Inferring phase transitions and critical exponents from limited observations with Thermodynamic Maps

Wang, Herron, Tiwary, PNAS (2022): e2203656119
Herron…Tiwary, arXiv:2308.14885

Pratyush Tiwary
University of Maryland

DEPARTMENT OF CHEMISTRY & BIOCHEMISTRY

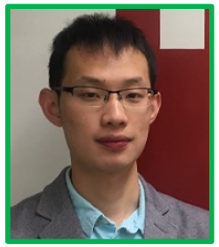
Institute for Physical Science & Technology
IPST

UNIVERSITY OF MARYLAND
INSTITUTE FOR HEALTH COMPUTING
MPOWERING THE STATE

Our team

Postdocs

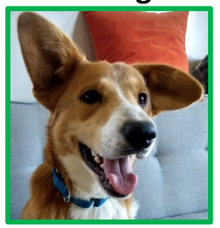
Eric Beyerle



Ruiyu Wang

Xinyu Gu

Postdog



Pakora the Dog

Undergrads

Michael Strobel
(CS+Statistics)

Disha Sanwal
(Chemistry+Appl Math)

Highschooler

Anurag Sodhi

PhD students

Dedi Wang
(Biophysics) → Genentech

Connor Zou
(Chemistry)
→ Fraser Lab@UCSF

Shams Mehdi
(Biophysics)

Alumni

- Joao Ribeiro (Asst Prof @St Josephs University NYC)
- Zachary Smith (Senior Scientist @Schrodinger)
- Bodhi Vani (ML Scientist @Prescient Design)
- Yihang Wang (Schmidt AI in Science Postdoc @UChicago)
- Sun-Ting Tsai (Postdoc @UMichigan)
- Luke Evans (Postdoc @Flatiron)

Lukas Herron
(Biophysics)

Akashnathan Aranganathan
(Biophysics)

Suemin Lee
(Biophysics)

Vanessa Meraz
(Chemical Physics)

Anjali Verma
(Biophysics)

Richard John
(Physics)

Venkata Adury
(Chemical Physics)

$ and  support

Current sources of financial support:

NSF CAREER CHE-2044165 (CTMC, 2021-2026)

NIH MIRA R35GM142719 (NIGMS, 2021-2026)

DOE BES SC0021009 (CPIMS, 2020-2026)

Sloan Research Fellowship (2022-2024)

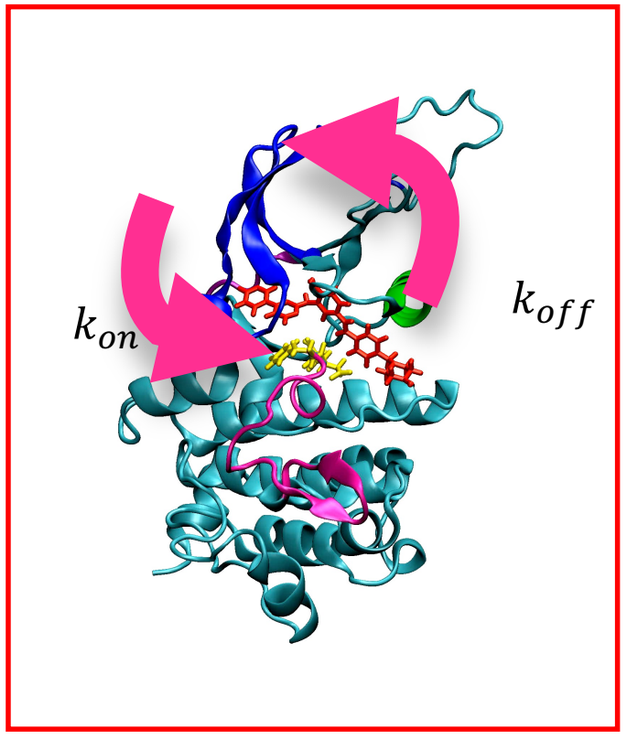
NCI-UMD partnership (2018-2025)

UM-IHC start-up grants

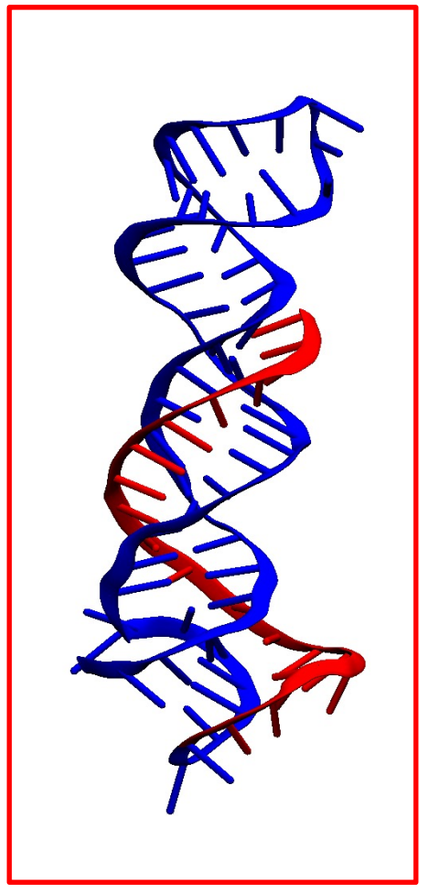
Millard and Lee Alexander Professorship funds

COI: P.T. is on Scientific Advisory Board of Schrödinger

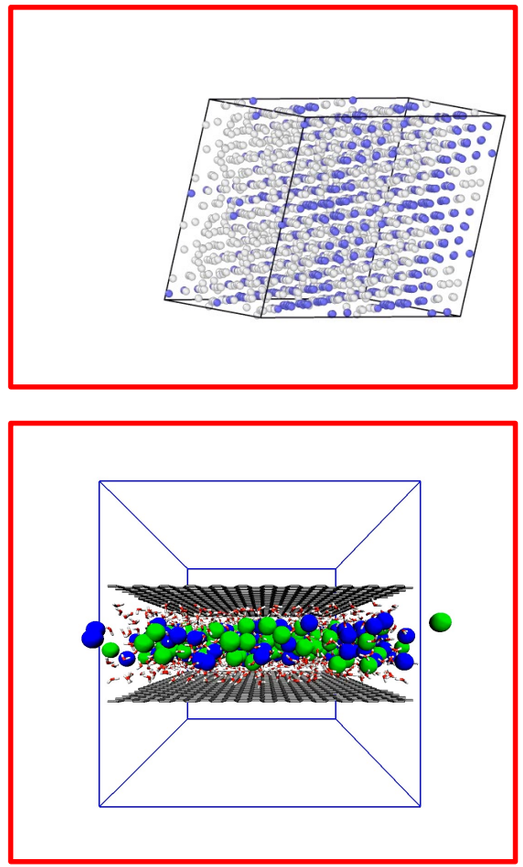
My lab combines Molecular Simulations and generative AI to tackle problems of human health and energy relevance guided by structure & dynamics



Protein-small molecules drug discovery



RNA therapeutics



Finite-temperature crystal polymorphs & phase transitions

github.com/tiwarylab

AI can now be used to easily, routinely predict structure

THANKS FOR YOUR ATTENTION,
THIS TALK IS NOW OVER

Google DeepMind

AlphaFold: a solution to a 50-year-old grand challenge in biology

November 30, 2020

AI can now be used to easily, routinely predict structure

THANKS FOR YOUR ATTENTION, THIS TALK IS NOW OVER

Not quite …

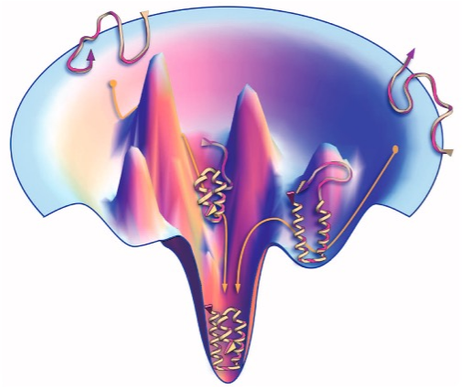
Google DeepMind

AlphaFold: a solution to a 50-year-old grand challenge in biology

November 30, 2020

Life is not about a single structure but an ensemble of structures with right kind and magnitude of fluctuations

Proteins

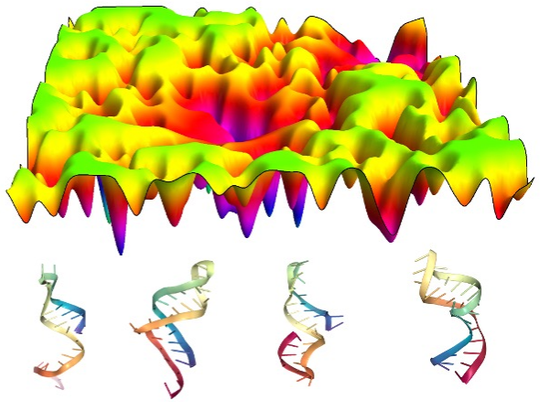


K. Dill, J. McCallum, Science (2012).
The Protein-Folding problem, 50 years on

Fluctuations between metastable states described by few slow modes & timescale separation

Order/Disorder spectrum

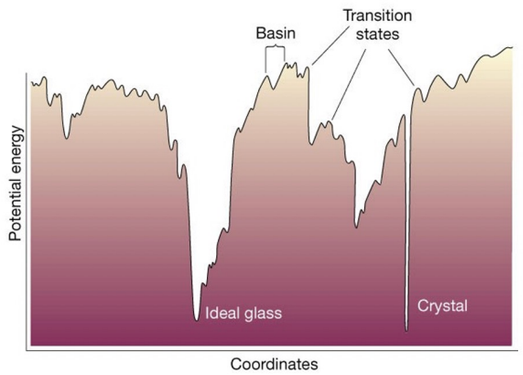
RNAs



D.J. Wales, Ann. Rev. Phys. Chem. (2017).
Exploring Energy Landscapes

No obvious timescale separation & no dominant driving fluctuations

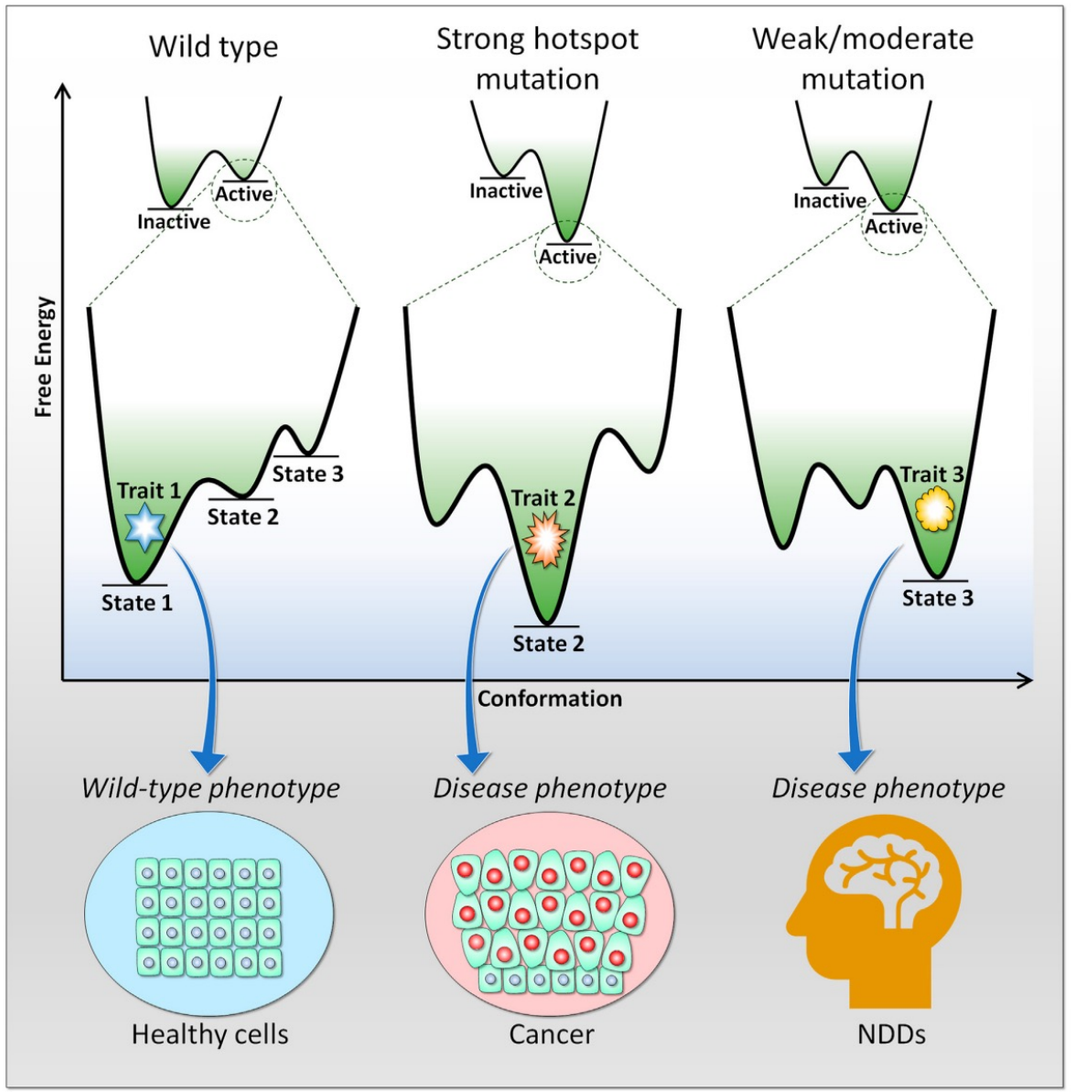
Somewhere on the order/disorder spectrum



Beyerle, Zou, Tsai, Tiwary PNAS 2023

Debenedetti, Stillinger, Nature (2001).
Supercooled Liquids & the Glass Transition

Proteins: Subtle fluctuations in state populations dictate disease phenotypes



Nussinov et al, Curr Op Struc Bio 2023

RNA: glassy landscapes and deep learning disaster

PDB Statistics: Overall Growth of Released Structures Per Year

Year	Total Number of Entries Available	Number of Structures Released Annually
2020	172934	14021

PDB Statistics: RNA-only Structures Released Per Year

Year	Total Number of Entries Available	Number of Structures Released Annually
2020	1523	102

Known human RNA >>
10 * Human Proteins
Number of RNA structures in PDB =
1% of protein structures

Number of FDA approved drugs
targeting human RNA =~ 2

RNA: glassy landscapes and deep learning disaster

PDB Statistics: Overall Growth of Released Structures Per Year

Year	Total Number of Entries Available	Number of Structures Released Annually
2020	172934	14021

PDB Statistics: RNA-only Structures Released Per Year

Year	Total Number of Entries Available	Number of Structures Released Annually
2020	1523	102

Known human RNA >>
10 * Human Proteins
Number of RNA structures in PDB =
1% of protein structures

Number of FDA approved drugs
targeting human RNA =~ 2

Received: 25 April 2023 | Revised: 21 August 2023 | Accepted: 7 September 2023
DOI: 10.1002/prot.26602

RESEARCH ARTICLE

PROTEINS WILEY

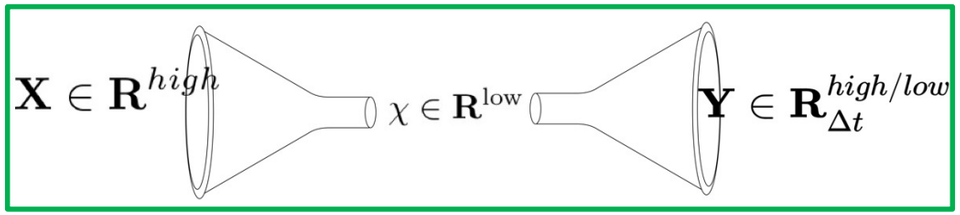
Assessment of three-dimensional RNA structure prediction in CASP15

Rhiju Das[1,2,3] | Rachael C. Kretsch[2] | Adam J. Simpkin[4] | Thomas Mulvaney[5,6] | Phillip Pham[1] | Ramya Rangan[2] | Fan Bu[7,8] | Ronan M. Keegan[4,9] | Maya Topf[5,6] | Daniel J. Rigden[4] | Zhichao Miao[10,11] | Eric Westhof[12]

Abstract

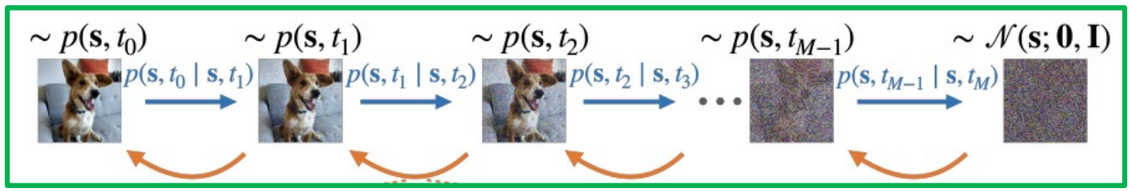
The prediction of RNA three-dimensional structures remains an unsolved problem. Here, we report assessments of RNA structure predictions in CASP15, the first CASP exercise that involved RNA structure modeling. Forty-two predictor groups submitted models for at least one of twelve RNA-containing targets. These models were evaluated by the RNA-Puzzles organizers and, separately, by a CASP-recruited team using metrics (GDT, lDDT) and approaches (Z-score rankings) initially developed for assessment of proteins and generalized here for RNA assessment. The two assessments independently ranked the same predictor groups as first (AIchemy_RNA2), second (Chen), and third (RNAPolis and GeneSilico, tied); predictions from deep learning approaches were significantly worse than these top ranked groups, which did not use deep learning. Further analyses based on direct comparison of predicted models to cryogenic electron microscopy (cryo-EM) maps and x-ray diffraction data support these rankings. With the exception of two RNA-protein complexes, models submitted by CASP15

My lab develops all-atom resolution Molecular Dynamics (MD) simulation methods integrated with machine learning to model molecular diversity and dynamics



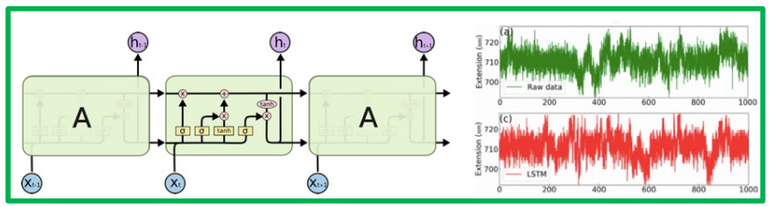
(1) Past-future information bottleneck

Y. Wang, Ribeiro, Tiwary ***Nature Comm. 2019; Angew. Chem. 2022; ACS Central Science 2022***



(2) Denoising diffusion probabilistic models for replica (no) exchange

Y. Wang, Herron, Tiwary ***PNAS 2022***



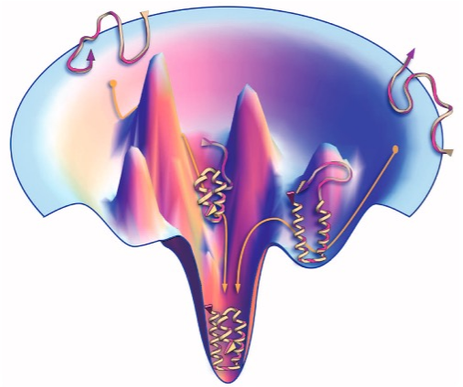
(3) Long-short term memory networks for learning constrained non-Markovian dynamics

Tsai, Kuo, Tiwary ***Nature Comm. 2020; Nature Comm. 2022***

All codes open-source @ github.com/tiwarylab

Today's talk will focus on Thermodynamic Maps approach for sampling RNA type energy landscapes

Proteins

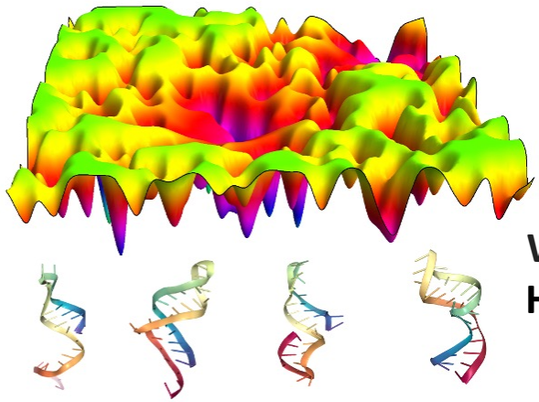


K. Dill, J. McCallum, Science (2012).
The Protein-Folding problem, 50 years on

Fluctuations between metastable states described by few slow modes & timescale separation

Order/Disorder spectrum

RNAs

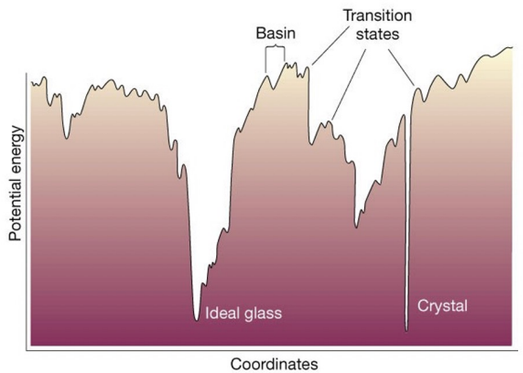


D.J. Wales, Ann. Rev. Phys. Chem. (2017).
Exploring Energy Landscapes

No obvious timescale separation & no dominant driving fluctuations

Wang, Herron, Tiwary, PNAS 2022
Herron…Tiwary, arXiv:2308.14885

Somewhere on the order/disorder spectrum



Beyerle, Zou, Tsai, Tiwary PNAS 2023

Debenedetti, Stillinger, Nature (2001).
Supercooled Liquids & the Glass Transition

Today's talk will focus on Thermodynamic Maps approach for sampling RNA type energy landscapes

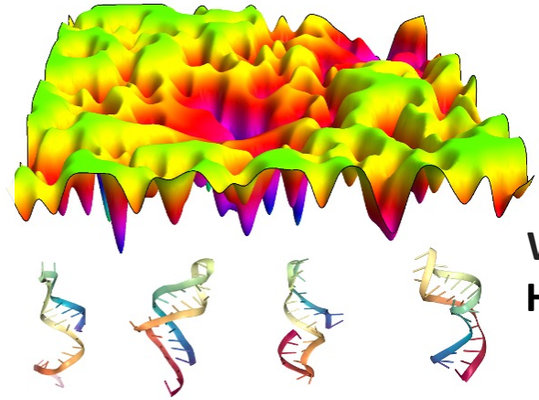
Proteins

K. Dill, J. McCallum, Science (2012).
The Protein-Folding problem, 50 years on

Fluctuations between metastable states described by few slow modes & timescale separation

Order/Disorder spectrum

RNAs



D.J. Wales, Ann. Rev. Phys. Chem. (2017).
Exploring Energy Landscapes

No obvious timescale separation & no dominant driving fluctuations

Wang, Herron, Tiwary, PNAS 2022
Herron…Tiwary, arXiv:2308.14885

Somewhere on the order/disorder spectrum

Debenedetti, Stillinger, Nature (2001).
Supercooled Liquids & the Glass Transition

Beyerle, Zou, Tsai, Tiwary PNAS 2023

Structure of this talk here onwards:

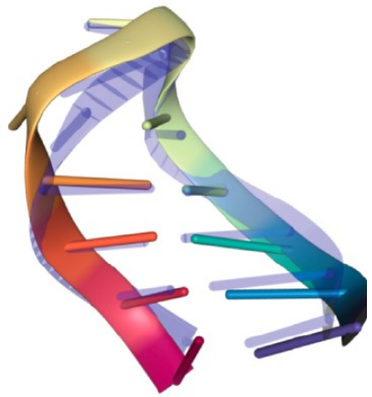
1. Motivating an ensemble view of molecular systems
2. A brief history of methods to estimate ratios of partition functions
3. Score-based models
4. Formulating Thermodynamic Maps
5. Inferring the Ising phase transition with Thermodynamic Maps
6. Thermodynamic Map-accelerated Molecular Dynamics (TM-aMD) applied to RNAs

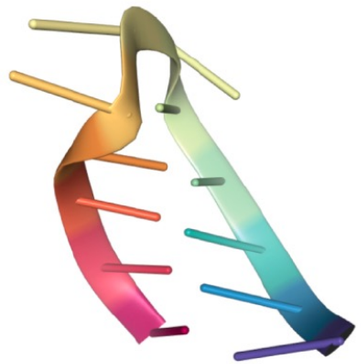
Wang…Tiwary, PNAS (2022): e2203656119
Herron…Tiwary, arXiv:2308.14885

For (bio)molecular systems an ensemble view is required

A single structure is insufficient.
We need to consider the probability distribution over all structures.

$\mathbf{x}$





$p(\mathbf{x})$ 80% 10% 9% 0.4%

$A(\mathbf{x}) =$	Energy $U(\mathbf{x})$	Magnetization $M(\mathbf{x})$	Radius of Gyration $R_g(\mathbf{x})$	NMR couplings $J(\mathbf{x})$

Ensemble-weighted observable can be computed from $p(\mathbf{x})$

$$\langle A(\mathbf{x})\rangle_p = \langle A(\mathbf{x})|p(\mathbf{x})\rangle$$

This framework encompasses widely studied observables across scales:

$A(\mathbf{x}) =$	Energy $U(\mathbf{x})$	Magnetization $M(\mathbf{x})$	Radius of Gyration $R_g(\mathbf{x})$	NMR couplings $J(\mathbf{x})$

Ensemble-weighted observable can be computed from $p(\mathbf{x})$

$$\langle A(\mathbf{x})\rangle_p = \langle A(\mathbf{x})|p(\mathbf{x})\rangle$$

This framework encompasses widely studied observables across scales:

$A(\mathbf{x}) =$	Energy $U(\mathbf{x})$	Magnetization $M(\mathbf{x})$	Radius of Gyration $R_g(\mathbf{x})$	NMR couplings $J(\mathbf{x})$

Ensemble-weighted observable can be computed from $p(\mathbf{x})$

$$\langle A(\mathbf{x})\rangle_p = \langle A(\mathbf{x})|p(\mathbf{x})\rangle$$

This framework encompasses widely studied observables across scales:

	Energy	Magnetization	Radius of Gyration	NMR couplings
$A(\mathbf{x}) =$	$U(\mathbf{x})$	$M(\mathbf{x})$	$R_g(\mathbf{x})$	$J(\mathbf{x})$

Three issues:

1. $\dim(\mathbf{x})$ is often very large, so $p(\mathbf{x})$ is computationally intractable.
2. $p(\mathbf{x})$ is really $p(\mathbf{x}\,|N,P,T)$. Can we infer the dependence of $p(\mathbf{x})$ across thermodynamic ensembles in finite size cases?
3. Exploration of $p(\mathbf{x})$ is usually slow (i.e. MD/MC simulation)

Addressing Point 1:

1. $\dim(\mathbf{x})$ is often very large, so $p(\mathbf{x})$ is computationally intractable.

Estimating the partition function $Z(\beta)$ is sufficient to characterize $p(\mathbf{x})$

$$p(\mathbf{x}) = \frac{e^{-\beta H(\mathbf{x})}}{Z(\beta)} \qquad Z(\beta) = \iiint e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x} \qquad F(\beta) = -\beta^{-1}\log Z(\beta)$$

If you are not convinced, moments (which uniquely describe a distribution) are <u>closely</u> related to the partition function:

$$\mu_1(\beta) = \iiint \mathbf{x}\, e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x} \qquad \mu_2(\beta) = \iiint \mathbf{x}^2\, e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x} \qquad \mu_3(\beta) = \iiint \mathbf{x}^3\, e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x}$$

Mean Variance Skew

Estimating the partition function $Z(\beta)$ is sufficient to characterize $p(\mathbf{x})$

$$p(\mathbf{x}) = \frac{e^{-\beta H(\mathbf{x})}}{Z(\beta)} \qquad Z(\beta) = \iiint e^{-\beta H(\mathbf{x})}\, \mathrm{d}\mathbf{x} \qquad F(\beta) = -\beta^{-1} \log Z(\beta)$$

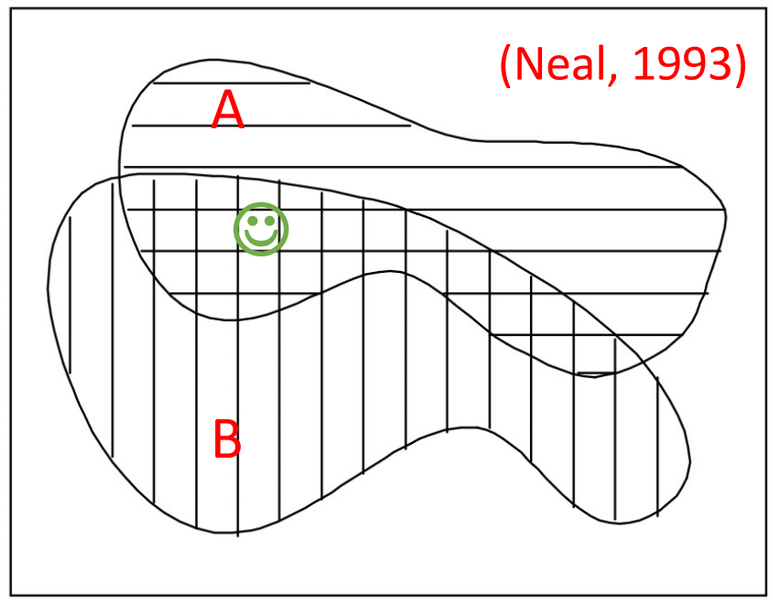
Key observations:

1. We are usually interested in free energy differences (not absolute free energies) and therefore the ratio of partition functions.

$$F_A(\beta) - F_B(\beta) = \log \frac{Z_B(\beta)}{Z_A(\beta)}$$

Central ideas behind Free Energy Perturbation

2. Configuration space overlap increases convergence of estimates of Z



Estimating the partition function $Z(\beta)$ is sufficient to characterize $p(\mathbf{x})$

$$p(\mathbf{x}) = \frac{e^{-\beta H(\mathbf{x})}}{Z(\beta)} \qquad Z(\beta) = \iiint e^{-\beta H(\mathbf{x})}\, d\mathbf{x} \qquad F(\beta) = -\beta^{-1} \log Z(\beta)$$

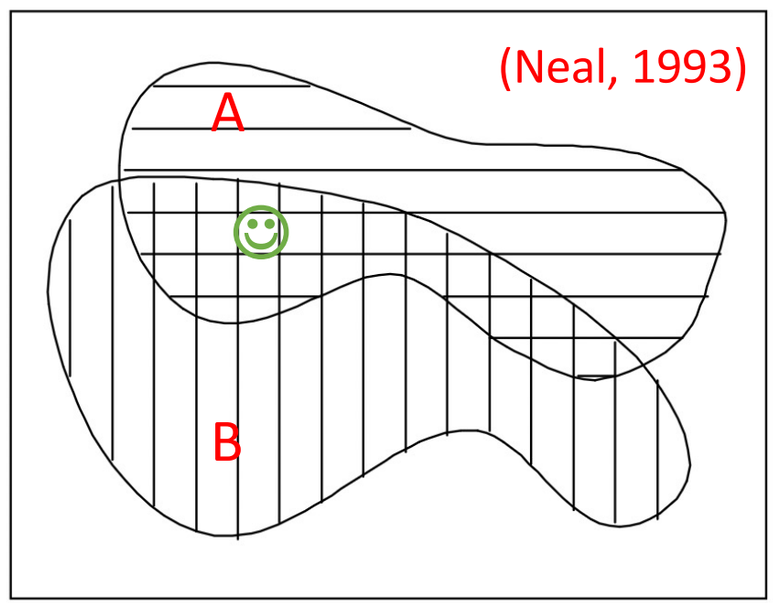
Key observations:

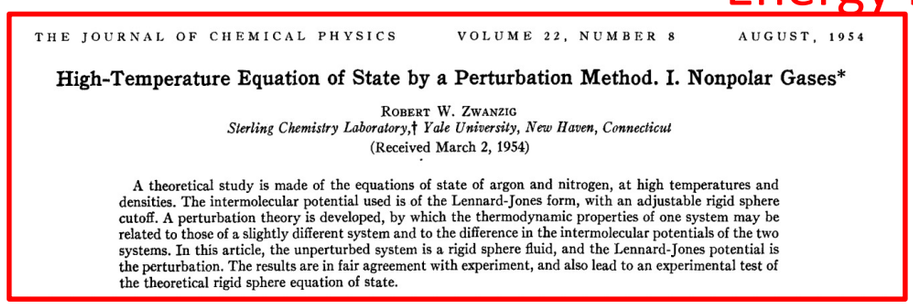
1. We are usually interested in free energy differences (not absolute free energies) and therefore the ratio of partition functions.

$$F_A(\beta) - F_B(\beta) = \log \frac{Z_B(\beta)}{Z_A(\beta)}$$

Central ideas behind Free Energy Perturbation

2. Configuration space overlap increases convergence of estimates of Z





THE JOURNAL OF CHEMICAL PHYSICS VOLUME 22, NUMBER 8 AUGUST, 1954

High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases*

ROBERT W. ZWANZIG
Sterling Chemistry Laboratory,† Yale University, New Haven, Connecticut
(Received March 2, 1954)

A theoretical study is made of the equations of state of argon and nitrogen, at high temperatures and densities. The intermolecular potential used is of the Lennard-Jones form, with an adjustable rigid sphere cutoff. A perturbation theory is developed, by which the thermodynamic properties of one system may be related to those of a slightly different system and to the difference in the intermolecular potentials of the two systems. In this article, the unperturbed system is a rigid sphere fluid, and the Lennard-Jones potential is the perturbation. The results are in fair agreement with experiment, and also lead to an experimental test of the theoretical rigid sphere equation of state.

A brief history of methods for estimating ratios of partition functions

Chapter 1: Markov-Chain Monte-Carlo (Teller, 1953)

Explore $p(\mathbf{x}) = e^{-f(\mathbf{x})}/Z$ while only knowing the energy function $f(\mathbf{x})$

(Z is unknown)

A brief history of methods for estimating ratios of partition functions

Chapter 1: Markov-Chain Monte-Carlo (Teller, 1953)

Explore $p(\mathbf{x}) = e^{-f(\mathbf{x})}/Z$ while only knowing the energy function $f(\mathbf{x})$

(Z is unknown)

1. Generate proposed move $\mathbf{x} \to \mathbf{x}'$

A brief history of methods for estimating ratios of partition functions

Chapter 1: Markov-Chain Monte-Carlo (Teller, 1953)

Explore $p(\mathbf{x}) = e^{-f(\mathbf{x})}/Z$ while only knowing the energy function $f(\mathbf{x})$

(Z is unknown)

1. Generate proposed move	$\mathbf{x} \to \mathbf{x}'$
2. Compute ratio of probabilities	$p(\mathbf{x})/p(\mathbf{x}') = e^{-(f(\mathbf{x})-f(\mathbf{x}'))} = e^{-\Delta f}$

A brief history of methods for estimating ratios of partition functions

Chapter 1: Markov-Chain Monte-Carlo (Teller, 1953)

Explore $p(\mathbf{x}) = e^{-f(\mathbf{x})}/Z$ while only knowing the energy function $f(\mathbf{x})$

(Z is unknown)

1. Generate proposed move	$\mathbf{x} \rightarrow \mathbf{x}'$
2. Compute ratio of probabilities	$p(\mathbf{x})/p(\mathbf{x}') = e^{-(f(\mathbf{x})-f(\mathbf{x}'))} = e^{-\Delta f}$
3. Accept or reject proposal	If $e^{-\Delta f} < 1$, then $p(\text{accept}) = 1$ If $e^{-\Delta f} > 1$, then $p(\text{accept}) = e^{-\Delta f}$

A brief history of methods for estimating ratios of partition functions

Chapter 1: Markov-Chain Monte-Carlo (Teller, 1953)

Explore $p(\mathbf{x}) = e^{-f(\mathbf{x})}/Z$ while only knowing the energy function $f(\mathbf{x})$

(Z is unknown)

1. Generate proposed move	$\mathbf{x} \to \mathbf{x}'$
2. Compute ratio of probabilities	$p(\mathbf{x})/p(\mathbf{x}') = e^{-(f(\mathbf{x})-f(\mathbf{x}'))} = e^{-\Delta f}$
3. Accept or reject proposal	If $e^{-\Delta f} < 1$, then $p(\text{accept}) = 1$ If $e^{-\Delta f} > 1$, then $p(\text{accept}) = e^{-\Delta f}$

Detailed balance criteria

Ensures microscopic reversibility + sampling equilibrium distribution of $f(\mathbf{x})$

Problem: Monte-Carlo sampling can be very slow to generate independent samples

A brief history of methods for estimating ratios of partition functions

Chapter 2: Simulated Tempering (Parisi, 1992)

Accelerate exploration of $p(\mathbf{x})$ by exploring different ensembles
(e.g. a hotter $p(\mathbf{x}) \propto e^{-\beta H(\mathbf{x})}$)

A brief history of methods for estimating ratios of partition functions

Chapter 2: Simulated Tempering (Parisi, 1992)

Accelerate exploration of $p(\mathbf{x})$ by exploring different ensembles
(e.g. a hotter $p(\mathbf{x}) \propto e^{-\beta H(\mathbf{x})}$)

1. Generate proposed move	$(\mathbf{x}, \beta) \rightarrow (\mathbf{x}', \beta')$

A brief history of methods for estimating ratios of partition functions

Chapter 2: Simulated Tempering (Parisi, 1992)

Accelerate exploration of $p(\mathbf{x})$ by exploring different ensembles
(e.g. a hotter $p(\mathbf{x}) \propto e^{-\beta H(\mathbf{x})}$)

1. Generate proposed move $(\mathbf{x}, \beta) \rightarrow (\mathbf{x}', \beta')$

2. Compute ratio of probabilities $\frac{p(\mathbf{x}|\beta)}{p(\mathbf{x}'|\beta')} = e^{-(\beta-\beta')(H(\mathbf{x})-H(\mathbf{x}'))} = e^{-\Delta H \Delta \beta}$

Partition function of the extended ensemble vanishes

$$\mathcal{Z} = \prod Z(\beta_i)$$

A brief history of methods for estimating ratios of partition functions

Chapter 2: Simulated Tempering (Parisi, 1992)

Accelerate exploration of $p(\mathbf{x})$ by exploring different ensembles
(e.g. a hotter $p(\mathbf{x}) \propto e^{-\beta H(\mathbf{x})}$)

1. Generate proposed move	$(\mathbf{x}, \beta) \rightarrow (\mathbf{x}', \beta')$
2. Compute ratio of probabilities	$\frac{p(\mathbf{x}\|\beta)}{p(\mathbf{x}'\|\beta')} = e^{-(\beta-\beta')(H(\mathbf{x})-H(\mathbf{x}'))} = e^{-\Delta H \Delta\beta}$
3. Accept or reject proposal	If $e^{-\Delta H \Delta\beta} < 1$, then $p(\text{accept}) = 1$ If $e^{-\Delta H \Delta\beta} > 1$, then $p(\text{accept}) = e^{-\Delta H \Delta\beta}$

Partition function of the extended ensemble vanishes

$$\mathcal{Z} = \prod Z(\beta_i)$$

Transitions *between* ensembles must also satisfy the detailed balance condition

A brief history of methods for estimating ratios of partition functions

Chapter 2: Simulated Tempering (Parisi, 1992)

Accelerate exploration of $p(\mathbf{x})$ by exploring different ensembles
(e.g. a hotter $p(\mathbf{x}) \propto e^{-\beta H(\mathbf{x})}$)

1. Generate proposed move	$(\mathbf{x}, \beta) \rightarrow (\mathbf{x}', \beta')$
2. Compute ratio of probabilities	$\frac{p(\mathbf{x}\|\beta)}{p(\mathbf{x}'\|\beta')} = e^{-(\beta-\beta')(H(\mathbf{x})-H(\mathbf{x}'))} = e^{-\Delta H \Delta\beta}$
3. Accept or reject proposal	If $e^{-\Delta H \Delta\beta} < 1$, then $p(\text{accept}) = 1$ If $e^{-\Delta H \Delta\beta} > 1$, then $p(\text{accept}) = e^{-\Delta H \Delta\beta}$

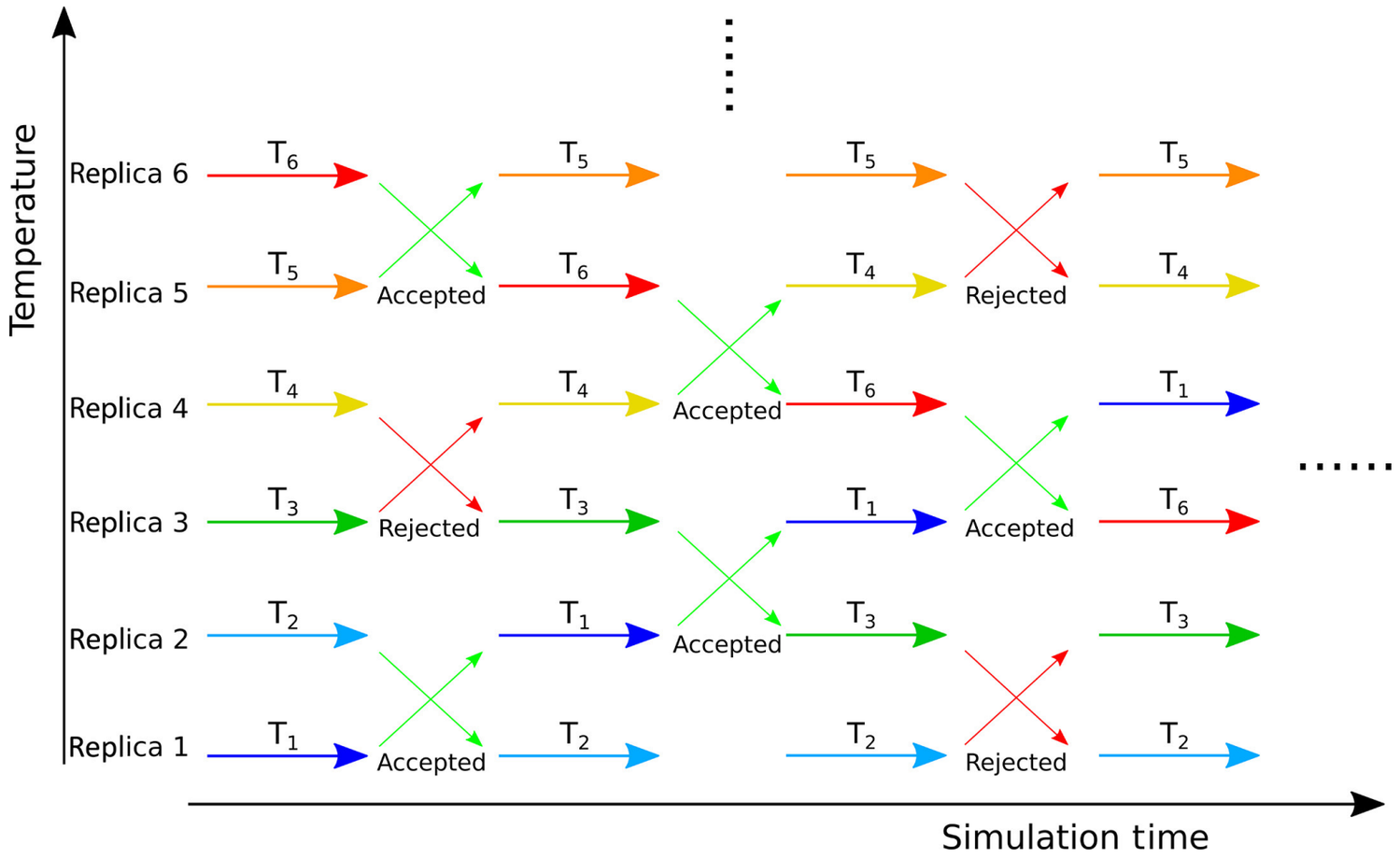
Partition function of the extended ensemble vanishes

$$\mathcal{Z} = \prod Z(\beta_i)$$

Transitions *between* ensembles must also satisfy the detailed balance condition

Core idea: Accelerate sampling by exploring a distributions other than the one of interest. In this case tempered distributions sharing $H(\mathbf{x})$.

Replica exchange is a simulated tempering protocol



Each rung on the temperature ladder is a tempered distribution

$H(\mathbf{x})$ is the MD force field

Use MBAR (Chodera, Shirts, 2008) to (approximately) recover

$\frac{Z(\beta_i)}{Z(\beta_j)}$ from $\mathcal{Z}$

Sugita, Okamoto. Chem. Phys. Lett. (1999)

A brief history of methods for estimating ratios of partition functions

Chapter 3: Annealed Importance Sampling (Neal, 1998*) (aka "The missing link")

Accelerate exploration of $p(\mathbf{x})$ by interpolating between energy functions $H_0(\mathbf{x})$ and $H_1(\mathbf{x})$

Intermediate energies:

$$H_\lambda(\mathbf{x}) = \lambda H_0(\mathbf{x}) + (1-\lambda) H_1(\mathbf{x}).$$

- Independently discovered by Gelman and Meng (1998) who called their method Thermodynamic Integration and very closely related to Jarzynski (1997)

A brief history of methods for estimating ratios of partition functions

Chapter 3: Annealed Importance Sampling (Neal, 1998*) (aka "The missing link")

Accelerate exploration of $p(\mathbf{x})$ by interpolating between energy functions $H_0(\mathbf{x})$ and $H_1(\mathbf{x})$

Intermediate energies:

$$H_\lambda(\mathbf{x}) = \lambda H_0(\mathbf{x}) + (1-\lambda) H_1(\mathbf{x}).$$

$$\frac{Z_0}{Z_1} = \frac{Z_0}{Z_{\lambda_1}} \frac{Z_{\lambda_1}}{Z_{\lambda_2}} \cdots \frac{Z_{\lambda_i}}{Z_{\lambda_{i+1}}} \cdots \frac{Z_{\lambda_{n-1}}}{Z_1}$$

This is the first situation where taking advantage of a prior distribution for $H_1(\mathbf{x})$ seems useful.

- Independently discovered by Gelman and Meng (1998) who called their method Thermodynamic Integration and very closely related to Jarzynski (1997)

A brief history of methods for estimating ratios of partition functions

Chapter 3: Annealed Importance Sampling (Neal, 1998*) (aka "The missing link")

$$\frac{Z_0}{Z_1} = \frac{Z_0}{Z_{\lambda_1}} \frac{Z_{\lambda_1}}{Z_{\lambda_2}} \cdots \frac{Z_{\lambda_i}}{Z_{\lambda_{i+1}}} \cdots \frac{Z_{\lambda_{n-1}}}{Z_1}$$

This is the first situation where taking advantage of a prior distribution for $H_1(\mathbf{x})$ seems useful.

A brief history of methods for estimating ratios of partition functions

Chapter 3: Annealed Importance Sampling (Neal, 1998*) (aka "The missing link")

$$\frac{Z_0}{Z_1} = \frac{Z_0}{Z_{\lambda_1}} \frac{Z_{\lambda_1}}{Z_{\lambda_2}} \cdots \frac{Z_{\lambda_i}}{Z_{\lambda_{i+1}}} \cdots \frac{Z_{\lambda_{n-1}}}{Z_1}$$

This is the first situation where taking advantage of a prior distribution for $H_1(\mathbf{x})$ seems useful.

Let $H_0(\mathbf{x}) =$ MD force field

$$H_1(\mathbf{x}) = \mathbf{x}^2$$

$$H_\lambda(\mathbf{x}) = \lambda H_0(\mathbf{x}) + (1-\lambda) H_1(\mathbf{x}).$$

A brief history of methods for estimating ratios of partition functions

Chapter 3: Annealed Importance Sampling (Neal, 1998*) (aka "The missing link")

$$\frac{Z_0}{Z_1} = \frac{Z_0}{Z_{\lambda_1}}\frac{Z_{\lambda_1}}{Z_{\lambda_2}} \cdots \frac{Z_{\lambda_i}}{Z_{\lambda_{i+1}}} \cdots \frac{Z_{\lambda_{n-1}}}{Z_1}$$ ⟵

This is the first situation where taking advantage of a prior distribution for $H_1(\mathbf{x})$ seems useful.

Let $H_0(\mathbf{x})$ = MD force field

$$H_1(\mathbf{x}) = \mathbf{x}^2$$

$$H_\lambda(\mathbf{x}) = \lambda H_0(\mathbf{x}) + (1-\lambda)H_1(\mathbf{x}).$$

$p_0(\mathbf{x}) \propto e^{-\beta H_0(\mathbf{x})}$ ⟵ Slow to generate i.i.d samples ☹

$p_1(\mathbf{x}) \propto e^{-\beta \mathbf{x}^2}$ ⟵ Every sample is i.i.d. ☺

A brief history of methods for estimating ratios of partition functions

Chapter 3: Annealed Importance Sampling (Neal, 1998*) (aka "The missing link")

$$\frac{Z_0}{Z_1} = \frac{Z_0}{Z_{\lambda_1}}\frac{Z_{\lambda_1}}{Z_{\lambda_2}} \cdots \frac{Z_{\lambda_i}}{Z_{\lambda_{i+1}}} \cdots \frac{Z_{\lambda_{n-1}}}{Z_1}$$ ⟵

This is the first situation where taking advantage of a prior distribution for $H_1(\mathbf{x})$ seems useful.

Let $H_0(\mathbf{x})$ = MD force field

$$H_1(\mathbf{x}) = \mathbf{x}^2$$

$$H_\lambda(\mathbf{x}) = \lambda H_0(\mathbf{x}) + (1-\lambda) H_1(\mathbf{x})$$

$p_0(\mathbf{x}) \propto e^{-\beta H_0(\mathbf{x})}$ ⟵ Slow to generate i.i.d samples ☹

$p_1(\mathbf{x}) \propto e^{-\beta \mathbf{x}^2}$ ⟵ Every sample is i.i.d. ☺

After lots of simulation, once all $\frac{Z_i}{Z_{i+1}}$ are estimated, samples from $\mathbf{x} \sim \mathcal{N}(0, \beta^{-1})$ can be used to compute free energy estimates for $p_0(\mathbf{x})$ *at virtually no cost*

Score-based models extend Annealed Importance Sampling to cases where the energy function is unknown.

Instead of interpolating energy functions,
interpolate the samples themselves.

Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv:2011.13456* (2020).

Score-based models extend Annealed Importance Sampling to cases where the energy function is unknown.

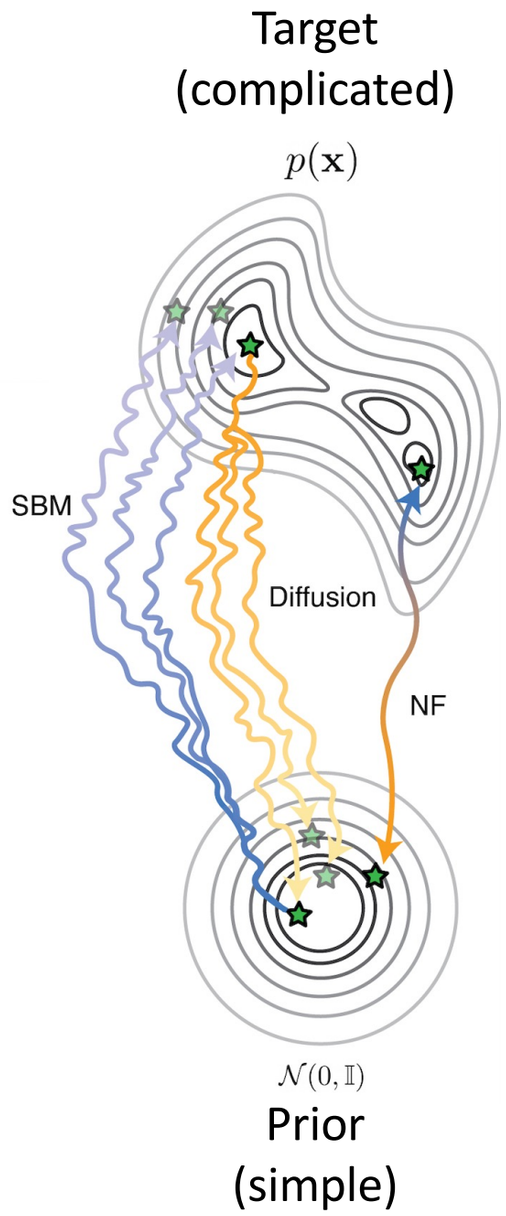
Instead of interpolating energy functions, interpolate the samples themselves.

If $\mathbf{x}_0 \sim e^{-H_0(\mathbf{x})}$ and $\mathbf{x}_1 \sim e^{-H_1(\mathbf{x})}$, then $\mathbf{x}_\lambda \sim e^{-H_\lambda(\mathbf{x})}$

Generally take $H_1(\mathbf{x}) = \mathbf{x}^2$, and calculate how

probability $p_\lambda(\mathbf{x}) \propto e^{-H_\lambda(\mathbf{x})}$ flows

$p_\lambda(\mathbf{x}) \propto e^{-H_\lambda(\mathbf{x})}$ is then a diffusion process as $\lambda: 0 \to 1$



Picture from
Mehdi, …,Tiwary Ann Rev Phys Chem vol 75

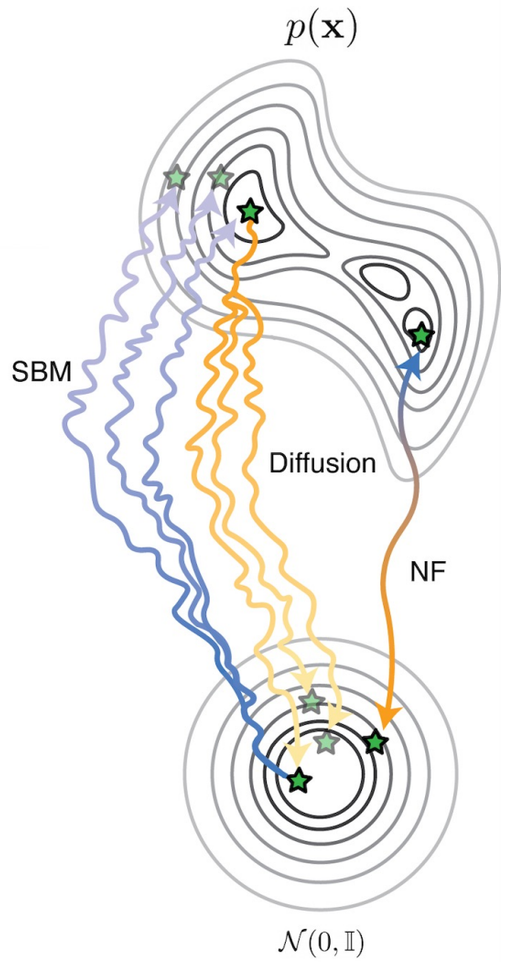
Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv:2011.13456* (2020).

A remarkable property of diffusion processes makes score-based models possible

(Anderson, 1982)

For any diffusion SDE of the form following $p(\mathbf{x}, t)$

$$d\mathbf{x} = \underbrace{-f(\mathbf{x}, t)dt}_{\text{Drift}} + \underbrace{g(t)dw}_{\text{Noise}}$$



Picture from
Mehdi, …,Tiwary Ann Rev Phys Chem vol 75

Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv:2011.13456* (2020).

A remarkable property of diffusion processes makes score-based models possible

(Anderson, 1982)

For any diffusion SDE of the form following $p(\mathbf{x}, t)$

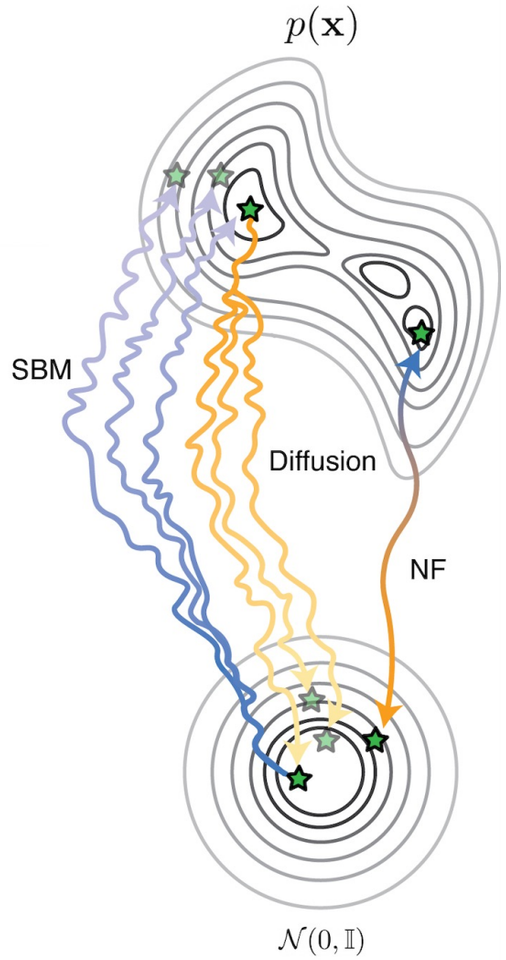
$$d\mathbf{x} = \underbrace{-f(\mathbf{x}, t)dt}_{\text{Drift}} + \underbrace{g(t)dw}_{\text{Noise}}$$

There exists a reverse diffusion of the form:

$$d\mathbf{x} = \underbrace{-[f(\mathbf{x}, t) + g(t)^2 \nabla_{\mathbf{x}} \log p(\mathbf{x}, t)]dt}_{\text{Drift}} + \underbrace{g(t)d\bar{w}}_{\text{Noise}}$$

Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv:2011.13456* (2020).

Target
(complicated)



Prior
(simple)

Picture from
Mehdi, …,Tiwary Ann Rev Phys Chem vol 75

A remarkable property of diffusion processes makes score-based models possible

(Anderson, 1982)

For any diffusion SDE of the form following $p(\mathbf{x}, t)$

$$d\mathbf{x} = \underbrace{-f(\mathbf{x}, t)dt}_{\text{Drift}} + \underbrace{g(t)dw}_{\text{Noise}}$$

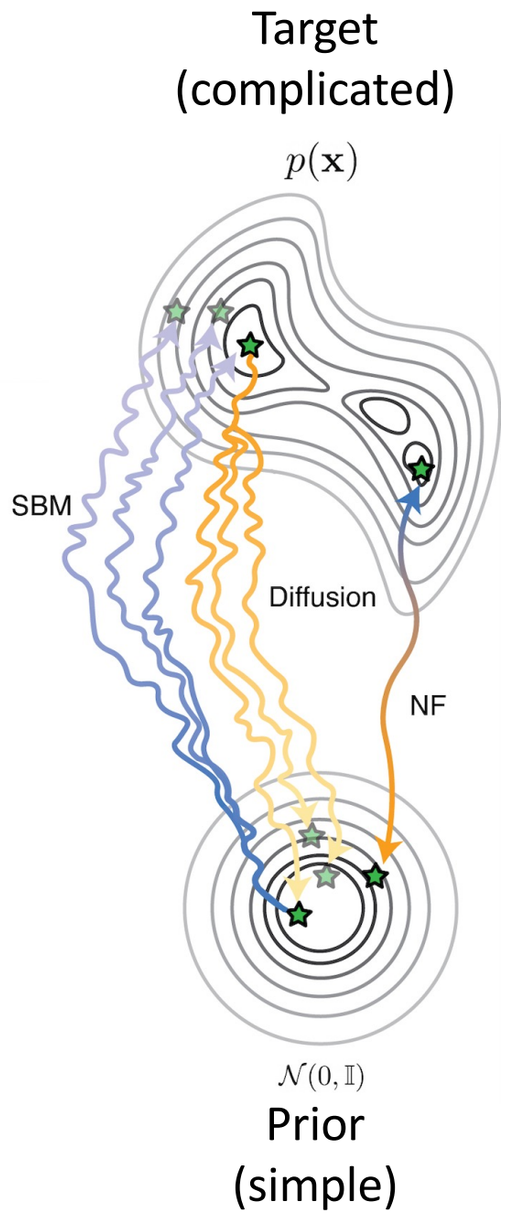
There exists a reverse diffusion of the form:

$$d\mathbf{x} = \underbrace{-[f(\mathbf{x}, t) + g(t)^2 \nabla_{\mathbf{x}} \log p(\mathbf{x}, t)]dt}_{\text{Drift}} + \underbrace{g(t)d\bar{w}}_{\text{Noise}}$$

Define the score: $s_\theta(\mathbf{x}) = \nabla_{\mathbf{x}} \log p(\mathbf{x}, t)$

The score is estimated by many evaluations of the forward SDE, and a Deep Neural Network is trained to estimate the score from $\mathbf{x}_t$

Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv:2011.13456* (2020).

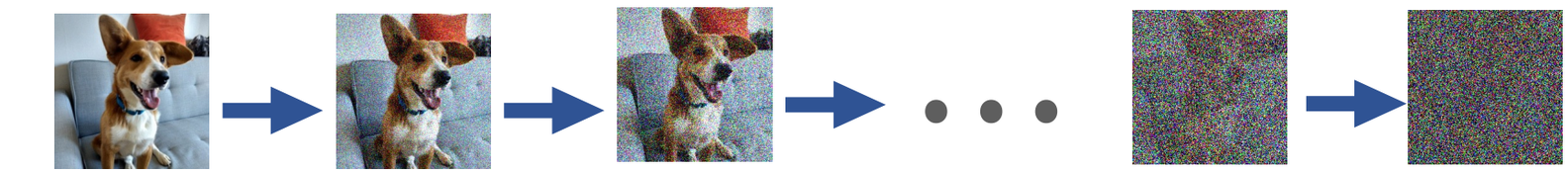


Picture from
Mehdi, …,Tiwary Ann Rev Phys Chem vol 75

Example: Denoising Diffusion Probabilistic Models (DDPMs)

Diffusion

$$q(\mathbf{x}^{(0)}) \longrightarrow q(\mathbf{x}^{(1)}) \longrightarrow \cdots \longrightarrow q(\mathbf{x}^{(T)})$$

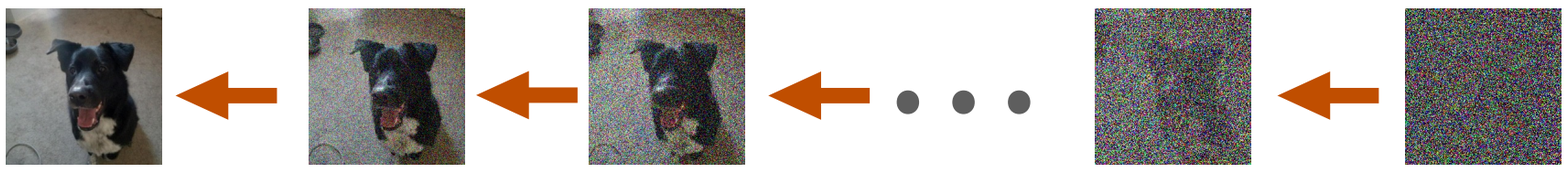


$\sim q(\mathbf{x}^{(0)})$ $\sim q(\mathbf{x}^{(1)})$ $\sim q(\mathbf{x}^{(2)})$ $\sim q(\mathbf{x}^{(T-1)})$ $\sim \mathcal{N}(\mathbf{x}^{(t)}; \mathbf{0}, \mathbf{I})$

Easier to sample noise
But how to convert noise back to data?

In reverse (denoising)

$$p(\mathbf{x}^{(0)}) \longleftarrow \cdots \longleftarrow p(\mathbf{x}^{(T-1)}) \longleftarrow p(\mathbf{x}^{(T)})$$



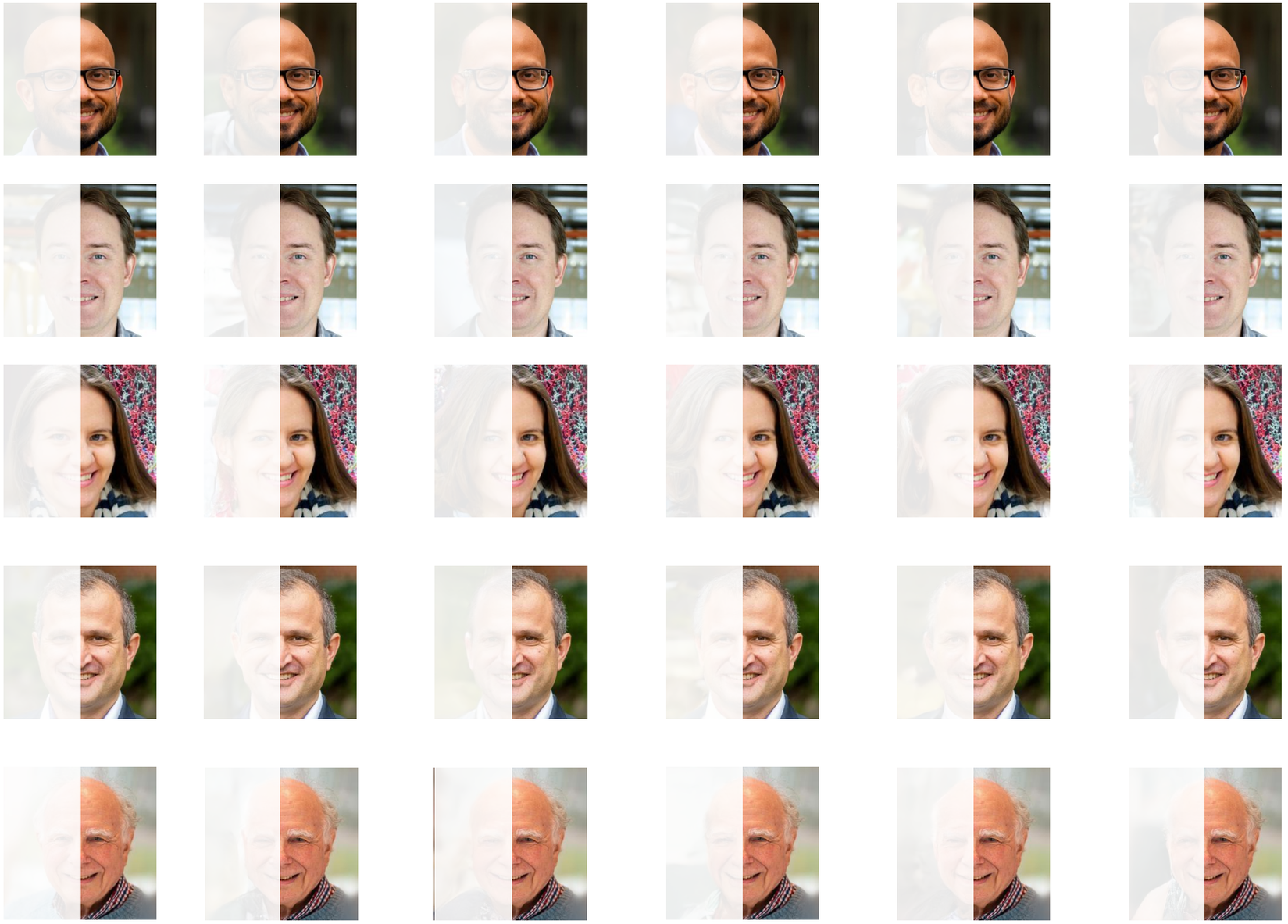
$\sim p(\mathbf{x}^{(0)})$ $\sim p(\mathbf{x}^{(1)})$ $\sim p(\mathbf{x}^{(2)})$ $\sim p(\mathbf{x}^{(T-1)})$ $\sim p(\mathbf{x}^{(T)})$

Backward diffusion kernel $p(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}) = \mathcal{N}\left(\mathbf{x}^{(t-1)}; \mathbf{f}_\mu(\mathbf{x}^{(t)}), \mathbf{f}_\Sigma(\mathbf{x}^{(t)})\right)$

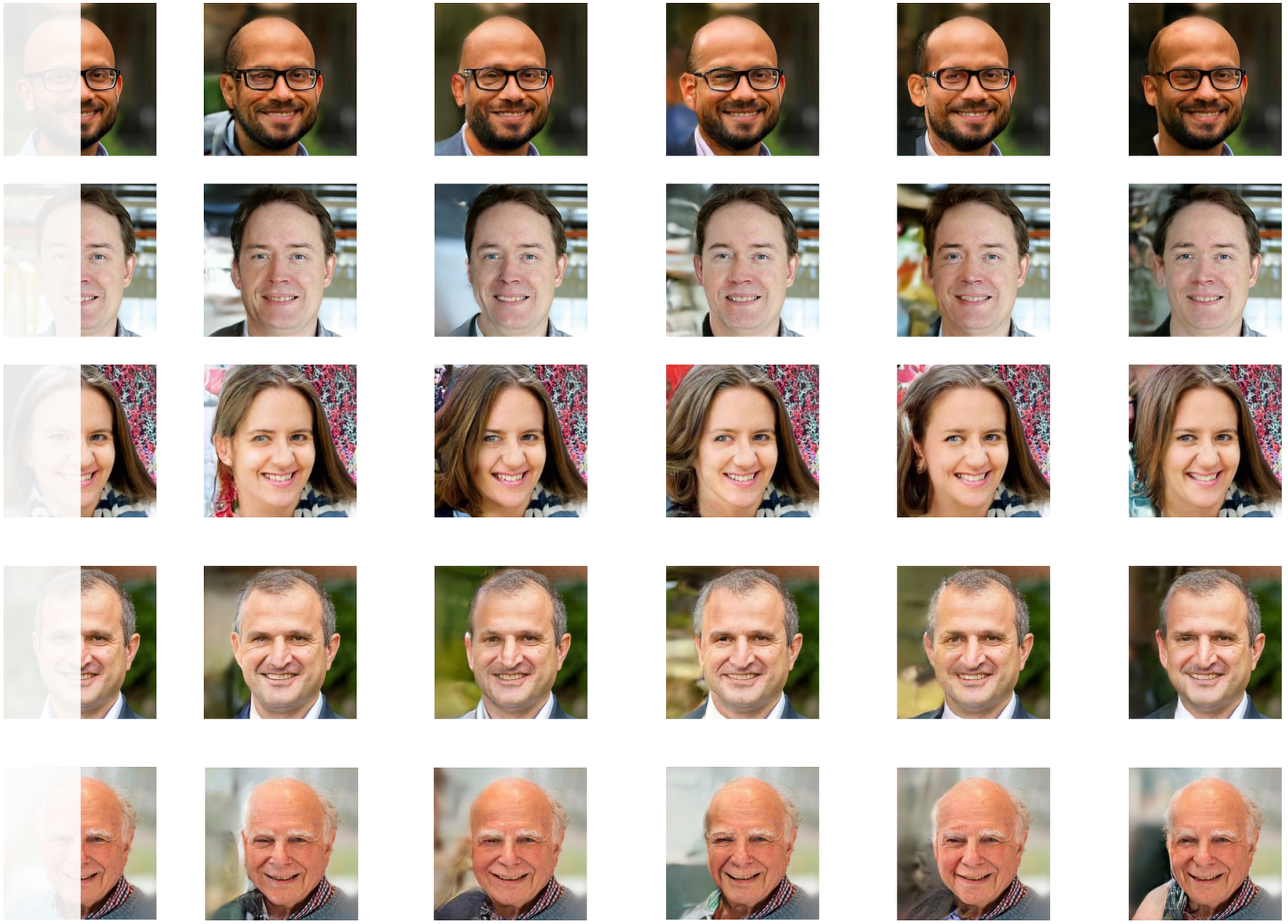
Wang, Herron, Tiwary PNAS 2022

Herron, Mondal, Schneekloth, Tiwary arXiv:2308.14885

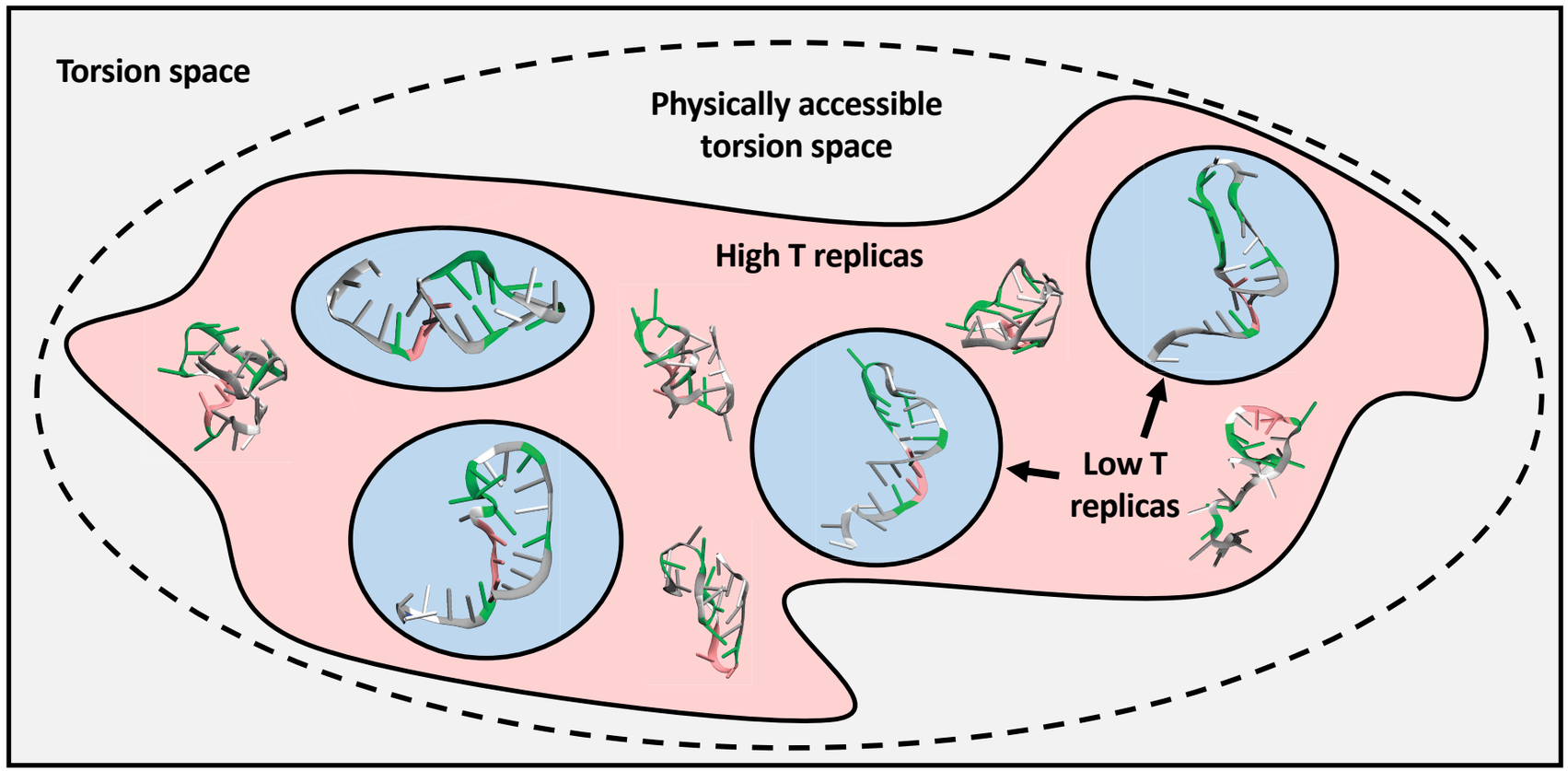
Conditional Generation: right side of face having seen only left side



Conditional Generation: right side of face having seen only left side



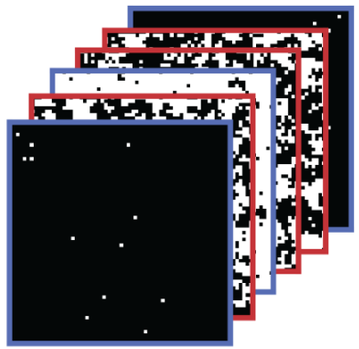
Original motivation:
Conditional generation of face halves →
Conditional generation of molecular conformations



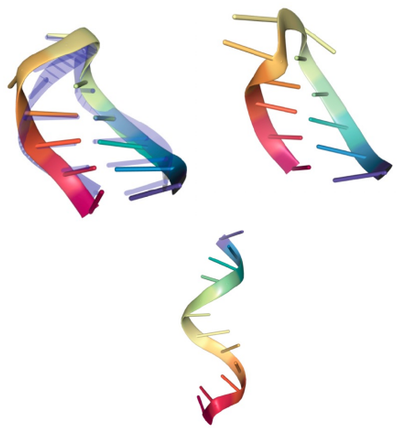
Wang…Tiwary, PNAS (2022): e2203656119
Herron…Tiwary, arXiv:2308.14885

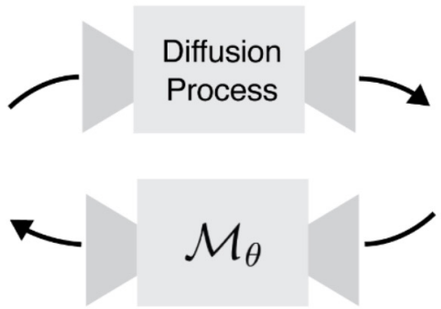
Thermodynamic Maps: Extended-ensemble Free Energy Perturbation (FEP) with score-based models

Samples from a complicated equilibrium distribution: $(\mathbf{x}, \beta)$

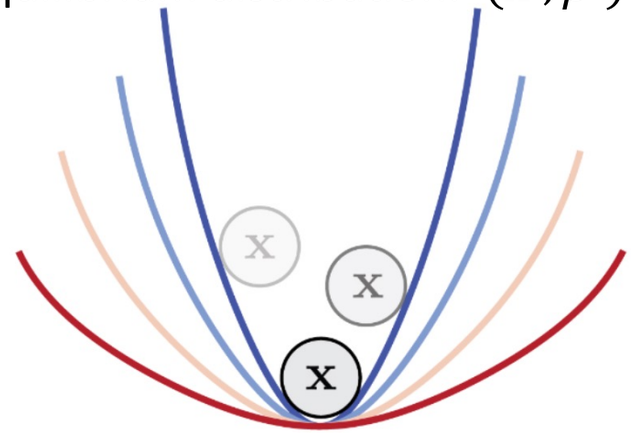


$p(\mathbf{x}, \beta) \propto e^{-\beta U(\mathbf{x})}$





Samples from a **simple** equilibrium distribution: $(\mathbf{x}', \beta')$



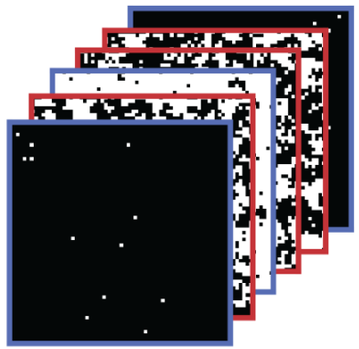
$q(\mathbf{x}', \beta') \propto e^{-\beta' \mathbf{x}'^2}$

$$Z(\beta') = \sqrt{\frac{2\pi}{\beta'}}$$

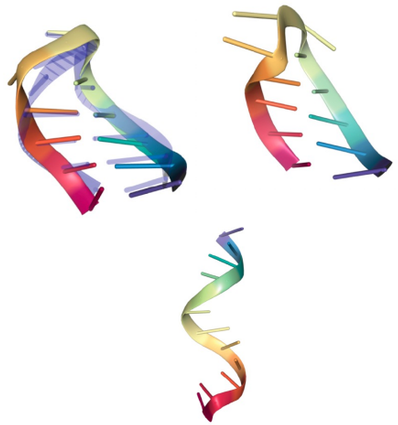
Herron, Mondal, Schneekloth, Tiwary arXiv:2308.14885

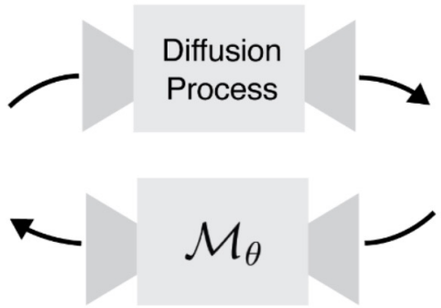
Thermodynamic Maps: Extended-ensemble Free Energy Perturbation (FEP) with score-based models

Samples from a complicated equilibrium distribution: $(\mathbf{x}, \beta)$

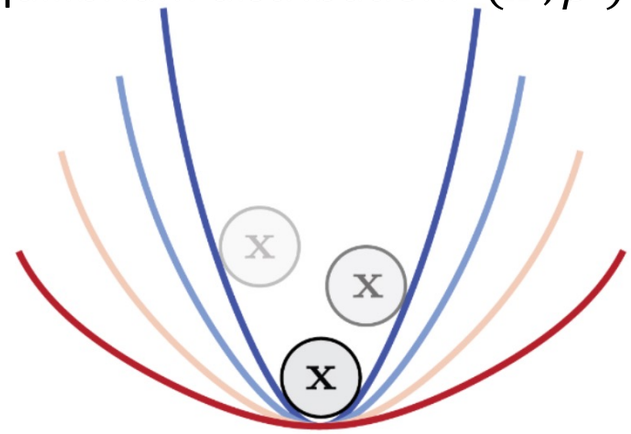


$p(\mathbf{x}, \beta) \propto e^{-\beta U(\mathbf{x})}$





Samples from a **simple** equilibrium distribution: $(\mathbf{x}', \beta')$



$q(\mathbf{x}', \beta') \propto e^{-\beta' \mathbf{x}'^2}$

$Z(\beta') = \sqrt{\frac{2\pi}{\beta'}}$

Forward diffusion:

$$\begin{pmatrix} \mathrm{d}\mathbf{x} \\ \mathrm{d}\boldsymbol{\beta}^{-1} \end{pmatrix} = -\frac{1}{2}\sigma(t) \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\beta}^{-1} \end{pmatrix} \mathrm{d}t + \sqrt{\sigma(t)} \begin{pmatrix} \sqrt{\boldsymbol{\beta}_0^{-1}} \\ \mathbf{1} \end{pmatrix} \mathrm{d}\mathbf{w}$$

Reverse diffusion:

$$\begin{pmatrix} \mathrm{d}\mathbf{x} \\ \mathrm{d}\boldsymbol{\beta}^{-1} \end{pmatrix} = -\frac{1}{2}\sigma(t) \left[\begin{pmatrix} \mathbf{x} \\ \boldsymbol{\beta}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{s}_\theta(\mathbf{x}, t) \\ \mathbf{s}_\theta(\boldsymbol{\beta}^{-1}, t) \end{pmatrix} \right] \mathrm{d}t + \sqrt{\sigma(t)} \begin{pmatrix} \sqrt{\boldsymbol{\beta}_0^{-1}} \\ \mathbf{1} \end{pmatrix} \mathrm{d}\mathbf{w}$$

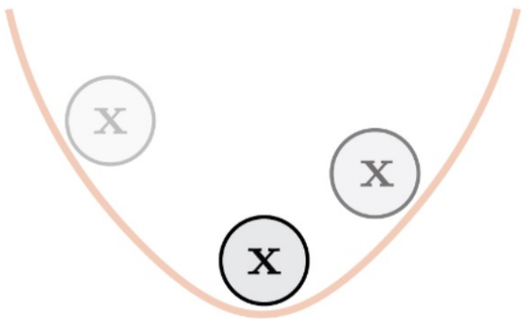
$\boldsymbol{\beta}^{-1}$ is statistical temperature – same dimension as $\mathbf{X}$ and calculated from variances

Herron, Mondal, Schneekloth, Tiwary arXiv:2308.14885

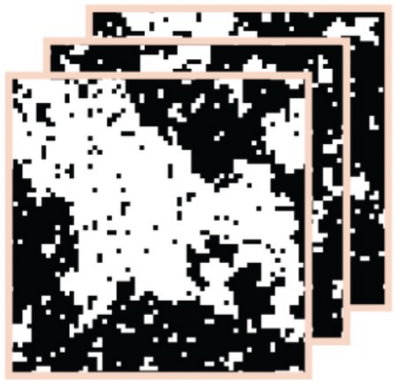
Thermodynamic Maps: Extended-ensemble FEP with score-based models

$$\begin{pmatrix} \mathrm{d}\mathbf{x} \\ \mathrm{d}\boldsymbol{\beta}^{-1} \end{pmatrix} = -\frac{1}{2}\sigma(t)\left[\begin{pmatrix} \mathbf{x} \\ \boldsymbol{\beta}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{s}_\theta(\mathbf{x}, t) \\ \mathbf{s}_\theta(\boldsymbol{\beta}^{-1}, t) \end{pmatrix}\right]\mathrm{d}t + \sqrt{\sigma(t)}\begin{pmatrix} \sqrt{\boldsymbol{\beta}_0^{-1}} \\ \mathbf{1} \end{pmatrix}\mathrm{d}\mathbf{w}$$

To generate samples at temperature β_c, sample from the corresponding prior system $\mathcal{N}(0, \beta_c^{-1})$
where β_c^{-1} is some temperature of interest



$q(\mathbf{x}', \beta_c') \propto e^{-\beta_c' \mathbf{x}'^2}$



$p(\mathbf{x}, \beta_c) \propto e^{-\beta_c U(\mathbf{x})}$

Herron, Mondal, Schneekloth, Tiwary arXiv:2308.14885

Thermodynamic Maps: Extended-ensemble FEP with score-based models

Ensemble-weighted observable can be computed from $p(\mathbf{x})$

$$\langle A(\mathbf{x})\rangle_p = \langle A(\mathbf{x})|p(\mathbf{x})\rangle$$

This framework encompasses widely studied observables across scales:

$A(\mathbf{x}) =$	Energy $U(\mathbf{x})$	Magnetization $M(\mathbf{x})$	Radius of Gyration $R_g(\mathbf{x})$	NMR couplings $J(\mathbf{x})$

Three issues:

1. $\dim(\mathbf{x})$ is often very large, so $p(\mathbf{x})$ is computationally intractable.
2. $p(\mathbf{x})$ is really $p(\mathbf{x}|N,P,T)$. Can we infer the dependence of $p(\mathbf{x})$ across thermodynamic ensembles in finite size cases?
3. Exploration of $p(\mathbf{x})$ is usually slow (i.e. MD/MC simulation)

Within the simple system, the dependence of the partition function on temperature is analytically tractable:

$$Z(\beta') = \sqrt{\frac{2\pi}{\beta'}} \qquad \mu_2(\beta') = \beta'^{-1}$$

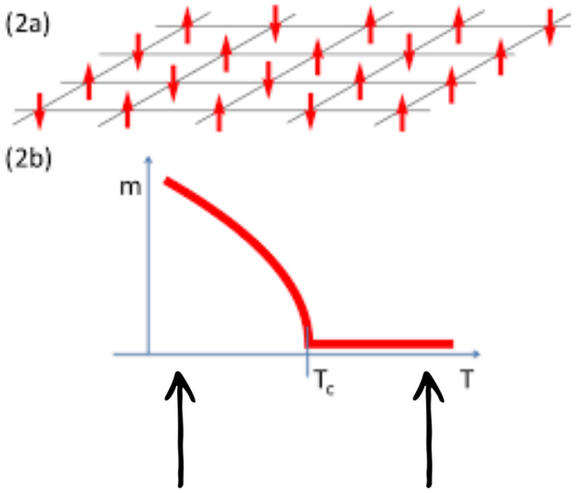
$$\begin{pmatrix} \mathrm{d}\mathbf{x} \\ \mathrm{d}\boldsymbol{\beta}^{-1} \end{pmatrix} = -\frac{1}{2}\sigma(t)\begin{pmatrix} \mathbf{x} \\ \boldsymbol{\beta}^{-1} \end{pmatrix}\mathrm{d}t + \sqrt{\sigma(t)}\begin{pmatrix} \sqrt{\boldsymbol{\beta}_0^{-1}} \\ \mathbf{1} \end{pmatrix}\mathrm{d}\mathbf{w}$$ has equilibrium distribution $\mathcal{N}(0,\beta_0^{-1})$

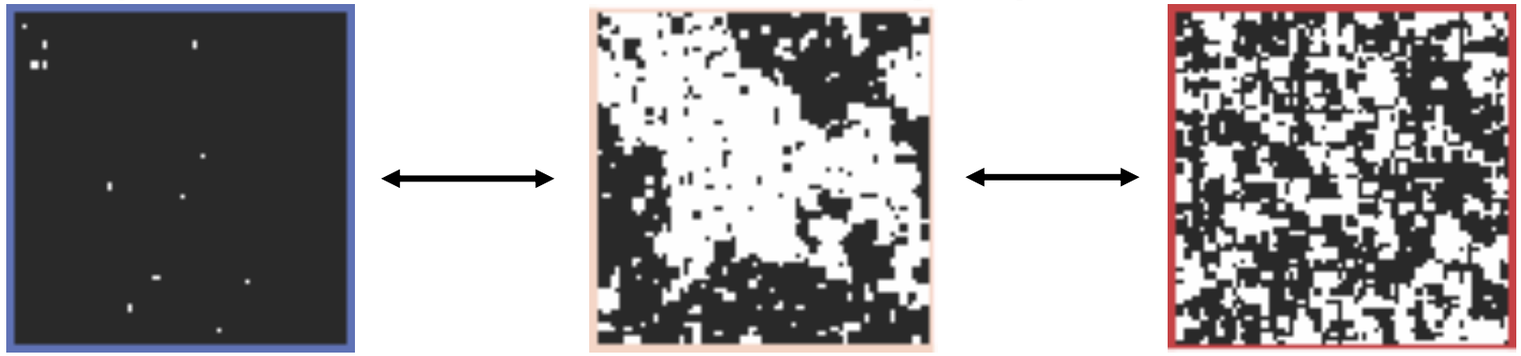
Addresses Point 2: Can we infer the dependence of $p(x)$ across thermodynamic ensembles?

Given samples x from p(x|β_1), p(x|β_2)… we can generate samples x at other p(x|β)

Can also do conditional on pressure, number, other constraints

Thermodynamic Maps on Ising Model: predict phase transitions without seeing any





Low T

Critical T

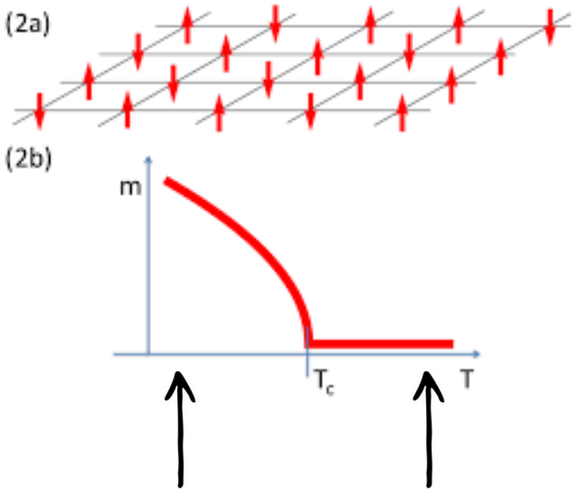
High T

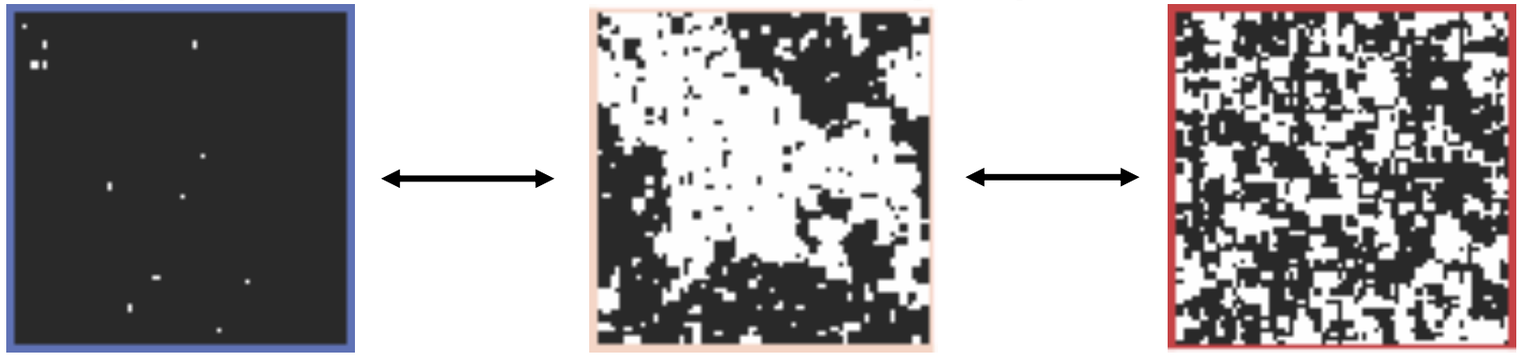
Can one predict critical temperature, heat capacity, critical exponents with samples from 2 temperatures deep within paramagnetic and ferromagnetic phases?

$$M \sim |\tau|^{\beta} \quad \text{and} \quad C \sim |\tau|^{-\alpha} \quad \text{where} \quad \tau = \frac{T - T_c}{T_c}.$$

Herron, Mondal, Schneekloth, Tiwary arXiv:2308.14885

Thermodynamic Maps on Ising Model: predict phase transitions without seeing any





Low T | Critical T | High T

Can one predict critical temperature, heat capacity, critical exponents with samples from 2 temperatures deep within paramagnetic and ferromagnetic phases?

YES!

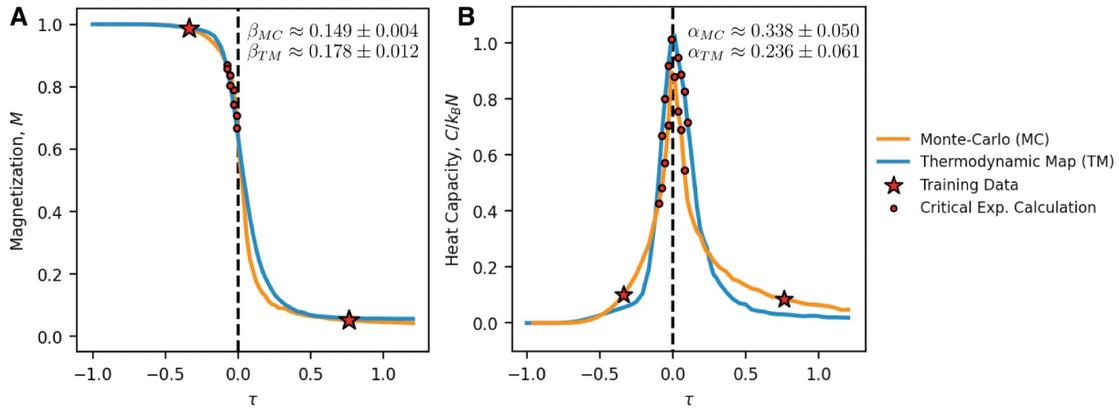


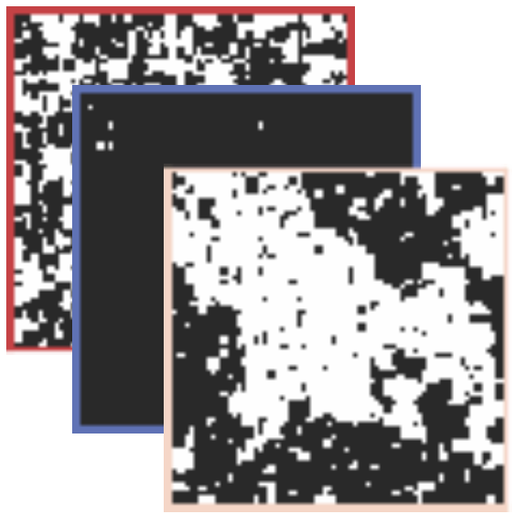
FIG. 2. **Inferring the phase transition of the 2D Ising model from limited sampling. A** The magnetization is plotted for samples of a 32×32 square Ising model generated through MC sampling (orange) and the thermodynamic map (blue). The thermodynamic map predicts change in magnetization at T_c when trained on samples generated at $T = 1.5$ and $T = 4$ (red stars). **B** The heat capacity of samples generated from MC sampling (orange) and the thermodynamic map (blue) is plotted. The thermodynamic map correctly infers the divergence in the heat capacity, numerically computed for the red dots, when trained on the same samples as panel A (red stars).

$$M \sim |\tau|^{\beta} \quad \text{and} \quad C \sim |\tau|^{-\alpha} \quad \text{where} \quad \tau = \frac{T - T_c}{T_c}$$

Herron, Mondal, Schneekloth, Tiwary arXiv:2308.14885

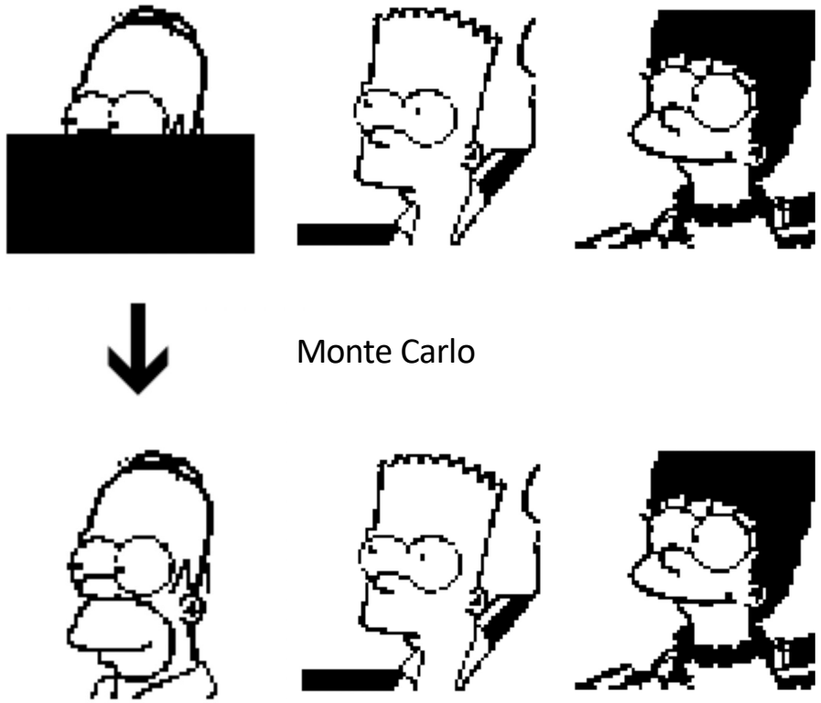
From spin glasses to RNAs

Ising Model

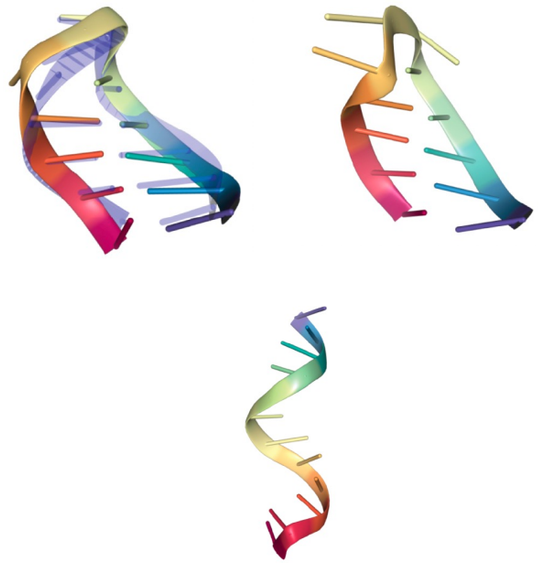


Homogeneous short-ranged J_{ij}
(2 phases)

Long range spin glasses



Biopolymers



Stable configurations (Phases) stored in
long-ranged heterogeneous J_{ij}

[Hopfield networks is all you need]

Take home so far: Thermodynamic Maps allows generating samples conditioned on some control thermodynamic parameter (so far, temperature)

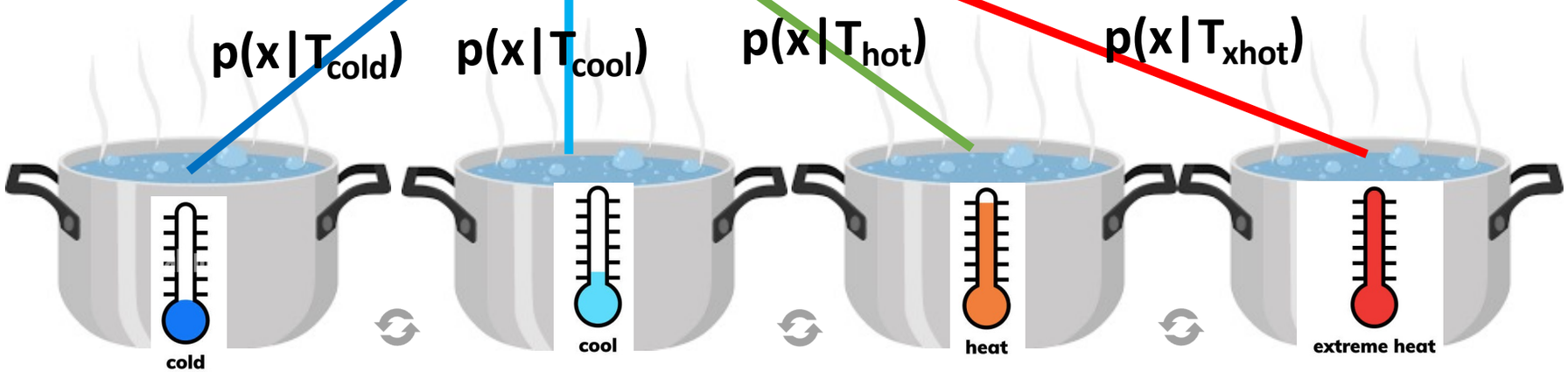
Post-processing

$p(x|T_{any})$

where T_{any} is any temperature, even < T_{cold} or > T_{xhot}

Unlike PES, FES(x) is temperature-dependent (dimensionality of x >> 1 but <<N)

$p(x|T_{cold})$ $p(x|T_{cool})$ $p(x|T_{hot})$ $p(x|T_{xhot})$



Wang, Herron, Tiwary, PNAS (2022): e2203656119

Herron…Tiwary, arXiv:2308.14885

Ensemble-weighted observable can be computed from $p(\mathbf{x})$

$$\langle A(\mathbf{x})\rangle_p = \langle A(\mathbf{x})|p(\mathbf{x})\rangle$$

This framework encompasses widely studied observables across scales:

$A(\mathbf{x}) =$	Energy $U(\mathbf{x})$	Magnetization $M(\mathbf{x})$	Radius of Gyration $R_g(\mathbf{x})$	NMR couplings $J(\mathbf{x})$

Three issues:

1. $\dim(\mathbf{x})$ is often very large, so $p(\mathbf{x})$ is computationally intractable.
2. $p(\mathbf{x})$ is really $p(\mathbf{x}\,|N,P,T)$. Can we infer the dependence of $p(\mathbf{x})$ across thermodynamic ensembles in finite size cases?
3. **Exploration of $p(\mathbf{x})$ is usually slow (i.e. MD/MC simulation)**

Thermodynamic Map-accelerated Molecular Dynamics (TM-aMD)

<u>Point 3</u>: Exploration of $p(\boldsymbol{x})$ is usually slow

Molecular Dynamics Simulation for different putative structures from Rosetta/AlphaFold2 at different

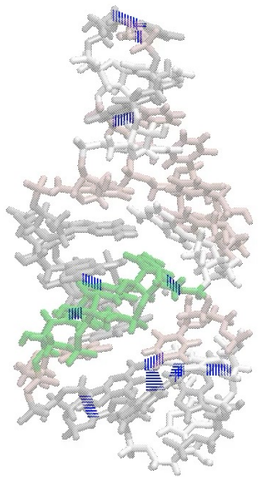
$\{\beta_i\}$

Thermodynamic Map-accelerated Molecular Dynamics (TM-aMD)

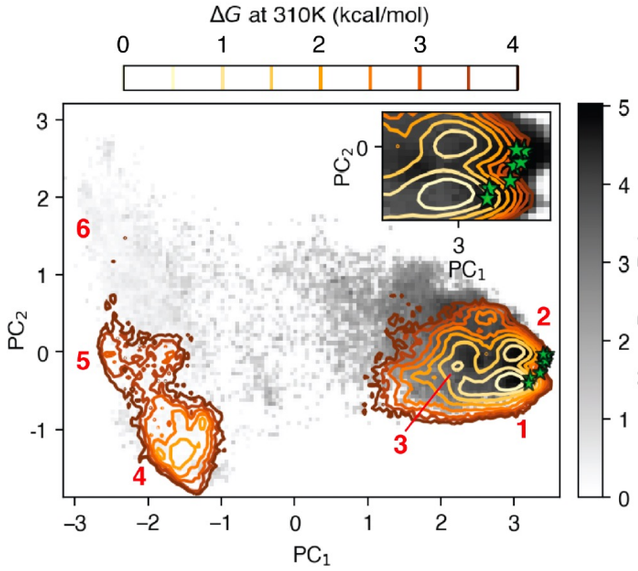
<u>Point 3</u>: Exploration of $p(\boldsymbol{x})$ is usually slow

Molecular Dynamics Simulation for different putative structures from Rosetta/AlphaFold2 at different

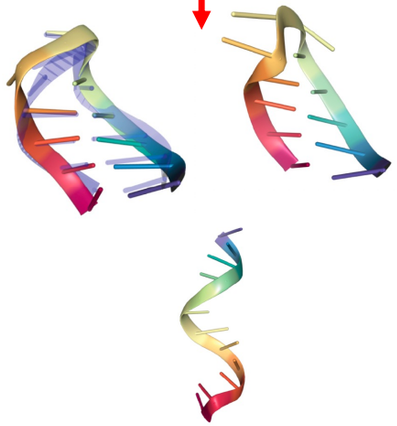
$\{\beta_i\}$



Sample new seeds for cluster-guided exploration

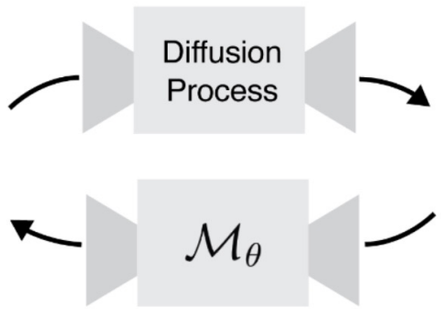


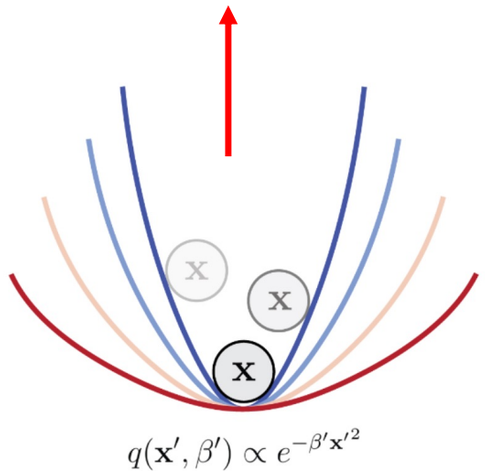
Cluster generated structures



Train Thermodynamic Map on MD structures

$\{(\mathbf{x}_i, \boldsymbol{\beta}_i)\}$

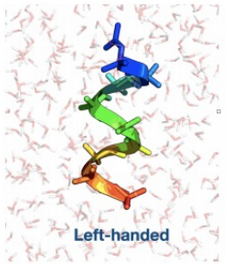




$q(\mathbf{x}', \beta') \propto e^{-\beta' \mathbf{x}'^2}$

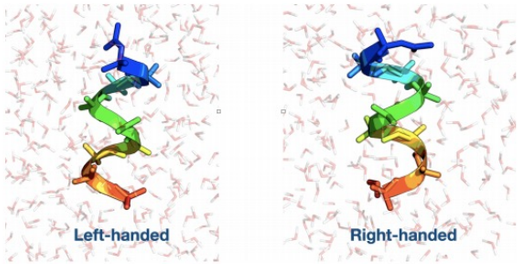
Sample from Thermodynamic Map at each T

Results for chirally symmetric peptide

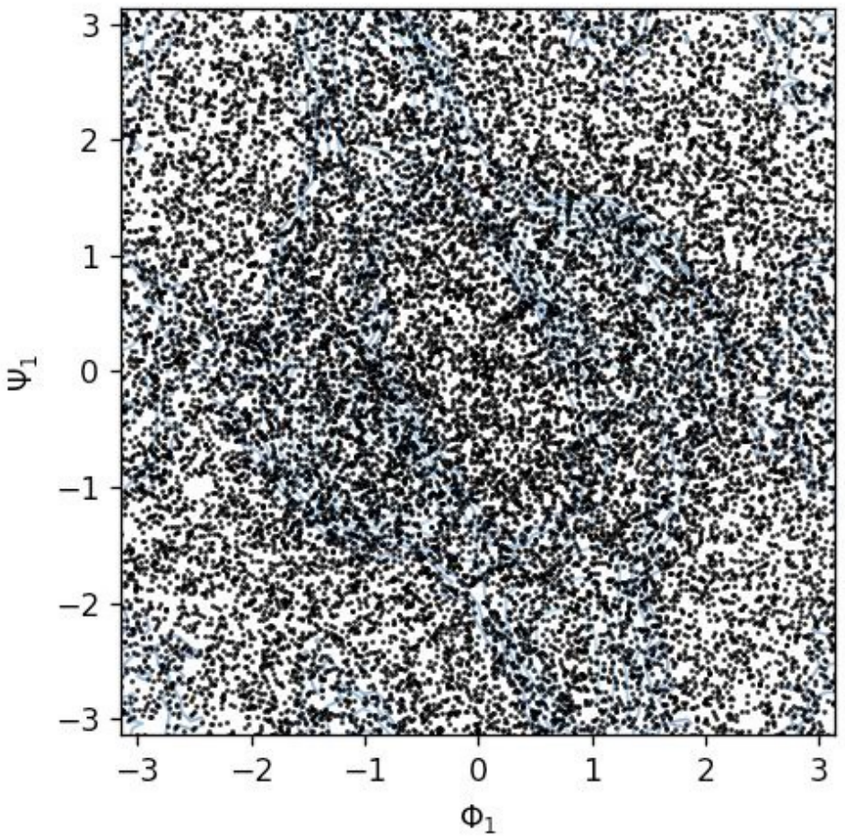


REMD performed with replicas at 400, 412,…500 K and 0% exchange (!!!)

Results for chirally symmetric peptide

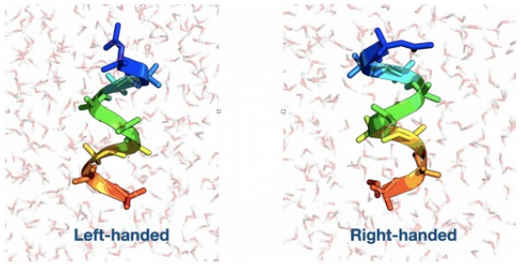


REMD performed with replicas at 400, 412,…500 K and 0% exchange (!!!)

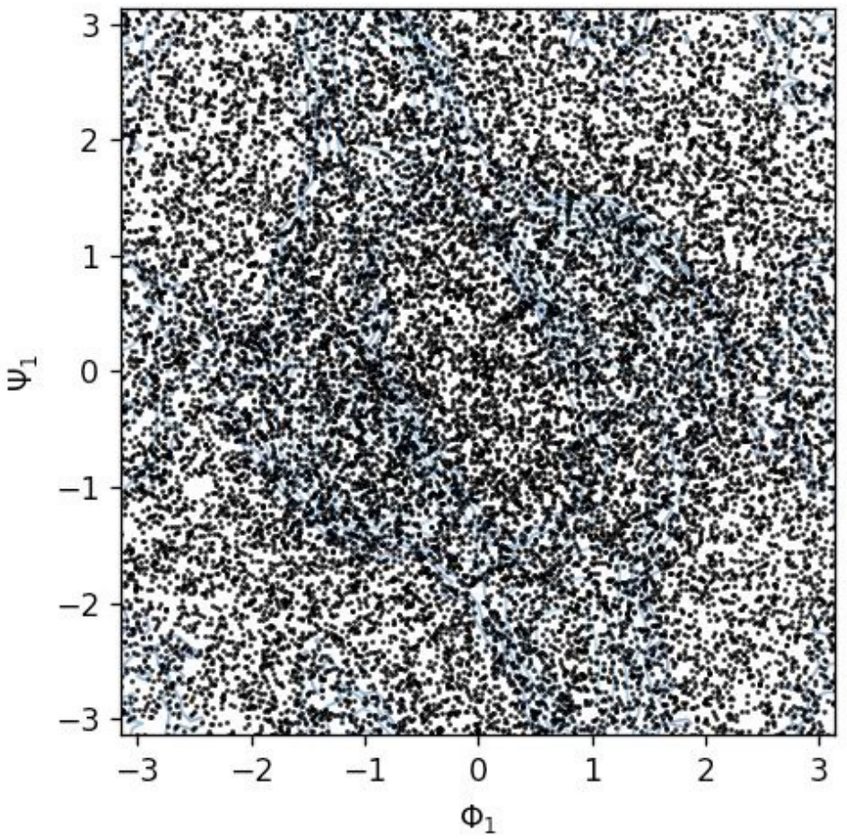


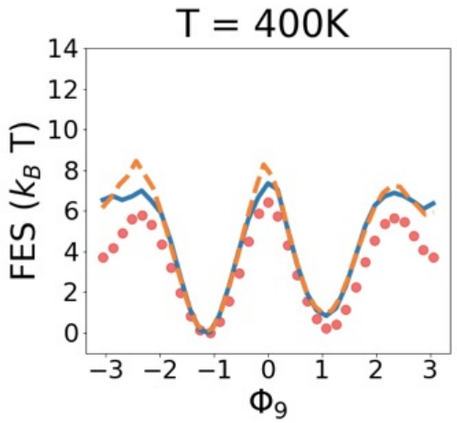
Input to TM: 18 dihedrals
Noise to data projected along 2 dihedrals

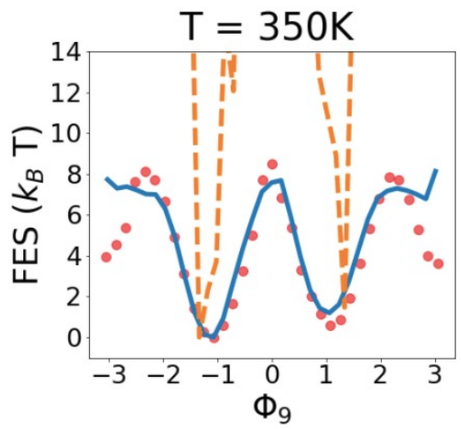
Results for chirally symmetric peptide

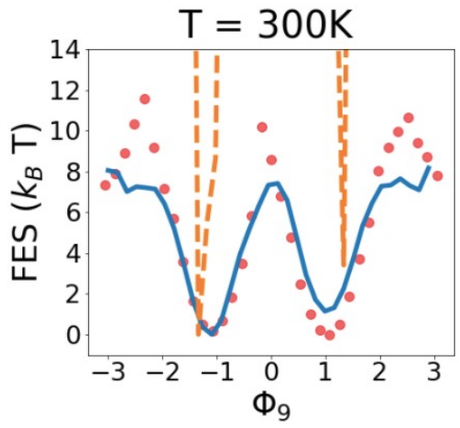


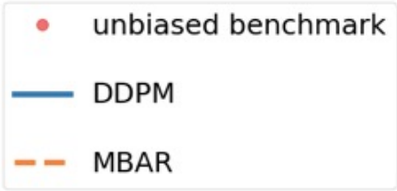
REMD performed with replicas at 400, 412,…500 K and 0% exchange (!!!)

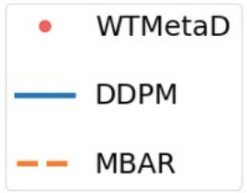








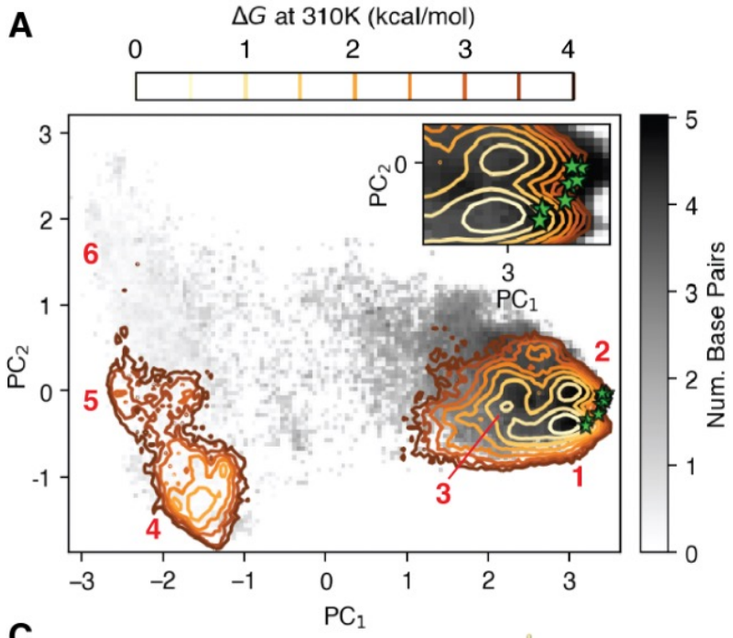




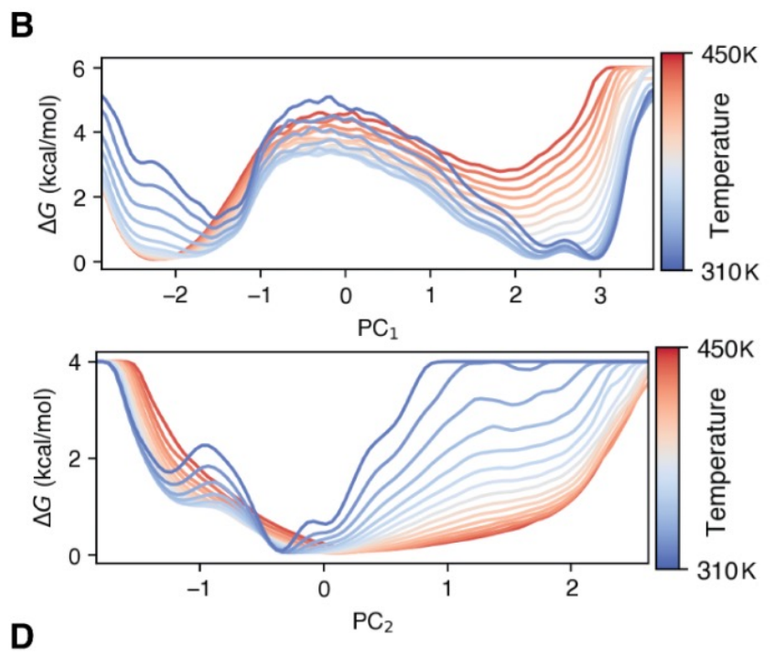
Input to TM: 18 dihedrals
Noise to data projected along 2 dihedrals

Results for GCAA tetraloop

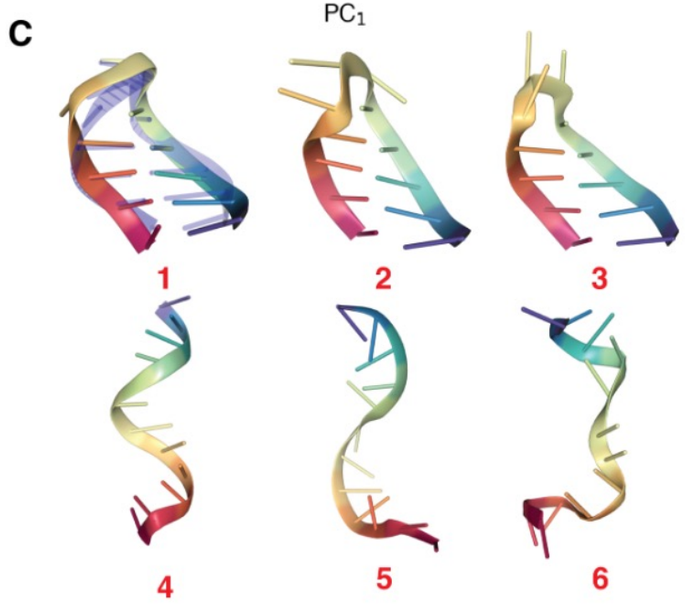
Free Energy (FE) landscape (NMR in green)



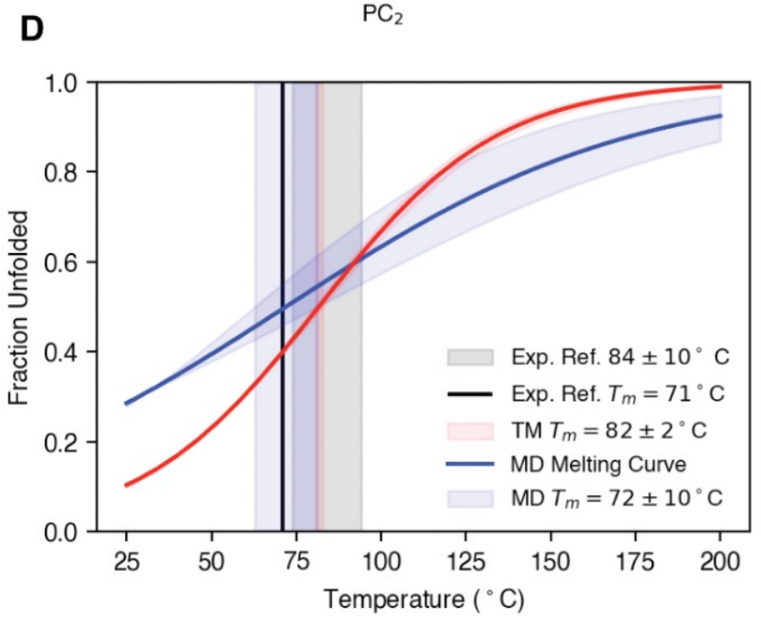
Temperature dependence of FE along PCA coordinates



Stable Conformations (NMR in blue)

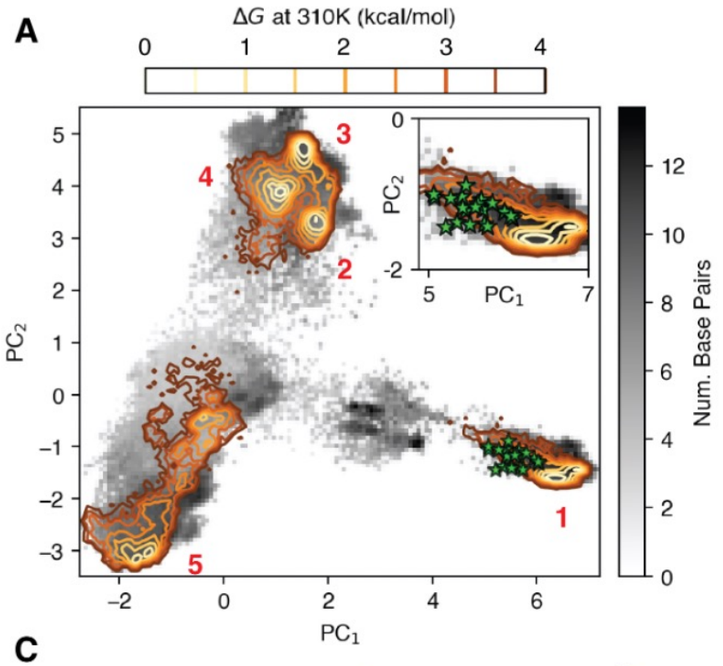


Converged melting temperature

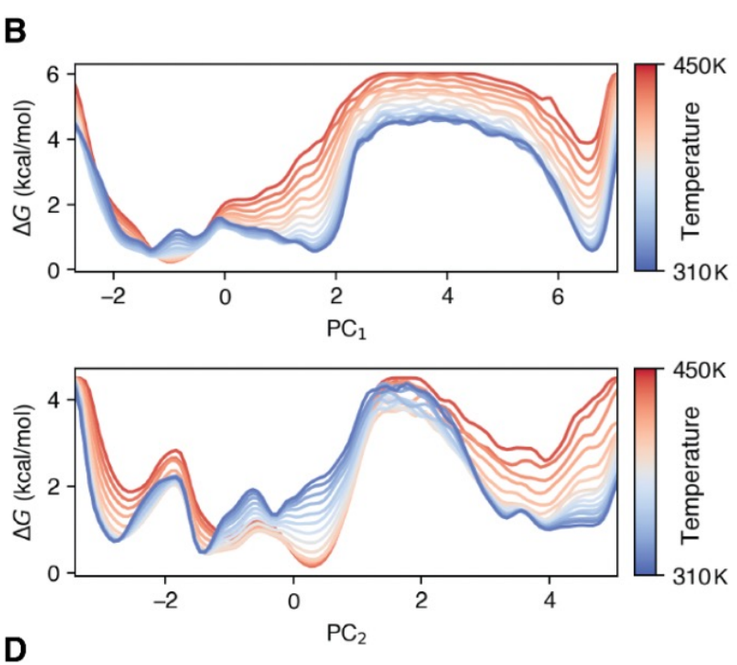


Results for HIV-TAR RNA

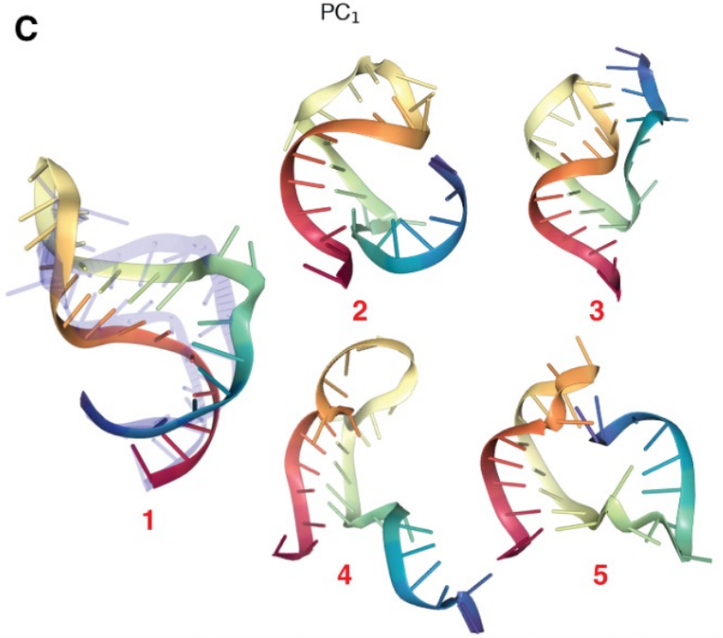
Free Energy (FE) landscape (NMR in green)



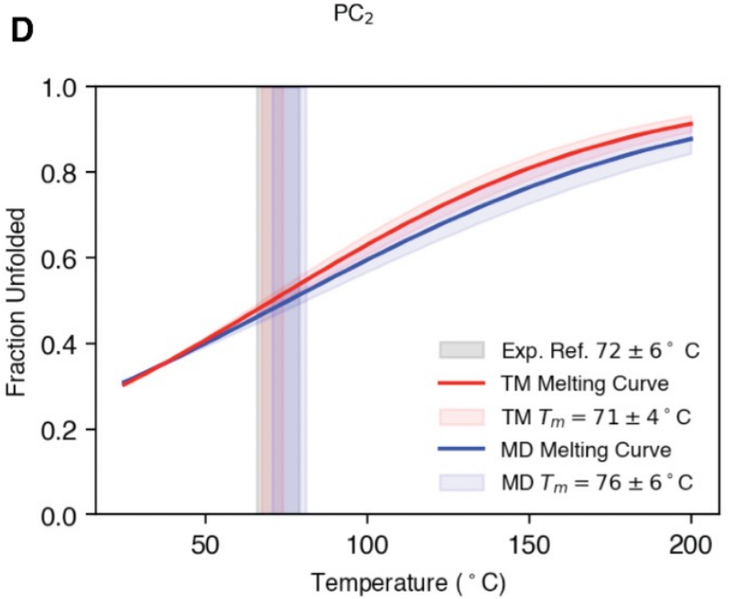
Temperature dependence of FE along PCA coordinates



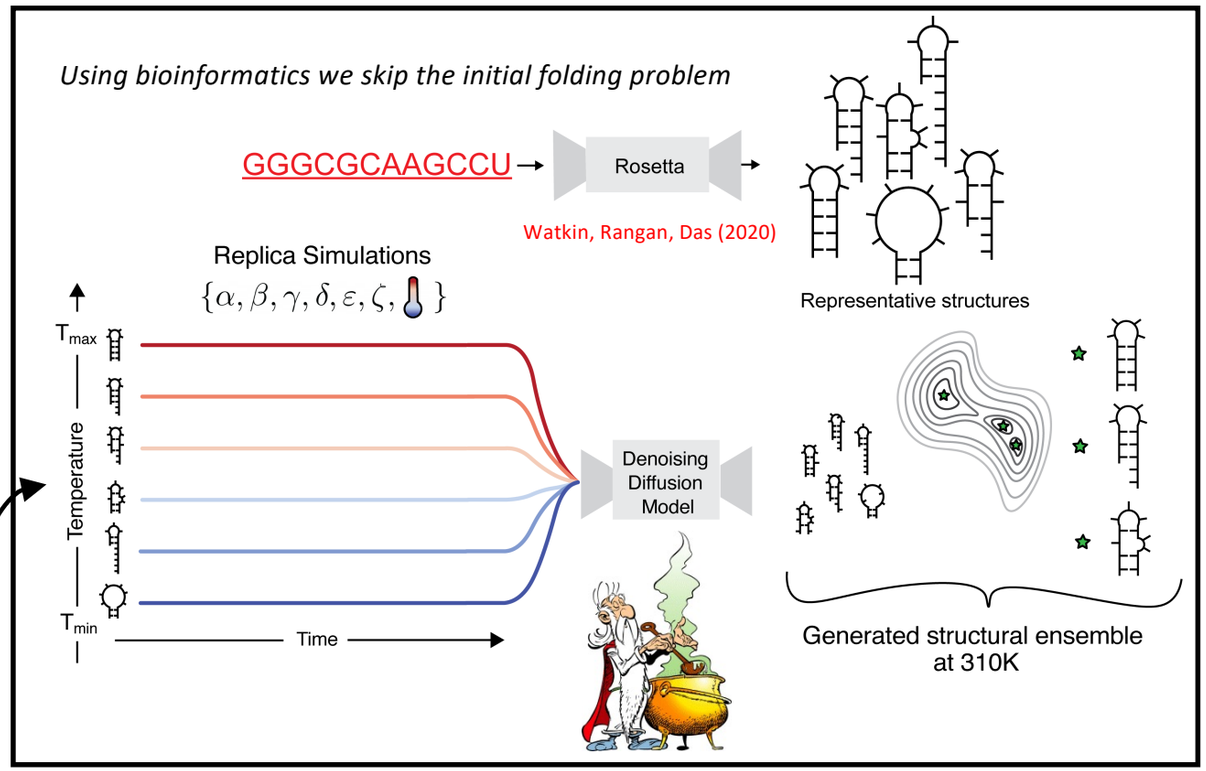
Stable Conformations (NMR in blue)



Unconverged melting temperature



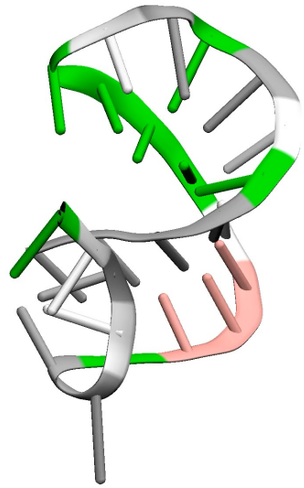
MALAT1 and Let-7f RNA
In-progress work joint with Schneekloth Lab @National Cancer Institute
RNA sequence and conformation specific drug discovery



- GROMACS 2020
- DESRES force field Tan et. al., (2018)
- NPT ensemble

- 10 replicas
- 5us total simulation time (500 ns/ replica)
- Torsion angles recorded every 0.2ps

Let-7f miRNA



- ✓ No experimental structure for free Let-7f
- ✓ No computational studies
- ✓ Oncogenic

Wang, Herron, Tiwary PNAS 2022

Herron, Mondal, Schneekloth, Tiwary arXiv

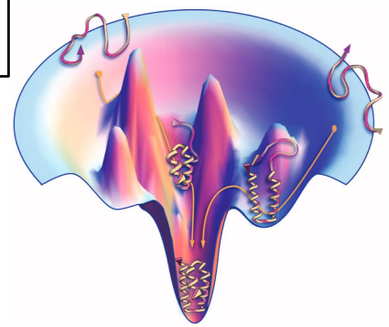
Wrapping up (1):
Proteins, RNA and crystals: Life on different energy landscapes can be sampled with different AI-stat mech integrated sampling schemes

AlphaFold2-RAVE: From Sequence to Boltzmann Ranking
Bodhi P. Vani, Akashnathan Aranganathan, Dedi Wang, and Pratyush Tiwary*
Cite This: *J. Chem. Theory Comput.* 2023, 19, 4351−4354 Read Online

nature COMMUNICATIONS
ARTICLE
https://doi.org/10.1038/s41467-019-11405-4 OPEN
Past-future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics
Yihang Wang[1], João Marcelo Lamim Ribeiro[2] & Pratyush Tiwary[2]

State predictive information bottleneck
Cite as: J. Chem. Phys. 154, 134111 (2021); doi: 10.1063/5.0038198
Submitted: 19 November 2020 • Accepted: 23 March 2021 •
Published Online: 5 April 2021
Dedi Wang[1] and Pratyush Tiwary[2,a)]

Proteins

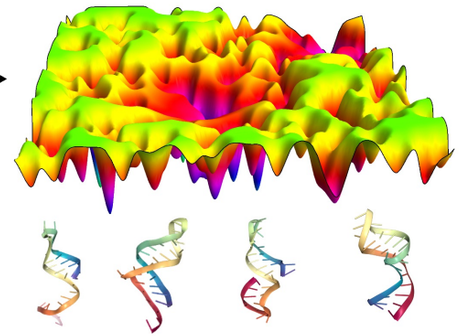


K. Dill, J. McCallum, Science (2012).
The Protein-Folding problem, 50 years on

Ordered → ∃ low-dimensional RC
∃ well-separated metastable domains

Order/Disorder spectrum

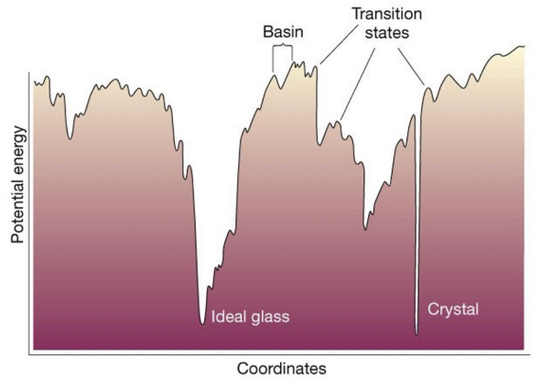
RNAs



D.J. Wales, Ann. Rev. Phys. Chem. (2017).
Exploring Energy Landscapes

Disordered → ∄ low dimensional RC
Metastability is suspect

ACS central science
http://pubs.acs.org/journal/acscii Research
Interrogating RNA−Small Molecule Interactions with Structure Probing and Artificial Intelligence-Augmented Molecular Simulations
Yihang Wang,# Shaifaly Parmar,# John S. Schneekloth,* and Pratyush Tiwary*
Cite This: *ACS Cent. Sci.* 2022, 8, 741−748 Read Online



Debenedetti, Stillinger, Nature (2001).
Supercooled Liquids & the Glass Transition

PNAS RESEARCH ARTICLE | APPLIED PHYSICAL SCIENCES
From data to noise to data for mixing physics across temperatures with generative artificial intelligence
Yihang Wang[a,b], Lukas Herron[a,b], and Pratyush Tiwary[b,c,1]
Edited by Sharon Glotzer, University of Michigan, Ann Arbor, MI; received March 2, 2022; accepted July 7, 2022

PNAS RESEARCH ARTICLE | APPLIED PHYSICAL SCIENCES
Driving and characterizing nucleation of urea and glycine polymorphs in water
Ziyue Zou[a,1], Eric R. Beyerle[b,1], Sun-Ting Tsai[c], and Pratyush Tiwary[a,b,2]
Edited by Pablo Debenedetti, Princeton University, Princeton, NJ; received September 20, 2022; accepted January 17, 2023

Inferring phase transitions and critical exponents from limited observations with Thermodynamic Maps

Lukas Herron and Kinjal Mondal
Biophysics Program and Institute for Physical Science and Technology, University of Maryland, College Park, MD, 20742, USA

John S. Schneekloth Jr.
Chemical Biology Laboratory, National Cancer Institute, Frederick, MD 21702, USA

Pratyush Tiwary*
Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD, 20742, USA

github.com/tiwarylab

Wrapping up (2): Diffusion Models, with origins in Zwanzig and Jarzynski are arguably the most extrapolative Generative AI models currently available

Deep unsupervised learning using nonequilibrium thermodynamics

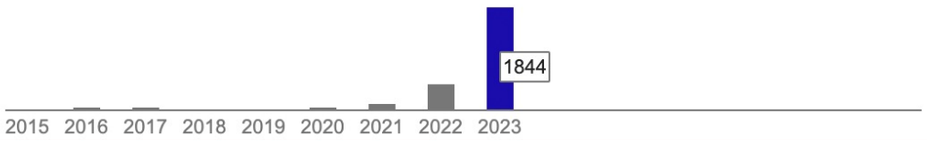
Authors: Jascha Sohl-Dickstein, Eric A Weiss, Niru Maheswaranathan, Surya Ganguli

Publication date: 2015/3/12

Journal: International Conference on Machine Learning

Description: A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable. Here, we develop an approach that simultaneously achieves both flexibility and tractability. The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data. This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model. We additionally release an open source reference implementation of the algorithm.

Total citations: Cited by 2502



GENERALIZATION IN DIFFUSION MODELS ARISES FROM GEOMETRY-ADAPTIVE HARMONIC REPRESENTATION

Zahra Kadkhodaie
Ctr. for Data Science, New York University
zk388@nyu.edu

Florentin Guth
Ctr. for Data Science, New York University
Flatiron Institute, Simons Foundation
florentin.guth@nyu.edu

Eero P. Simoncelli
New York University
Flatiron Institute, Simons Foundation
eero.simoncelli@nyu.edu

Stéphane Mallat
Collège de France
Flatiron Institute, Simons Foundation
stephane.mallat@ens.fr

ABSTRACT

High-quality samples generated with score-based reverse diffusion algorithms provide evidence that deep neural networks (DNN) trained for denoising can learn high-dimensional densities, despite the curse of dimensionality. However, recent reports of memorization of the training set raise the question of whether these networks are learning the "true" continuous density of the data. Here, we show that two denoising DNNs trained on non-overlapping subsets of a dataset learn nearly the same score function, and thus the same density, with a surprisingly small number of training images. This strong generalization demonstrates an alignment of powerful inductive biases in the DNN architecture and/or training algorithm with properties of the data distribution. We analyze these, demonstrating that the denoiser performs a shrinkage operation in a basis adapted to the underlying image. Examination of these bases reveals oscillating harmonic structures along contours and in homogeneous image regions. We show that trained denoisers are inductively biased towards these geometry-adaptive harmonic representations by demonstrating that they arise even when the network is trained on image classes such as low-dimensional manifolds, for which the harmonic basis is suboptimal. Additionally, we show that the denoising performance of the networks is near-optimal when trained on regular image classes for which the optimal basis is known to be geometry-adaptive and harmonic.

Stat Mech and Theoretical Chemistry have a lot to teach AI for practical, data-sparse applications to natural sciences

References

All about Monte Carlo

Simulated Tempering

Annealed Importance Sampling

Score-Based Generative modeling with Stochastic Differential Equations

Thermodynamic Maps