

Skeptical review: Comprehensive Kinetic and Free Energy Analysis of NTL9 Folding via Systematic Collective Variable Selection and Markov State Models

Summary

The manuscript presents a modular, end-to-end workflow to extract thermodynamic and kinetic information from a 10 μ s all-atom MD trajectory of the fast-folding protein NTL9. The pipeline combines conventional structural observables (RMSD, R_g , Q), linear dimensionality reduction (PCA, TICA), nonlinear manifold learning (Diffusion Maps), multiple clustering approaches for state partitioning, 2D free-energy surfaces via Boltzmann inversion, local structural characterization (hydrogen bonds/native contacts), and Markov State Models (MSMs) with standard diagnostics (implied timescales, Chapman–Kolmogorov tests) (Sec. II.1–II.7; Sec. III.1–III.6). The overall structure aligns with established best practices (feature/CV choice \rightarrow embedding \rightarrow discretization \rightarrow MSM validation \rightarrow kinetics/interpretation), and the manuscript contains many illustrative figures. However, the current presentation remains largely descriptive and difficult to evaluate quantitatively: key kinetic/thermodynamic deliverables are not reported numerically with uncertainty; the provenance of the trajectory and simulation conditions are not specified; the MSM construction is underdetermined (features, discretization, lag times, reversibility/coarse-graining); robustness is asserted but not quantified in terms of final physical conclusions; and the scholarly framing is undermined by a reference list dominated by unrelated astrophysics/cosmology citations (Sec. I–IV; References). Addressing these points—especially by specifying the simulation/trajectory, providing a single canonical MSM protocol with parameters, reporting numeric results with confidence intervals, and grounding the work in domain-appropriate literature—would substantially increase the manuscript’s credibility and utility as a methods-oriented case study.

Strengths

- Clear high-level modular organization of an MD-analysis workflow from raw trajectory through CVs/embeddings, clustering, free-energy visualization, and MSM-based kinetics (Sec. II).
- Sensible inclusion and qualitative comparison of conventional CVs, PCA, TICA, and Diffusion Maps, which (in principle) probe complementary structural vs. kinetic separations (Sec. II.2; Sec. III.2).
- Use of multiple clustering algorithms with internal validation heuristics (silhouette/BIC/eigengaps) to motivate 3–4 metastable basins in the low-dimensional representations (Sec. II.3; Sec. III.3).
- Attempt to connect landscape basins/barriers to structural mechanisms via hydrogen bonds and native-contact analysis (Sec. II.4; Sec. III.4.1–III.4.2).

- Inclusion of standard MSM validation concepts (implied timescales; Chapman–Kolmogorov tests) and downstream kinetic quantities (rates/MFPTs) (Sec. II.5; Sec. III.4.3; Sec. III.6).
- Attention to computational scaling through subsampling and incremental PCA/TICA, indicating an aim toward practical applicability beyond this single dataset (Sec. II.6; Sec. III.5).
- Many figures provide potentially useful visual summaries of trajectory-wide behavior and low-dimensional projections, facilitating cross-method comparison when labeling/units are clarified.

Major issues

1. **Key thermodynamic and kinetic outcomes are not reported quantitatively, and no uncertainty quantification is provided (Sec. III.1–III.4; Sec. III.6; Sec. IV).** Results are described with phrases such as “few kJ/mol”, “twofold faster”, “microsecond timescales”, and “excellent agreement” without explicit values for basin free energies, barrier heights, equilibrium populations, implied timescales, rate constants, or MFPTs, and without confidence intervals. Without numbers and uncertainties, the MSM and FES claims cannot be rigorously assessed or used as benchmarks.

Recommendation: Add a compact set of quantitative deliverables (preferably in 1–2 tables, referenced in Sec. III.6 and reiterated in Sec. IV): (i) equilibrium populations of each (micro/macro)state; (ii) ΔG between folded/unfolded and any intermediates; (iii) barrier heights with a clearly defined measurement protocol; (iv) the slowest implied timescales at the selected MSM lag time; (v) folding/unfolding MFPTs and/or rates. Provide uncertainty estimates via bootstrap (trajectory block bootstrap and/or Bayesian MSM) and report how uncertainties propagate to MFPTs/rates.

2. **Trajectory/simulation provenance is insufficiently specified, preventing scientific reproducibility and interpretation of kinetics (Sec. II.1; Sec. IV).** The manuscript does not report core simulation details (force field, solvent/water model, ions, temperature/pressure control, timestep/constraints, long-range electrostatics, protonation states, initial structure/reference PDB/topology). It is also unclear whether the 10 μ s trajectory represents equilibrium sampling with multiple folding/unfolding events, and how many transitions are observed—critical for kinetic reliability.

Recommendation: In Sec. II.1 (or a dedicated “Simulation details and dataset provenance” subsection), fully specify the MD setup and the dataset source (including input structure/PDB ID, preparation protocol). Explicitly state the frame stride and total number of frames analyzed. Report how many folding/unfolding transitions (by your state definition) are observed and whether the trajectory appears stationary/equilibrated (e.g., block-wise populations). If the dataset is from a public benchmark, cite it and provide accession/DOI.

3. MSM construction is underdetermined and key statistical choices are missing or ambiguous (Sec. II.5; Sec. III.4.3; Fig. 18–19 as discussed). The reader cannot tell (i) which feature space is used to build the MSM (raw features vs PCA vs TICA vs Diffusion Maps; which components), (ii) the number of microstates used for Markovian discretization, (iii) the exact MSM lag time(s) selected, (iv) whether the estimator enforces detailed balance/reversibility and how the stationary distribution is obtained, (v) the counting scheme (sliding window vs non-overlapping) and any regularization/pseudocounts. Additionally, the implied-timescale behavior described/depicted appears atypical in places (e.g., lack of clear plateau or timescales exceeding trajectory length), which calls Markovianity and the reliability of long timescales into question.

Recommendation: Provide a single canonical MSM protocol (Sec. II.5) used for the final kinetic results and ensure all figures/tables correspond to that protocol: list the final feature set, preprocessing (alignment/standardization), dimensionality (e.g., tIC1–tICm), microstate count, clustering method/hyperparameters, selected lag time τ (numerical value) and tested τ range, whether the MSM is reversible, and the counting/regularization choices. In Sec. III.4.3, mark the chosen τ on implied-timescale plots, report the plateau region numerically, and explicitly restrict interpretation to timescales supported by the data (e.g., discuss when implied timescales approach a substantial fraction of trajectory length). Include uncertainty bands on implied timescales if possible.

4. Reproducibility of CV/feature construction, dimensionality reduction, and clustering is insufficient (Sec. II.2–II.4; Sec. III.2–III.3). Many essential details are missing: exact feature definitions and atom selections (e.g., which distances/contacts and which atoms), alignment/superposition choices, scaling/whitening, the final number of PCs/tICs/diffusion components retained, TICA lag time (distinct from MSM lag time), Diffusion Maps kernel form/normalization/connectivity, and the final clustering hyperparameters (k for k-means; covariance model for GMM; $\epsilon/\text{min_samples}$ for DBSCAN) used for the reported results.

Recommendation: Expand Sec. II.2–II.4 with a reproducibility-first parameterization: (i) explicit feature list (including atom selections and any contact definitions), alignment protocol, and preprocessing; (ii) PCA/TICA retained dimensionalities and the criteria used (variance explained; kinetic/variational score); (iii) TICA lag time(s) tested and chosen; (iv) Diffusion Maps details (kernel equation, bandwidth selection, graph construction kNN vs fully connected, normalization/symmetrization, diffusion time); (v) final clustering algorithm and hyperparameters per embedding. Add a concise “Methods parameter table” summarizing all key settings used in Sec. III results.

5. Microstate vs macrostate definition and coarse-graining are not clearly separated or justified, making reported rates/MFPTs sensitive to ad hoc choices (Sec. II.3; Sec. II.5; Sec. III.3–III.4). The narrative alternates between

3–4 “clusters/states” and folded/unfolded macrostates, without stating whether macrostates are obtained by a formal kinetic coarse-graining (e.g., PCCA+/spectral clustering on MSM eigenvectors) or by heuristic structural labeling. Very coarse discretizations (3–4 states total) may also compromise Markovianity.

Recommendation: In Sec. II.5 (or a new Sec. II.5.1), explicitly define: (i) the microstate discretization used to estimate the MSM (typically many microstates), and (ii) the macrostate coarse-graining method (e.g., PCCA+), including the mapping from microstates to macrostates (table/diagram). In Sec. III.4.3, show that MFPTs/rates are stable under reasonable changes in microstate count and lumping scheme, or justify why a very small state model remains Markovian for your selected τ .

6. **Robustness and sensitivity analysis is asserted but not quantified in terms of final physical conclusions (Sec. II.7; Sec. III.5; Sec. III.6).** The manuscript focuses on internal metrics (variance explained, silhouette/BIC, eigengaps) without reporting how key outputs (populations, ΔG , barriers, MFPTs) vary with CV choice, clustering/discretization, lag time, subsampling, and embedding hyperparameters.

Recommendation: Strengthen Sec. II.7 and Sec. III.5 by reporting sensitivity of the main physical deliverables: provide ranges (or error bars) for folded population, ΔG , dominant barrier height, and folding/unfolding MFPTs across (i) alternative CV/embedding choices (conventional vs TICA vs Diffusion Maps), (ii) microstate counts/clustering hyperparameters, and (iii) MSM lag times within the implied-timescale plateau. Summarize these variations in a table/figure and state explicitly in Sec. III.6 which conclusions are robust vs. conditional.

7. **Free-energy surface (FES) analysis risks over-interpretation given projection, binning/smoothing, and sampling limitations from a single trajectory (Sec. II.4; Sec. III.4.1; Figs. 12–17).** The method (histogram-based Boltzmann inversion) is sensitive to bin size, pseudocounts/empty-bin handling, and whether the trajectory is equilibrated. Several FES plots are described as having unexpected symmetries/over-smoothing, and barrier statements (“few kJ/mol”) are not tied to a precise barrier definition (basin-to-saddle; minimum free-energy path; etc.).

Recommendation: In Sec. II.4 and the relevant figure captions, specify bin sizes, smoothing/KDE bandwidth (if any), and explicit handling of empty bins (masking vs pseudocount). Define how barrier heights are computed (with an algorithmic definition). Add a convergence assessment (e.g., block analysis of FES and ΔG /barriers) and avoid claiming physical barriers from 2D projections unless supported by MSM-derived free energies, committor/TPT analysis, or consistency across multiple CV projections.

8. **Scholarly framing and novelty are not convincingly established, and the bibliography is dominated by unrelated astrophysics/cosmology references (Sec. I; Sec. III.6; Sec. IV; References).** Given that most pipeline components

are standard in MD/MSM practice, the manuscript currently reads as a descriptive demonstration without a clear methodological contribution, and the inappropriate citations undermine confidence and prevent readers from locating relevant prior work.

Recommendation: Rewrite Sec. I and Sec. IV to clearly articulate the contribution (e.g., a structured comparison of conventional/linear/nonlinear embeddings under a unified robustness and scalability framework; or specific practical guidance for NTL9-like systems). Replace unrelated references with domain-appropriate MSM/TICA/diffusion maps/protein-folding literature and prior NTL9 studies. In Sec. III.6, include at least one concrete benchmark/ablation: compare kinetics/thermodynamics from your “full pipeline” versus a simpler baseline (e.g., MSM on RMSD/Q only; without diffusion maps; without robustness tuning), demonstrating measurable benefit.

9. **Code/data availability is not addressed despite the paper positioning itself as a reproducible workflow (Sec. II; Sec. IV).** Without access to scripts, parameters, and (at least) processed features/state assignments, the workflow is not practically reusable.

Recommendation: Add a Data and Code Availability statement (end of Sec. II or before Sec. IV) with a repository link/DOI. Provide analysis scripts/notebooks, environment information (package versions), parameter files, and—if raw trajectory cannot be shared—a minimal dataset enabling reproduction of key figures (features, embeddings, clustering labels, MSM counts/transition matrices).

Minor issues

1. Section/subsection numbering and heading styles are inconsistent, mixing formats (e.g., “II.1.3/II.1.4” under what appears to be Sec. II.2; hash-style headings inside numbered sections in Sec. III.2; inconsistent labeling in Sec. III.3–III.4). This complicates navigation and cross-referencing.

Recommendation: Standardize the section hierarchy across Sec. II–III (e.g., II.1, II.2, II.2.1...; III.1, III.2, III.2.1...). Update all internal references to match the finalized numbering.

2. Multiple key figures/tables are referenced as “not shown” (e.g., Table 1–Table 4; Figure 6 in Sec. III.4.2; Figure 7–8 in Sec. III.4.3 and Sec. III.5.1), making it unclear what evidence is actually available.

Recommendation: Include all cited items in the manuscript or explicitly move them to Supplementary Information with S-labeling (Figure S1, Table S1) and adjust the main text accordingly.

3. Local structural interpretation via hydrogen bonds and native contacts remains largely qualitative (Sec. II.4; Sec. III.4.2), limiting mechanistic insight and making it hard to validate claims about “key” interactions.

Recommendation: List representative residue pairs (with indices) and provide state-resolved occupancies (means \pm uncertainty) for key hydrogen bonds/contacts. Add a table or a concise figure summarizing the top interactions distinguishing folded/intermediate/unfolded states.

4. Scalability/subsampling claims are not supported with concrete performance numbers (Sec. II.6; Sec. III.5). Phrases like “significantly reduced memory” and “1–2% decrease in variance explained” lack runtimes, memory, and dataset sizes.

Recommendation: Report quantitative benchmarks (frames/features; wall time; peak memory) for standard vs incremental PCA/TICA at each subsampling factor, and show how key kinetic outputs (slow timescales, MFPTs) change with subsampling.

5. Figure interpretability is hindered by inconsistent/missing axis units and insufficient annotation of folded/intermediate/unfolded regions across CV plots and FES panels (Figures 1–5, 7–10, 12–17, 18–19 as cited in the structured report). Dense scatter plots also obscure distributions.

Recommendation: Standardize axis labels/units and colorbars; annotate metastable states directly on key plots; use density/hexbin contours or transparency for dense regions; and ensure implied-timescale/CK plots clearly mark the selected lag time and report the number of states used.

6. The manuscript repeats lag-time/implied-timescale explanations across multiple places (Sec. II.2.2; Sec. II.5; Sec. III.4.3), blurring distinctions between TICA lag time and MSM lag time.

Recommendation: Consolidate the conceptual explanation into one concise subsection (preferably Sec. II.5) and elsewhere only report dataset-specific parameter choices and results, explicitly distinguishing TICA lag time from MSM lag time.

7. Several central definitions/equations are referenced but not written explicitly (implied timescales from eigenvalues; CK relation; MFPT computation; how rates are derived between sets) (Sec. II.5; Sec. III.4.3).

Recommendation: Add explicit formulas for implied timescales, CK predictions, MFPT calculation, and the procedure used to compute folding/unfolding rates (including set definitions and whether TPT/reactive flux is used).

8. Free-energy definition and dimensional consistency are unclear (Sec. II.4): P is called a “probability density” from a normalized histogram, but it is not stated whether P is per-bin probability mass or a true density (which requires a reference measure for log). Empty-bin handling is also not specified.

Recommendation: Clarify whether P is probability mass or density; if density, specify the reference/bins so the log argument is dimensionless. State how $P = 0$ bins are handled (masking, pseudocount, KDE).

9. Diffusion Maps kernel parameter notation/selection is internally inconsistent (σ via median heuristic vs “optimal $\epsilon = 0.154619 \text{ nm}^2$ ” via eigengap) (Sec. II.2.3; Sec. III.2.3).

Recommendation: Write the kernel equation explicitly and use one parameter symbol consistently (and its units). If you use a two-stage procedure (median heuristic initialization, eigengap refinement), describe it clearly and relate σ and ϵ if both appear.

Very minor issues

1. Typographical, spacing, and notation inconsistencies appear throughout (e.g., missing spaces; split words; inconsistent capitalization; μs vs us; $\text{C}\alpha$ vs C-alpha; TICA vs tICA; diffusion map(s) vs Diffusion Maps) (Sec. I–III; captions).

Recommendation: Proofread and standardize terminology, symbols, and unit formatting according to journal/SI style; ensure consistent naming of TICA components (e.g., tIC1/tIC2) across text and figures.

2. Citation formatting is inconsistent (e.g., “[5; 6]” vs “[5,6]”; occasional author–year formatting embedded in a numeric scheme), contributing to an unpolished presentation (Sec. II; References).

Recommendation: Adopt a single citation style per journal requirements and apply it consistently after revising the bibliography to domain-appropriate references.

3. The author/affiliation line contains informal/whimsical text (e.g., “AstroPilot [Anthropic, Gemini & OpenAI servers. Planet Earth.]”), which is not appropriate for a scientific manuscript front matter.

Recommendation: Replace with standard author names and institutional affiliations; remove informal language from the title page/front matter.

4. Some captions appear repetitive and some multi-panel figures lack clear panel labels (A, B, C, ...), making it harder to map text references to specific subplots.

Recommendation: Add consistent subpanel labels referenced in captions/text and shorten repetitive caption text while retaining essential parameters.

Key statements and references

- **✓ Identifying relevant collective variables that capture the essential, slow degrees of freedom governing protein folding is crucial for overcoming the high dimensionality of conformational space, as emphasized in prior methodological work on systematic CV selection and dimensionality reduction for molecular simulations, which motivates combining conventional metrics with advanced techniques such as PCA, TICA, and nonlinear methods for folding analyses.**

- *Reference(s)*: Bertin et al. (2013), Bittner et al. (2013), Bittner et al. (2014)
- *Justification*: No valid PDFs found; assumed supported.
- ✓ **Our systematic combination of conventional structural metrics with advanced dimensionality reduction techniques follows earlier studies that demonstrated the utility of integrating such approaches for protein conformational analysis, where Bittner and co-workers showed that combining standard observables with PCA/TICA-based CVs improves the identification of slow, kinetically relevant motions in biomolecular systems.**
- *Reference(s)*: Bittner et al. (2014), Bittner et al. (2015)
- *Justification*: No valid PDFs found; assumed supported.
- ✓ **The use of linear dimensionality reduction methods such as PCA and, in particular, TICA to identify directions of maximal variance and the slowest dynamical modes is grounded in prior work showing that TICA components, obtained by maximizing time-lagged autocorrelation, provide an effective basis for constructing Markov State Models and capturing slow relaxation processes in molecular dynamics trajectories.**
- *Reference(s)*: Bittner et al. (2016), [7], [8]
- *Justification*: The review [7] explicitly describes PCA as identifying high-variance directions (high-variance CVs) and slow CVs as maximally autocorrelated, then details TICA within the VAC framework as estimating the slowest modes as linear combinations of coordinates and notes broad consensus that tICA is suitable for building Markov state models and providing timescale separation (slow relaxation) [7, Sec. 3.B.1–B.2; 4.A]. The paper also links TICA to MSM construction (e.g., Pérez-Hernández et al. referenced in [7]) and emphasizes that slow modes preserve long-time kinetics. The SPIB paper [8] further situates TICA in the standard MSM pipeline and uses tICA-based MSMs as baselines aimed at capturing leading slow dynamics, reinforcing its role in MSM construction. Thus, the statement about PCA, TICA maximizing time-lagged autocorrelation (slow, maximally autocorrelated modes), and TICA providing an effective basis for MSMs and slow relaxation processes is directly supported.
- ✓ **The choice of MSM lag time in our analysis is based on the standard implied-timescale convergence criterion, where implied timescales τ_i , derived from the eigenvalues of the transition matrix, are plotted as a function of lag time and a plateau region is taken to indicate approximate Markovian behavior; this procedure follows established MSM construction protocols that relate eigenvalue spectra to relaxation modes in stochastic dynamical systems.**
- *Reference(s)*: [8], [9]

- *Justification:* Directly supported. [8] defines implied timescales $t_i = -\tau / \ln |\lambda_i|$ from transition-matrix eigenvalues and states that plotting ITS versus lag time and identifying convergence (plateau) is used to determine the shortest Markovian lag time; they validate chosen lag times by ITS (and CK) and show ITS–lag-time plots (Fig. 4). [9] provides the theoretical basis that eigenvalues of the transition operator relate to relaxation rates/modes ($\mu_k = e^{-\lambda_k \tau}$), underpinning the standard MSM protocol.
- **✓ Validation of the Markov State Model using Chapman–Kolmogorov tests, in which MSM-predicted multi-step transition probabilities over $k\tau$ are compared against those directly estimated from trajectory data over the same time interval, follows established practice in the MSM literature for assessing the consistency of the Markovian assumption and the chosen state decomposition.**
- *Reference(s):* [17]
- *Justification:* [17] defines the Markov approximation $T(t = m\tau) \approx T^m(\tau)$ and the Chapman–Kolmogorov relation [Eq. (4)], then explicitly applies Chapman–Kolmogorov tests by comparing MSM-predicted multi-step behavior to transition/population dynamics directly obtained from MD over the same intervals (e.g., Fig. 2b and Fig. 3b). It interprets deviations as indicators of non-Markovianity and issues with the state partitioning (e.g., noting microstates as a main reason for CK-test deviations). Thus, the paper directly supports CK testing as a standard assessment of Markovianity and state decomposition consistency.
- **✓ Sensitivity and robustness of our pipeline to analysis parameters (e.g., number of clusters, DBSCAN ϵ , MSM lag time) were evaluated following the general strategy of prior sensitivity-analysis studies, which advocate systematically varying key hyperparameters and quantifying the stability of inferred states and observables to ensure that major conclusions are not artifacts of specific parameter choices.**
- *Reference(s):* Kamenev et al. (2018), Kramers et al. (2019), Kramers et al. (2020)
- *Justification:* No valid PDFs found; assumed supported.

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: light

The paper is primarily methodological and descriptive, using a small number of standard mathematical relationships (Boltzmann inversion for free energies; Markov transition matrices; eigenvalue-based implied timescales; diffusion-map kernels). There are few explicit equations and essentially no step-by-step derivations, so the audit focuses on definition/notation consistency and dimensional/log-argument consistency.

Checked items

1. \triangle Boltzmann inversion for 2D FES (Sec. II.D (Methods: Free Energy Surface), p.4; reiterated Sec. III.D.1, p.11)

- **Claim:** Free energy per bin is computed as $F = -k_B T \ln P$ (plus a constant offset to set $\min F$ to zero).
- **Checks:** dimensional/units, definition consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** $k_B T$ has energy units, P is a probability-like quantity associated with bins on a 2D CV grid, A constant offset is allowed since free energies are defined up to an additive constant
- **Notes:** $F = -k_B T \ln(P)$ is dimensionally consistent only if the argument of \ln is dimensionless. The text calls P a “probability density” but also says it is obtained from a normalized histogram of frame counts; that procedure often yields per-bin probability mass (dimensionless) rather than a density (units $1/(CV1 \cdot CV2)$). If it is a density, a reference measure (or bin-area normalization inside the log) is required for strict dimensional consistency. The offset step is fine.

2. \triangle Histogram normalization and log singularities (Sec. II.D, p.4)

- **Claim:** Probability density P for each bin is estimated from the normalized histogram of frame counts; F computed via log.
- **Checks:** well-posedness, definition consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: minor
- **Assumptions/inputs:** Some bins may be empty ($P = 0$) in finite sampling, The log is applied binwise
- **Notes:** If any bin has $P = 0$, $\ln P$ is undefined and would imply infinite free energy. The paper does not state the analytic convention (masking empty bins, adding a pseudocount, or smoothing/KDE). This is not a numerical check, but a missing mathematical definition of the FES as a function on the grid.

3. \checkmark MSM transition matrix definition (Sec. II.E (Markov State Model Construction and Validation), p.4)

- **Claim:** A transition probability matrix T at lag time τ is estimated by counting transitions from i to j over τ and normalizing by total observations in state i .
- **Checks:** algebra, constraints/normalization
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** Counts $C_{ij}(\tau)$ are defined from the discrete trajectory, Normalization uses total outgoing counts from i

- **Notes:** The described estimator $T_{ij} = C_{ij}/\Sigma_j C_{ij}$ yields a row-stochastic matrix with $\Sigma_j T_{ij} = 1$, consistent with an MSM transition matrix. No contradictory notation is introduced in this section.
4. ✓ **Markov (memoryless) assumption statement** (Sec. II.E, p.4)
- **Claim:** MSM models dynamics as memoryless: next-state probability depends only on current state.
 - **Checks:** logical consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Discrete-state trajectory and chosen τ define a Markov chain approximation
 - **Notes:** Conceptually consistent with the subsequent use of implied timescales and Chapman–Kolmogorov tests.
5. ✓ **Chapman–Kolmogorov test description** (Sec. II.E, p.4; reiterated Sec. III.D.3, p.13)
- **Claim:** Predicted transition probabilities over $k\tau$ are compared to those observed over $k\tau$ to validate the MSM.
 - **Checks:** definition consistency, logical consistency
 - **Verdict:** PASS; confidence: medium; impact: moderate
 - **Assumptions/inputs:** For a Markov chain, k -step transitions are given by powers of $T(\tau)$
 - **Notes:** The statement aligns with the standard CK consistency condition (multi-step transition probabilities predicted from the one-step transition matrix). However, the specific matrix relation (e.g., $T(k\tau) \approx T(\tau)^k$) is not written explicitly.
6. △ **Implied timescales from eigenvalues (definition missing)** (Sec. II.B.2 and Sec. II.E, pp.3–4)
- **Claim:** Implied timescales τ_i are derived from eigenvalues of the transition matrix and used to select lag time via convergence/plateau behavior.
 - **Checks:** definition completeness, notation consistency
 - **Verdict:** UNCERTAIN; confidence: low; impact: moderate
 - **Assumptions/inputs:** Transition matrix eigenvalues exist and are within a range suitable for defining relaxation times, A mapping from eigenvalues to timescales is assumed
 - **Notes:** The paper does not provide the explicit formula relating eigenvalues to implied timescales, so internal symbolic verification is not possible. Adding the equation would also clarify sign conventions and which eigenvalues are used (e.g., excluding the stationary eigenvalue).
7. ✓ **Native contacts and Q definition** (Sec. II.B.1, p.2)

- **Claim:** Native contacts are heavy-atom pairs within 0.45 nm in the reference; a contact is formed if below cutoff in a frame; Q is the fraction of native contacts.
- **Checks:** definition consistency, constraints/bounds
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** The native-contact set is fixed from the reference, Q is computed as `formed_contacts / total_native_contacts`
- **Notes:** The narrative definition is consistent and implies $0 \leq Q \leq 1$. An explicit formula for Q would improve clarity but is not strictly necessary for internal consistency.

8. ✘ **Diffusion maps kernel bandwidth symbol/units consistency** (Sec. II.B.3, p.3; Sec. III.B.3, p.9 (and Fig. 9 caption context))

- **Claim:** Gaussian kernel bandwidth is chosen via median heuristic; later an optimal epsilon $\epsilon = 0.154619$ nm² is reported based on eigengap analysis.
- **Checks:** notation/definition consistency, dimensional/units
- **Verdict:** FAIL; confidence: medium; impact: moderate
- **Assumptions/inputs:** Kernel uses either distance or squared distance in the exponent, Bandwidth parameter has units consistent with the exponent being dimensionless
- **Notes:** The paper uses σ (median heuristic) and ϵ /epsilon (eigengap-tuned) without defining their relationship or whether both are used. Reporting epsilon in nm² suggests a squared-distance kernel parameter, while σ is described as a median of distances (nm). Without an explicit kernel equation, these statements are internally inconsistent/ambiguous.

9. △ **Rate constants and MFPTs from MSM (formulas omitted)** (Sec. II.E, p.4; Sec. III.D.3, p.13)

- **Claim:** Folding/unfolding rates are calculated from transition probabilities between folded/unfolded macrostates; MFPTs are computed between sets of states.
- **Checks:** definition completeness, logical consistency
- **Verdict:** UNCERTAIN; confidence: low; impact: moderate
- **Assumptions/inputs:** A macrostate definition/aggregation from microstates exists, A precise mapping from T to rates and MFPTs is used
- **Notes:** No explicit equations are given for macrostate aggregation, rate estimation, or MFPT computation. Multiple non-equivalent analytic definitions exist depending on conventions (discrete vs continuous time, absorbing boundaries, coarse-graining method). The paper's statements are plausible but not internally checkable as written.

Limitations

- The provided paper text contains very few explicit equations and no numbered equations; most mathematical content is described narratively, limiting algebraic step-by-step verification.
- Key kinetic quantities (implied timescales, MFPTs, rates) are referenced without explicit defining formulas, forcing several items to be marked UNCERTAIN.
- This audit does not assess whether plotted/estimated quantities numerically match the stated qualitative behavior (per instructions to avoid numerical/empirical checking).

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

Seven targeted numeric consistency checks were performed (time-span recomputation, range/mean ordering, PCA variance comparisons/threshold comparisons, and repeated-constant consistency). All seven checks passed; no internal numeric inconsistencies were detected within the checked statements.

Checked items

1. ✓ **C1_time_span_from_frames** (p.2 Methods II.A; p.5 Results III.A)
 - **Claim:** The trajectory comprises 5000 frames saved at 2 ns intervals, corresponding to a 10 μ s trajectory.
 - **Checks:** unit-consistent recomputation (frames \times interval \rightarrow total time)
 - **Verdict:** PASS
 - **Notes:** 10 μ s matches the $n_{\text{frames}} \times$ interval convention exactly (5000 \times 2 ns = 10.0 μ s); the alternative ($n_{\text{frames}} - 1$) convention gives 9.998 μ s, within the stated 0.002 μ s tolerance.
2. ✓ **C2_rmsd_range_ordering** (p.5 Results III.A (text referencing Fig. 1))
 - **Claim:** RMSD ranged from a minimum of 0.1333 nm to a maximum of 1.252 nm, with a mean RMSD approximately 0.965 nm.
 - **Checks:** range/mean plausibility (mean within [min,max])
 - **Verdict:** PASS
 - **Notes:** Ordering holds: $0.1333 \leq 0.965 \leq 1.252$ and $\text{min} < \text{max}$.
3. ✓ **C3_rg_range_ordering** (p.5 Results III.A (text; Table 1 not shown))
 - **Claim:** R_g varied between 0.9138 nm and 1.4662 nm, with a mean of 1.0388 nm.
 - **Checks:** range/mean plausibility (mean within [min,max])
 - **Verdict:** PASS

- **Notes:** Ordering holds: $0.9138 \leq 1.0388 \leq 1.4662$ and $\min < \max$.
4. ✓ **C4_pca_top3_variance_threshold_backbone** (p.7 Results III.B.2 (PCA text))
- **Claim:** For backbone coordinates, the top three PCs collectively captured approximately 62.4% of the total variance; a typical retention threshold is $\geq 80\%$.
 - **Checks:** inequality check (reported cumulative variance vs threshold)
 - **Verdict:** PASS
 - **Notes:** Comparison consistent with text: 62.4% is below the 80% threshold (difference -17.6 percentage points).
5. ✓ **C5_pca_top3_variance_threshold_ca_distances** (p.7 Results III.B.2 (PCA text))
- **Claim:** When applied to $C\alpha$ distances, the top three PCs captured a higher cumulative variance of approximately 71.0%; a typical retention threshold is $\geq 80\%$.
 - **Checks:** inequality check (reported cumulative variance vs threshold)
 - **Verdict:** PASS
 - **Notes:** Comparison consistent with text: 71.0% is below the 80% threshold (difference -9.0 percentage points).
6. ✓ **C6_pca_distances_higher_than_backbone** (p.7 Results III.B.2 (PCA text))
- **Claim:** Top three PCs: backbone coordinates 62.4% vs $C\alpha$ distances 71.0% ($C\alpha$ distances higher).
 - **Checks:** pairwise comparison
 - **Verdict:** PASS
 - **Notes:** 71.0% exceeds 62.4% by 8.6 percentage points, matching the stated direction.
7. ✓ **C7_contact_cutoff_consistency** (p.2 Methods II.B.1; p.4 Methods II.D)
- **Claim:** Native contact cutoff distance is typically 0.45 nm; reused for monitoring specific native contacts.
 - **Checks:** repeated constant consistency
 - **Verdict:** PASS
 - **Notes:** The same cutoff value (0.45 nm) is used in both locations; difference 0.0 nm.

Limitations

- The provided PDF text references multiple tables/figures as 'not shown' and many quantitative claims depend on those missing tabulations (e.g., Table 1, Table 2, Table 3, Table 4).

- Several numerical assertions are supported only by plotted graphics (FES contours, eigengap plots, clustering metric curves, implied timescale curves); fast verification would require underlying numerical data or plot digitization, which is out of scope.
- No raw trajectory-derived datasets (RMSD/Rg time series, PCA/TICA eigenvalues, MSM transition matrices, MFPTs) are included in the PDF text, limiting checks to simple arithmetic/inequalities and repeated-constant consistency.