Deep-learning techniques for identifying collective variables of molecular dynamics

Wei Zhang

Zuse Institute Berlin and Free University Berlin

Inria-ECDF Partnership Kick-Off Workshop, 5-7 June, Berlin

Collaborations

- Tony Lelièvre (ENPC/Inria, Paris)
- Gabriel Stoltz (ENPC/Inria, Paris)
- Christof Schütte (ZIB/FUB, Berlin)

Collaborations

Joint Publications:

- Effective dynamics of stochastic dynamics

  T. Lelièvre and W. Zhang (2019). "Pathwise estimates for effective dynamics: The case of nonlinear vectorial reaction coordinates". In: *Multiscale Model. Simul.* 17.3, pp. 1019–1051

- Monte Carlo on submanfolds

  T. Lelièvre, G. Stoltz, and W. Zhang (2023). "Multiple projection MCMC algorithms on submanifolds". In: *IMA J. Numer. Anal.* 43.2, pp. 737–788

- Autoencoders

  T. Lelièvre, T. Pigeon, G. Stoltz, and W. Zhang (2024). "Analyzing multimodal probability measures with autoencoders". In: *J. Phys. Chem. B*

Outline

Applications

Understanding **structural** and **dynamical** properties of proteins are crucial for:

- disease prevention
- drug discovery

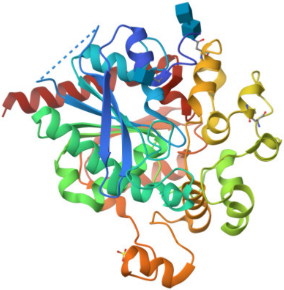


Figure: 6V7N in PDB, 6241 atoms.

Approaches

- Experiments
- Computational approach, i.e. molecular dynamics (MD) simulations
  - efficient numerical methods: integration schemes, free energy calculations, coarse-graining methods...
  - high-performance software packages, i.e. Gromacs, NAMD, LAMPS, OpenMM...
  - significant improvement of hardware
- AI-based methods, e.g. AlphaFold

MD simulations by D.E. Shaw Research (2011):

Science
Current Issue   First release pa[p]

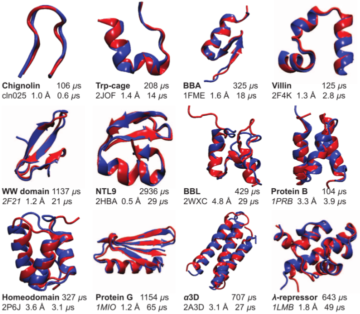
HOME > SCIENCE > VOL. 334, NO. 6055 > HOW FAST-FOLDING PROTEINS FOLD

REPORT

How Fast-Folding Proteins Fold

KRESTEN LINDORFF-LARSEN, STEFANO PIANA, RON O. DROR, AND DAVID E. SHAW   Authors Info & Affiliations

SCIENCE • 28 Oct 2011 • Vol 334, Issue 6055 • pp. 517-520 • DOI: 10.1126/science.1208351



Protein folding

Prediction of protein structures by AlphaFold (2021):

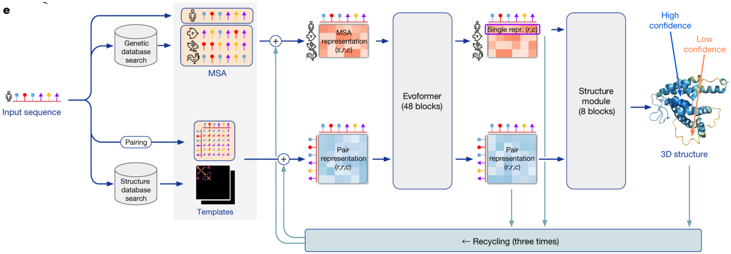
Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, ... Demis Hassabis

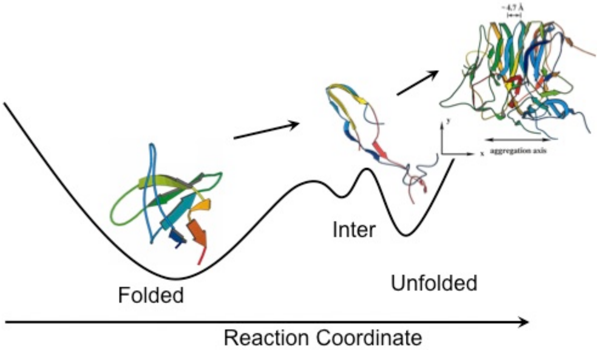
+ Show authors

Nature **596**, 583–589 (2021) | Cite this article

1.59m Accesses | **8815** Citations | **3591** Altmetric | Metrics



Conformational changes of bio-molecular systems



Challenge: metastability

Protein dynamics

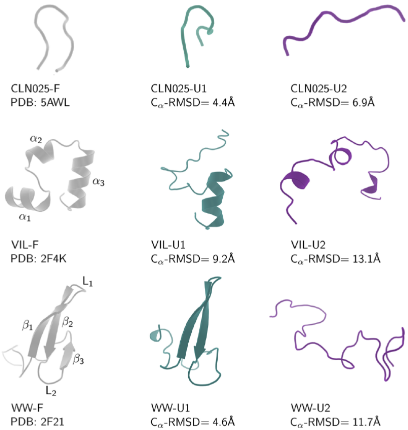
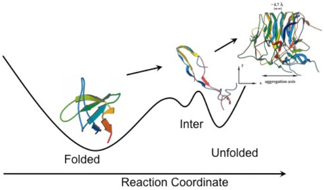


Figure: Folded and unfolded structures of proteins (Kamenik, Handle, Hofer, Kahler, Kraml, and Liedl, 2020).

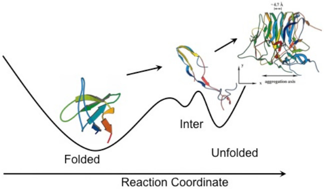
Main questions and challenges



Research topics:

- identify metastable conformations
- understand transition mechanism, or find transition pathways
- estimate transition rates
- build simpler (low-dimensional) surrogate models

Main questions and challenges



Research topics:

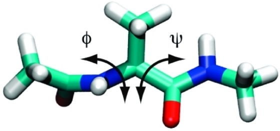
- identify metastable conformations
- understand transition mechanism, or find transition pathways
- estimate transition rates
- build simpler (low-dimensional) surrogate models

Major challenges:

- many atoms $\Longrightarrow$ high dimensions
- many integration steps due to disparity between small step-size (10^{-15}s) and large simulation time (from 10^{-6}s to 10^{-3}s).

Collective variables (CVs)

- **High-dimensional** conformations of molecular systems can often be represented by **a few variables**.



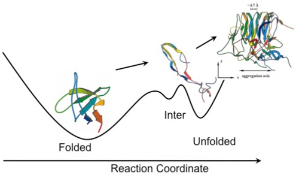
- In general, a set of collective variables, or reaction coordinates, is a map

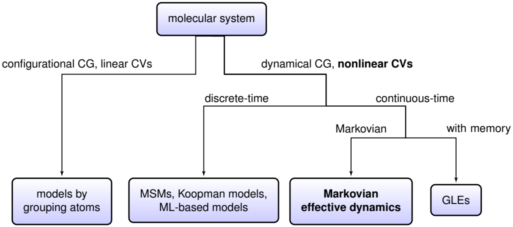
$$\xi = (\xi_1, \ldots, \xi_m)^T : \mathbb{R}^d \to \mathbb{R}^m, \quad \text{where } 1 \leq m < d.$$

Applications of CVs

CVs are useful in:

1. **enhanced sampling**, e.g. adaptive biasing force (ABF), metadynamics.
2. **coarse-graining**





Outline

1. Motivation
2. Eigenfunctions
   - motivation and variational principle
   - training algorithm
   - application to transfer operator
3. Autoencoders

Overdamped Langevin dynamics

- SDE on $\mathbb{R}^d$

$$dx(s) = -\nabla V(x(s))\, ds + \sqrt{2\beta^{-1}}\, dw(s)\,, \quad s \geq 0\,,$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a potential, $w(s)$ is a d-dimensional Brownian motion, $\beta > 0$ specifies the strength of the noise.

For MD systems, $d = 3N_{\mathrm{atom}}$ and $\beta = (k_B T)^{-1}$.

- Invariant measure: $d\mu(x) = \frac{1}{Z} e^{-\beta V(x)}\, dx$.

- Generator $\mathcal{L}f = -\nabla V \cdot \nabla f + \frac{1}{\beta}\Delta f$ is self-adjoint.

Overdamped Langevin dynamics

- SDE on $\mathbb{R}^d$

$$dx(s) = -\nabla V(x(s))\,ds + \sqrt{2\beta^{-1}}\,dw(s)\,, \quad s \geq 0\,,$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a potential, $w(s)$ is a d-dimensional Brownian motion, $\beta > 0$ specifies the strength of the noise.

For MD systems, $d = 3N_{\mathrm{atom}}$ and $\beta = (k_B T)^{-1}$.

- Invariant measure: $d\mu(x) = \frac{1}{Z} e^{-\beta V(x)}\,dx$.

- Generator $\mathcal{L}f = -\nabla V \cdot \nabla f + \frac{1}{\beta}\Delta f$ is self-adjoint.

Remark:

1. $f(x,t) = \mathbf{E}(g(x(t))|x(0) = x)$ solves $\frac{\partial f}{\partial t} = \mathcal{L}f$.
2. Eigenvalues of $\mathcal{L}$ encode the speed of convergence to equilibrium.

Effective dynamics by conditional expectation

Assume $\xi : \mathbb{R}^d \to \mathbb{R}^m$ is given.

- Ito's formula implies

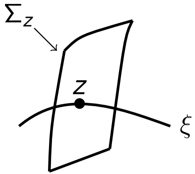
$$d\xi_l(x(s)) = \mathcal{L}\xi_l(x(s))ds + \sqrt{2\beta^{-1}}\sum_{i=1}^{d} \frac{\partial \xi_l}{\partial x_i}(x(s))\, dw_i(s)\,. \tag{1}$$

- (Legoll and Lelièvre, 2010) proposed a **Markovian** approximation of (1):

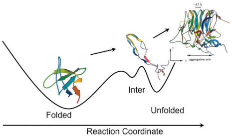
$$dz(s) = \widetilde{b}(z(s))\, ds + \sqrt{2\beta^{-1}}\widetilde{\sigma}(z(s))\, d\widetilde{w}(s)\,, \tag{2}$$

where, for $z \in \mathbb{R}^m$,

$$\begin{aligned}\widetilde{b}_l(z) =& \mathbf{E}_{\mu_z}(\mathcal{L}\xi_l)\,,\\ \widetilde{a}(z) =& (\widetilde{\sigma}\widetilde{\sigma}^T)(z) = \mathbf{E}_{\mu_z}\left(\nabla\xi\nabla\xi^T\right)\,.\end{aligned}$$



Optimality



Question: How good is the effective dynamics (2)?

We compare small eigenvalues of PDEs $-\mathcal{L}\varphi_i = \lambda_i \varphi_i$ and $-\widetilde{\mathcal{L}}\widetilde{\varphi}_i = \widetilde{\lambda}_i \widetilde{\varphi}_i$.

Proposition (Zhang, Hartmann, and Schütte, 2016)

Define $\mathcal{E}(f) = \frac{1}{\beta}\mathbf{E}_\mu\big(|\nabla f|^2\big)$. *For* $i = 1, 2, \ldots,$ *we have*

$$\lambda_i \leq \widetilde{\lambda}_i \leq \lambda_i + \mathcal{E}(\varphi_i - \widetilde{\varphi}_i \circ \xi)\,. \tag{3}$$

In particular, when $\xi(x) = \big(\varphi_1(x), \varphi_2(x), \cdots, \varphi_m(x)\big) \in \mathbb{R}^m$, *we have* $\widetilde{\lambda}_i = \lambda_i$, *for* $0 \leq i \leq m$.

Variational principle for the first K eigenpairs

Theorem

Let $K \in \mathbb{N}$ and $\omega_1 \geq \cdots \geq \omega_K > 0$. $\mathcal{E}(f) = \frac{1}{\beta}\mathbf{E}_\mu\big(|\nabla f|^2\big)$. We have

$$\sum_{i=1}^{K} \omega_i \lambda_i = \min_{f_1,\dots,f_K \in \mathcal{H}^1} \sum_{i=1}^{K} \omega_i \mathcal{E}(f_i)\,, \tag{4}$$

where the minimum is over all $f_1, f_2, \dots, f_K \in \mathcal{H}^1$ such that

$$\langle f_i, f_j \rangle_\mu = \delta_{ij}\,, \quad \forall i, j \in \{1, \dots, K\}\,. \tag{5}$$

Moreover, the minimum in (4) is achieved when $f_i = \varphi_i$ for $1 \leq i \leq K$.

Remark: See Zhang and Schütte (2017) and Zhang, Li, and Schütte (2022) for proofs.

Training task

- Optimization task

$$\min_{f_1,\dots,f_K\in\mathcal{H}^1} \sum_{i=1}^{K} \omega_i \mathcal{E}(f_i)\,, \qquad \text{s.t.,} \quad \langle f_i, f_j\rangle_\mu = \delta_{ij}, \quad \forall i,j \in \{1,\dots,K\}. \tag{6}$$

- Loss in practice

$$\mathrm{Loss}(f_1, f_2, \dots, f_K) = \sum_{i=1}^{K} \omega_i \widetilde{\mathcal{E}}^{\mathrm{data}}(f_i) + \alpha \sum_{1\le i_1\le i_2\le K} \Big(\mathrm{Cov}^{\mathrm{data}}(f_{i_1}, f_{i_2}) - \delta_{i_1 i_2}\Big)^2,$$

where α is a penalty constant and

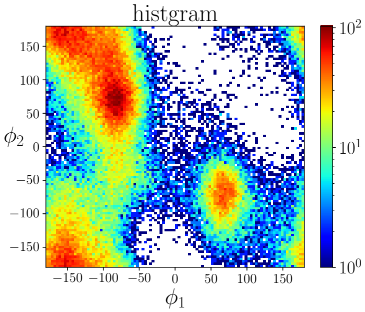
$$\widetilde{\mathcal{E}}^{\mathrm{data}}(f) := \frac{1}{\beta}\frac{\mathbf{E}^{\mathrm{data}}(|\nabla f|^2)}{\mathrm{Var}^{\mathrm{data}} f}\,.$$

Zhang, Li, and Schütte (2022)

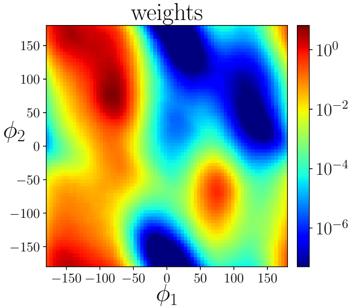
Example: Alanine dipeptide in vacuum

22 atoms, dimension $d = 66$.

- NAMD package, Langevin dynamics, damping $\gamma = 1\,\mathrm{ps}^{-1}$, temperature 300 K
- ABF, with fixed biasing potential $V_{\mathrm{bias}} = 0.7\,V_{\mathrm{PMF}}$.
- step-size 1 fs, trajectory length $T = 100$ ns.
- Training data: $n = 10^5$ states.



(a) training data

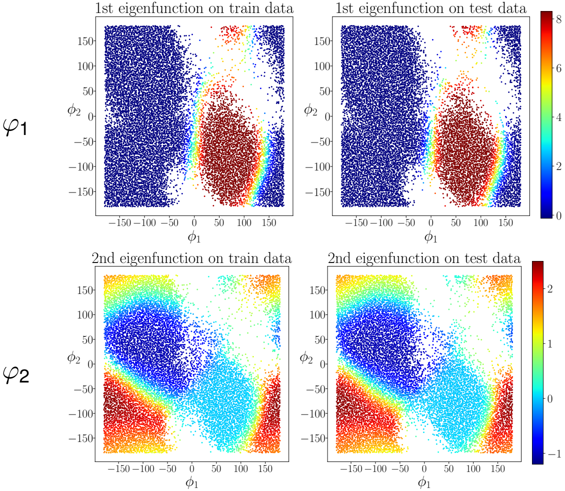


(b) weights as a function of angles

Results

In the last 4000 steps,

$$\lambda_1 = 0.047 \pm 0.005 \text{ns}^{-1}, \quad \lambda_2 = 23.92 \pm 0.60 \text{ns}^{-1}.$$



Application to transfer operator

- Given lag-time τ, recall the transfer operator

$$(P_\tau u)(y) = \frac{1}{\pi(y)} \int_{\mathbb{R}^d} p_\tau(y|x) u(x) \pi(x)\, dx, \tag{7}$$

$\implies \langle P_\tau f, g \rangle_\mu = \langle f, P_\tau g \rangle_\mu$.

For overdamped Langevin: $P_\tau = \mathrm{e}^{-\tau \mathcal{L}}$, $\forall\, \tau \geq 0$.

Application to transfer operator

- Given lag-time τ, recall the transfer operator

$$(P_\tau u)(y) = \frac{1}{\pi(y)} \int_{\mathbb{R}^d} p_\tau(y|x) u(x) \pi(x)\, dx, \tag{7}$$

$$\Longrightarrow \langle P_\tau f, g\rangle_\mu = \langle f, P_\tau g\rangle_\mu.$$

For overdamped Langevin: $P_\tau = \mathrm{e}^{-\tau\mathcal{L}}$, $\forall\, \tau \geq 0$.

- Loss function:

$$\begin{aligned}\mathrm{Loss}(f_1, f_2, \ldots, f_K) =& \frac{1}{2} \sum_{i=1}^{K} \omega_i \mathbf{E}^{\mathrm{data}} \left|f_i(x_{\cdot+\tau}) - f_i(x_\cdot)\right|^2 \\ &+ \alpha \sum_{1\leq i_1 \leq i_2 \leq K} \left(\mathrm{Cov}^{\mathrm{data}}(f_{i_1}, f_{i_2}) - \delta_{i_1 i_2}\right)^2.\end{aligned}$$

Application to transfer operator

- Given lag-time τ, recall the transfer operator

$$(P_\tau u)(y) = \frac{1}{\pi(y)} \int_{\mathbb{R}^d} p_\tau(y|x)u(x)\pi(x)\,dx, \tag{7}$$

$$\Longrightarrow \langle P_\tau f, g\rangle_\mu = \langle f, P_\tau g\rangle_\mu.$$

  For overdamped Langevin: $P_\tau = \mathrm{e}^{-\tau\mathcal{L}}$, $\forall\, \tau \geq 0$.

- Loss function:

$$\begin{aligned}\mathrm{Loss}(f_1, f_2, \dots, f_K) =& \frac{1}{2}\sum_{i=1}^{K} \omega_i \mathbf{E}^{\mathrm{data}} |f_i(x_{\cdot+\tau}) - f_i(x_\cdot)|^2 \\ &+ \alpha \sum_{1\le i_1 \le i_2 \le K} \Big(\mathrm{Cov}^{\mathrm{data}}(f_{i_1}, f_{i_2}) - \delta_{i_1 i_2}\Big)^2.\end{aligned}$$

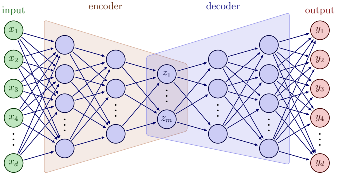
- Compared to VAMPnets (Mardt, Pasquali, Wu, and Noé, 2018; Chen, Sidky, and Ferguson, 2019):
  - eigenfunctions v.s. basis of eigenspace.
  - constraint: explicit v.s. implicit.
  - simple objective v.s. matrix (eigenvalue) problem.

Outline

Autoencoders

Given bottleneck dimension m, where $1 \leq m < d$.

- Autoencoder: $f = f_{dec} \circ f_{enc}$, where $f_{enc} : \mathbb{R}^d \to \mathbb{R}^m$ and $f_{dec} : \mathbb{R}^m \to \mathbb{R}^d$.



- **Reconstruction loss** is often used

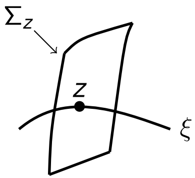
$$\mathcal{L}(f_{enc}, f_{dec}) = \int_{\mathbb{R}^d} \|f_{dec} \circ f_{enc}(x) - x\|^2 \, d\mu \,.$$

- The encoder is used as CVs, i.e. $\xi = f_{enc}$.

A simple characterization

We have derived:

$$\min_{f_{enc}} \min_{f_{dec}} \int_{\mathbb{R}^d} \|f_{dec} \circ f_{enc}(x) - x\|^2 \, d\mu = \min_{f_{enc}} \mathbf{E}_{\widetilde{\mu}} \left[Var_{\mu_z}(x) \right]$$



It implies:

1. optimal encoder f_{enc} minimizes the conditional variances on its level-sets.
2. optimal decoder: $f_{dec}(z) = \mathbf{E}_{\mu_z} x$.

Autoencoder loss does **not** contain dynamical information!

Decoder may predict **wrong** transiton path

Example: Müller-Brown potential

$\beta = 1.0$. Training data 10^5. Batch-size 70000, 2000 epochs, learning rate 0.005. Activation $\tanh$. Encoder: $(2, 30, 30, 30, 30, 1)$. Decoder: $(1, 30, 30, 30, 2)$. Random seed 2046.

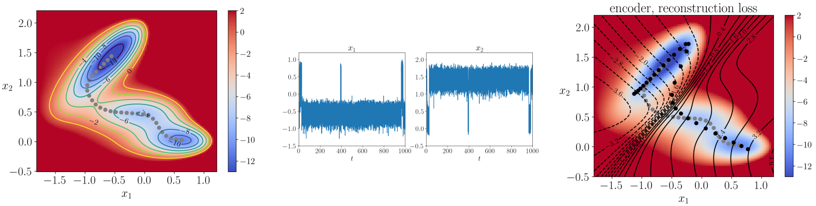
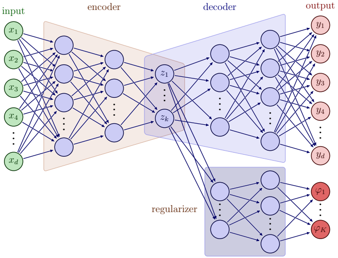


Figure: potential, data, encoder (contour lines) and decoder (black dots).

Dynamical regularization in autoencoders

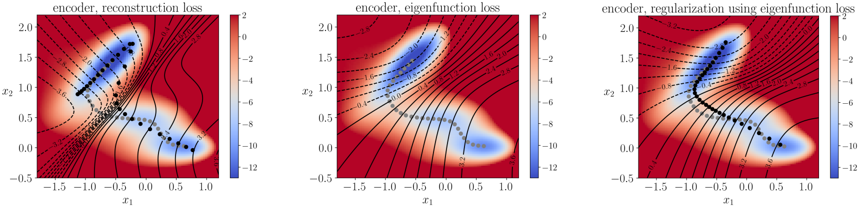
- Regularization by representing eigenfunctions (Lelièvre, Pigeon, Stoltz, and Zhang, 2024):

$$\min_{\xi} \gamma_1 \left[\min_{\eta} \int_{\mathbb{R}^d} \left| \eta(\xi(x)) - x \right|^2 \mu(dx) \right] + \gamma_2 \mathcal{R}(\xi) \,. \tag{8}$$



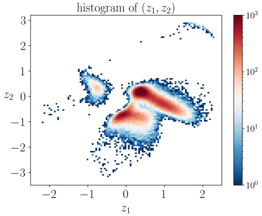
Remark: Also see the **extended autoencoder** by Frassek, Arjun, and Bolhuis, 2021 and **time-lagged autoencoder** by Wehmeyer and Noé, 2018.

The same parameters, regularization helps!

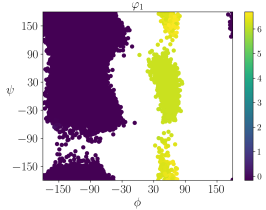


(a) without regularization (b) loss (8), $\gamma_1 = 0$, $\gamma_2 = 1$. (c) loss (8), $\gamma_1 = 0.1$, $\gamma_2 = 1$.

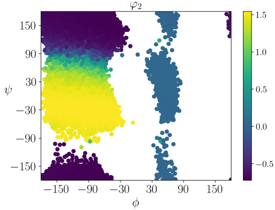
Example: Alanine dipeptide in water



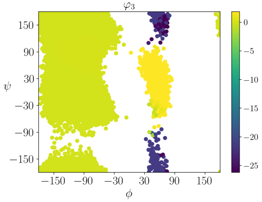
(a) encoder



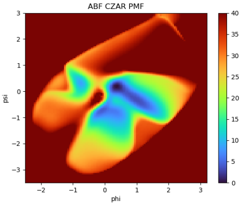
(b) φ_1



(c) φ_2



(d) φ_3



(e) PMF by Gromacs+Colvars

Conclusion

Today's talk:

- Motivations from MD, and usefulness of CVs
- Two approaches:
  - eigenfunctions: **spectral, dynamical** point of view.
  - autoencoder: **static** point of view, dimension reduction method, nonlinear counterpart of PCA.
- Training algorithms: AEs with regularization

Future work:

- Methods for jointly learning CVs and reduced dynamics
- Combine learning CVs and enhance sampling
- Applications to complex MD systems

References

1. W. Zhang, C. Hartmann, and C. Schütte (2016). "Effective dynamics along given reaction coordinates, and reaction rate theory". In: *Faraday Discuss.* 195, pp. 365–394
2. W. Zhang and C. Schütte (2017). "Reliable approximation of long relaxation timescales in molecular dynamics". In: *Entropy* 19.7. DOI: 10.3390/e19070367
3. W. Zhang, T. Li, and C. Schütte (2022). "Solving eigenvalue PDEs of metastable diffusion processes using artificial neural networks". In: *J. Comput. Phys.* 465, p. 111377
4. T. Lelièvre, T. Pigeon, G. Stoltz, and W. Zhang (2024). "Analyzing multimodal probability measures with autoencoders". In: *J. Phys. Chem. B*
5. W. Zhang and C. Schütte (2023). "Understanding recent deep-learning techniques for identifying collective variables of molecular dynamics". In: *PAMM*, e202300189

Thank you!