Unlocking Guidance for Discrete State-Space Diffusion and Flow Models

Hunter Nisonoff*, Junhao (Bear) Xiong*, Stephan Allenspach*, Jennifer Listgarten
University of California, Berkeley

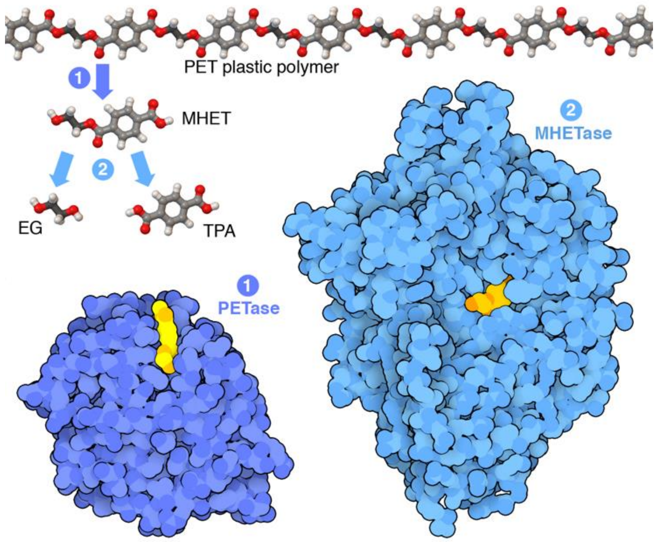
ICLR (2025)

*Equal Contribution

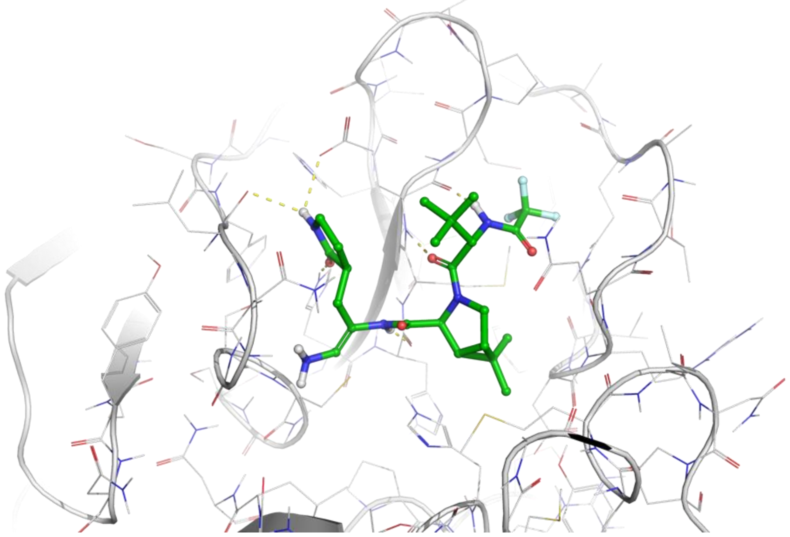
https://openreview.net/forum?id=XsgHl54yO7

Berkeley
UNIVERSITY OF CALIFORNIA

Conditional generation tasks are ubiquitous in the sciences



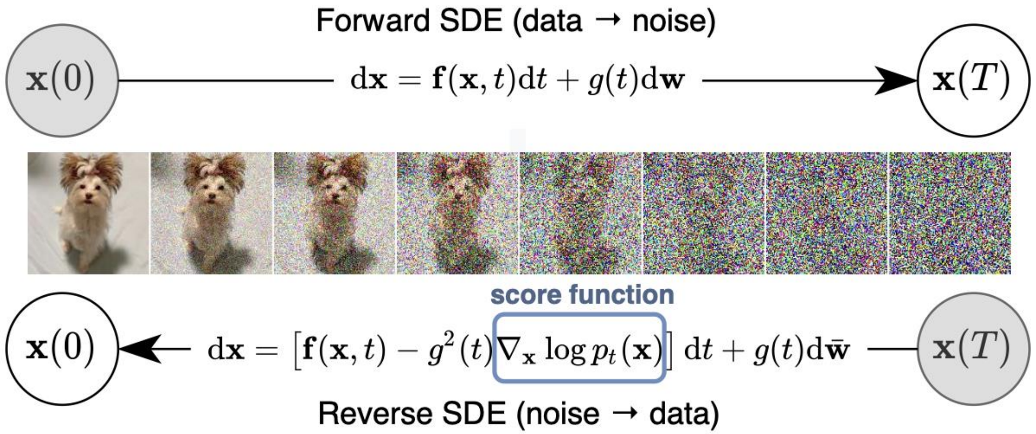
Designing plastic-degrading enzymes



SARS-CoV-2 Main Protease Inhibitor

DNA promoter engineering

Diffusion models can be guided to produce conditional samples



$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)}$$

$$\nabla_{\mathbf{x}} \log p_t(\mathbf{x} \mid y) = \underbrace{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}_{\text{Unconditional model}} + \underbrace{\nabla_{\mathbf{x}} \log p_t(y \mid \mathbf{x})}_{\text{Predictor}}$$

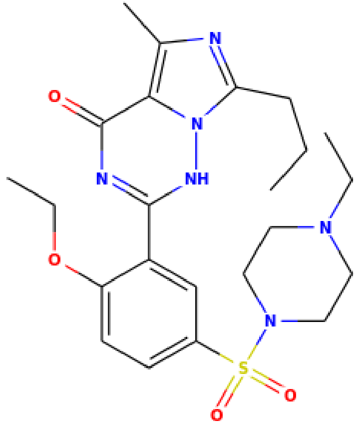
P. Dhariwal* and A. Nichol*, NeurIPS (2021).

Can we extend guidance to discrete objects?

Protein Sequence

```
MYTWTGALITPCAAEESKLPINPLSNSLLRHH
YDTRCFDSTVTESDIRVEESIYQCCDLAPEEA
LTERLYIGGPLTNSKGQNCGYRRCRASGVLTT
SCGNTLTCYLKATAACRAAKLQDCTMLVNGDD
LVVICESAGTQEDAAALRAFTEAMTRYSAPPG
```

Small molecule Graphs



DNA Sequences

```
TGGGATTCAGGTCTCGGAGCCGTGGTGGTGCT
GAACAACCGCGTGATTGGCGAAGGCTGGAACC
GCGCGATTGGCCTGCATGATCCGACCGCGCAT
GCGGAATTATGGCGCTGCGCCAGGCGGGCCTG
GTGATGCAGAACTATCGCATTATTGATGCGAC
```

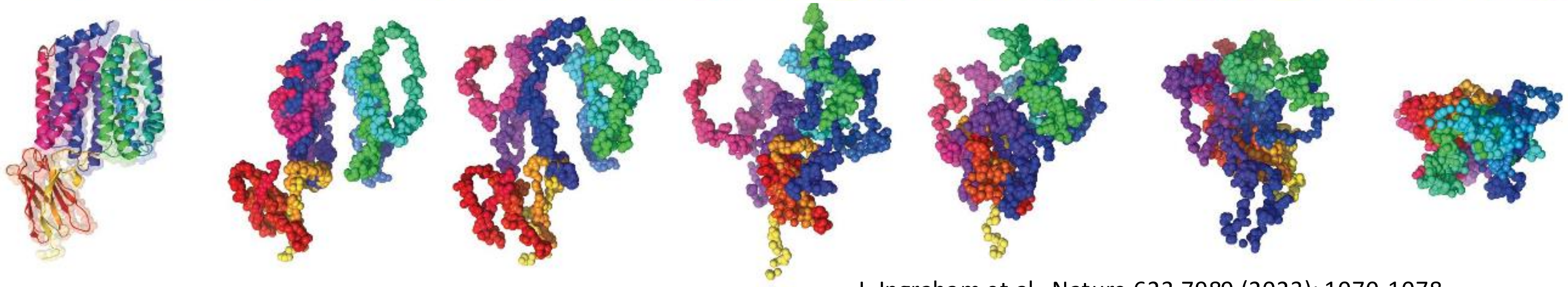
$$\nabla_x \log p(x \mid y) = \underbrace{\nabla_x \log p(x)}_{\text{Unconditional model}} + \underbrace{\nabla_x \log p(y \mid x)}_{\text{Predictor}}$$

Challenge: for **sequences, graphs, text**, etc., $\nabla_x \log p_\theta(x)$ not useful.

Consequence I: Standard diffusion/score doesn't work. (other work)

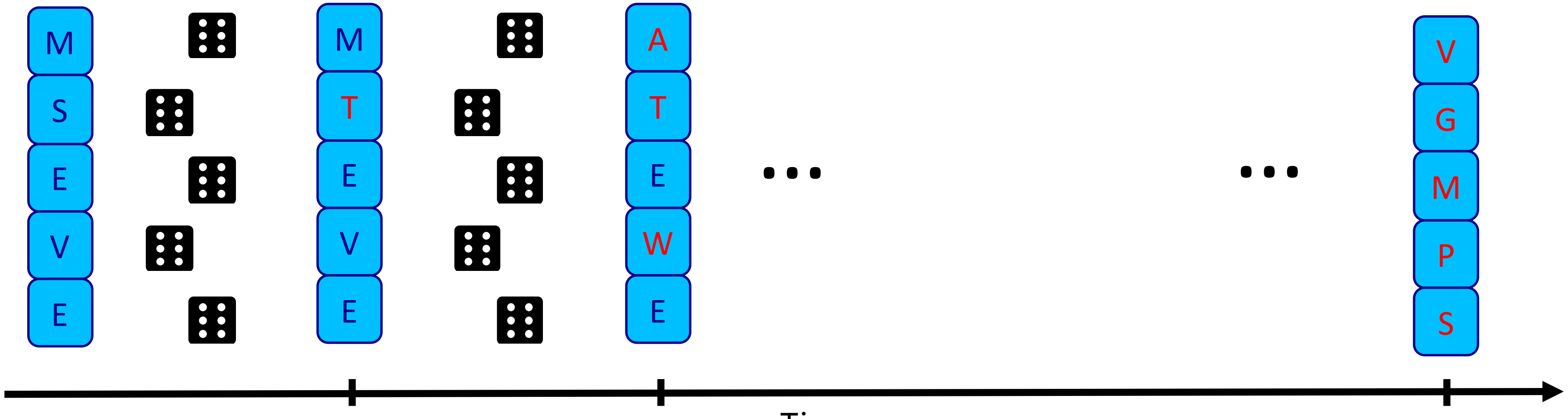
Consequence II: Cannot use guidance, $\nabla_x \log p_\phi(y|x)$. (our work)

Discrete time diffusion on discrete spaces



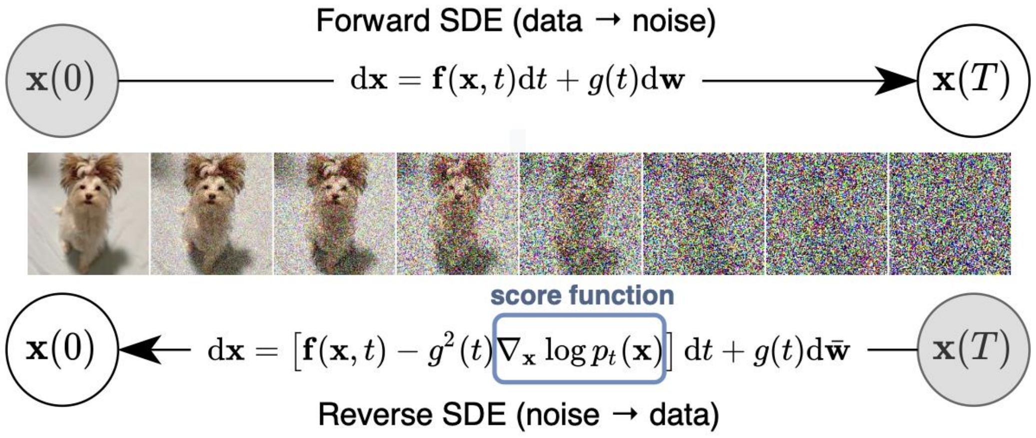
J. Ingraham et al. Nature 623.7989 (2023): 1070-1078.

Mutation-based noise process



Austin et al. NeurIPS (2021).

Diffusion models can be guided to produce conditional samples



$$\boxed{p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)}}$$

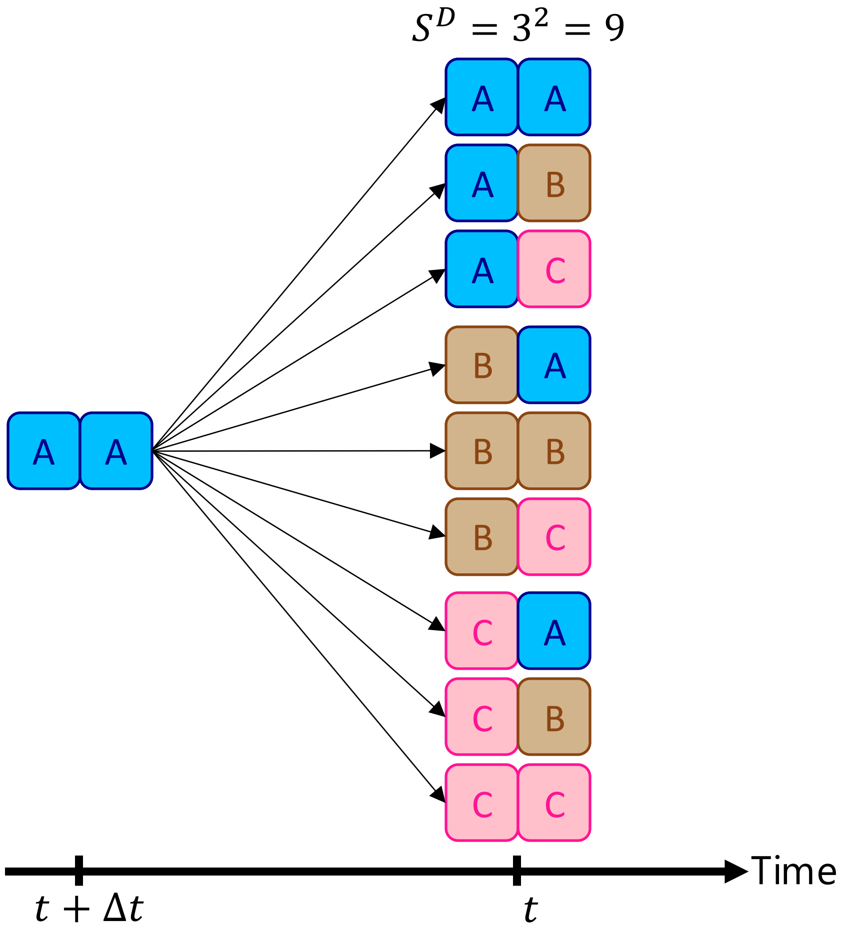
$$\cancel{\nabla_x \log p(x \mid y) = \underbrace{\nabla_x \log p(x)}_{\text{Unconditional model}} + \underbrace{\nabla_x \log p(y \mid x)}_{\text{Predictor}}}$$

Guidance in discrete spaces seems intractable

$$\cancel{\nabla_x \log p(x \mid y) = \underbrace{\nabla_x \log p(x)}_{\text{Unconditional model}} + \underbrace{\nabla_x \log p(y \mid x)}_{\text{Predictor}}}$$

Bayes' theorem:

$$p(x_t \mid y, x_{t+\Delta t}) = \frac{p(y \mid x_t, x_{t+\Delta t})\, p(x_t \mid x_{t+\Delta t})}{\sum_{x'_t} p(y \mid x'_t, x_{t+\Delta t})\, p(x'_t \mid x_{t+\Delta t})}$$



denoising process toward $t = 0$

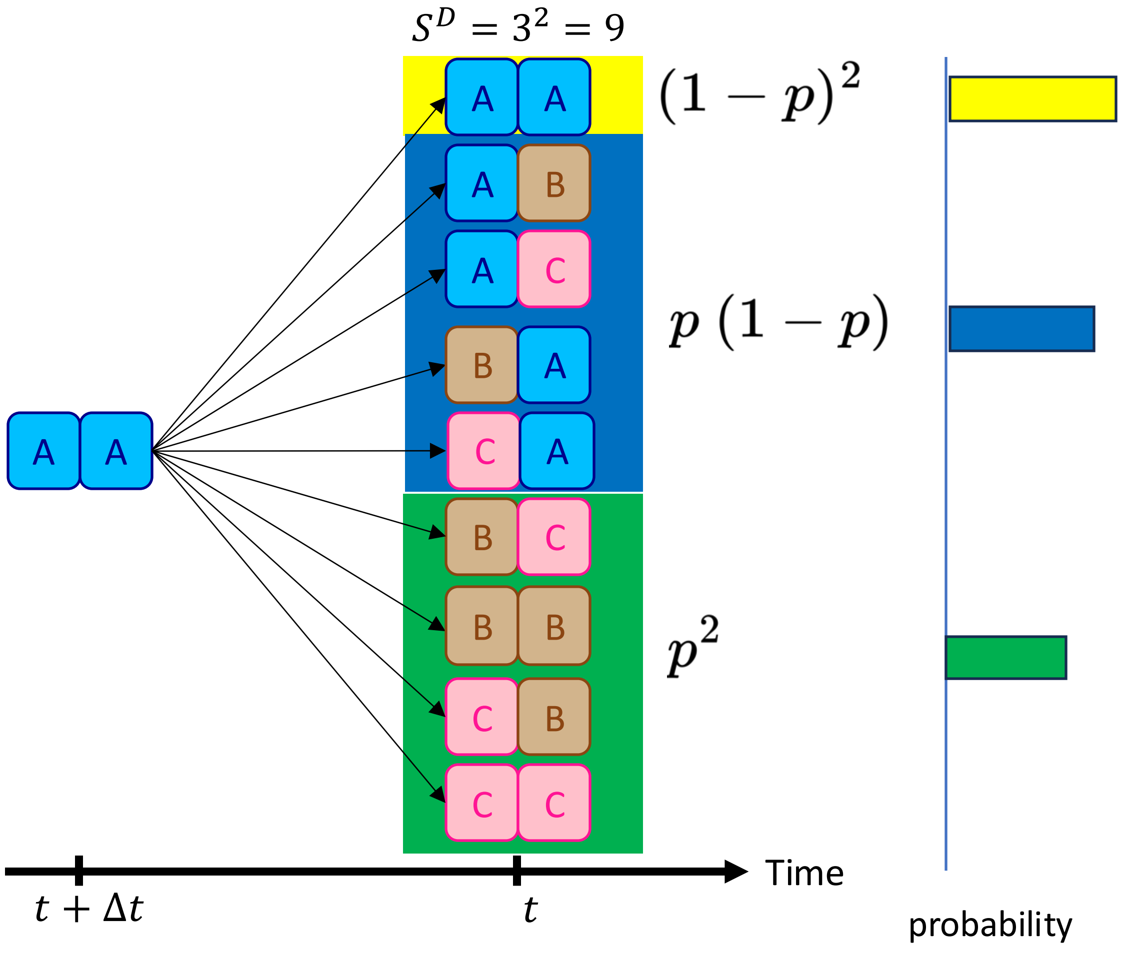
Guidance in discrete spaces seems intractable

$$\cancel{\nabla_x \log p(x \mid y) = \underbrace{\nabla_x \log p(x)}_{\text{Unconditional model}} + \underbrace{\nabla_x \log p(y \mid x)}_{\text{Predictor}}}$$

Bayes' theorem:

$$p(x_t|y, x_{t+\Delta t}) = \frac{p(y|x_t, x_{t+\Delta t})p(x_t|x_{t+\Delta t})}{\sum_{x'_t} p(y|x'_t, x_{t+\Delta t})p(x'_t|x_{t+\Delta t})}$$

Mutate with probability, $p \propto \Delta t$



denoising process toward $t = 0$

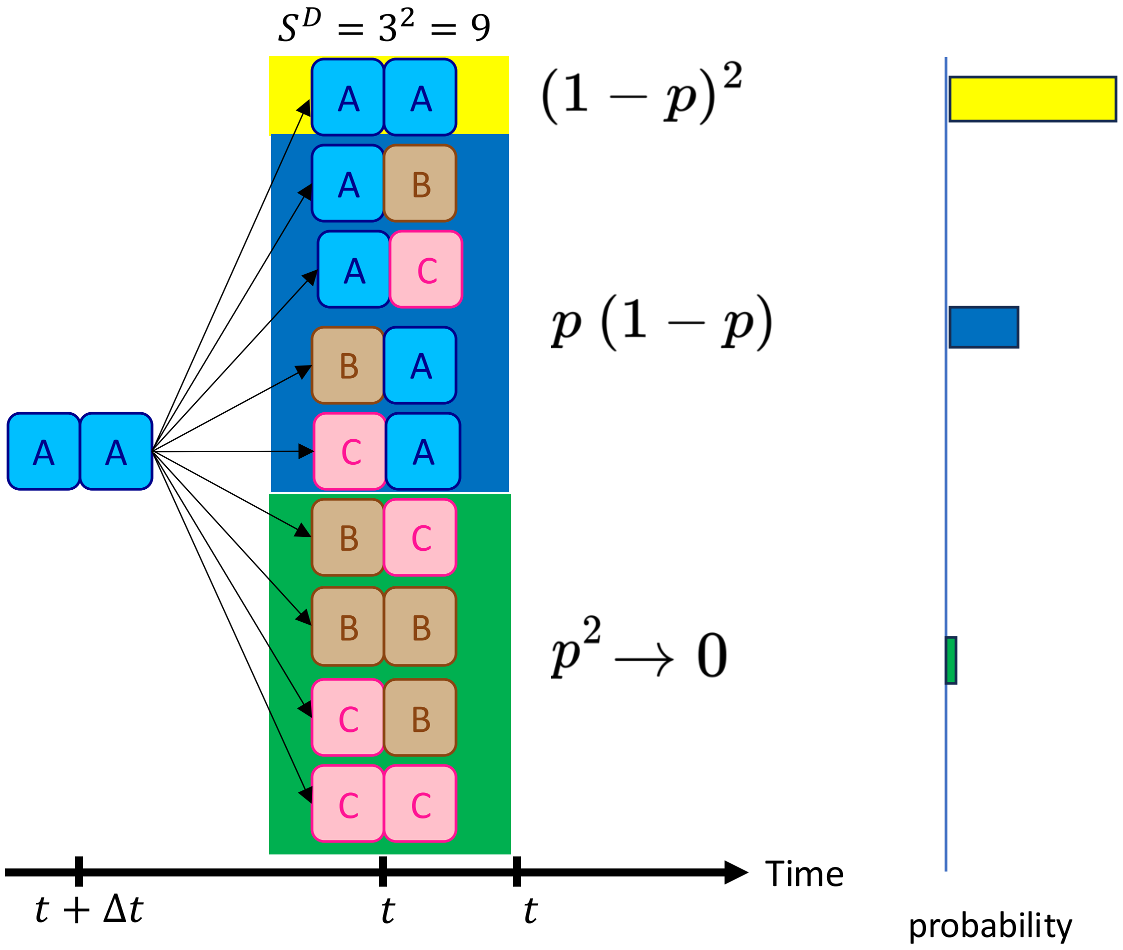
Guidance in discrete spaces seems intractable

$$\cancel{\nabla_x \log p(x \mid y) = \underbrace{\nabla_x \log p(x)}_{\text{Unconditional model}} + \underbrace{\nabla_x \log p(y \mid x)}_{\text{Predictor}}}$$

Bayes' theorem:

$$p(x_t|y, x_{t+\Delta t}) = \frac{p(y|x_t, x_{t+\Delta t})p(x_t|x_{t+\Delta t})}{\sum_{x_t'} p(y|x'_t, x_{t+\Delta t})p(x'_t|x_{t+\Delta t})}$$

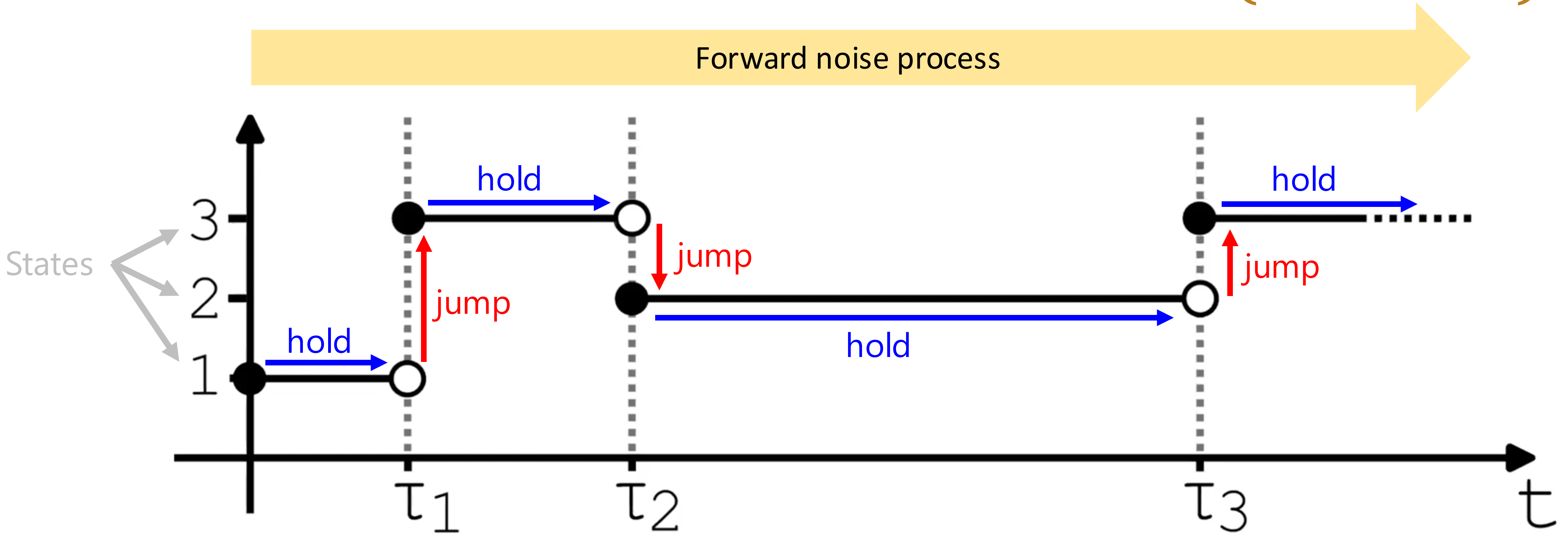
Mutate with probability, $p \propto \Delta t$



denoising process toward $t = 0$

Berkeley UNIVERSITY OF CALIFORNIA

Diffusion models for discrete spaces with continuous-time Markov chains (CTMCs)

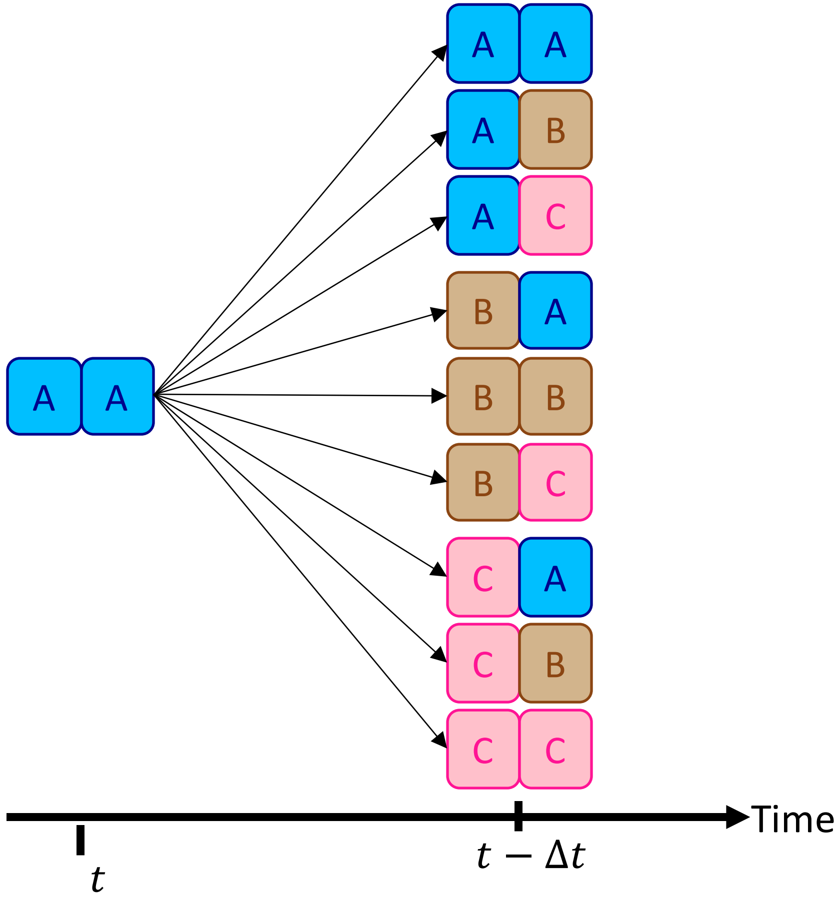


$$p(x_{t+dt}=\tilde{x}|x_t=x) = \delta_{x,\tilde{x}} + R_t(x,\tilde{x})dt$$

Campbell *et al.*, NeurIPS (2022).
Campbell*, J. Yim* *et al.*, ICML (2024).

Unlocking guidance in discrete spaces

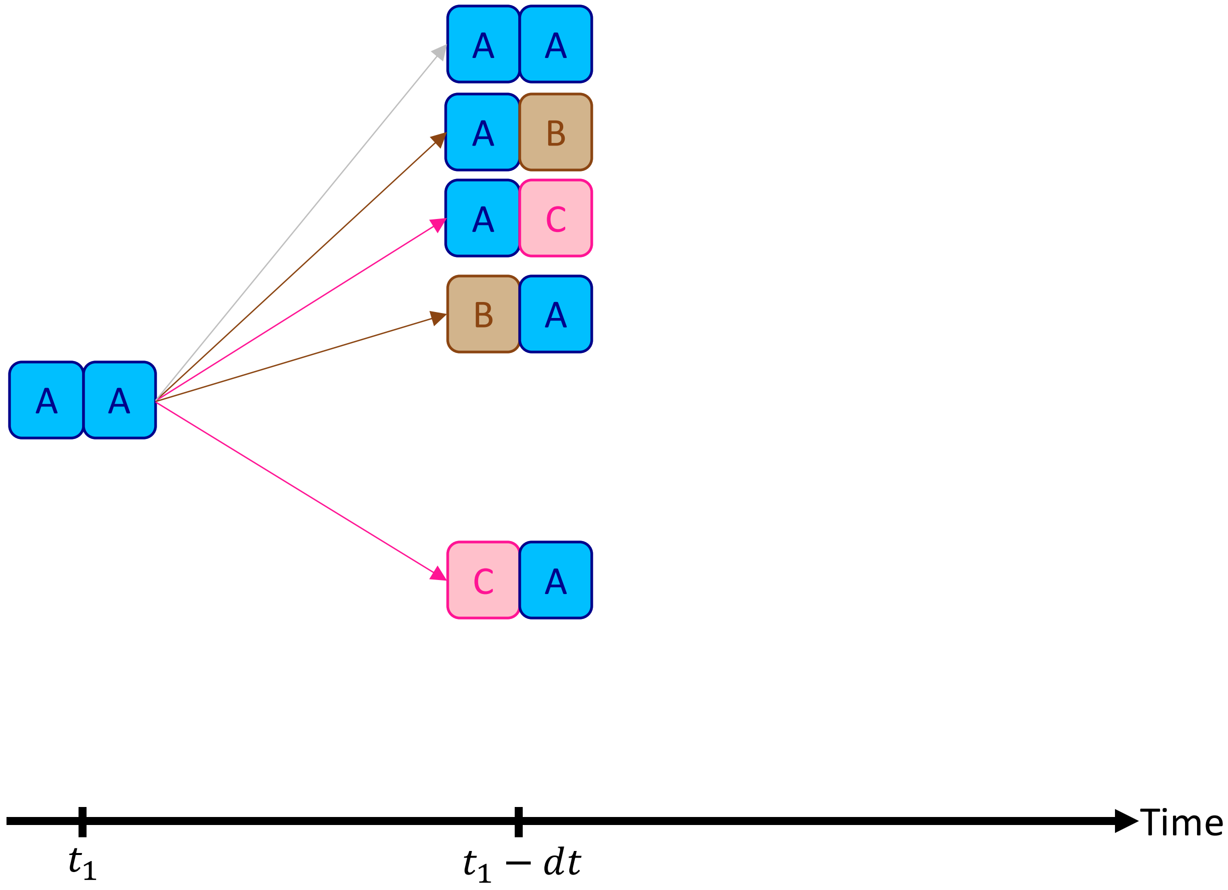
$S^D = 3^2 = 9$



Bayes' theorem:

$$p(x_t|y, x_{t+\Delta t}) = \frac{p(y|x_t, x_{t+\Delta t})p(x_t|x_{t+\Delta t})}{\sum_{x'_t} p(y|x'_t, x_{t+\Delta t})p(x'_t|x_{t+\Delta t})}$$

$(S-1) \times D + 1 = 5$



In continuous time diffusion:

- Only single dimensional changes can occur in infinitesimal time!
- This enables tractable guidance!

Guidance in continuous and discrete spaces

	Predictor-guidance (PG)	Predictor-free guidance (PFG)
Continuous state-space	unconditional, predictor $\nabla_{x_t} \log p^{(\gamma)}(x_t \mid y) = \nabla_{x_t} \log \underbrace{p(x_t)}_{\text{unconditional}} + \gamma \nabla_{x_t} \log \underbrace{p(y \mid x_t)}_{\text{predictor}}$ J. Sohl-Dickstein *et al.*, ICML (2015). P. Dhariwal* and A. Nichol*, NeurIPS (2021). Y. Song *et al.*, ICLR (2021).	conditional, unconditional $\nabla_{x_t} \log p^{(\gamma)}(x_t \mid y) = \gamma \nabla_{x_t} \log \underbrace{p(x_t \mid y)}_{\text{conditional}} + (1-\gamma) \nabla_{x_t} \log \underbrace{p(x_t)}_{\text{unconditional}}$ J. Ho and T. Salimans, NeurIPS - Workshop (2021).
Discrete state-space	unconditional, predictor $\log R_t^{(\gamma)}(x, \tilde{x} \mid y) = \log \underbrace{R_t(x, \tilde{x})}_{\text{unconditional}} + \gamma \Big(\log \underbrace{p(y \mid \tilde{x}, t)}_{\text{predictor}} - \log \underbrace{p(y \mid x, t)}_{\text{predictor}} \Big)$ H.N., J.X., S.A. and J.L., ICLR (2025). [This work]	conditional, unconditional $\log R_t^{(\gamma)}(x, \tilde{x} \mid y) = \gamma \log \underbrace{R_t(x, \tilde{x} \mid y)}_{\text{conditional}} + (1-\gamma) \log \underbrace{R_t(x, \tilde{x})}_{\text{unconditional}}$ H.N., J.X., S.A. and J.L., ICLR (2025). [This work]

Berkeley UNIVERSITY OF CALIFORNIA

Accurate approximation for further speed-up

- *Predictor-guidance* requires $O(D \times S)$ forward passes of the predictor model for each sampling step.
- *e.g.,* state-space of CIFAR10:

$D = 32 \times 32 \times 3, S = 256 \Longrightarrow D \times S \approx \mathbf{7.8 \times 10^5}$

- Solution: **Taylor-Approximated Guidance (TAG)** requires one forward pass (but still $D \times S$ computations):

$$\boxed{\log p^{\phi}(y|\tilde{x}, t) - \log p^{\phi}(y|x, t) \approx (\tilde{x} - x)^T \nabla_x \log p^{\phi}(y|x, t)}$$

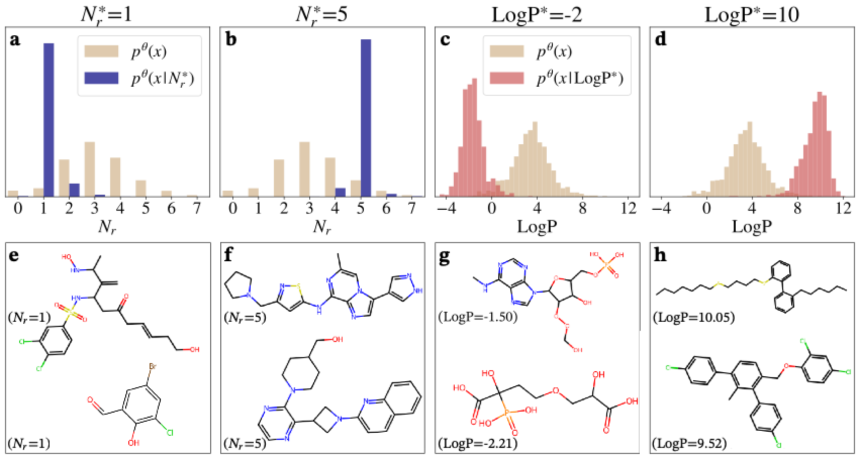
Bayes' theorem:

$$p(x_t|y, x_{t+\Delta t}) = \frac{p(y|x_t, x_{t+\Delta t})p(x_t|x_{t+\Delta t})}{\sum_{x'_t} p(y|x'_t, x_{t+\Delta t})p(x'_t|x_{t+\Delta t})}$$

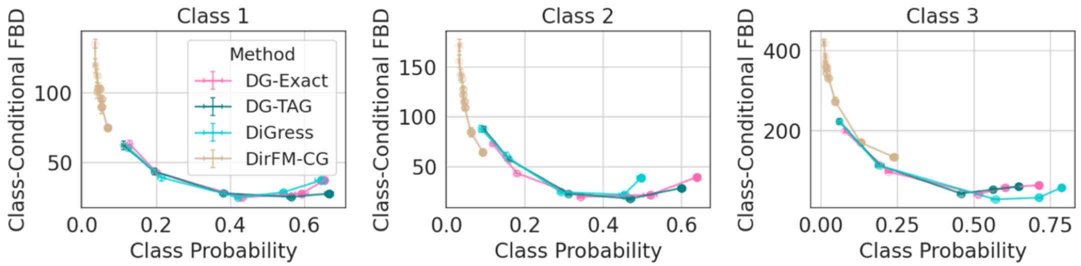
Oops I Took A Gradient: Scalable Sampling for Discrete Distributions

Will Grathwohl [1,2] Kevin Swersky [2] Milad Hashemi [2] David Duvenaud [1,2] Chris J. Maddison [1]

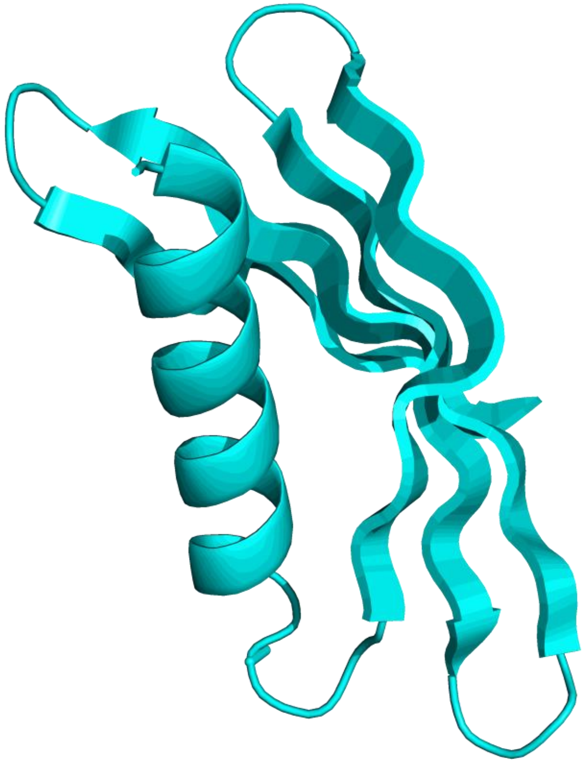
Conditional generation tasks in the sciences



Small molecules



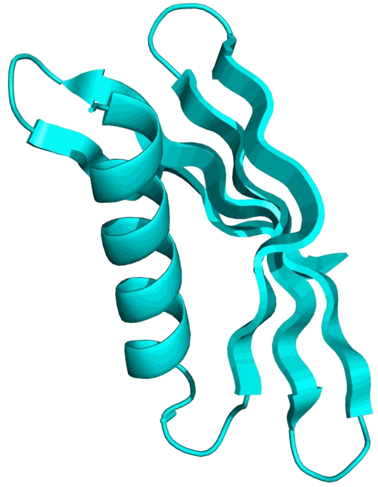
Enhancer DNA



Proteins

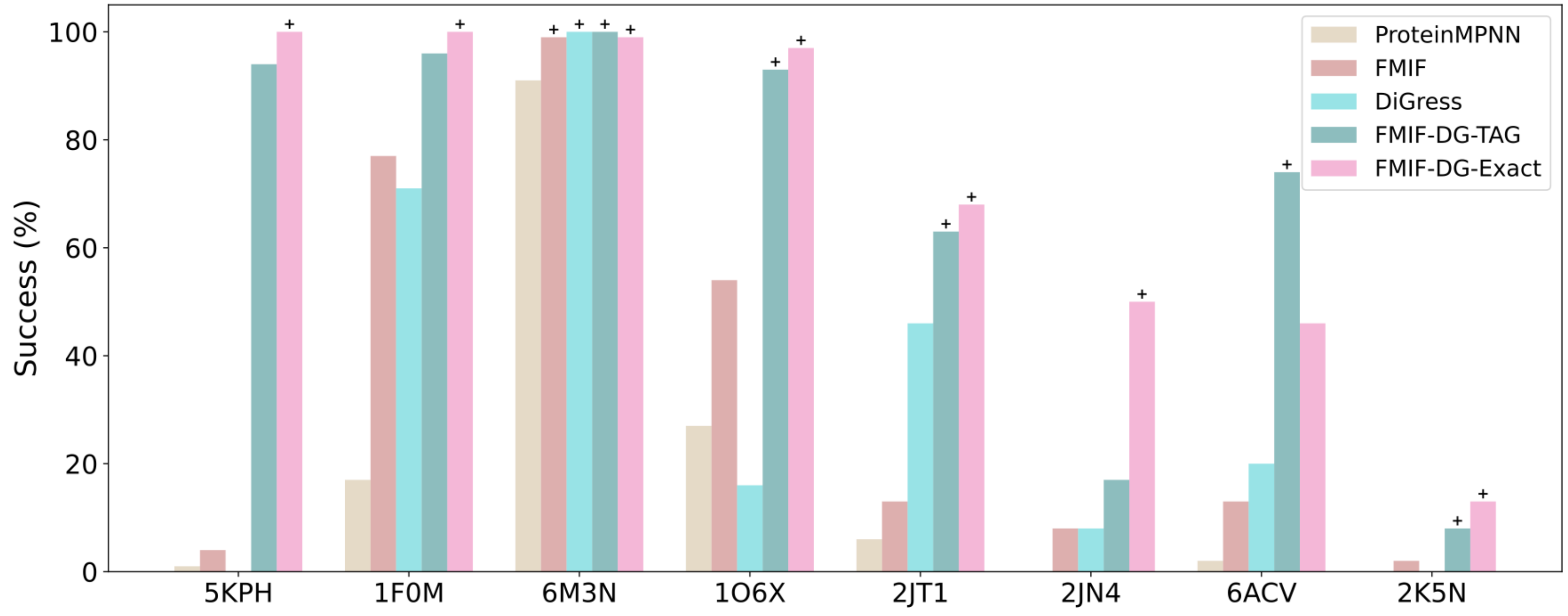
Stability-conditioned inverse folding

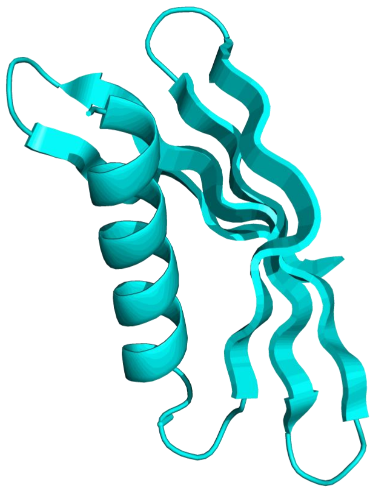
$$\text{P(sequence | structure, }\Delta\Delta G \geq 0) \propto \underbrace{\text{P(sequence | structure)}}_{\substack{\text{Inverse Folding} \\ \text{(FMIF)}}} \underbrace{\text{P(}\Delta\Delta G \geq 0 \text{ | sequence, structure)}}_{\text{Stability predictor}}$$



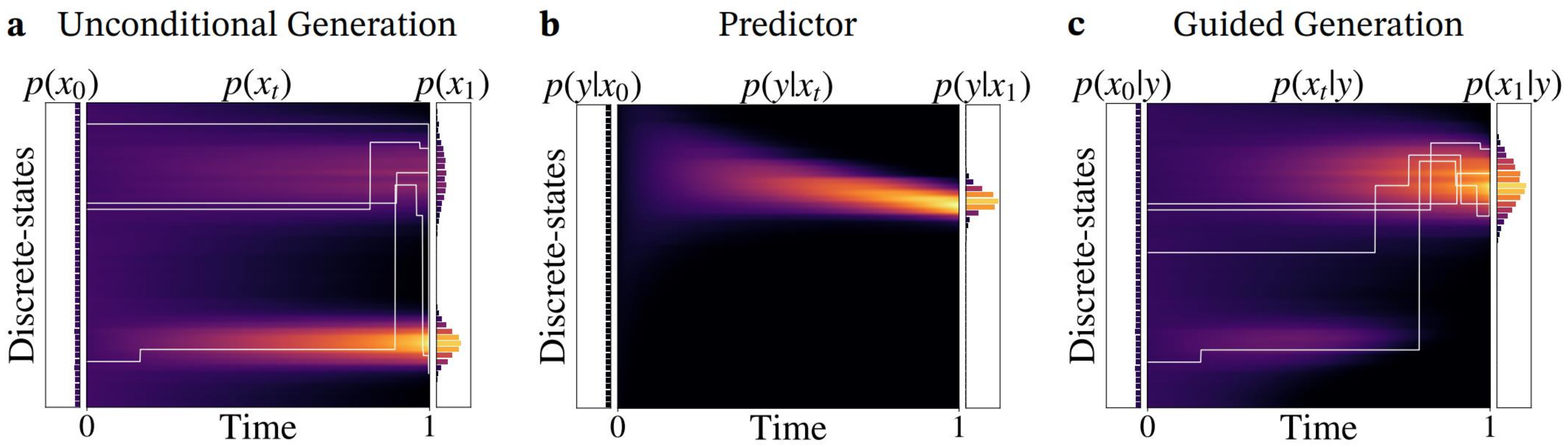
Stability-conditioned inverse folding

$$\text{P(sequence} \mid \text{structure}, \Delta\Delta G \geq 0) \propto \underbrace{\text{P(sequence} \mid \text{structure)}}_{\substack{\text{Inverse Folding} \\ \text{(FMIF)}}} \underbrace{\text{P}(\Delta\Delta G \geq 0 \mid \text{sequence, structure)}}_{\text{Stability predictor}}$$





Thank You!



a Unconditional Generation
$p(x_0)$
$p(x_t)$
$p(x_1)$
Discrete-states
0
Time
1
b Predictor
$p(y|x_0)$
$p(y|x_t)$
$p(y|x_1)$
Discrete-states
0
Time
1
c Guided Generation
$p(x_0|y)$
$p(x_t|y)$
$p(x_1|y)$
Discrete-states
0
Time
1