ICML
International Conference
On Machine Learning

Protein Language Model Embeddings Improve Generalization of Implicit Transfer Operators

Panagiotis Antoniadis [1]

Beatrice Pavesi [2]

Simon Olsson* [2]

simonols@chalmers.se

Ole Winther* [1 3]

ole.winther@bio.ku.dk

[1] University of Copenhagen

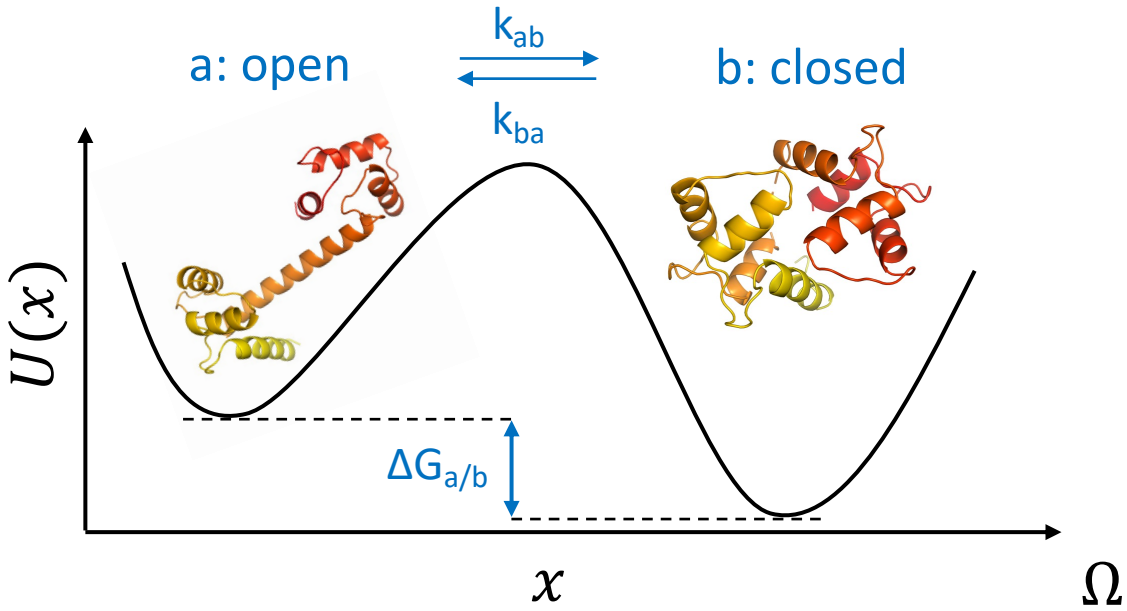
[2] Chalmers University of Technology

[3] Technical University of Denmark

*Equal contribution, ordered alphabetically.

Predicting molecular observables reveals function

- Drug design:
  - How well a drug binds to a target (affinity)
  - How long does a drug stay bound (k)
- Biophysics:
  - What is the stability of a protein (ΔG)
  - How does the stability change with mutations (ΔΔG)



Observables are computed as expectations over high-dimensional molecular distributions:

- Stationary: $$O_{\mathrm{f}} = \mathbb{E}_{\mu}[\mathrm{f}(\mathbf{x})]$$

- Dynamic: $$O_{\mathrm{f,h}}(\mathrm{t}, \mathrm{t}+\mathrm{N}\tau) = \mathbb{E}_{\mathrm{x_t}\sim\mu}\left[\mathbb{E}_{\mathrm{x_{t+N\tau}}\sim \mathrm{p_\tau}(\mathrm{x_{t+N\tau}}|\mathrm{x_t})}[f(\mathrm{x_t})h(\mathrm{x_{t+N\tau}})]\right]$$

Computing molecular observables

$$O_{\mathrm{f}} = \mathbb{E}_{\mu}[\mathrm{f}(\mathbf{x})] \qquad O_{\mathrm{f,h}}(\mathrm{t}, \mathrm{t} + \mathrm{N}\tau) = \mathbb{E}_{\mathrm{x_t} \sim \mu}\left[\mathbb{E}_{\mathrm{x_{t+\Delta t}} \sim \mathrm{p}_\tau(\mathrm{x_{t+N\tau}}|\mathrm{x_t})}[f(\mathrm{x_t})h(\mathrm{x_{t+N\tau}})]\right]$$

Boltzmann distribution:

$$\mu(\mathbf{x}) = \mathcal{Z}^{-1} \exp\left(-\beta U(\mathbf{x})\right)$$

equilibrium distribution of molecular dynamics.

Transition Density:

$$p_\tau(\mathbf{x_{t+N\tau}}|\mathbf{x_t})$$

encodes the physical dynamics of the system.

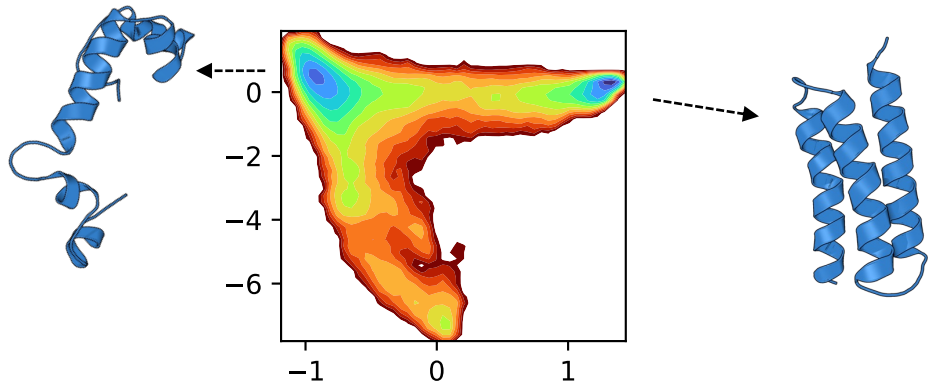
Limitations:

- Sampling i.i.d. samples from $\mu(\mathbf{x})$ is challenging.
- Integration time steps required for stable simulations are tiny compared to time-scales of interest.

How to efficiently compute molecular observables?

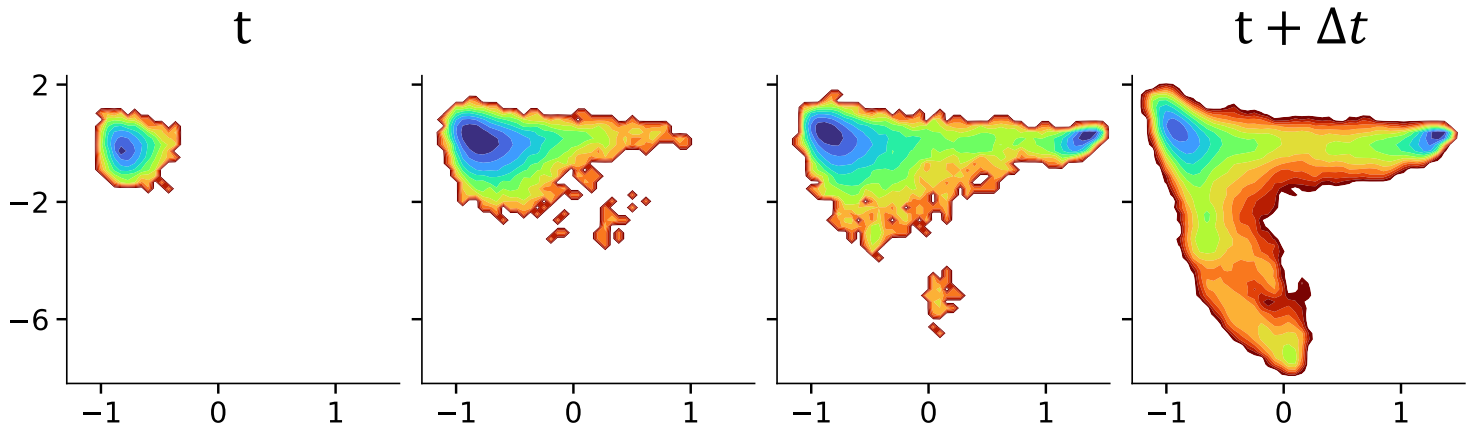
Boltzmann Generators/Emulators: approximate the Boltzmann distribution $\mu(\mathbf{x})$

...LAAFEKEI...



Implicit Transfer Operators: approximate the transition density $p_{\Delta t}(\mathbf{x_{t+\Delta t}}|\mathbf{x_t})$

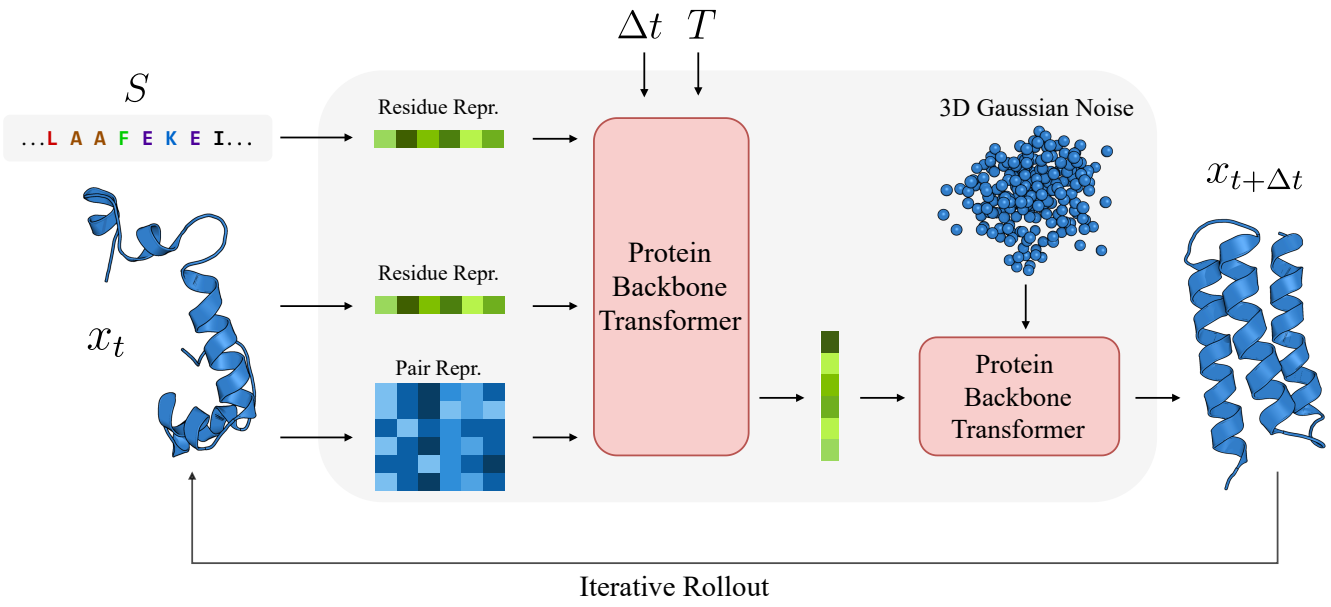
...LAAFEKEI...

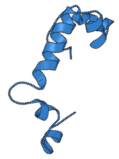
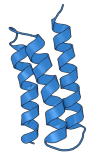


ITO learning with multiple conditioning

TITO: a flow-based model that approximates

$$p(x_{t+\Delta t} \mid x_t, \Delta t, S, T)$$



x_t	$x_{t+\Delta t}$	Δt (ns)	S	T (K)
		50	. . L K E I . .	320
		100	. . L S D E . .	370
		30	. . K N E L . .	420

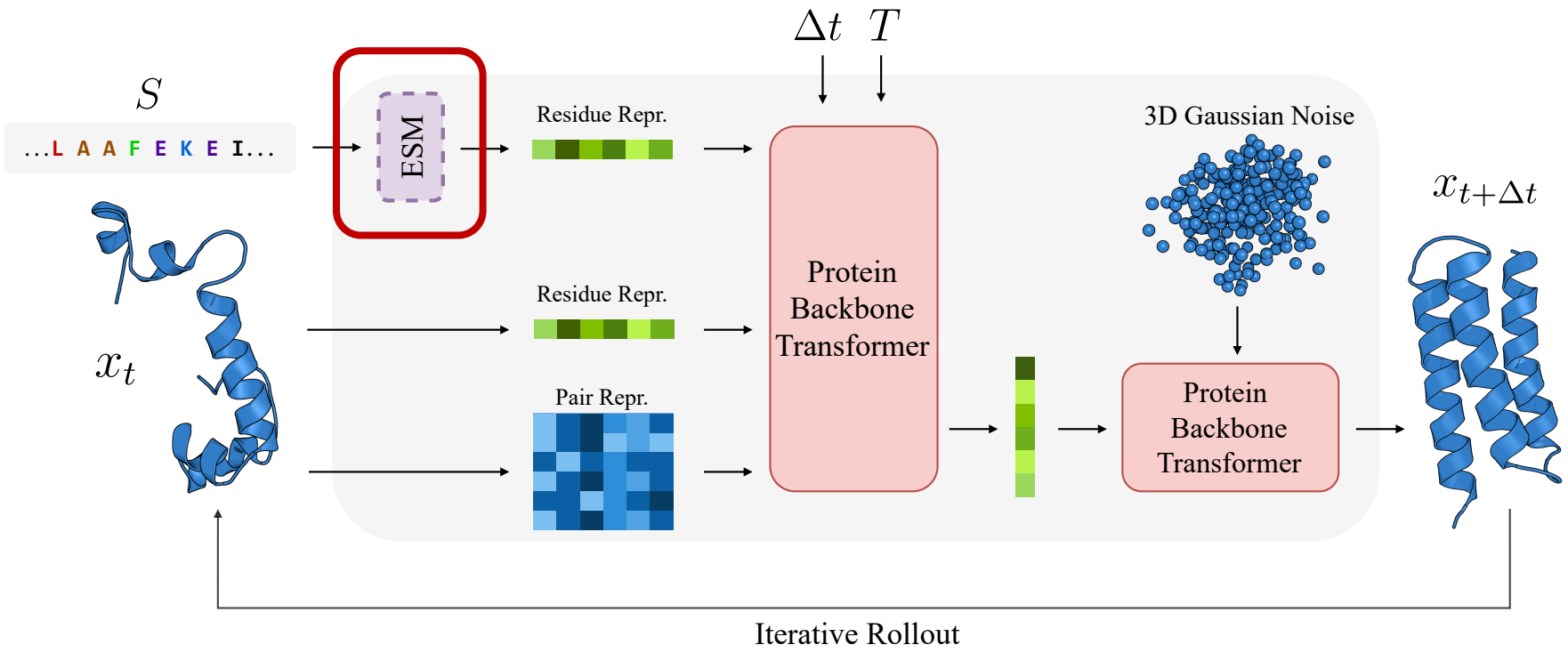
train across multiple
- Proteins
- Temperatures
- Time steps

Incorporating external representations: PLaTITO

Pretrained protein LMs capture evolutionary, structural and functional information in their latent representations.

$$e_{seq} = \varphi_{PLM}(S)$$

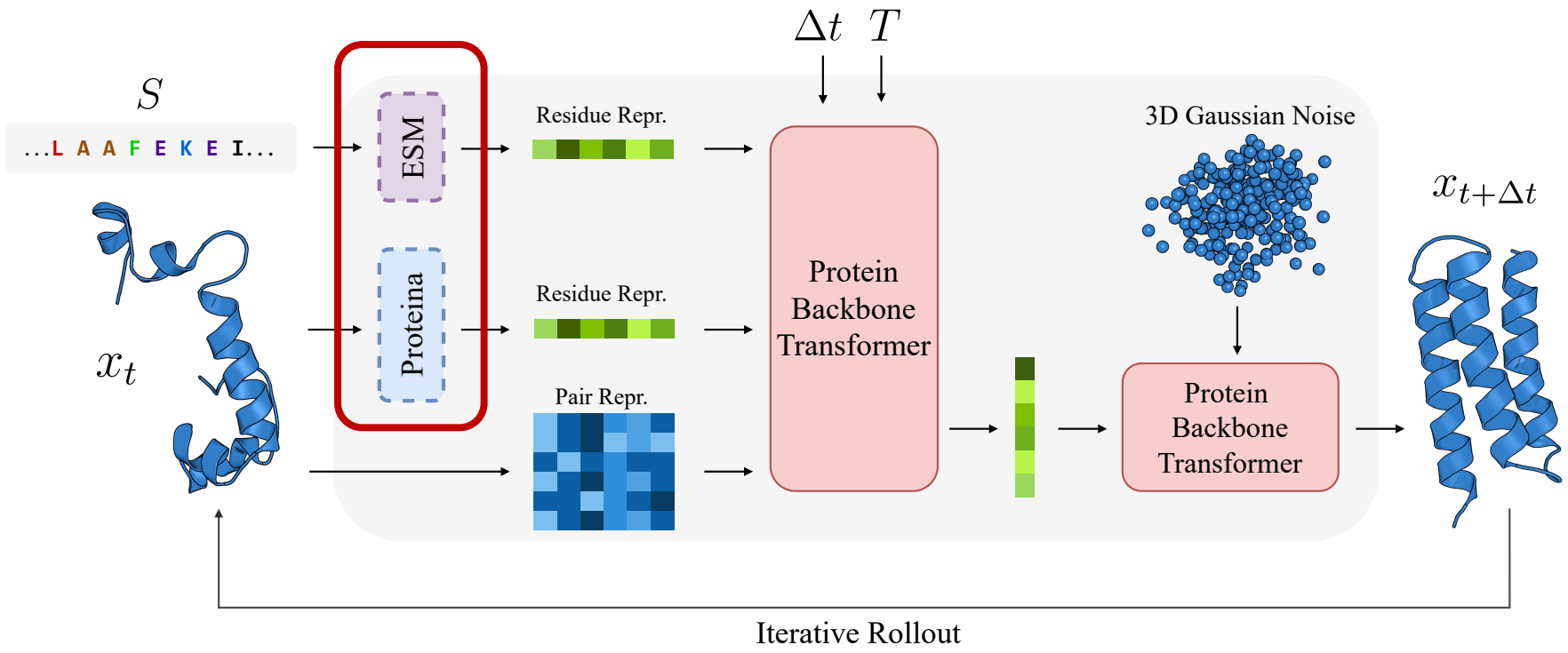
$$x_{t+\Delta t} \sim p(x_{t+\Delta t} \mid x_t, \Delta t, \cancel{S}, T)$$



Incorporating external representations: PLaTITO+Struct

Learned representations from structure-aware models like protein backbone generators contain useful local and global geometric features.

$$x_{t+\Delta t} \sim p(x_{t+\Delta t} \mid x_t, \Delta t, e_{seq}, e_{struct}, T)$$

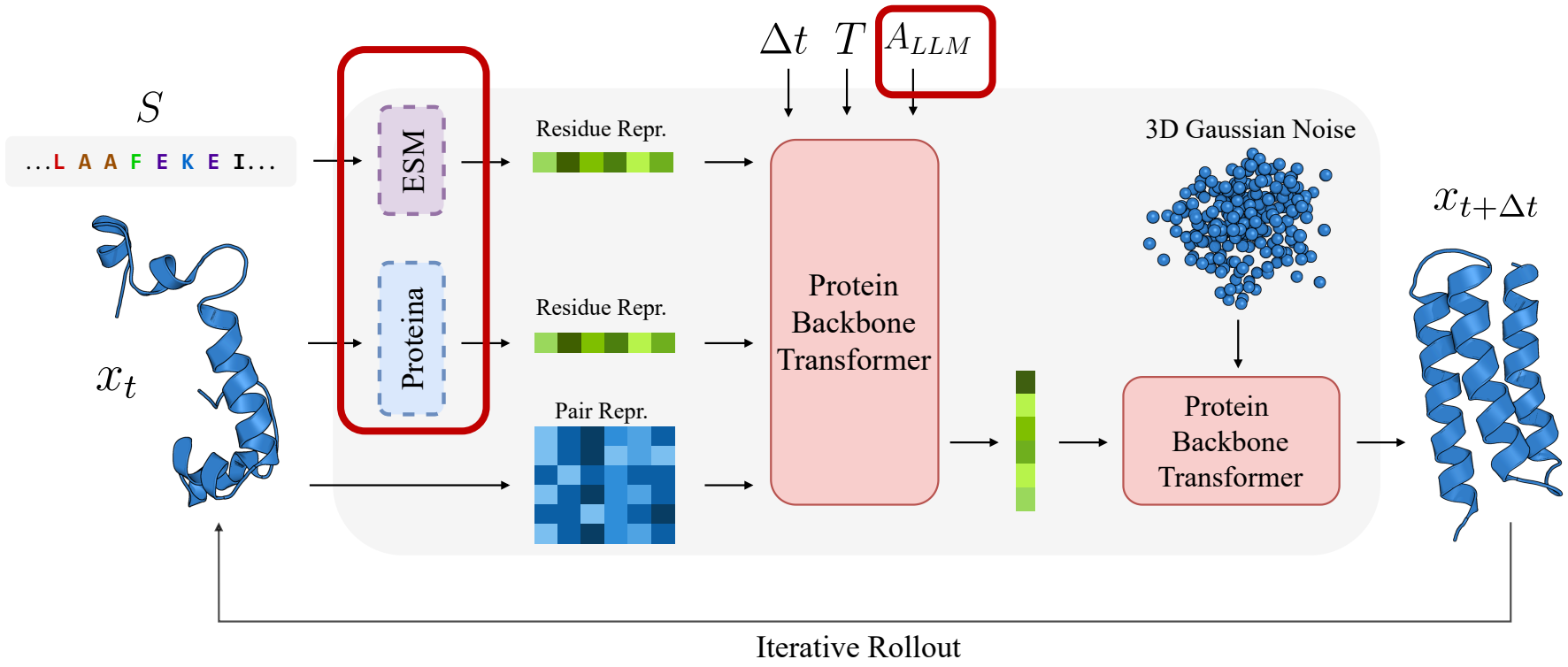


Incorporating external representations: PLaTITO+Struct+LLM

Protein domains are often extracted from their environment and simulated in isolation

- Unphysical behavior due to missing binding partners or cofactors.
- Use LLM as a heuristic to assess suitability of trajectories for training.

$$x_{t+\Delta t} \sim p(x_{t+\Delta t} \mid x_t, \Delta t, e_{seq}, e_{struct}, T, A_{LLM})$$

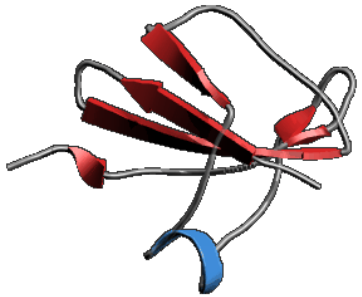


Molecular Dynamics (MD) datasets

Training: mdCATH

Short off-equilibrium simulations across 5 temperatures and 4,471 systems.

1dtdB00 from mdCATH

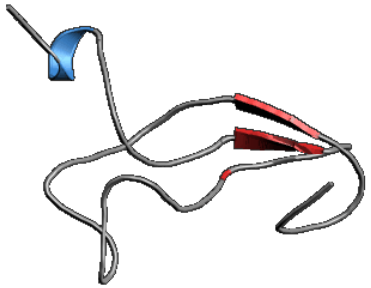


t = 0 ns

Evaluation: Fast-folding proteins

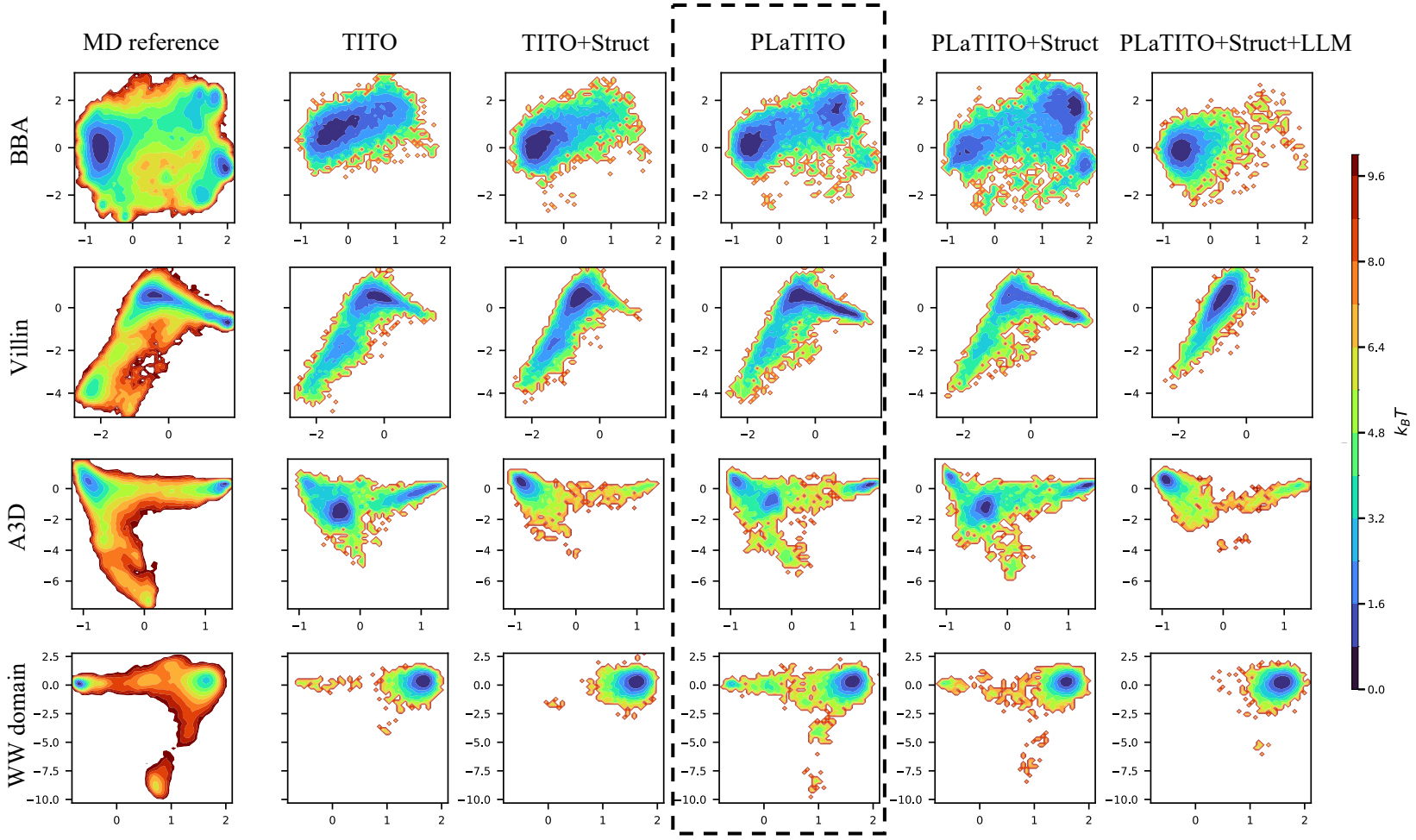
Long simulations at a single temperature for 12 systems that contain multiple folding and unfolding events.

α3D from fast-folding proteins



t = 0.0 µs

Pretrained representations improve generalization



Test-time predictions of free energy landscapes of four fast-folders.

PLaTITO-19M achieves SOTA on equilibrium sampling

MODEL	MAE (↓)	RMSE (↓)	COVERAGE (↑)	GPU HOURS[1]	MD DATA	PARAMETERS
TITO	1.068±0.272	1.382±0.302	0.590±0.111	1100	56 ms	$3M$
TITO+STRUCT	1.004±0.290	1.310±0.350	0.560±0.134	1100	56 ms	$3M$
PLATITO	0.949±0.269	1.228±0.328	0.651±0.151	1100	56 ms	$3M$
PLATITO+STRUCT	0.938±0.321	1.213±0.348	0.655±0.158	1100	56 ms	$3M$
PLATITO+STRUCT+LLM	1.066±0.270	1.346±0.292	0.570±0.087	1100	56 ms	$3M$
PLATITO-19M	**0.824±0.170**	**1.099±0.212**	**0.666±0.136**	1100	56 ms	$19M$
EMU	1.305±0.378	1.639±0.406	0.529±0.112	1100	56 ms	$3M$
BIOEMU[2]	1.110±0.292	1.389±0.346	0.594±0.175	9216	216 ms	$31M$

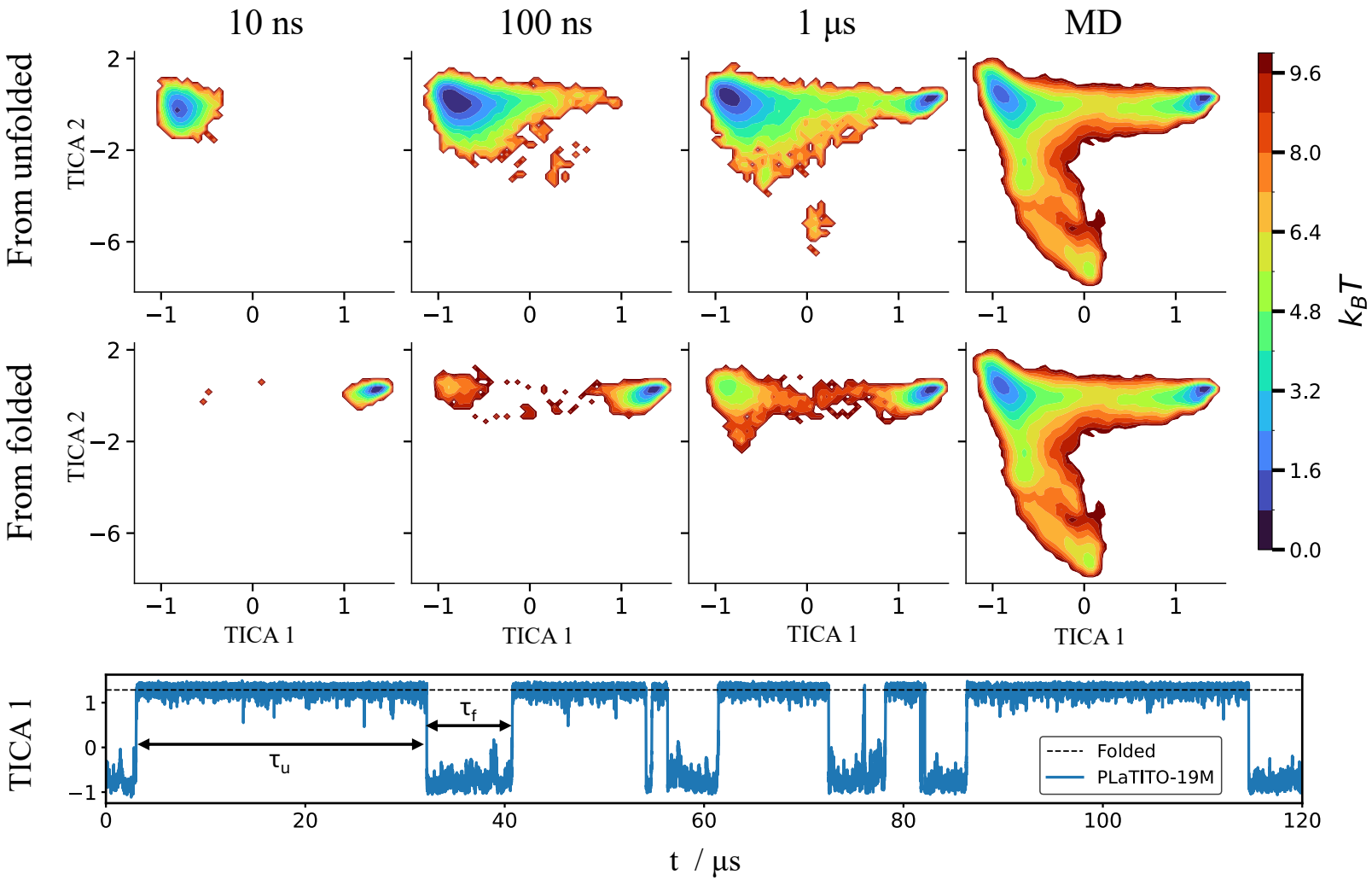
[1] GPU hours for training measured on a single NVIDIA A100 80GB GPU.
[2] BioEmu was additionally trained on 131k AFDB structures and 502k experimental ΔG measurements.

Boltzmann emulator (Lewis, 2025)
approximates $p(x \mid S) \sim \mu$

PLaTITO-19M outperforms BioEmu across all equilibrium sampling evaluation metrics despite being trained with substantially less data and a lower compute budget.

TITO models are substantially more data-efficient than Boltzmann Emulators

PLaTITO captures the kinetics of protein folding



	MD	PLaTITO-19M
$\langle \tau_f \rangle$	27 ± 8 μs	5.5 ± 0.9 μs
$\langle \tau_u \rangle$	31 ± 9 μs	14.8 ± 1.2 μs

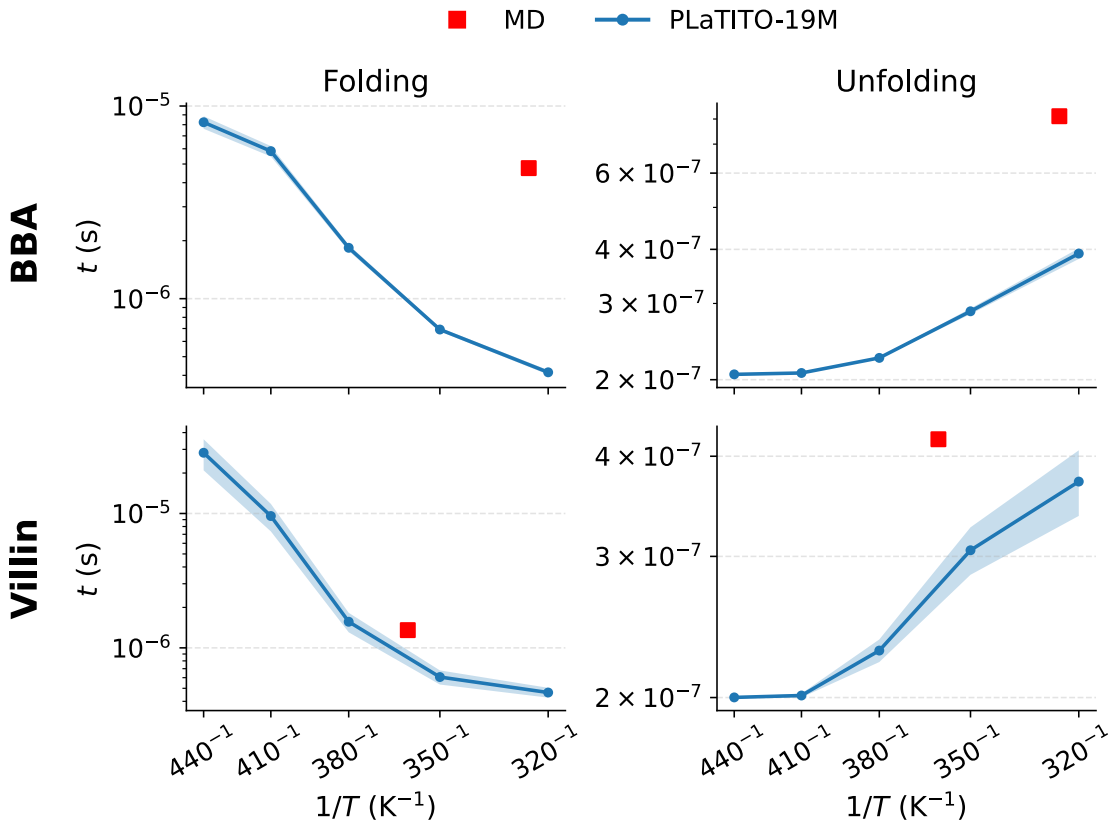
PLaTITO has learned physically meaningful dynamics

Protein folding presents deviations from the Arrhenius equation:

$$k(T) = A \exp\left(-\frac{E_a}{k_B T}\right)$$

folding/unfolding timescales

simulation temperature

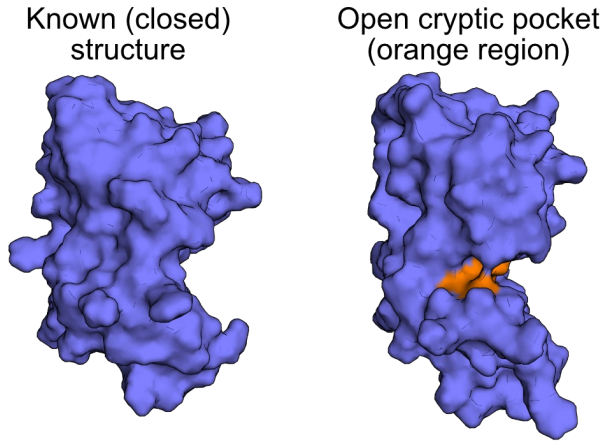


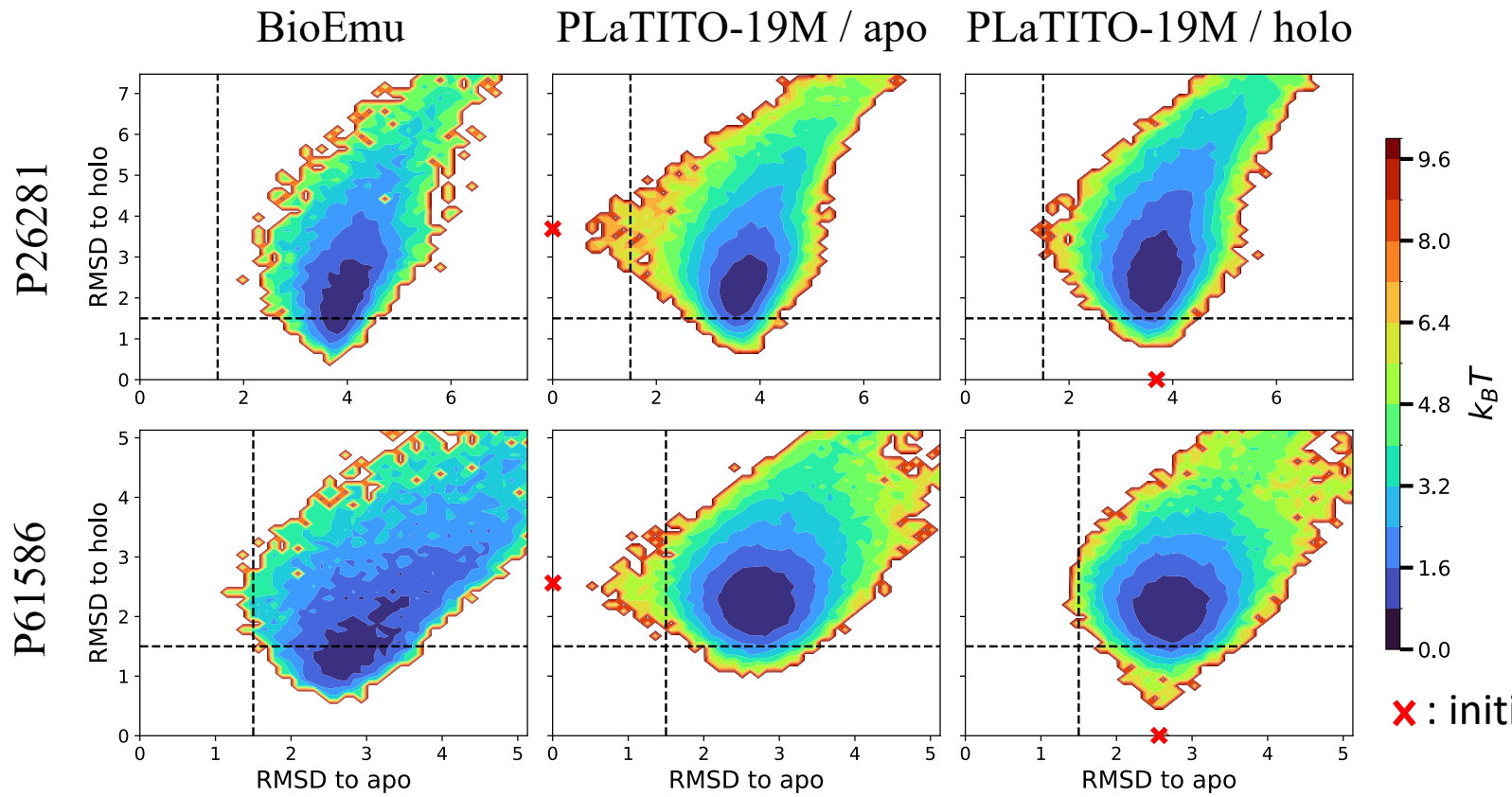
PLaTITO exhibits a non-Arrhenius temperature dependence, consistent with prior studies of protein folding kinetics

PLaTITO samples conformations connected to cryptic pockets

Cryptic pockets: lowly populated states targeted by therapeutics.

- apo state: unbound pocket
- holo state: ligand-bound pocket





× : initial state

ICML
International Conference
On Machine Learning

Thank you for listening

UNIVERSITY OF COPENHAGEN

CHALMERS
UNIVERSITY OF TECHNOLOGY

BBA | Villin | Trp-cage | BBL | α3D | WW domain

Panagiotis Antoniadis [1]

Beatrice Pavesi [2]

Simon Olsson* [2]

simonols@chalmers.se

Ole Winther* [1 3]

ole.winther@bio.ku.dk

*Equal contribution, ordered alphabetically.