fixed across transformer layers. SupraTITO does not use triangle attention or iterative pair updates.

Table 3: Architecture of the reported translation-invariant, molecule-consistent nonuniform-64 model.

Component	Condition network	Velocity network
Transformer layers	4	6
Site-representation width	64	128
Conditioning width	64	128
Pair width	32	64
Attention heads	8	8
Distance categories	64	64
Relative directions	Yes	Yes
Directed topology categories	3	3
Initial site representation	Zero vector	$W_z\mathbf{z}_u$
Pair-value messages	Yes	Yes
Pair conditioning	None	Flow time

Training uses AdamW with learning rate 10^{-4}, zero weight decay, and gradient-norm clipping at 0.1. Physical lags are sampled uniformly from the integers 1 through 1000 saved frames, and sequences

are sampled uniformly. Each epoch contains 500,000 transition pairs. Both frames of a pair receive the same random wrapped translation, and the Gaussian perturbation of the flow path is $\sigma = 0.01$.

The model is trained for 1,000 epochs on NVIDIA H100 GPUs using bfloat16 mixed precision and distributed data parallelism. The global batch size is 768, giving 651 optimizer updates per epoch. The reported checkpoint is saved after epoch 999 at optimization step 651,000.

The nonuniform-64 distance representation used for the main results has 63 ordered boundaries (in nm), compactly written as

$$\begin{aligned}\{\beta_k\}_{k=1}^{63} = \{&0.100{:}0.025{:}0.600\} \cup \{0.650{:}0.050{:}1.000\} \\ &\cup \{1.100{:}0.100{:}2.000\} \cup \{2.200{:}0.200{:}4.000\} \\ &\cup \{4.250{:}0.250{:}6.000\} \cup \{6.500, 7.000, 7.500, 8.000, 9.000, 10.500\}.\end{aligned} \tag{23}$$

Here, $a{:}h{:}b$ denotes the inclusive arithmetic sequence from a to b with spacing h. Distances below the first boundary and above the last boundary occupy the first and final categories, respectively.

D.3 Trajectory Generation

Reported SupraTITO trajectories use a lag of 1000 saved MD frames, corresponding to $2.5\,\mathrm{ns}$ per generated transition. The flow ODE for each transition is integrated using 100 explicit Euler steps. Applying the operator 500 times produces a $1.25\,\mu\mathrm{s}$ trajectory with 501 stored configurations, including the initial state. Five trajectories are initialized from the first frame of the corresponding MD trajectory and generated with independent flow noise. Scalar late-time observables are averaged over the final $125\,\mathrm{ns}$. During inference, the repeatedly evaluated velocity network is compiled using `torch.compile` in its default mode.

D.4 Computational Cost

Training the final model required approximately 1,300 H100 GPU-hours. After compilation warm-up, one lag-1000 transition requires $\approx 0.25\,\mathrm{s}$ for one system and $\approx 6.4\,\mathrm{s}$ for a batch of 80 systems on one H100 GPU. A complete $1.25\,\mu\mathrm{s}$ single-system trajectory is estimated to require approximately $2\,\mathrm{min}$ of inference compute.

Table 4: **Computational cost of reference simulation and SupraTITO.** Inference timings exclude compilation, file output, and post-processing.

Task	Hardware	Computational cost
MD production, $1.25\,\mu\mathrm{s}$	Intel Xeon Platinum 8480+ and GPU Max 1550	$\approx 2\,\mathrm{h/system}$
SupraTITO training	32 NVIDIA H100 GPUs	$\approx 1{,}300$ GPU-hours
Lag-1000 transition, one system	One NVIDIA H100 GPU	$\approx 0.25\,\mathrm{s}$
Lag-1000 transition, 80 systems	One NVIDIA H100 GPU	$\approx 6.4\,\mathrm{s}$

E Metric Definitions

The evaluation spans three physical scales: solvent exposure and intramolecular structure at the residue level, association and morphology at the peptide and cluster levels, and temporal persistence along the trajectory. All coordinate-derived quantities are evaluated frame by frame using the orthorhombic minimum-image operation in equation (20). Unless stated otherwise, distances are in nanometers, areas are in square nanometers, and order parameters are dimensionless. We do not use a binary assembly label as a primary metric because thresholding would discard structural and temporal information.

E.1 Evaluation Hierarchy and Notation

Let a system contain M dipeptides and $N = 2M$ residue beads. For sequence s, $\mathbf{x}_{s,t}^{\mathrm{MD}}$ denotes the available MD trajectory and $\mathbf{x}_{s,t}^{\mathrm{Gen}}$ denotes the corresponding generated trajectory. A coordinate-based observable is first evaluated independently for every frame. Temporal summaries are then

formed independently for each of five generated trajectories. Their mean is the SupraTITO estimate and their sample standard deviation describes variation across rollout seeds. The resulting sequence-level MD and generated values are the paired observations used for cross-sequence statistics.

This order matters. Frames from one trajectory are temporally correlated, and peptides or bonds from the same frame are structurally correlated. Neither is treated as an independent replicate when cross-sequence statistics are computed. Autocorrelation is likewise evaluated independently within each sequence before curves are summarized across sequences.

Table 5: Numerical parameters used for trajectory analysis unless stated otherwise in the corresponding figure caption.

Quantity	Symbol	Default	Role
SASA bead radius	r_b	$0.29\,\mathrm{nm}$	Effective radius of one residue bead
Probe radius	r_p	$0.14\,\mathrm{nm}$	Solvent-probe contribution
Sphere directions	P	20	Surface quadrature
Contact cutoff	r_c	$0.875\,\mathrm{nm}$	Peptide-center contact graph
Late interval	T_{late}	$125\,\mathrm{ns}$	Physical-time averaging window
Contact-memory interval	T_{pers}	$25\,\mathrm{ns}$	Corrected persistence AUC
Maximum ACF lag	$\ell_{\max}$	$50\,\mathrm{ns}$	Maximum displayed temporal offset

E.2 Coarse-Grained Solvent Exposure

The solvent-accessible surface area (SASA) used here is a PBC-aware, Shrake–Rupley-style estimate computed from the representative residue beads. It should not be interpreted as atomistic SASA. Each bead is assigned an effective radius $r_b = 0.29\,\mathrm{nm}$, and a solvent probe of radius $r_p = 0.14\,\mathrm{nm}$ gives the sampled surface radius $r_s = r_b + r_p = 0.43\,\mathrm{nm}$. For bead center $\mathbf{x}_i$, the surface is represented by $P = 20$ approximately uniform Fibonacci-sphere directions $\mathbf{u}_p$:

$$\mathbf{y}_{ip} = \mathbf{x}_i + r_s \mathbf{u}_p, \qquad p = 1, \ldots, P. \tag{24}$$

A surface point is accessible if it is not inside the sampled surface sphere of any other bead. We denote this accessibility indicator by

$$\chi_{ip} = \mathbb{I}\left[\min_{j \neq i} \left\|\mathrm{MIC}_{\mathbf{L}}(\mathbf{y}_{ip} - \mathbf{x}_j)\right\|_2 \geq r_s\right]. \tag{25}$$

The total exposure of a frame is therefore

$$\mathrm{SASA}(\mathbf{x}) = \frac{4\pi r_s^2}{P} \sum_{i=1}^{N} \sum_{p=1}^{P} \chi_{ip}. \tag{26}$$

Lower values indicate that a larger fraction of the bead surfaces is buried, as generally occurs upon aggregation. The effective bead radius was selected using training-reference frames only by minimizing RMSE to matched GROMACS reference SASA computed from the original MD representation. A 20-direction calibration gave an optimum near $0.2825\,\mathrm{nm}$; the radius was rounded and fixed at $0.29\,\mathrm{nm}$. Generated and MD bead trajectories are evaluated with identical radii, directions, and PBC handling.

E.3 Peptide Centers and Contact Graph

For peptide q, let $\mathbf{n}_q$ and $\mathbf{c}_q$ denote its N- and C-terminal residue beads. The minimum-image terminal vector, wrapped peptide center, and unit orientation are

$$\mathbf{v}_q = \mathrm{MIC}_{\mathbf{L}}(\mathbf{c}_q - \mathbf{n}_q), \quad \mathbf{p}_q = \mathrm{wrap}_{\mathbf{L}}(\mathbf{n}_q + \tfrac{1}{2}\mathbf{v}_q), \quad \widehat{\mathbf{v}}_q = \frac{\mathbf{v}_q}{\|\mathbf{v}_q\|_2}. \tag{27}$$

This construction prevents a peptide crossing a periodic boundary from acquiring an artificial center near the middle of the box.

We define an undirected peptide contact graph $G = (V, E)$ with one node per peptide and adjacency matrix

$$A_{qr} = \mathbb{I}\big[q \neq r \ \wedge \ \|\mathrm{MIC}_{\mathbf{L}}(\mathbf{p}_q - \mathbf{p}_r)\|_2 < r_c\big], \qquad r_c = 0.875\,\mathrm{nm}. \tag{28}$$

The framewise contact count is

$$C = \sum_{q=1}^{M-1} \sum_{r=q+1}^{M} A_{qr}, \tag{29}$$

so each peptide pair contributes at most one contact, irrespective of how its two residue beads are arranged. The raw count is suitable for the present fixed-M benchmark; comparisons across different system sizes should instead use a normalized contact density.

The cutoff was fixed from the first minimum following the first peak of the training-MD peptide-center radial distribution function. Calibration used $0.01\,\mathrm{nm}$ radial shells, shell-volume and instantaneous-box-volume normalization, equal weighting of sequences, and Gaussian smoothing with $0.025\,\mathrm{nm}$ width. Held-out trajectories were not used to select the cutoff.

E.4 Contact Memory

Estimating long-time molecular diffusion requires unwrapped trajectories, or equivalently the winding numbers that record successive crossings of the periodic box. These are not retained by the processed trajectories, so we do not report diffusion coefficients. Instead, baseline-corrected contact persistence characterizes relative molecular dynamics directly from the PBC-aware contact graph.

Contact persistence measures whether the same peptide pairs remain associated over time, rather than only comparing the total number of contacts. For contact adjacency $A_{qr}(n)$ at frame n, the intermittent contact persistence at lag ℓ is

$$P_{\mathrm{c}}(\ell) = \frac{\sum_n \sum_{q<r} A_{qr}(n) A_{qr}(n+\ell)}{\sum_n \sum_{q<r} A_{qr}(n)}, \tag{30}$$

where the sums include time origins for which both frames lie within the selected temporal window. Thus, $P_{\mathrm{c}}(\ell)$ is the probability that a peptide pair in contact at one time is also in contact after lag ℓ. Because this is an intermittent definition, a contact may break and reform during the interval.

Under independent, uniformly mixed contacts, the persistence baseline is the mean probability of contact. For a window containing N_w frames and M peptides, we estimate this trajectory-specific baseline as

$$p_{\mathrm{c}} = \frac{1}{N_w \binom{M}{2}} \sum_{n=0}^{N_w-1} \sum_{q<r} A_{qr}(n), \tag{31}$$

and define the corrected persistence

$$\widetilde{P}_{\mathrm{c}}(\ell) = \frac{P_{\mathrm{c}}(\ell) - p_{\mathrm{c}}}{1 - p_{\mathrm{c}}}. \tag{32}$$

The baseline is computed separately for the MD trajectory and for every SupraTITO rollout before averaging across rollouts. Consequently, $\widetilde{P}_{\mathrm{c}}(0) = 1$ when contacts are present, and zero denotes persistence equal to the window-average occupancy. This is a random-mixing reference, not an assumed long-lag limit for a nonstationary assembly. We summarize contact memory by

$$\widetilde{A}_{\mathrm{c}} = \int_0^{25\,\mathrm{ns}} \widetilde{P}_{\mathrm{c}}(t)\, dt. \tag{33}$$

A larger $\widetilde{A}_{\mathrm{c}}$ indicates longer retention of peptide-pair identity, not simply a larger number of contacts. Because the definition is intermittent and generated configurations are separated by $2.5\,\mathrm{ns}$, it is a contact-memory diagnostic rather than an uninterrupted contact lifetime or a rate estimate below the output resolution.

E.5 Largest-Cluster Size and Morphology

Clusters are the connected components of the peptide contact graph. Let $\mathcal{C}_\star$ be the component with the greatest number of peptides and $M_\star = |\mathcal{C}_\star|$. Ties are resolved deterministically by the lowest-indexed component. The largest-cluster fraction is

$$f_{\mathrm{LC}} = \frac{M_\star}{M}, \tag{34}$$

which ranges from $1/M$ for a fully disconnected frame to one when every peptide belongs to a single connected component.

Cluster shape is evaluated from peptide centers rather than residue beads. Because wrapped coordinates cannot be averaged directly, centers in the largest cluster are first placed near an anchor $q_0 \in \mathcal{C}_\star$,

$$\widetilde{\mathbf{p}}_q = \mathbf{p}_{q_0} + \mathrm{MIC}_{\mathbf{L}}(\mathbf{p}_q - \mathbf{p}_{q_0}), \qquad q \in \mathcal{C}_\star. \tag{35}$$

With $\overline{\mathbf{p}}$ the mean unwrapped center, the gyration tensor is

$$\mathbf{S} = \frac{1}{M_\star} \sum_{q \in \mathcal{C}_\star} (\widetilde{\mathbf{p}}_q - \overline{\mathbf{p}})(\widetilde{\mathbf{p}}_q - \overline{\mathbf{p}})^\top. \tag{36}$$

For eigenvalues $0 \le \lambda_1 \le \lambda_2 \le \lambda_3$, the cluster radius of gyration and relative shape anisotropy are

$$R_g = \sqrt{\lambda_1 + \lambda_2 + \lambda_3}, \qquad \kappa^2 = \frac{3}{2} \frac{\sum_{\alpha=1}^{3} (\lambda_\alpha - R_g^2/3)^2}{R_g^4}. \tag{37}$$

R_g measures cluster extent and has units of nanometers. The dimensionless κ^2 is zero for an isotropic eigenvalue spectrum and approaches one for an ideal rod. For a singleton cluster, the implementation assigns $R_g = 0$ and $\kappa^2 = 0$; orientational order remains undefined. Anchor-based unwrapping is reliable for compact, non-percolating clusters but can be ambiguous for a cluster spanning more than half the box relative to its anchor. Shape metrics for such frames should therefore be interpreted cautiously.

E.6 Cluster Nematic Order

Nematic order measures whether the molecular axes in the largest cluster share a common orientation, while treating parallel and antiparallel axes as equivalent. For an index set $\mathcal{I}$, define the traceless alignment tensor

$$\mathbf{Q}(\mathcal{I}) = \frac{1}{|\mathcal{I}|} \sum_{q \in \mathcal{I}} \left(\frac{3}{2} \widehat{\mathbf{v}}_q \widehat{\mathbf{v}}_q^\top - \frac{1}{2} \mathbf{I} \right), \qquad S(\mathcal{I}) = \lambda_{\max}(\mathbf{Q}(\mathcal{I})). \tag{38}$$

Because $\mathbf{Q}$ is a real symmetric 3×3 matrix, it has three real eigenvalues and mutually orthogonal eigenvectors. The largest eigenvalue $\lambda_{\max}$ is the alignment strength S, and its eigenvector is the nematic director: the spatial axis along which the peptide orientations are most strongly aligned. The outer product $\widehat{\mathbf{v}}_q \widehat{\mathbf{v}}_q^\top$ is unchanged when $\widehat{\mathbf{v}}_q$ is replaced by $-\widehat{\mathbf{v}}_q$, so parallel and antiparallel peptides contribute identically.

We evaluate equation (38) with $\mathcal{I} = \mathcal{C}_\star$. Values near zero indicate that no common orientation dominates, whereas values approaching one indicate that nearly all molecular axes follow a common director. This orientation metric is independent of whether the peptides are spatially close: a compact aggregate can have low nematic order when its constituent peptides are orientationally disordered. The metric is undefined for a singleton cluster and is omitted from summaries for that frame.

E.7 Intramolecular Bond Fidelity

The indexed residue pair $(2q-1, 2q)$ defines peptide q throughout the MD and generated trajectories. Its minimum-image intramolecular distance is

$$d_q^{\mathrm{bond}} = \|\mathrm{MIC}_{\mathbf{L}}(\mathbf{c}_q - \mathbf{n}_q)\|_2. \tag{39}$$

For a predeclared temporal window, bond lengths are pooled over frames and peptide indices within each sequence and rollout.

E.8 Temporal Summaries and Autocorrelation

For a framewise observable y_n with $n = 0, \ldots, T-1$, we retain the complete time series and report three scalar summaries: the trajectory mean, the final value, and the mean over a fixed physical late interval. If output frames are separated by Δt, the number of late frames is

$$W = \min\left(T, \left\lceil \frac{T_{\mathrm{late}}}{\Delta t} \right\rceil\right), \qquad T_{\mathrm{late}} = 125\,\mathrm{ns}, \tag{40}$$

and

$$\overline{y}_{\mathrm{late}} = \frac{1}{W} \sum_{n=T-W}^{T-1} y_n. \tag{41}$$

Using physical time rather than a fixed number of output frames makes the summary comparable across rollout lags and output cadences. At the reported lag, this selects 50 frames with $1125 < t \leq 1250\,\mathrm{ns}$ (native MD frame indices 451000 through 500000 at stride 1000, indexed from zero). The same frames are used in the transfer, coverage, and DEP comparisons. Summaries are formed independently within each trajectory.

For a finite trajectory with mean $\bar{y}$ and variance $s_y^2 = T^{-1} \sum_{n=0}^{T-1} (y_n - \bar{y})^2$, the normalized autocorrelation at output lag ℓ is

$$\rho(\ell) = \frac{1}{(T-\ell) s_y^2} \sum_{n=0}^{T-\ell-1} (y_n - \bar{y})(y_{n+\ell} - \bar{y}). \tag{42}$$

We evaluate equation (42) separately for SASA, contact count, largest-cluster fraction, and cluster nematic order within the first and final $250\,\mathrm{ns}$ of each trajectory, using temporal offsets up to $50\,\mathrm{ns}$. Each generated rollout is evaluated separately before its autocorrelation is summarized across rollout seeds. Self-assembly trajectories are generally nonstationary, so $\rho(\ell)$ is interpreted as a descriptive comparison of temporal persistence rather than as an equilibrium correlation function or an independent kinetic-rate estimate. For a constant observable, the normalized ACF is undefined; no nonzero-lag curve is shown, and undefined values are excluded from rollout summaries.

E.9 Across-Sequence Agreement

Let a_s and $\widehat{a}_s$ be the MD and generated summaries for observable a on sequence $s \in \{1, \ldots, N_{\mathrm{seq}}\}$. We report the mean absolute error

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{seq}}} \sum_{s=1}^{N_{\mathrm{seq}}} |\widehat{a}_s - a_s|. \tag{43}$$

Pearson correlation measures linear agreement across sequences,

$$r = \frac{\sum_s (a_s - \bar{a})(\widehat{a}_s - \overline{\widehat{a}})}{\sqrt{\sum_s (a_s - \bar{a})^2} \sqrt{\sum_s (\widehat{a}_s - \overline{\widehat{a}})^2}}, \tag{44}$$

where $\bar{a}$ and $\overline{\widehat{a}}$ are the corresponding across-sequence means. For comparisons between observables with different units, figure 4 reports the range-normalized MAE

$$\mathrm{nMAE} = \frac{\mathrm{MAE}}{\max_s a_s - \min_s a_s}. \tag{45}$$

This normalization is used only for the visual comparison across observables; figure 5b reports each MAE in its physical units. For figure 4, agreement is computed using the mean of the five generated summaries for each sequence, then averaged over the six observables. Its 95% confidence bands use 5,000 bootstrap resamples of the 80 ordered sequences and five generation indices, both with replacement. The same indices are used across observables within each resample, with the MD normalization range held fixed. Band limits are the 2.5th and 97.5th percentiles. In contrast, figure 5b reports the mean and sample standard deviation of agreement computed separately for each of the five generated ensembles.

For distributional comparisons, let F_s^{MD} and F_s^{Gen} be the empirical cumulative distributions of an observable for sequence s. We use the one-dimensional Wasserstein and Kolmogorov–Smirnov distances

$$W_{1,s} = \int_{-\infty}^{\infty} \left| F_s^{\mathrm{Gen}}(a) - F_s^{\mathrm{MD}}(a) \right| \, da, \qquad D_{\mathrm{KS},s} = \sup_a \left| F_s^{\mathrm{Gen}}(a) - F_s^{\mathrm{MD}}(a) \right|. \tag{46}$$

The distributional values in figure 5b are arithmetic means of these per-sequence distances. Wasserstein distance retains the units of the observable, whereas the Kolmogorov–Smirnov distance is dimensionless.

F Complete Results for Held-Out Sequences

The main text summarizes agreement across sequences and highlights three representative trajectories. Here we report every held-out sequence to expose the full range of outcomes. All figures use the SupraTITO checkpoint and lag-1000 protocol in appendix D.3. Unless noted otherwise, generated curves and markers show the mean across five rollouts, and bands and bars show one standard deviation. Static configurations are taken from one representative rollout because averaging coordinates across trajectories does not produce a physical configuration.

F.1 Final-Frame Agreement Across Sequences

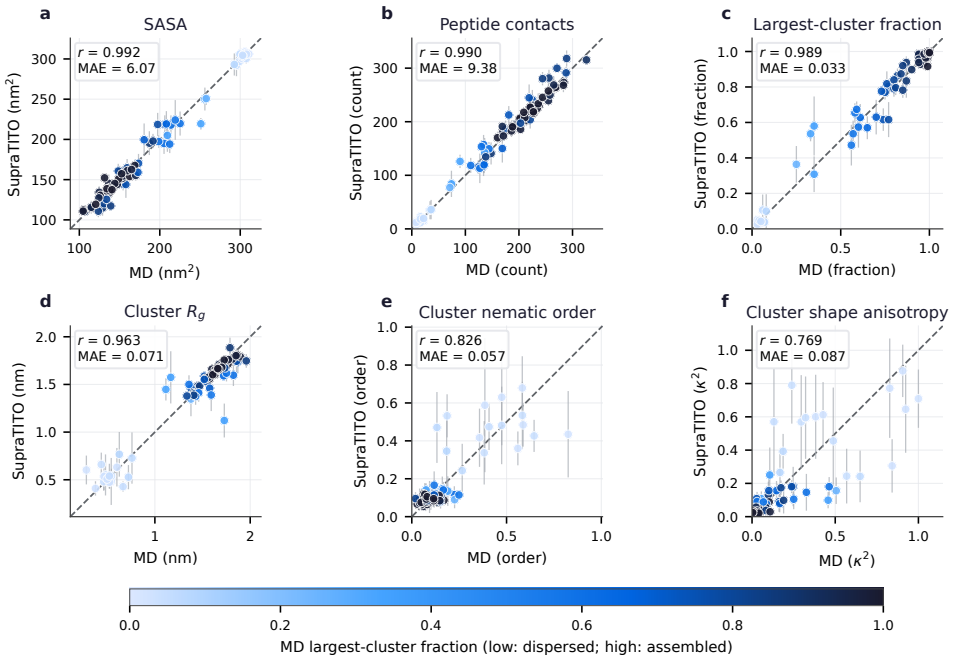


Figure 7: **Final-frame transfer to held-out dipeptide sequences.** SupraTITO is compared with MD using the instantaneous final configuration of each held-out trajectory. Each point is one ordered sequence, vertical bars show the rollout standard deviation, and color denotes the MD final-frame largest-cluster fraction. Dashed lines indicate exact agreement. Relative to the late-window averages in figure 2, the final-frame values retain the principal sequence trends but exhibit greater dispersion because no temporal averaging suppresses structural fluctuations or differences between metastable realizations.

F.2 TEMPORAL TRAJECTORIES

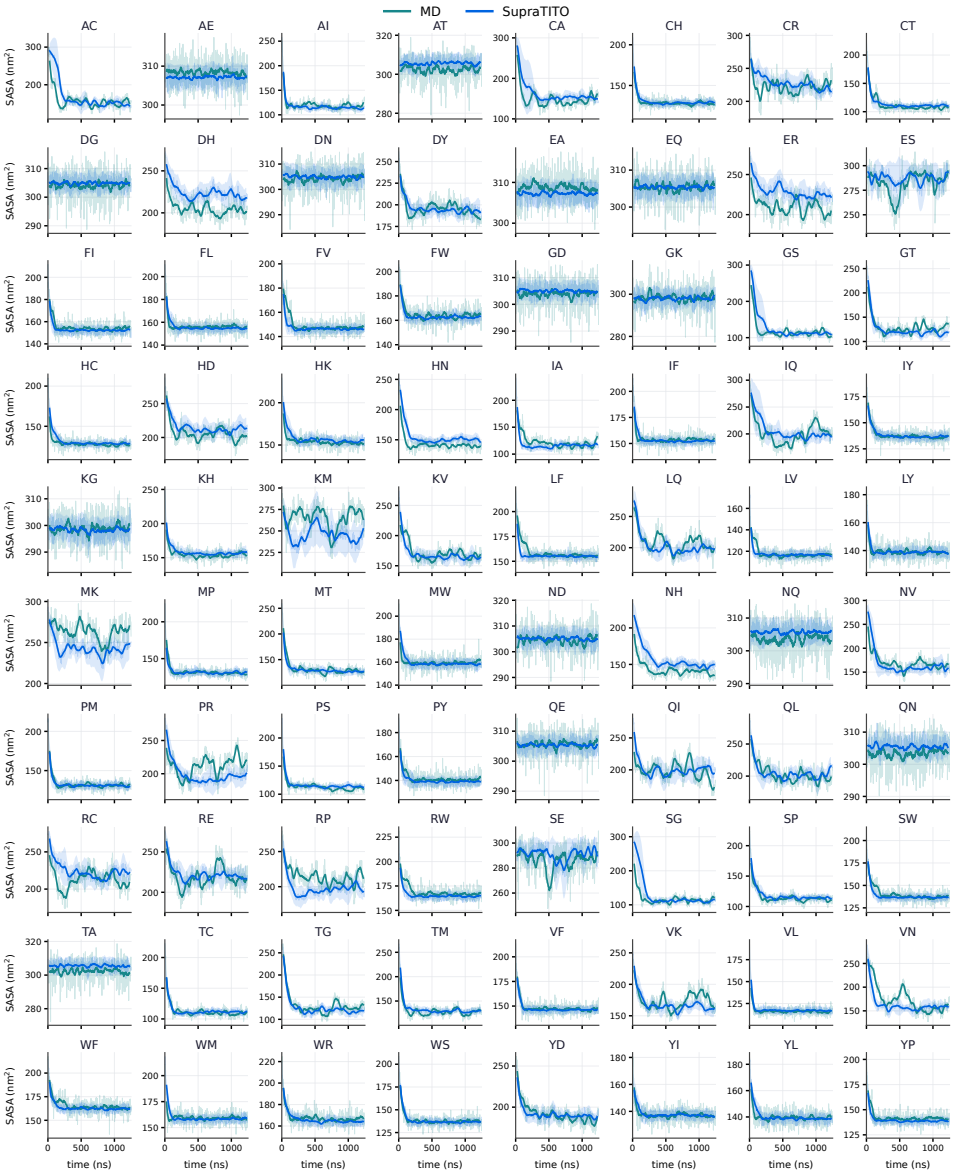


Figure 8: **SASA trajectories for all held-out sequences.** Each panel compares MD (green) and SupraTITO (blue) for one ordered sequence. Faint green curves show raw MD values, faint blue curves show framewise means across rollouts, solid curves show 21-frame moving averages, and blue bands show one standard deviation across rollouts. Each panel retains its own SASA scale.

To complement the scalar trajectories, figures 12 to 14 show matched three-dimensional configurations for the dispersed, intermediate, and assembled examples in the main text. Each sphere is one retained residue site, gray sticks connect the two indexed sites of each peptide, and the wireframe marks the periodic box. All panels use the same camera orientation and box scale. For visualization, the largest PBC-connected peptide cluster is unwrapped and centered, and the remaining peptides are reimaged around it. This operation changes only the displayed periodic images and preserves all minimum-image distances.

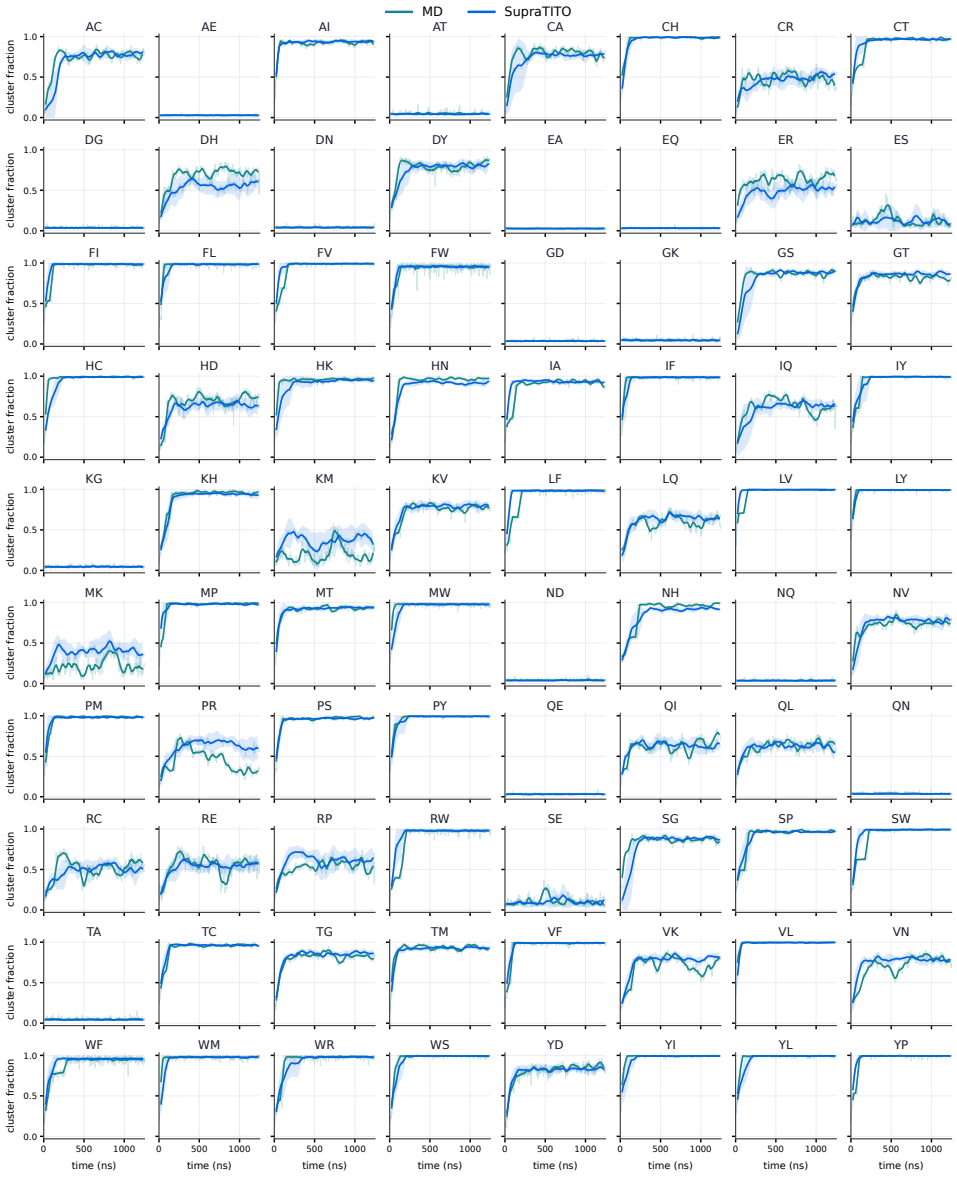


Figure 9: **Largest-cluster-fraction trajectories for all held-out sequences.** Green curves show MD and blue curves show SupraTITO. Faint curves show raw MD values and framewise SupraTITO means, solid curves show 21-frame moving averages, and blue bands show one standard deviation across rollouts. The observable is computed from the PBC-aware peptide-center contact graph.

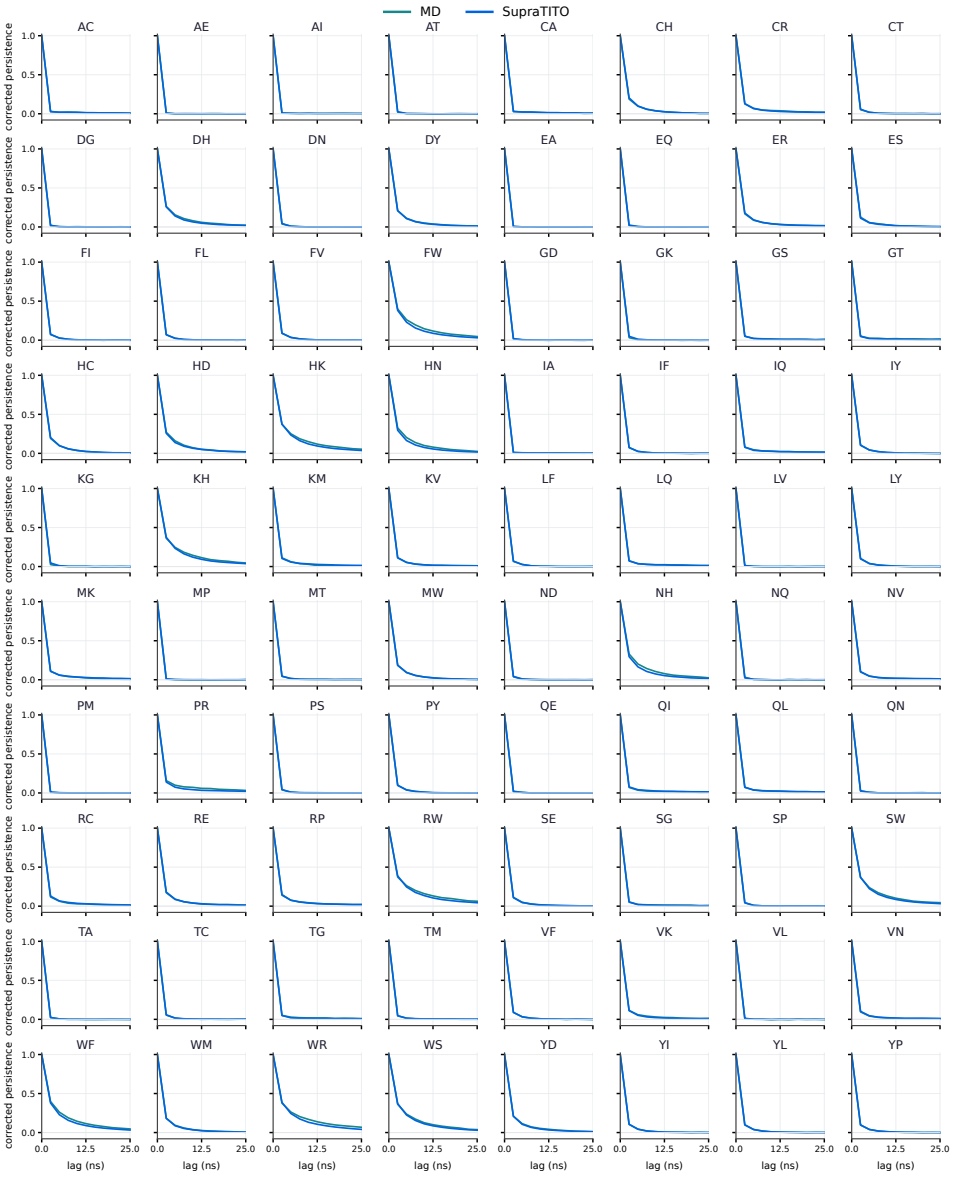


Figure 10: **Baseline-corrected contact persistence for all held-out sequences.** Each panel compares peptide-pair contact memory from MD and SupraTITO over the final 250 ns, with MD in green and SupraTITO in blue. The sequence- and trajectory-specific contact-occupancy baseline is removed before averaging the generated curves. The blue band shows one standard deviation across rollouts. Persistence is evaluated for lags up to 25 ns using the peptide-center contact definition in equation (28).

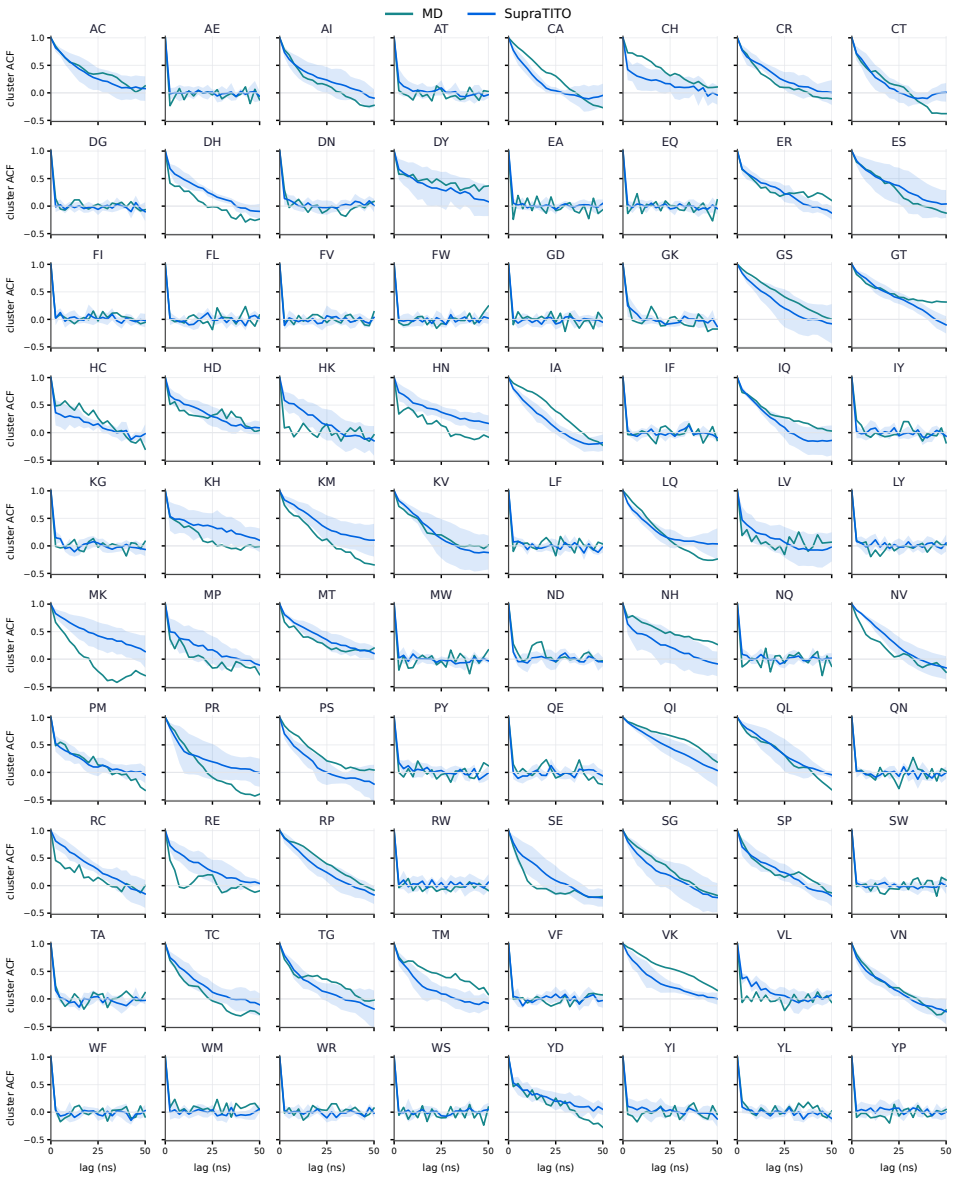


Figure 11: **Largest-cluster-fraction autocorrelation for all held-out sequences.** Each panel compares MD (green) and SupraTITO (blue) autocorrelations computed over the final 250 ns, using offsets up to 50 ns. The blue band shows one standard deviation across independently evaluated rollouts. These curves complement the contact-memory comparison in figure 3d.

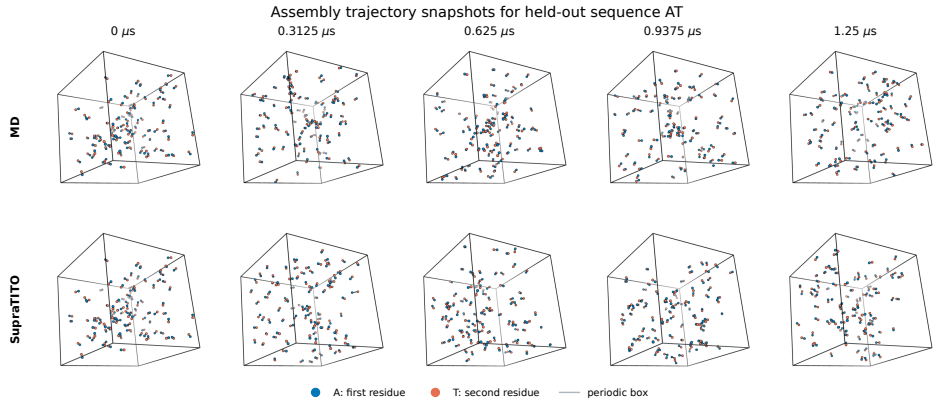


Figure 12: **Matched assembly snapshots for the representative dispersed sequence AT.** MD configurations are shown above the corresponding SupraTITO configurations at five physical times. Both trajectories retain a predominantly dispersed peptide population throughout the rollout. The first and second sequence positions are colored blue and coral, respectively.

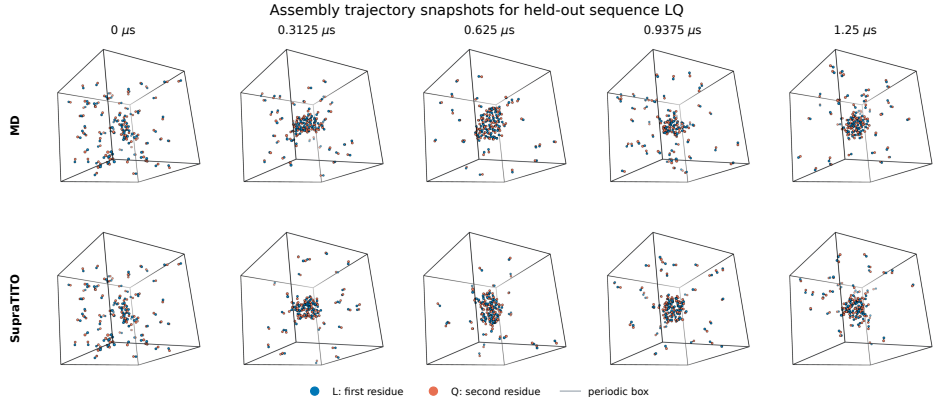


Figure 13: **Matched assembly snapshots for the representative intermediate sequence LQ.** MD and SupraTITO both form a substantial cluster while retaining a population of unassociated peptides, consistent with the partial and fluctuating assembly dynamics in figure 3b.

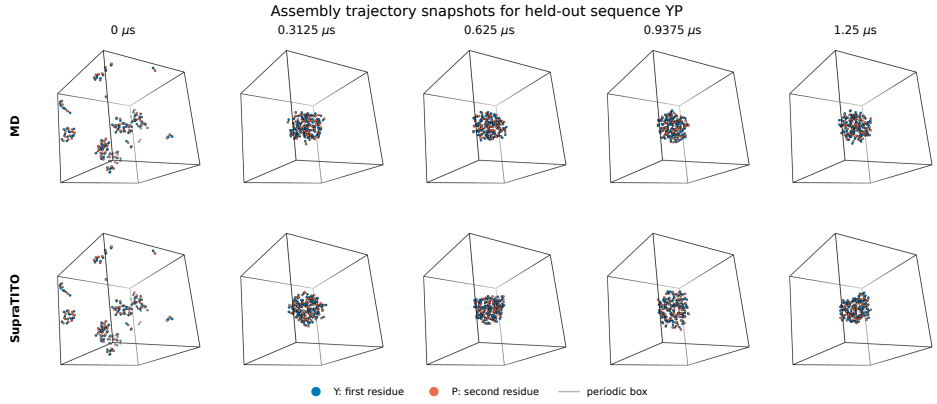


Figure 14: **Matched assembly snapshots for the representative held-out assembled sequence YP.** The two stochastic trajectories follow different microscopic pathways but both rapidly form a compact aggregate, consistent with figure 3c.

F.3 Final Configurations and Molecular Connectivity

Figure 15 compares the final MD and generated morphologies and shows the indexed intramolecular connectivity of the SupraTITO configurations for every held-out sequence.

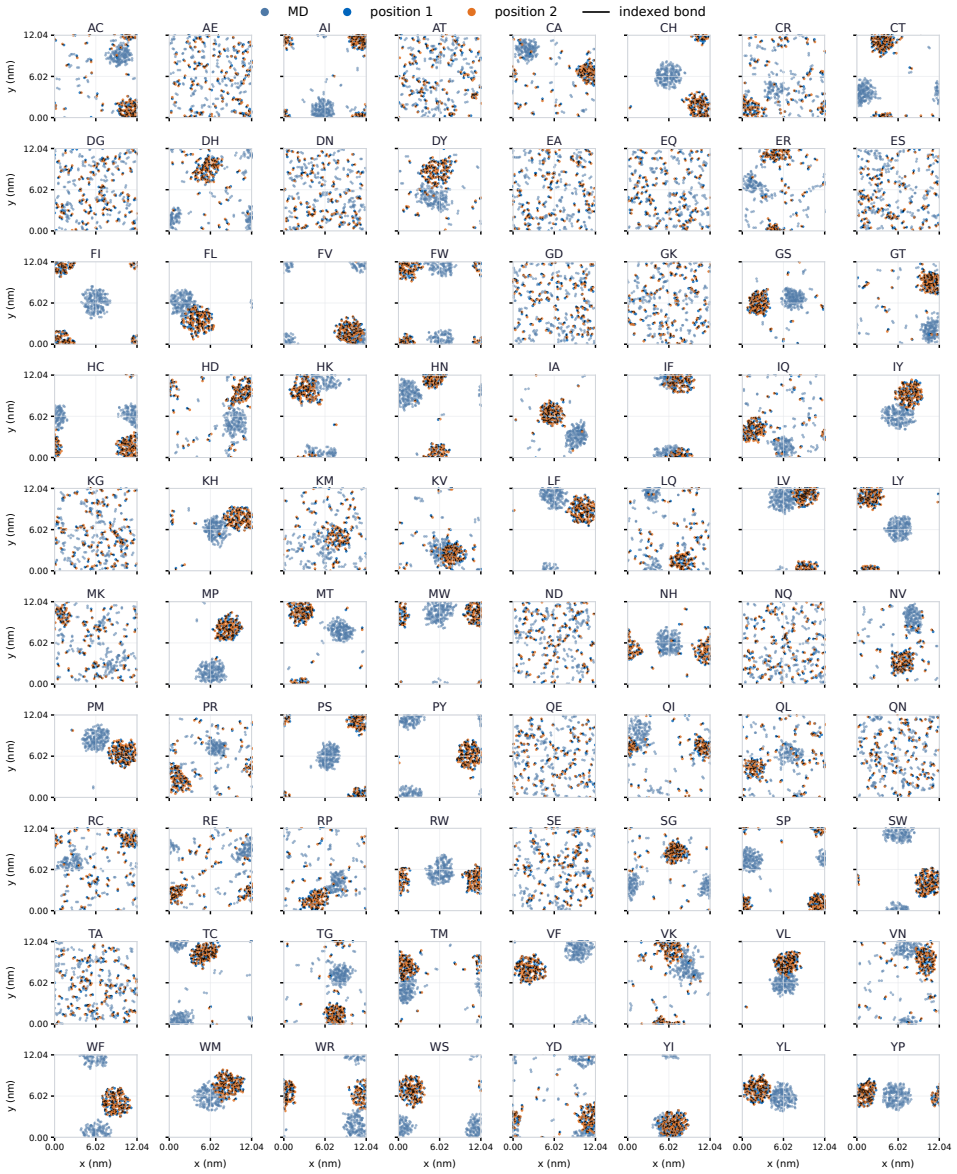


Figure 15: **Final configurations for all held-out sequences.** Muted blue points show the MD reference. Blue and orange points show the first and second residue positions in SupraTITO, respectively, and black lines connect the indexed sites of each peptide using minimum-image geometry. Generated coordinates are from one representative rollout. The XY projection supports visual comparison of morphology and molecular connectivity but can hide separation along the third coordinate.

F.4 Intramolecular Bond Fidelity

The complete bond-distance comparison in figure 16 tests whether the agreement in collective assembly is accompanied by physically consistent intramolecular separations for every held-out sequence.



Figure 16: **Intramolecular bond-distance distributions for all held-out sequences.** Distances are computed for fixed indexed residue pairs with the full three-dimensional minimum-image convention and pooled over analyzed frames and peptides within each rollout. Each panel compares the MD distribution with the SupraTITO mean and one standard deviation across rollouts.

G Direct Ensemble Prediction Baseline

The direct ensemble prediction (DEP) baseline learns the sequence-conditioned late-structure density

$$p_\phi(\mathbf{x}_{\text{late}} \mid s, \mathcal{G}, \mathbf{L}) \tag{47}$$

with flow matching along a linear probability path. Unlike SupraTITO, DEP receives neither a current configuration nor a physical lag. Its periodic Gaussian source draws each residue coordinate from an isotropic Gaussian with standard deviation $1\,\text{nm}$ around a uniformly sampled box origin. Because peptides of the same sequence are exchangeable, source and target peptide centers are matched by linear-sum assignment. The flow interpolates minimum-image peptide-center displacements and center-relative residue offsets, preserving peptide membership during the periodic interpolation.

Targets are sampled uniformly from the final 50,000 saved frames, corresponding to the final $125\,\text{ns}$ of each training trajectory. This window supplies multiple configurations per sequence and retains structural variation within the late ensemble; using only one endpoint per sequence would leave too few targets for conditional flow matching. DEP and SupraTITO therefore use the same MD trajectory collection, with DEP trained on the late ensemble and SupraTITO trained on finite-lag transitions throughout the trajectories.

The baseline uses a single seven-layer Proteína transformer with site width 128, pair width 64, and eight attention heads. Its inputs comprise flow time, amino-acid identity, terminal role, directed bond topology, and minimum-image pair-distance categories and relative directions. It is trained for 1,000 epochs with 500,000 sampled examples per epoch, AdamW at learning rate 10^{-4}, and flow noise $\sigma = 0.01$. Each independent sample is generated using 100 explicit Euler steps. For each held-out sequence, the matched evaluation uses the same 50 lag-strided frames from the MD late window as the transfer analysis, 250 independent DEP samples (50 from each of five inference seeds), and 250 configurations sampled evenly from the final $125\,\text{ns}$ of five lag-1000 SupraTITO rollouts. Agreement metrics are computed separately for each generated ensemble and reported as the mean and sample standard deviation across the five ensembles. The independent DEP samples support structural-distribution comparisons but do not define a trajectory or an autocorrelation function.

The aggregate comparison in figure 5 is complemented by two sequence-resolved views. Figure 17 shows how the SASA error is distributed across the held-out set, while figure 18 compares the complete generated and MD SASA distributions for every sequence.

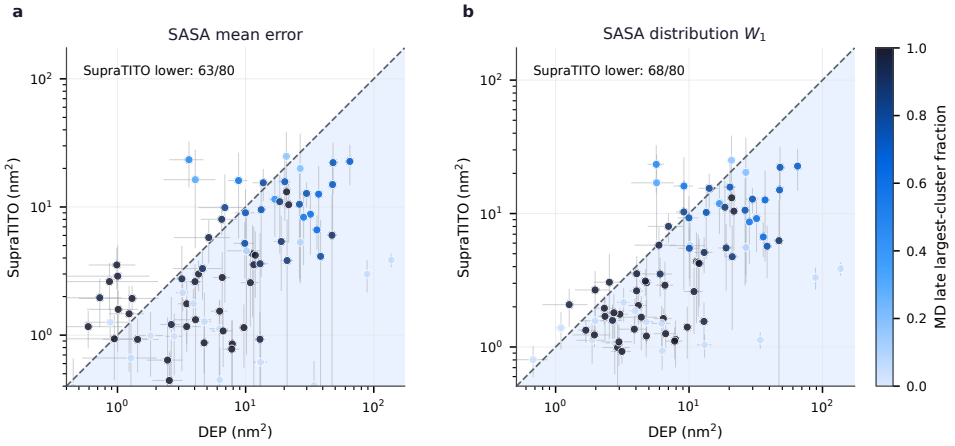


Figure 17: **Sequence-wise SASA errors for DEP and SupraTITO.** Each point is the mean for one held-out sequence, with horizontal and vertical bars showing one standard deviation across five DEP generation seeds and five SupraTITO rollout seeds, respectively. Points below the equal-error diagonal favor SupraTITO. Color denotes the MD late-window largest-cluster fraction, and logarithmic axes retain both low-error systems and outliers.



Figure 18: **Complete sequence-resolved SASA distributions for the DEP comparison.** Each panel contains 50 MD samples, 250 independent DEP predictions, and 250 SupraTITO late-window configurations. Per-panel axes retain the sequence-specific SASA range and probability-density scale.

H CONDITION AND SYSTEM-SIZE TRANSFER

H.1 CONCENTRATION-RESOLVED EXPOSURE DYNAMICS

The main text summarizes concentration transfer using the late largest-cluster fraction. Here, each concentration is shown separately so that solvent exposure and assembly timing can be compared without overlapping curves. SASA is normalized by peptide count. SupraTITO curves summarize five independent lag-1000 rollouts from the same initial MD configuration.

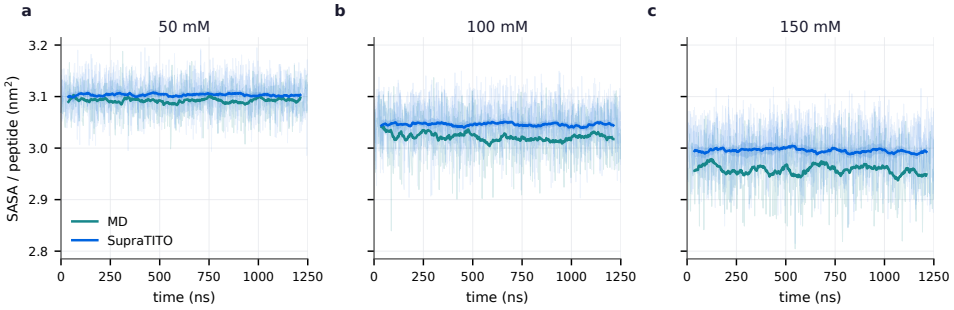


Figure 19: **Concentration-resolved SASA for AG.** Each panel is one 100-peptide system at the indicated concentration. The green lines show the raw MD values and their 31-frame moving average. Faint blue lines show the individual rollouts; the solid blue line and shaded band show the mean and one standard deviation of their 31-frame moving averages.

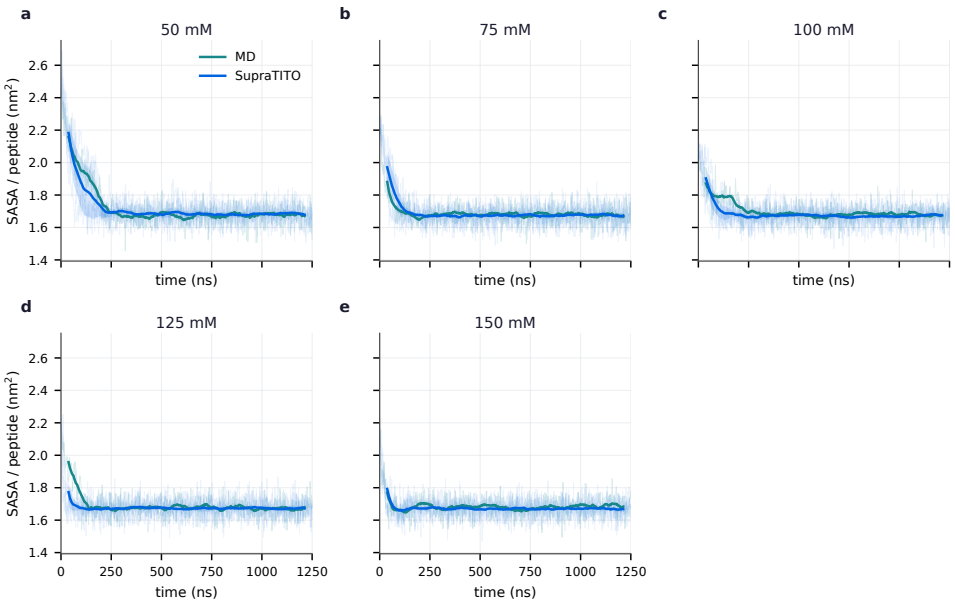


Figure 20: **Concentration-resolved SASA for FF.** Plotting conventions follow figure 19.

Across the 22 fixed-peptide-count conditions, the SASA trajectories in figures 19 to 24 support the late cluster-fraction trends in figure 6. SupraTITO preserves the distinct responses of dispersed, assembled, and concentration-sensitive sequences, including at the most dilute conditions. For QI, the concentration response is recovered even though the peptide sequence is held out from training.

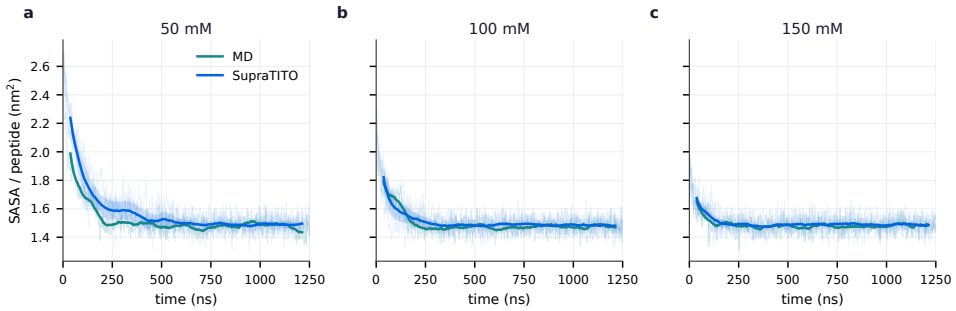


Figure 21: **Concentration-resolved SASA for FH.** Plotting conventions follow figure 19.

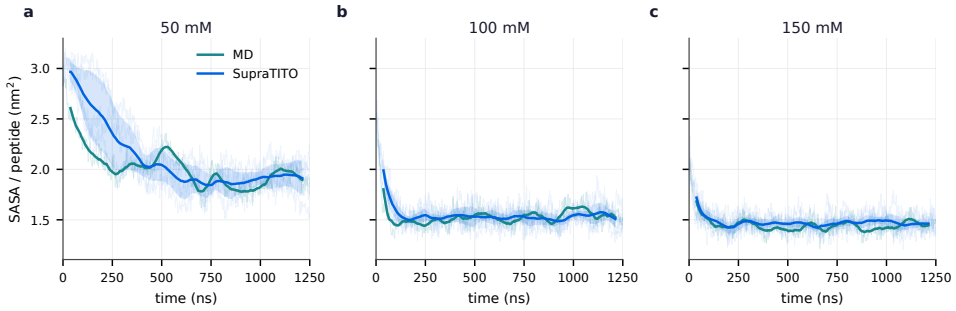


Figure 22: **Concentration-resolved SASA for IK.** Plotting conventions follow figure 19.

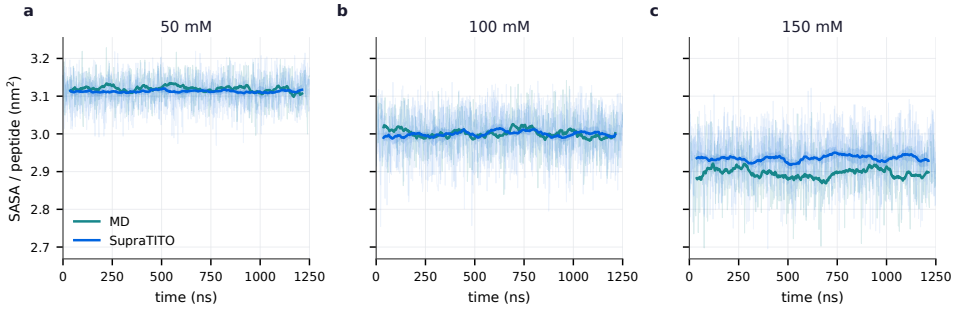


Figure 23: **Concentration-resolved SASA for KR.** Plotting conventions follow figure 19.

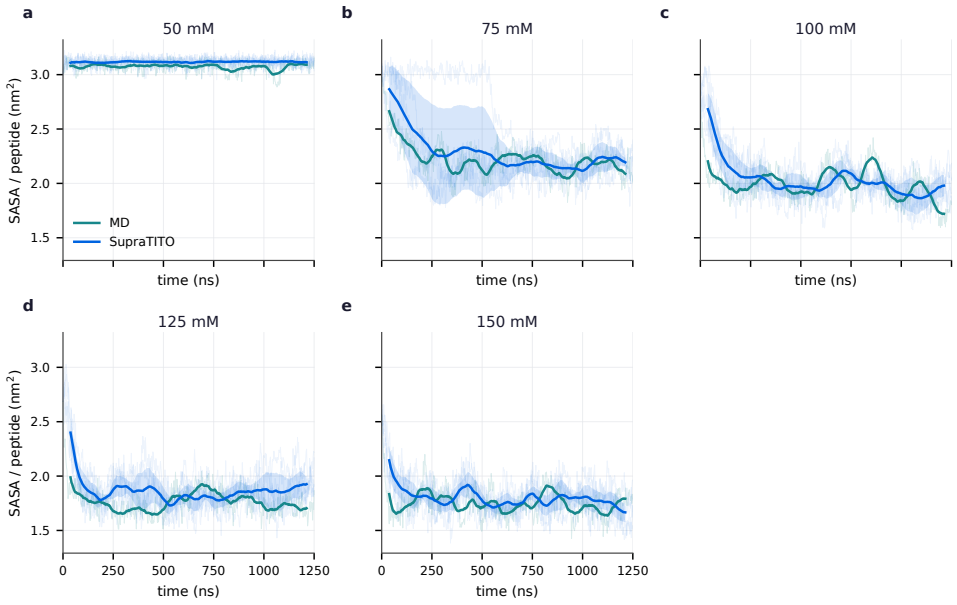


Figure 24: **Concentration-resolved SASA for QI.** Plotting conventions follow figure 19.

H.2 Evaluation at New System Sizes

The fixed-concentration systems change peptide count and box size together, whereas the fixed-box systems change peptide count at constant box length. SupraTITO accepts the resulting variable number of retained sites, although it is trained at only one system size. Figure 25 summarizes the late structural response, and figure 26 shows the corresponding time-resolved SASA.

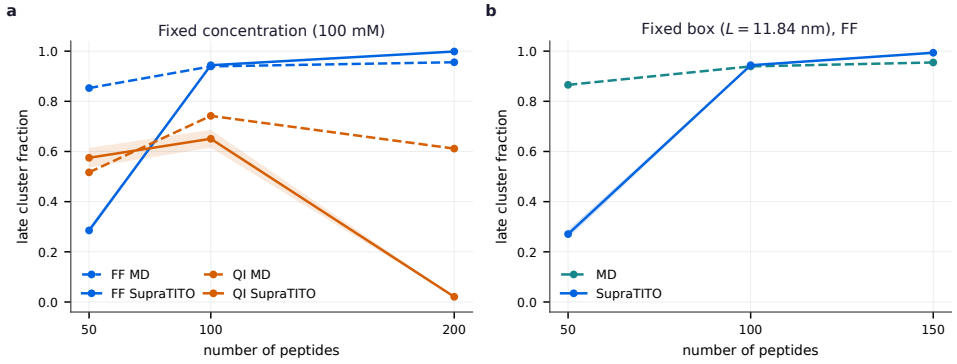


Figure 25: **Late structural response at new system sizes. a**, late largest-cluster fraction when peptide count and box size are changed to keep concentration near 100 mM. **b**, FF peptide-count sweep at fixed box size. Dashed green lines show MD; solid blue lines and bands show the SupraTITO mean and one standard deviation across five independent rollouts.

Transfer across system size is less consistent than transfer across concentration at fixed peptide number. At the smaller size, SupraTITO reproduces the partial assembly of QI but leaves FF too dispersed. At the larger size, it forms a nearly complete FF aggregate but does not recover the partially assembled QI reference. The fixed-box FF sweep shows a similar sensitivity to peptide number. These results separate architectural compatibility with a variable number of sites from learned physical transfer: SupraTITO can be evaluated at new sizes, but training at a single size does not determine all finite-size effects on collective assembly.

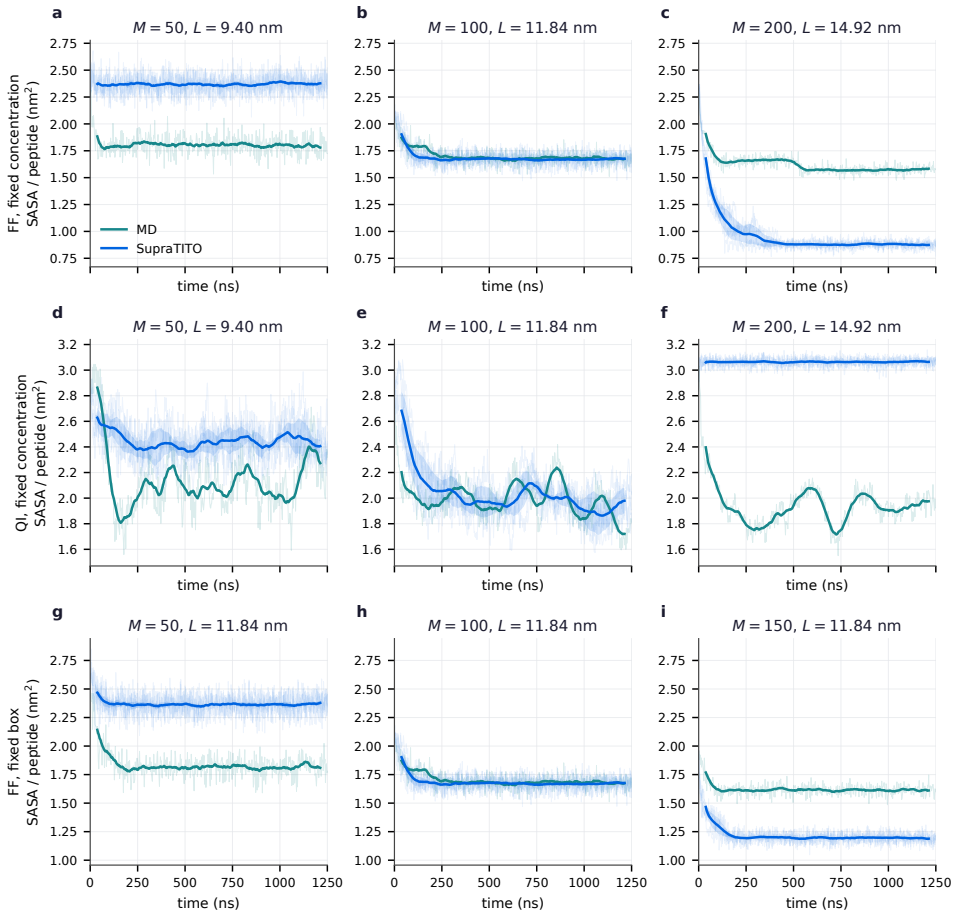


Figure 26: **Condition-specific SASA trajectories at new system sizes.** The top and middle rows vary peptide count and box size together at approximately $100\,\mathrm{mM}$ for FF and QI. The bottom row varies FF peptide count in a fixed $11.84\,\mathrm{nm}$ box. Every panel is one physical system. Line and uncertainty conventions follow figure 19.