Enhanced sampling with auxiliary models: from coarse-graining to rare events

Grant M. Rotskoff

with Shriram Chennakesavalu and David Toomer

University of Maryland

2 March 2023

arxiv:2205.01205 + GitHub

Stanford University

https://statmech.stanford.edu

State of the *art*

- *LLMs:*
  - Data acquisition: entire internet
  - Training costs: ~1m GPU hours
  - Achievement: *Seinfeld Forever*

State of the *art*

an image of coarse grain

- *LLMs:*
  - Data acquisition: entire internet
  - Training costs: ~1m GPU hours
  - Achievement: *Seinfeld Forever*
- *Computer Vision:*
  - Data acquisition: 3-5 billion images
  - Training costs: ~150k GPU hours
  - Achievement: *See lefthand side*

State of the *art*

an image of coarse grain

- *LLMs:*
  - Data acquisition: entire internet
  - Training costs: ~1m GPU hours
  - Achievement: *Seinfeld Forever*
- *Computer Vision:*
  - Data acquisition: 3-5 billion images
  - Training costs: ~150k GPU hours
  - Achievement: *See lefthand side*
- *AlphaFold:*
  - Data acquisition: 50 years of beam time
  - Training costs: ? ("about a week" + fine-tuning)
  - Achievement: *Real scientific progress*

Acknowledgements

Shriram Chennakesavalu
Graduate Student
Machine Learning, Nonequilibrium Control

Grant M. Rotskoff
Assistant Professor of Chemistry
Nonequilibrium Dynamics, Biophysics, Machine Learning, theory and practice

Andy Mitchell
Graduate Student
Driven Sampling, Transition States and Committors, Machine Learning

Clay Batton
Postdoctoral Researcher
Coarse Graining, Nonequilibrium Control

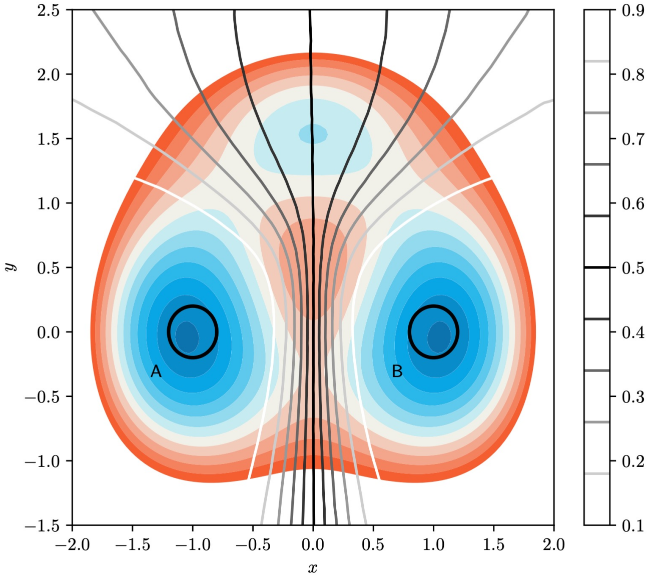
David Toomer
Undergraduate Researcher
Machine Learning

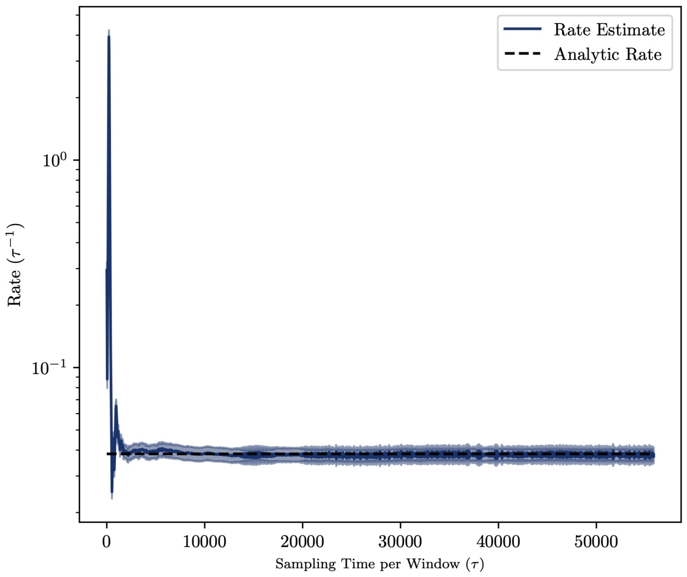
Emmit Pert
Graduate Student
Molecular Dynamics, Importance Sampling

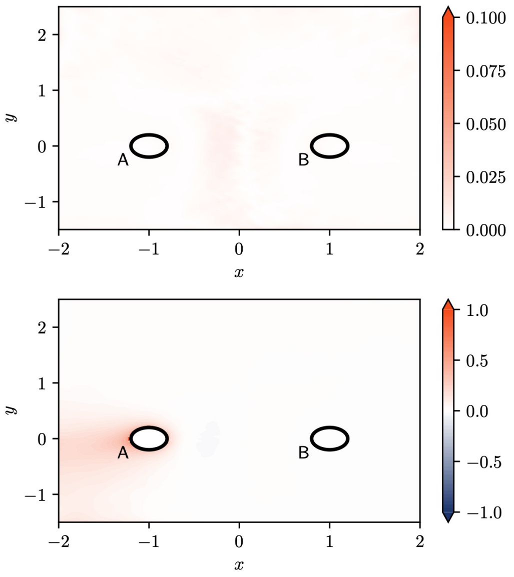
Isaac Applebaum
Undergraduate Researcher
Machine Learning, CARTs, (joint with Waymouth Group)

Sherry Li
Graduate Student
Machine Learning, Enhanced Sampling Methods

Opportunities afforded by high-d learning



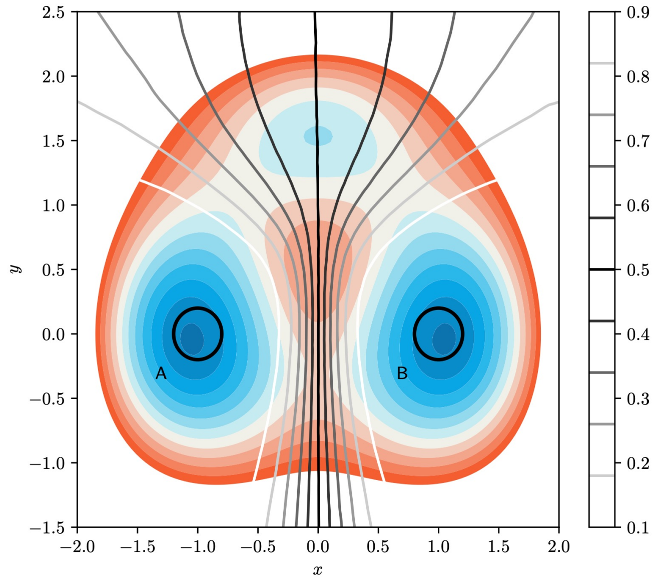


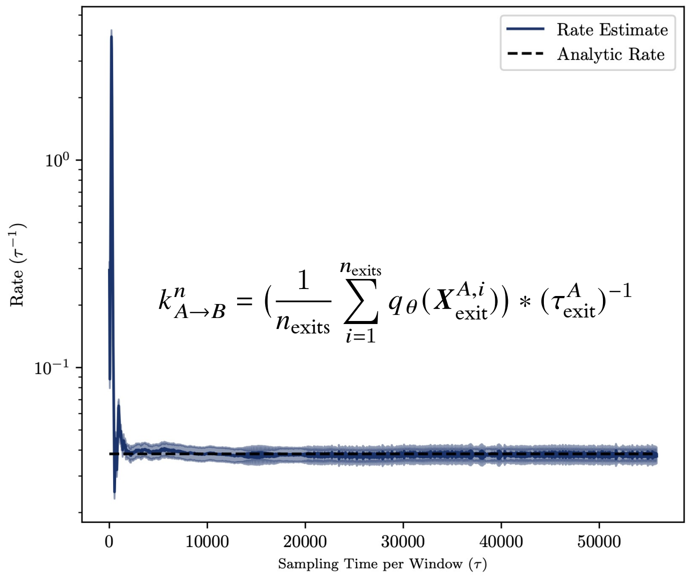


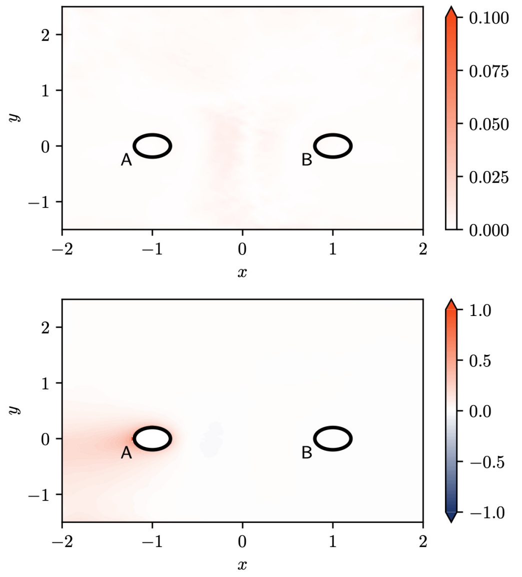
$$\mathcal{P}_t q = \mathbb{E}_{\boldsymbol{X}_t}[q(\boldsymbol{X}_t)|\boldsymbol{X}_0 = \boldsymbol{x}] \longrightarrow q_\star = \min_\theta \left(\frac{1}{2}\mathbb{E}_{\boldsymbol{X}_t;\boldsymbol{X}_0\sim\rho_s}\left(q_\theta(\boldsymbol{X}_0) - q_\theta(\boldsymbol{X}_t)\right)^2 \right)$$

GMR, Mitchell, Vanden-Eijnden PMLR 145:757-780, 2022. **Mitchell, GMR,** *In preparation.*

Opportunities afforded by high-*d* learning







$$\mathcal{P}_t q = \mathbb{E}_{\boldsymbol{X}_t}[q(\boldsymbol{X}_t)|\boldsymbol{X}_0 = \boldsymbol{x}] \xrightarrow{\textit{Variational problem}} q_\star = \min_\theta \left(\frac{1}{2}\mathbb{E}_{\boldsymbol{X}_t;\boldsymbol{X}_0\sim\rho_s}\left(q_\theta(\boldsymbol{X}_0) - q_\theta(\boldsymbol{X}_t)\right)^2\right)$$

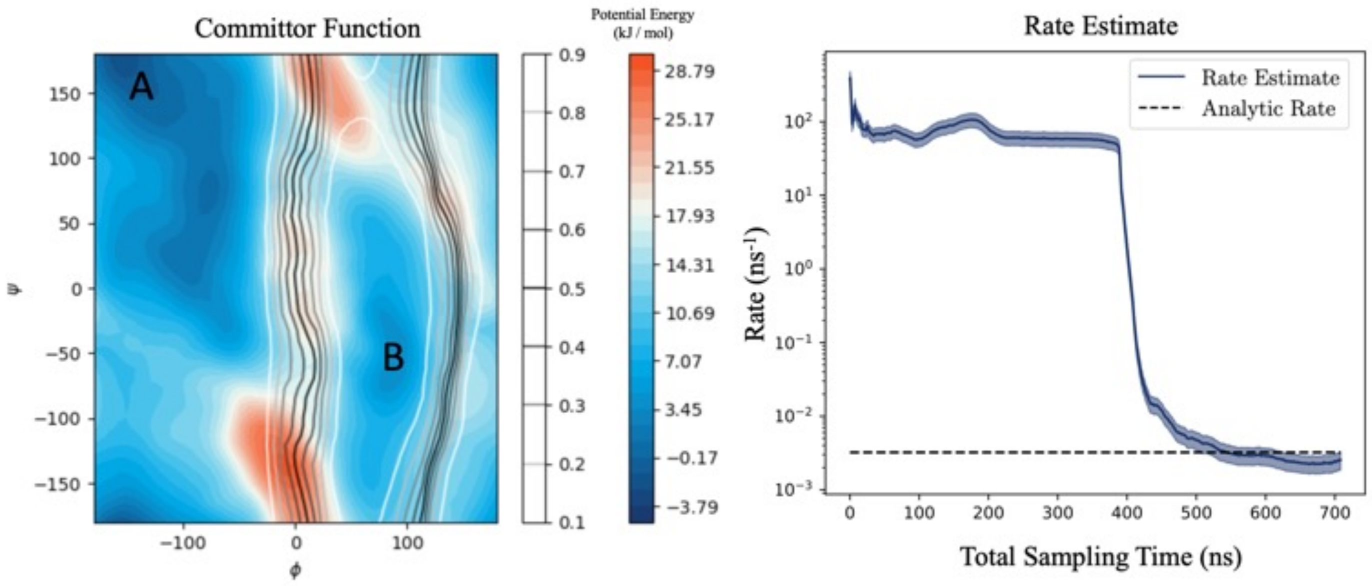
- High-dimensional committor
- On the fly data acquisition
- *NO* collective variables
- Accurate rates (with bias ☹)

Adaptive Importance Sampling

$$\boldsymbol{X}_i^n \sim e^{-\beta_{\text{sampling}}(U(\boldsymbol{X}_i^n) + \frac{k}{2}(q_\theta(\boldsymbol{X}_i^n) - q_i^{\text{target}})^2)}$$

Opportunities afforded by high-*d* learning

$$q : \mathbb{R}^{66} \to \mathbb{R}$$



$$\mathcal{P}_t q = \mathbb{E}_{\boldsymbol{X}_t}[q(\boldsymbol{X}_t) | \boldsymbol{X}_0 = \boldsymbol{x}] \longrightarrow q_\star = \min_\theta \left(\frac{1}{2} \mathbb{E}_{\boldsymbol{X}_t; \boldsymbol{X}_0 \sim \rho_s} \left(q_\theta(\boldsymbol{X}_0) - q_\theta(\boldsymbol{X}_t) \right)^2 \right)$$

- High-dimensional committor
- On the fly data acquistion
- *NO* collective variables
- Accurate rates (with bias ☹)

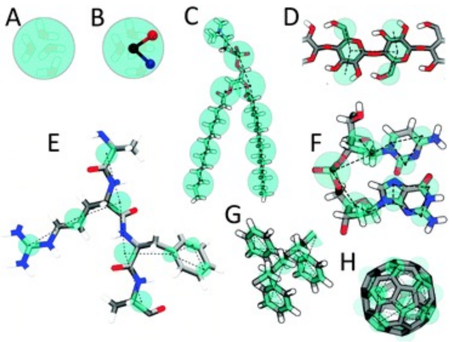
Cf. Strahan, John, Justin Finkel, Aaron R. Dinner, and Jonathan Weare. "Forecasting Using Neural Networks and Short-Trajectory Data." arXiv, August 2, 2022. http://arxiv.org/abs/2208.01717.

GMR, Mitchell, Vanden-Eijnden PMLR 145:757-780, 2022. **Mitchell, GMR,** *In preparation.*

Embeddings / nonlinear dimensionality reduction / ansatzë

Limitations of coarse-graining in biomolecular systems

Representation

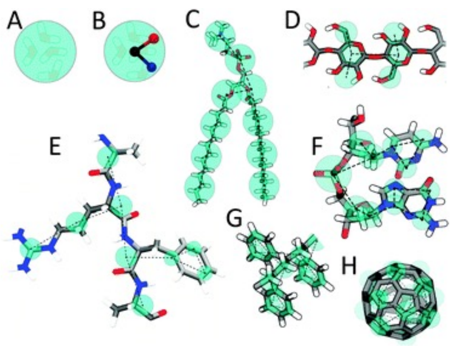


Marrink and Tieleman *Chem. Soc. Rev.*, 2013, **42**, 6801-6822

Linear projections
Independent of fine-grained state
Empirical potential (or delta ML)

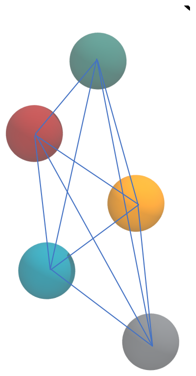
Limitations of coarse-graining in biomolecular systems

Representation



Marrink and Tieleman *Chem. Soc. Rev.*, 2013, **42**, 6801-6822

Interpretation

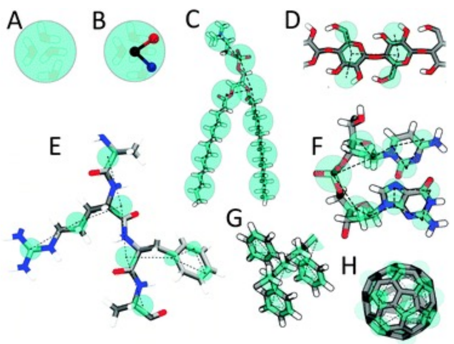


Linear projections
Independent of fine-grained state
Empirical potential (or delta ML)

No access to fine-grained state
Imperfect recovery
Dynamics difficult to map

Limitations of coarse-graining in biomolecular systems

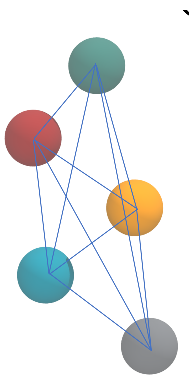
Representation



Marrink and Tieleman *Chem. Soc. Rev.*, 2013, **42**, 6801-6822

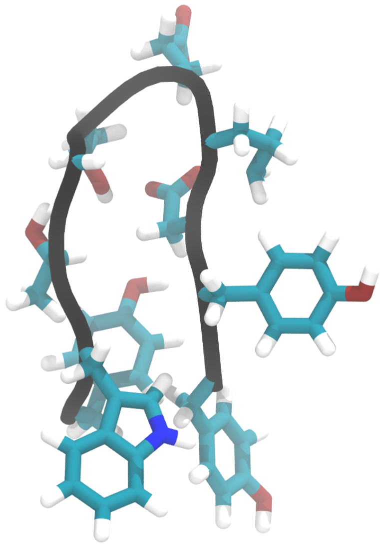
Linear projections
Independent of fine-grained state
Empirical potential (delta ML)

Interpretation



No access to fine-grained state
Imperfect recovery
Dynamics difficult to map

Transferability



Quality of generalization?
Need relevant rare configurations
Limited opportunities for feedback

Long list of efforts to address these issues

Representation

- M. Stieffenhofer, M. Wand, and T. Bereau, Mach. Learn.: Sci. Technol. **1**, 045014 (2020).
- A. E. P. Durumeric and G. A. Voth, J. Chem. Phys. **151**, 124110 (2019).
- M. Giulini, M. Rigoli, G. Mattiotti, R. Menichetti, T. Tarenzi, R. Fiorentini, and R. Potestio, Front. Mol. Biosci. **8**, 676976 (2021).
- G. Sivaraman and N. E. Jackson, J. Chem. Theory Comput. **18**, 1129 (2022).
- E. Pretti and M. S. Shell, J. Chem. Phys. **155**, 094102 (2021).
- J. C. Maier and N. E. Jackson, J. Chem. Phys. **157**, 174102 (2022).
- M. Giulini, R. Menichetti, M. S. Shell, and R. Potestio, J. Chem. Theory Comput. **16**, 6795 (2020).
- M. Chakraborty, C. Xu, and A. D. White, The Journal of Chemical Physics **149**, 134106 (2018).

Interpretation

- J. W. Wagner, J. F. Dama, A. E. P. Durumeric, and G. A. Voth, The Journal of Chemical Physics **145**, 044108 (2016).
- K. M. Kidder, R. J. Szukalo, and W. G. Noid, Eur. Phys. J. B **94**, 153 (2021).
- K. M. Lebold and W. G. Noid, J. Chem. Phys. **151**, 164113 (2019).
- Z. Zhang, L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. A. Voth, Biophysical Journal **95**, 5073 (2008).
- J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth, J. Chem. Theory Comput. **9**, 2466 (2013).
- C. Clementi, Current Opinion in Structural Biology **18**, 10 (2008).

Transferability

- M. Stieffenhofer, M. Wand, and T. Bereau, Mach. Learn.: Sci. Technol. **1**, 045014 (2020).
- A. E. P. Durumeric and G. A. Voth, J. Chem. Phys. **151**, 124110 (2019).
- E. Pretti and M. S. Shell, J. Chem. Phys. **155**, 094102 K.
- M. Schöberl, N. Zabaras, and P.-S. Koutsourelakis, J. Chem. Phys. **150**, 024109 (2019).
- T. T. Foley, K. M. Kidder, M. S. Shell, and W. G. Noid, Proc. Natl. Acad. Sci. U.S.A. **117**, 24061 (2020).
- Z. Zhang, L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. A. Voth, Biophysical Journal **95**, 5073 (2008).
- W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).
- H. Kenzaki, N. Koga, N. Hori, R. Kanada, W. Li, K. Okazaki, X.-Q. Yao, and S. Takada, J. Chem. Theory Comput. **7**, 1979 (2011).

Long list of efforts to address these issues

Representation

- M. Stieffenhofer, M. Wand, and T. Bereau, Mach. Learn.: Sci. Technol. **1**, 045014 (2020).
- A. E. P. Durumeric and G. A. Voth, J. Chem. Phys. **151**, 124110 (2019).
- M. Giulini, M. Rigoli, G. Mattiotti, R. Menichetti, T. Tarenzi, R. Fiorentini, and R. Potestio, Front. Mol. Biosci. **8**, 676976 (2021).
- G. Sivaraman and N. E. Jackson, J. Chem. Theory Comput. **18**, 1129 (2022).
- E. Pretti and M. S. Shell, J. Chem. Phys. **155**, 094102 (2021).
- J. C. Maier and N. E. Jackson, J. Chem. Phys. **157**, 174102 (2022).
- M. Giulini, R. Menichetti, M. S. Shell, and R. Potestio, J. Chem. Theory Comput. **16**, 6795 (2020).
- M. Chakraborty, C. Xu, and A. D. White, The Journal of Chemical Physics **149**, 134106 (2018).

Interpretation

- J. W. Wagner, J. F. Dama, A. E. P. Durumeric, and G. A. Voth, The Journal of Chemical Physics **145**, 044108 (2016).
- K. M. Kidder, R. J. Szukalo, and W. G. Noid, Eur. Phys. J. B **94**, 153 (2021).
- K. M. Lebold and W. G. Noid, J. Chem. Phys. **151**, 164113 (2019).
- Z. Zhang, L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. A. Voth, Biophysical Journal **95**, 5073 (2008).
- J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth, J. Chem. Theory Comput. **9**, 2466 (2013).
- C. Clementi, Current Opinion in Structural Biology **18**, 10 (2008).

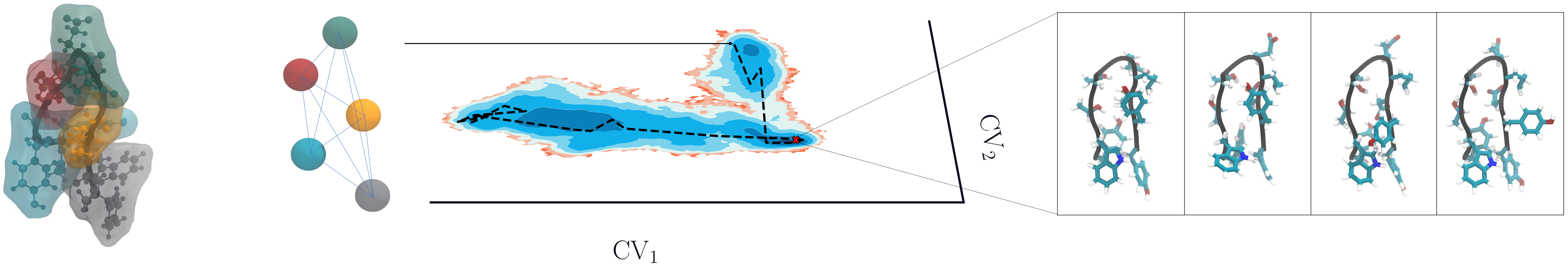
Transferability

- M. Stieffenhofer, M. Wand, and T. Bereau, Mach. Learn.: Sci. Technol. **1**, 045014 (2020).
- A. E. P. Durumeric and G. A. Voth, J. Chem. Phys. **151**, 124110 (2019).
- E. Pretti and M. S. Shell, J. Chem. Phys. **155**, 094102 K.
- M. Schöberl, N. Zabaras, and P.-S. Koutsourelakis, J. Chem. Phys. **150**, 024109 (2019).
- T. T. Foley, K. M. Kidder, M. S. Shell, and W. G. Noid, Proc. Natl. Acad. Sci. U.S.A. **117**, 24061 (2020).
- Z. Zhang, L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. A. Voth, Biophysical Journal **95**, 5073 (2008).
- W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).
- H. Kenzaki, N. Koga, N. Hori, R. Kanada, W. Li, K. Okazaki, X.-Q. Yao, and S. Takada, J. Chem. Theory Comput. **7**, 1979 (2011).

Few integrated strategies… few statistical guarantees

Closing the loop on coarse-grained modeling

$\boldsymbol{x}_0$ $\quad\quad$ $\boldsymbol{z}_0 = \mathcal{M}(\boldsymbol{x}_0)$ $\quad\quad$ $\boldsymbol{z} \sim e^{-\beta \hat{U}(\boldsymbol{z},\boldsymbol{\theta})}$ $\quad\quad$ $\boldsymbol{x}_i \sim p_{\text{inv}}(\boldsymbol{x}|\boldsymbol{z}_i)$



1. Fine-Grained Configuration

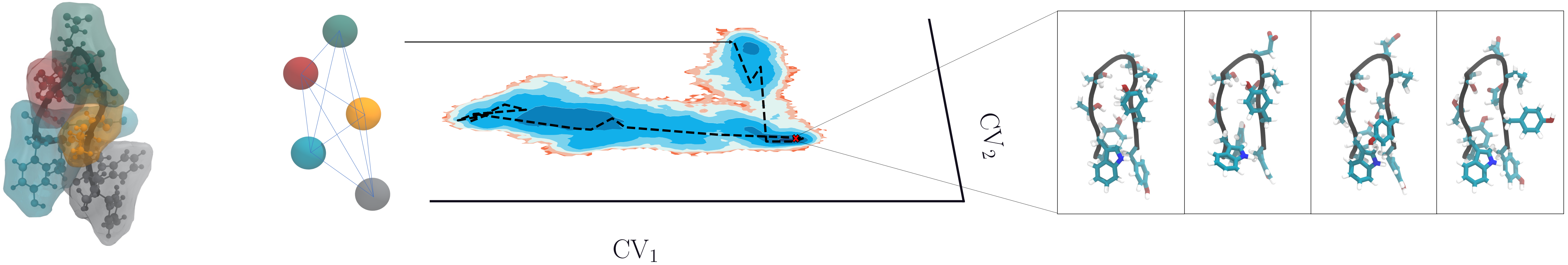
2. Coarse-Grained Dynamics

3. Generated Fine-Grained Configurations

Closing the loop on coarse-grained modeling

Learned *Learned* *Learned*

$\boldsymbol{x}_0$ $\boldsymbol{z}_0 = \mathcal{M}(\boldsymbol{x}_0)$ $\boldsymbol{z} \sim e^{-\beta \hat{U}(\boldsymbol{z},\boldsymbol{\theta})}$ $\boldsymbol{x}_i \sim p_{\text{inv}}(\boldsymbol{x}|\boldsymbol{z}_i)$



1. Fine-Grained Configuration

2. Coarse-Grained Dynamics

3. Generated Fine-Grained Configurations

Closing the loop on coarse-grained modeling

Graph autoencoder

Force-matching GNN

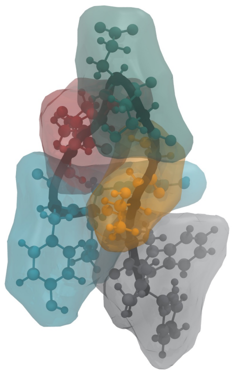
Conditional Normalizing Flow

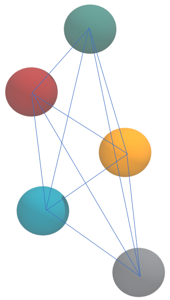
$\boldsymbol{x}_0$

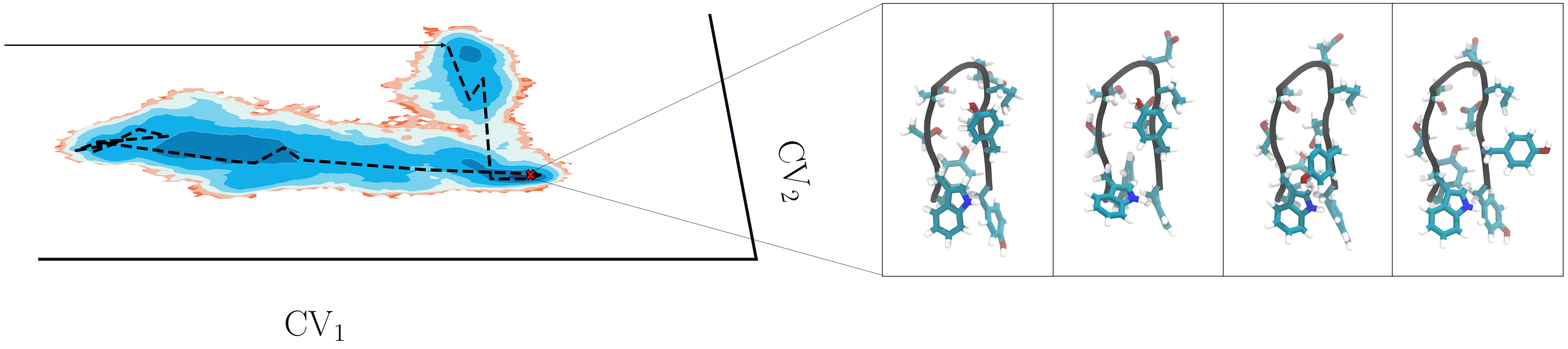
$\boldsymbol{z}_0 = \mathcal{M}(\boldsymbol{x}_0)$

$\boldsymbol{z} \sim e^{-\beta \hat{U}(\boldsymbol{z}, \boldsymbol{\theta})}$

$\boldsymbol{x}_i \sim p_{\text{inv}}(\boldsymbol{x}|\boldsymbol{z}_i)$







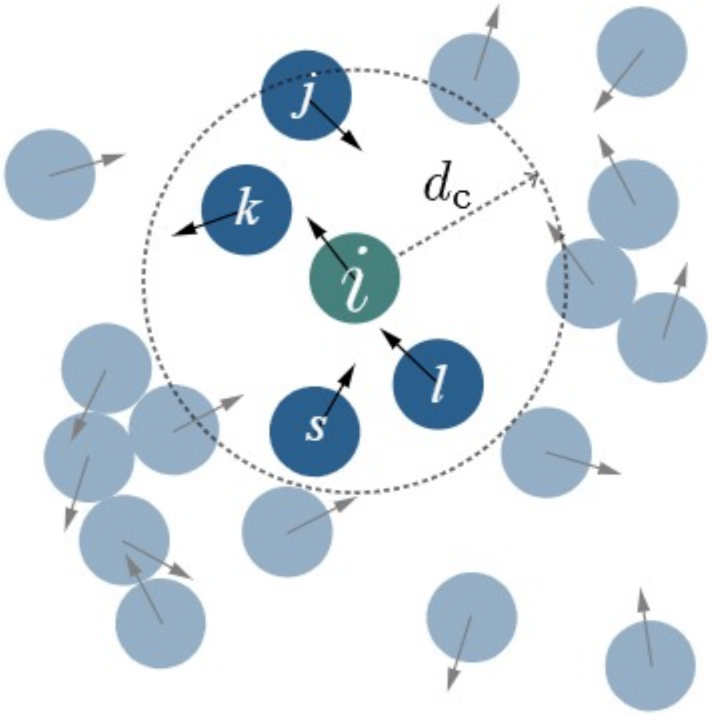
1. Fine-Grained Configuration

2. Coarse-Grained Dynamics

3. Generated Fine-Grained Configurations

Specific models not important for framework… pick your poison

Learning $\mathcal{M}$ --- physical inductive bias



Translation invariance

$$m_{ij} = f_e(\|\boldsymbol{x}_j - \boldsymbol{x}_i\|) \cdot \phi(\|\boldsymbol{x}_j - \boldsymbol{x}_i\|)$$

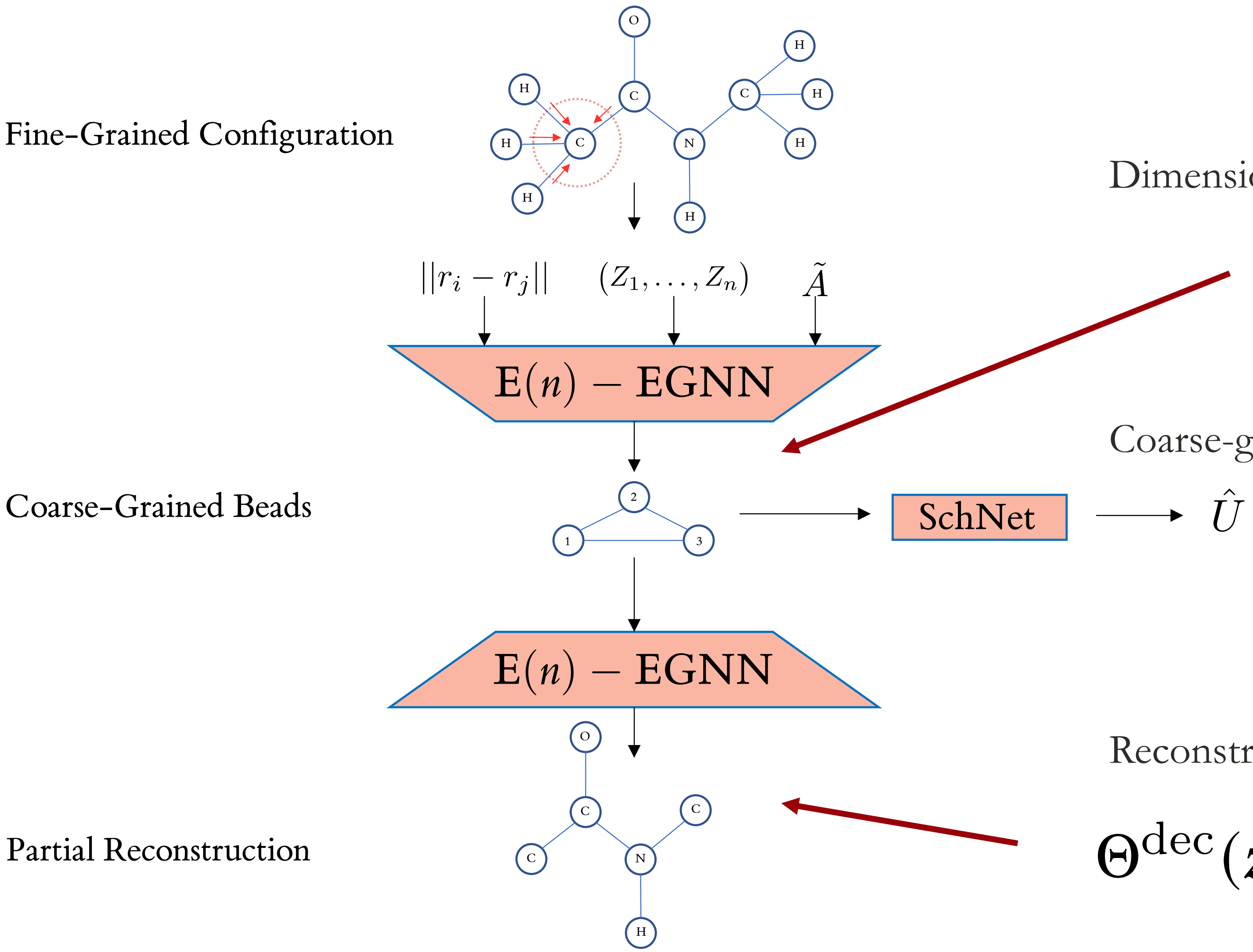
Permutation invariance

$$|\boldsymbol{u}_i| = f_g(\sum_j m_{ij})$$

Rotation equivariance

$$\frac{\boldsymbol{u}_i}{|\boldsymbol{u}_i|} = \sum_j R\left[f_\theta(\|\boldsymbol{x}_j - \boldsymbol{x}_i\|)\right] \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|}$$

Learning $\mathcal{M}$ --- building embeddings



Dimensionality reduction

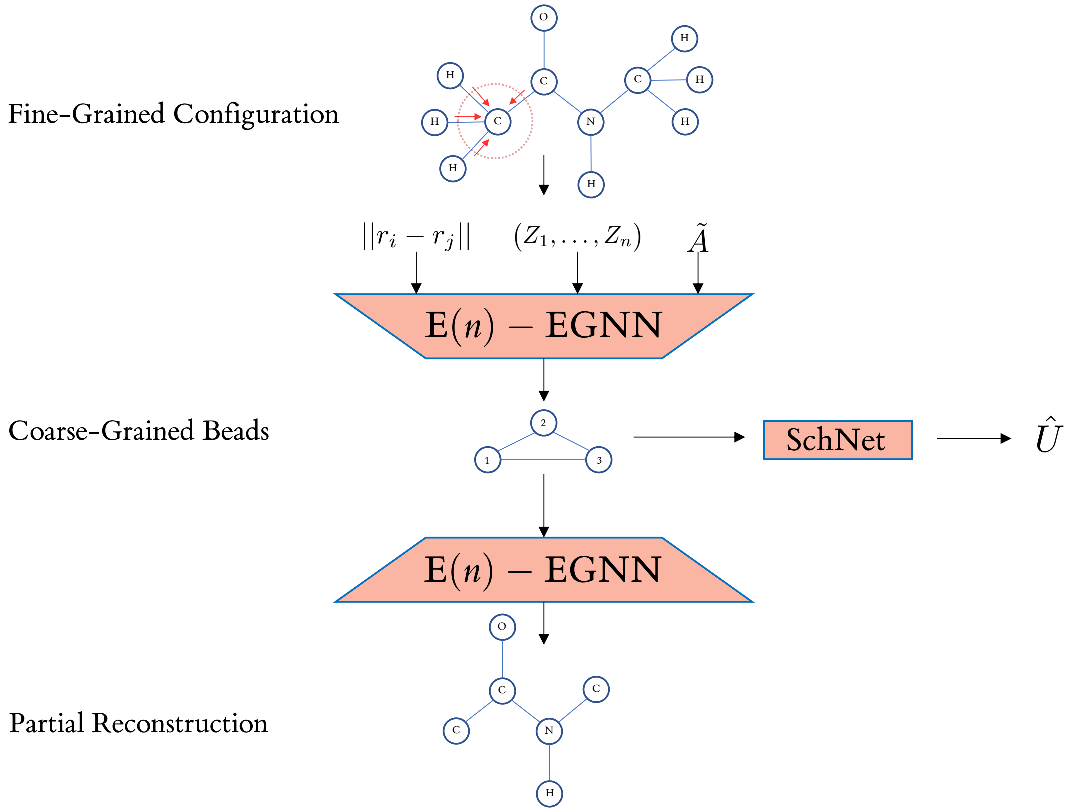
$$P_{\boldsymbol{x}} = \Theta(\boldsymbol{x})$$

Coarse-grained potential energy

Reconstruction map

$$\Theta^{\mathrm{dec}}(z_i) = P_{z_i} z_i$$

Learning $\mathcal{M}$ --- building embeddings



Optimizing the map:

$$\mathcal{L}_\Theta = \mathcal{L}_\text{r} + \lambda\left[\mathcal{L}_\text{link} + \mathcal{L}_\text{ent} + \mathcal{L}_\text{assgn} + \lambda_\text{mf}\mathcal{L}_\text{mf}\right]$$

Locality

Force-matching

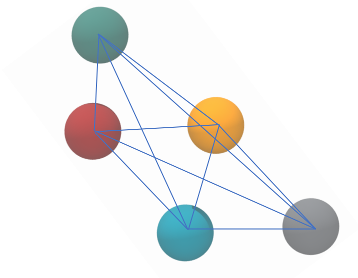
Optimizing the energy:

$$\mathcal{L}_{\hat{U}} = \left[\nabla_z \hat{U}(z) - F_\text{inst}(z)\right]^2$$

"Force-matching"

Interpreting $\mathcal{M}$ --- necessary sacrifices

Interpretability of the CG representation



Flexible, learned embeddings (potentially nonlinear)

Dynamics in CG <> Dynamics in FG

Quantitative accuracy of inverted Boltzmann averages

- *Noid & Voth* (CG-space criterion) :

$$\hat{F}(z) \equiv -\beta^{-1} \log Z^{-1} \int_{\Omega} e^{-\beta U(\boldsymbol{x})} \delta(\Theta(\boldsymbol{x}) - z) d\boldsymbol{x} \leftrightarrow \hat{U}(z).$$

Potential of mean force *Coarse-grained potential*

- *Us* (FG-space criterion):

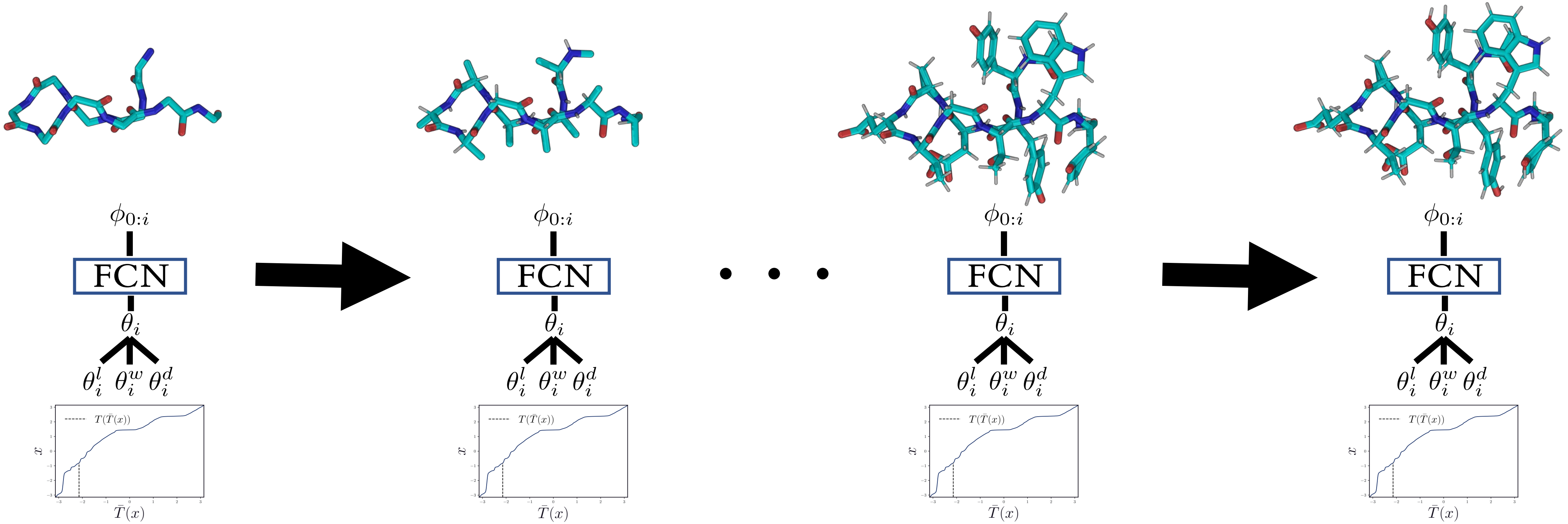
if for every observable $f \in \mathcal{F}$,

$$\int f(\boldsymbol{x}) p_{\text{inv}}(\boldsymbol{x}|z) \hat{\rho}(z, \boldsymbol{\theta}) \, d\boldsymbol{x} dz \longrightarrow \int f(\boldsymbol{x}) \rho(\boldsymbol{x}) \, d\boldsymbol{x}.$$

Inverted CG samples *Boltzmann*

$\mathcal{F}$ – weak thermodynamic consistency

Learning p_{inv} --- rigorously sampling FG space



$$T\sharp\varrho(\boldsymbol{x}) = \varrho(T^{-1}(\boldsymbol{x}))|\nabla T^{-1}(\boldsymbol{x})|$$

Rigorously inverting the CG sampling

Rational quadratic neural spline flow

Compute ϕ^{seed} from $\tilde{\boldsymbol{x}}_i = \Theta^{\text{dec}}(\boldsymbol{z}_i)$

Sample $\phi_{b_i} \sim \varrho$

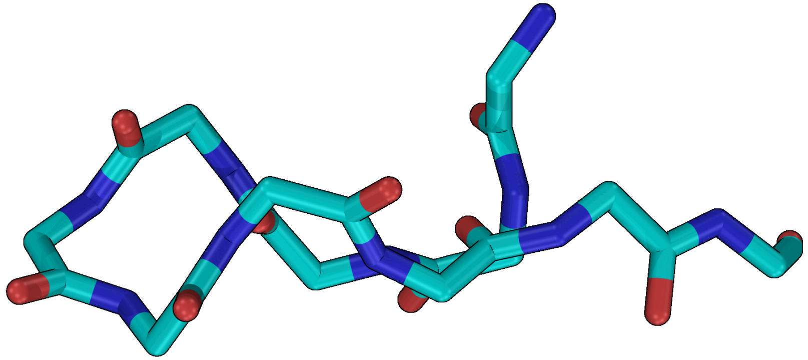
for $j = 0 \ldots m$ **do**

Compute $\theta_i^j = \text{FCN}(\phi_{\text{i}}^{\text{0:j}-1}, \phi^{\text{seed}})$

Compute $\phi_i^j = g_{\theta_i^j}(\phi_{b_i}^j)$

end for

Reconstruct $\boldsymbol{x}_i$ from $\tilde{\boldsymbol{x}}_i$ and ϕ_i



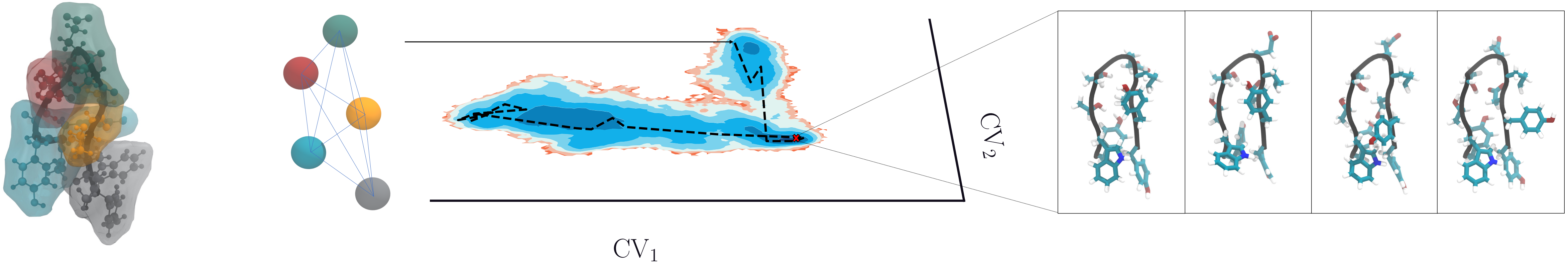
$$T\sharp\varrho(\boldsymbol{x}) = \varrho(T^{-1}(\boldsymbol{x}))|\nabla T^{-1}(\boldsymbol{x})|$$

Gabrié, M.; **GMR**.; Vanden-Eijnden, E. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (10), e2109420119. https://doi.org/10.1073/pnas.2109420119.

Closing the loop on coarse-grained modeling

Optimize with starting configurations → *Sample with Langevin* → *Invert!*

$\boldsymbol{x}_0$ $\boldsymbol{z}_0 = \mathcal{M}(\boldsymbol{x}_0)$ $\boldsymbol{z} \sim e^{-\beta \hat{U}(\boldsymbol{z},\boldsymbol{\theta})}$ $\boldsymbol{x}_i \sim p_{\text{inv}}(\boldsymbol{x}|\boldsymbol{z}_i)$



1. Fine-Grained Configuration

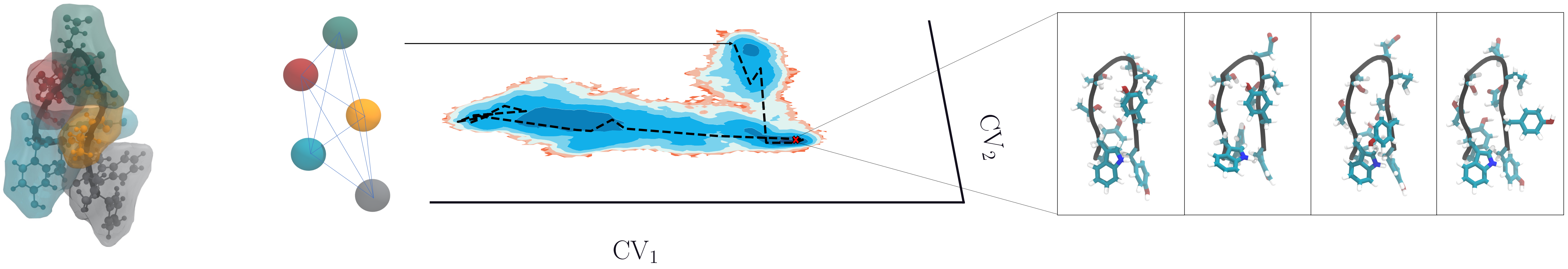
2. Coarse-Grained Dynamics

3. Generated Fine-Grained Configurations

Closing the loop on coarse-grained modeling

Optimize with starting configurations → *Sample with Langevin* → *Invert!*

$\boldsymbol{x}_0$ $\quad \boldsymbol{z}_0 = \mathcal{M}(\boldsymbol{x}_0)$ $\quad \boldsymbol{z} \sim e^{-\beta \hat{U}(\boldsymbol{z},\boldsymbol{\theta})}$ $\quad \boldsymbol{x}_i \sim p_{\text{inv}}(\boldsymbol{x}|\boldsymbol{z}_i)$

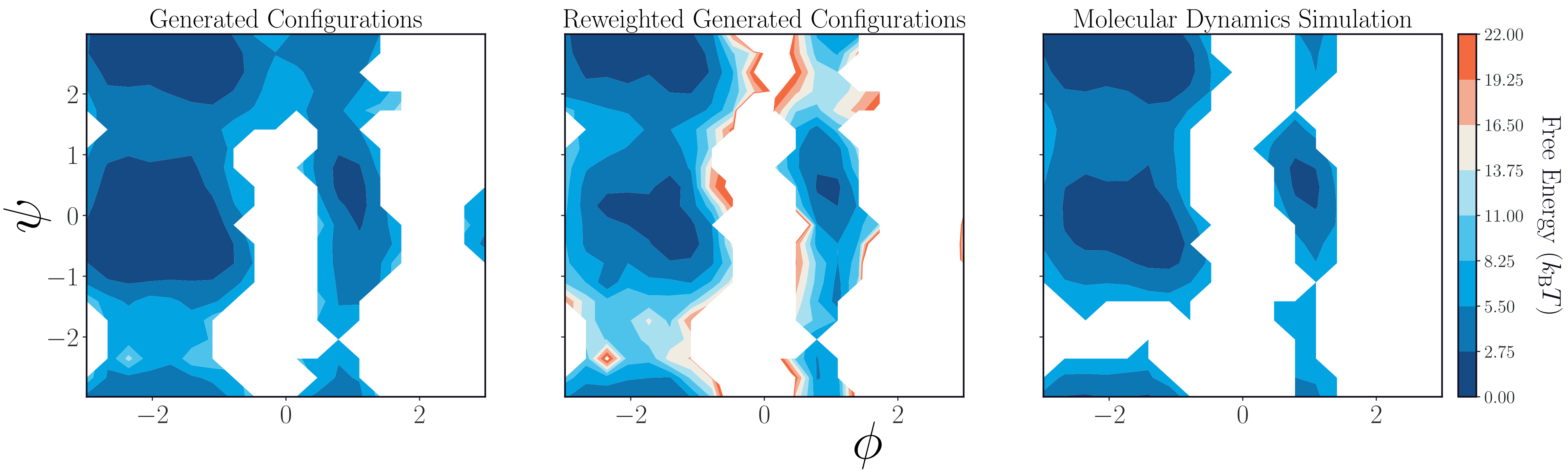


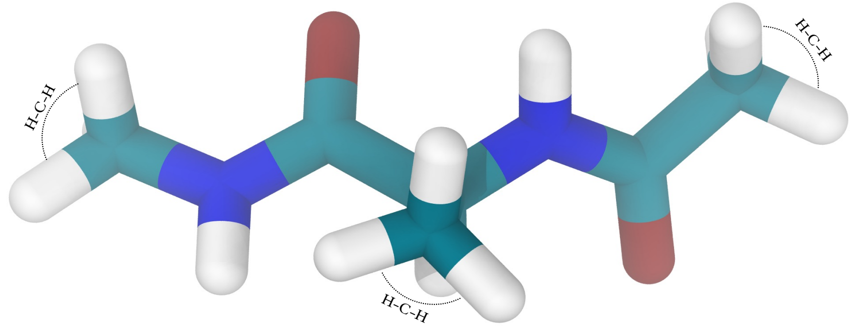
1. Fine-Grained Configuration

2. Coarse-Grained Dynamics

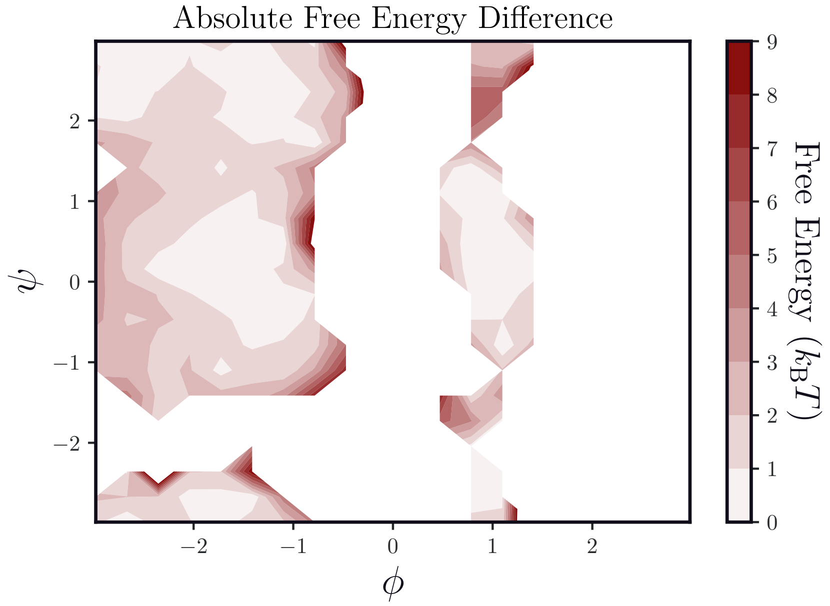
3. Generated Fine-Grained Configurations

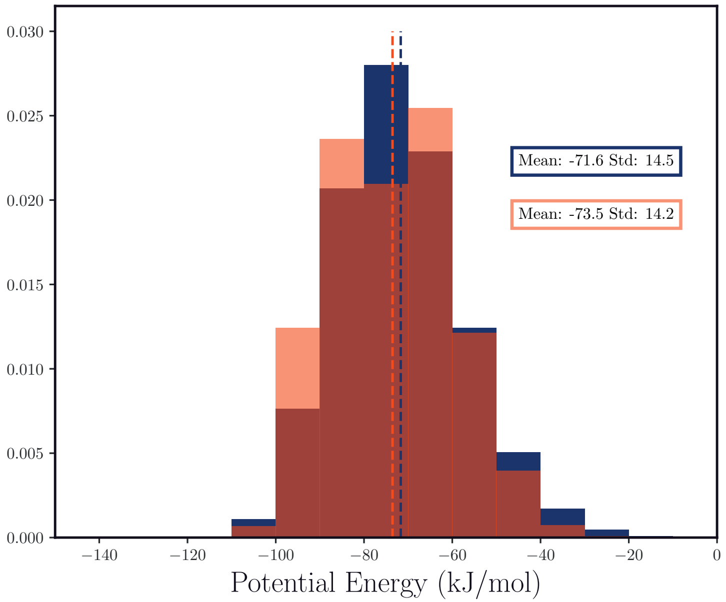
Compulsory example: alanine dipeptide

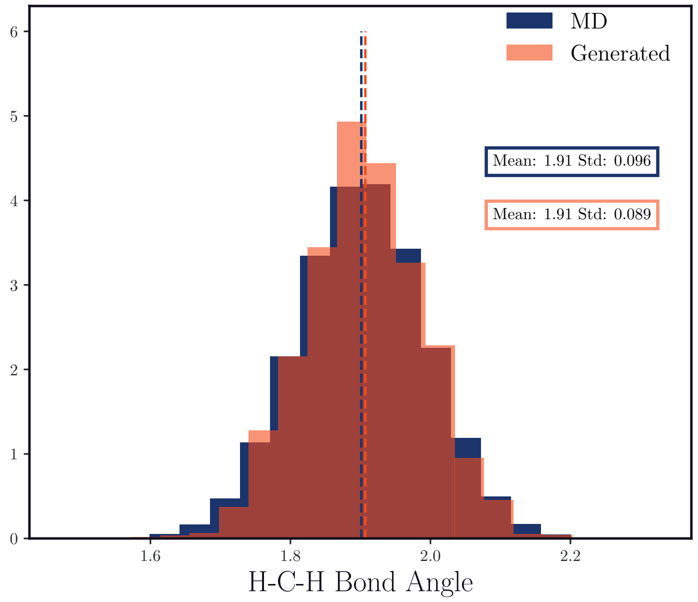




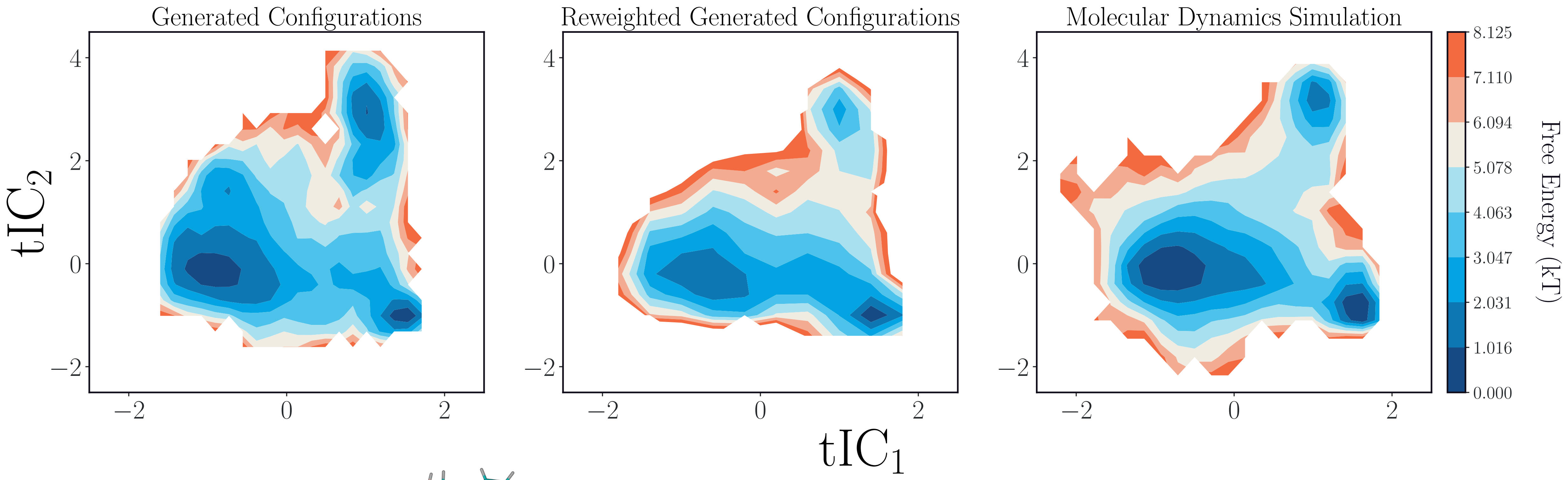
Error in basins is low, mixing is fast, expectations ☺

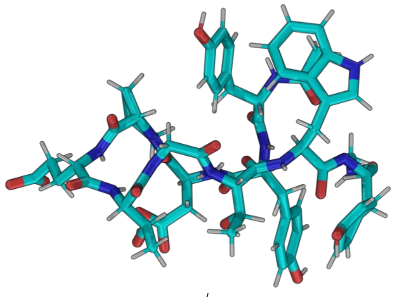




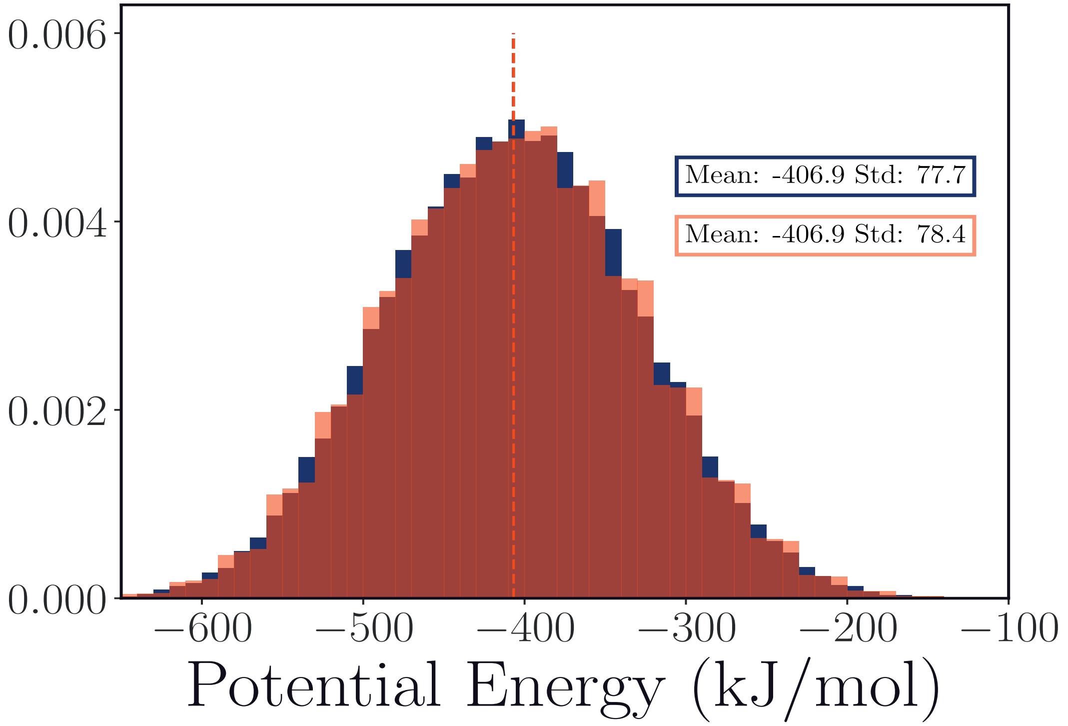


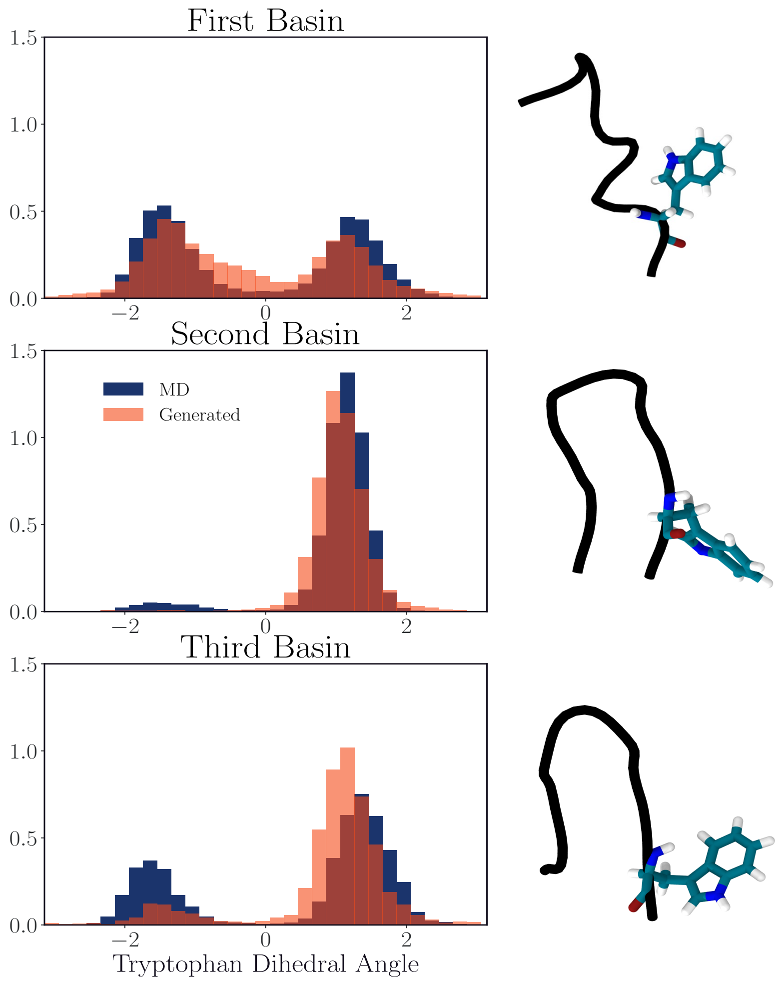
Folding of chignolin





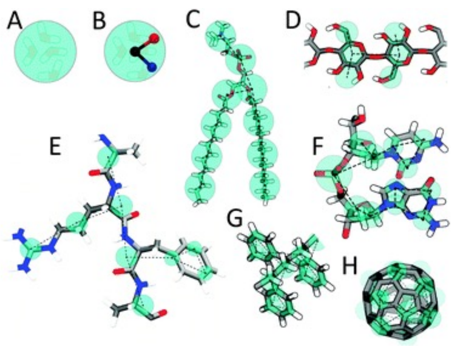
Large- and small-scale observables well-captured





Challenges in coarse-graining biomolecular systems

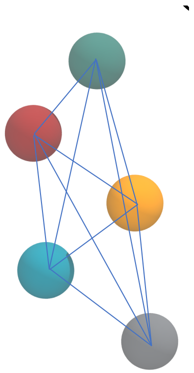
Representation



Marrink and Tieleman *Chem. Soc. Rev.*, 2013, **42**, 6801-6822

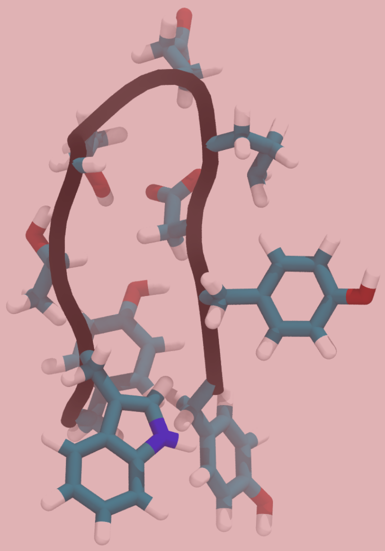
Linear projections
Independent of fine-grained state
Empirical potential (delta ML)

Interpretation



No access to fine-grained state
Imperfect PMF recovery / bias
Practical Mori-Zwanzig

Computation



Quality of generalization?
Need relevant rare configurations
Limited opportunities for feedback

Thanks!

U.S. DEPARTMENT OF ENERGY | Office of Science

Google Research

Shriram Chennakesavalu
Graduate Student

Grant M. Rotskoff
Assistant Professor of Chemistry

Andy Mitchell
Graduate Student

Clay Batton
Postdoctoral Researcher

David Toomer
Undergraduate Researcher

Emmit Pert
Graduate Student

Isaac Applebaum
Undergraduate Researcher

Sherry Li
Graduate Student