

# Machine Learning Based Enhanced Sampling Methods: Learning Collective Variables

Ming Chen

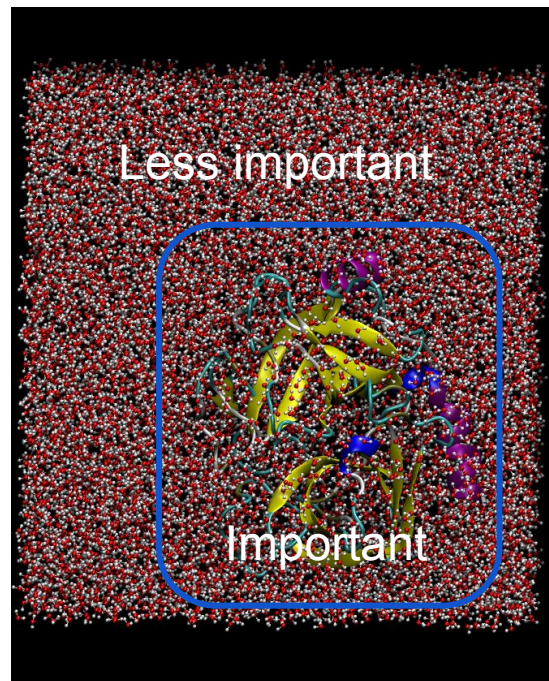
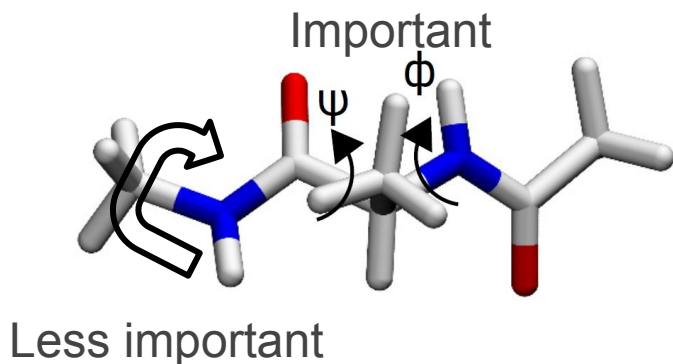
University of California, Berkeley

# Collective Variables (CV): Important Degrees of Freedom

A few degrees of freedom are important to capture the physical essences.

System Coordinates:  $\mathbf{X}$

Collective Variables:  $\mathbf{q} = (q_1(\mathbf{x}), \dots, q_{N_s}(\mathbf{x}))$

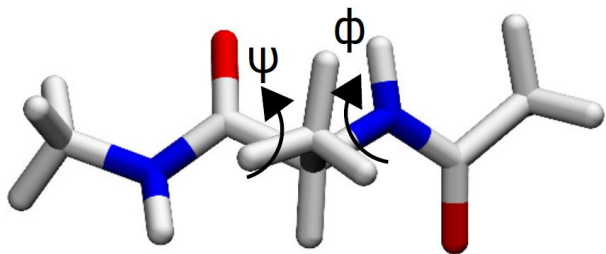


# Free Energy Surface and CVs

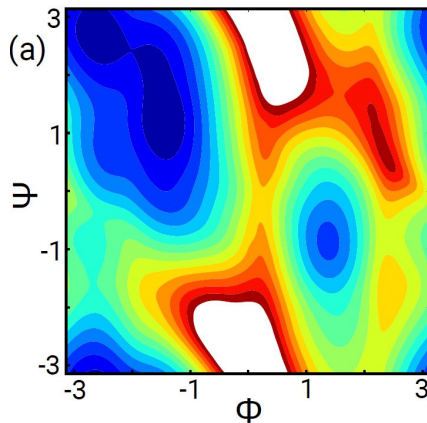
Fixing  $\mathbf{q} = \mathbf{s} = (s_1, \dots, s_{N_s})$

Marginal Distribution:  $P(\mathbf{s}) = C \int d\mathbf{x} e^{-\beta U(\mathbf{x})} \prod_{\alpha=1}^{N_s} \delta(q_{\alpha}(\mathbf{x}) - s_{\alpha})$

Free Energy Surface:  $A(\mathbf{s}) = -kT \log P(\mathbf{s}) + \tilde{C}$



J. Zhang and M. Chen, *arXiv:1705.07414*



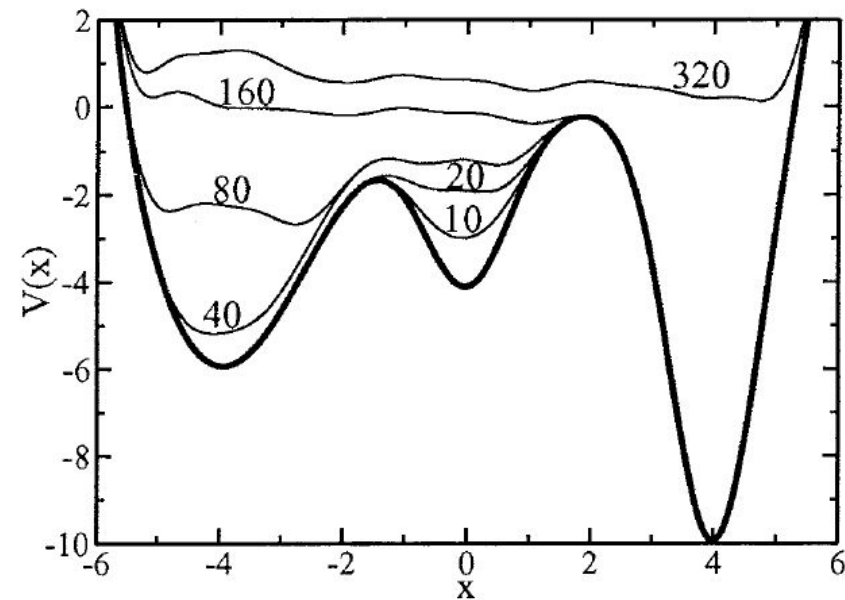
# Enhanced Sampling with CVs

Many Sampling methods enhance transitions along CVs. For example:

- Blue Moon Method
  - E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, *Chem. Phys. Lett.*, **156**, 472 (1989)
- Umbrella Sampling
  - G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977)
- Metadynamics/Well-tempered Metadynamics
  - A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002)
  - A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008)
- Driven Adiabatic Free Energy Dynamics/Temperature Accelerate Molecular Dynamics
  - J. B. Abrams and M. E. Tuckerman, *J. Phys. Chem. B* **112**, 15742 (2008).
  - L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.* **426**, 168 (2006).
- Adaptive Biasing Force
  - E. Darve and A. Pohorille, *J. Chem. Phys.* **115**, 9169 (2001).
- Many others...

# One Example: **W**ell-**t**empered **M**etadynamics

$$U(\mathbf{q}(\mathbf{x}), t) = \sum_{n, t_n = n\Delta t < t} W e^{-U(\mathbf{q}(\mathbf{x}(t_n)), t_n) / \Delta T} \exp \left\{ - \sum_{\alpha=1}^{N_s} \frac{(q_{\alpha}(\mathbf{x}) - q_{\alpha}(\mathbf{x}(t_n)))^2}{2\sigma_{\alpha}^2} \right\}$$

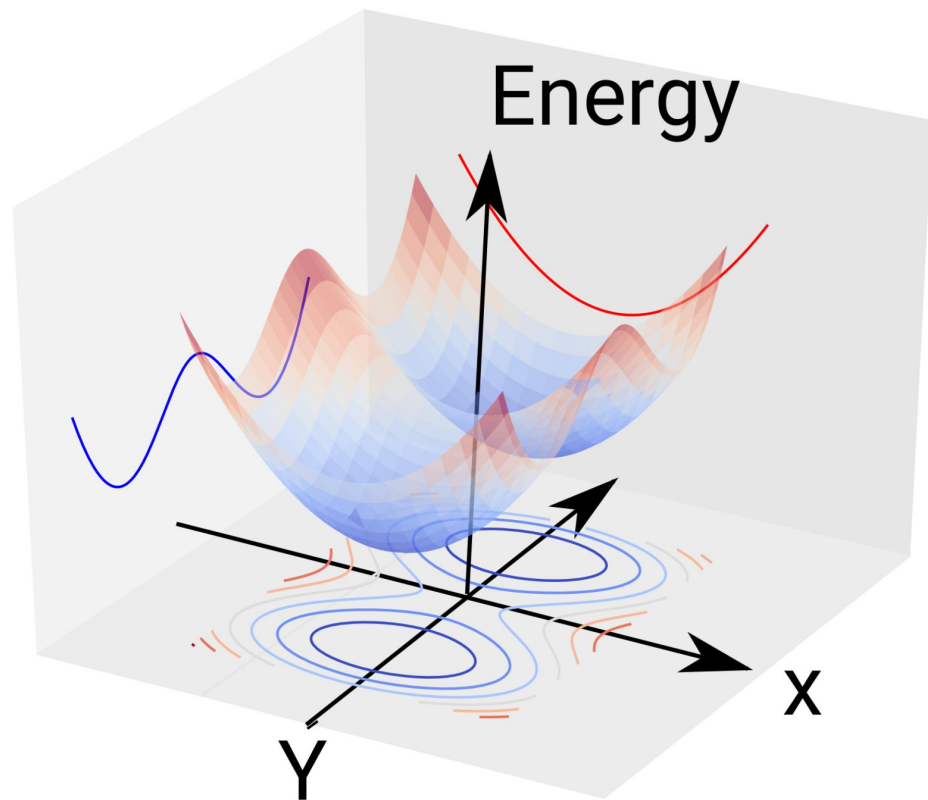


When  $t \rightarrow \infty$

$$U(\mathbf{q}(\mathbf{x}), t) \rightarrow -\frac{\Delta T}{T + \Delta T} A(\mathbf{q}(\mathbf{x}))$$

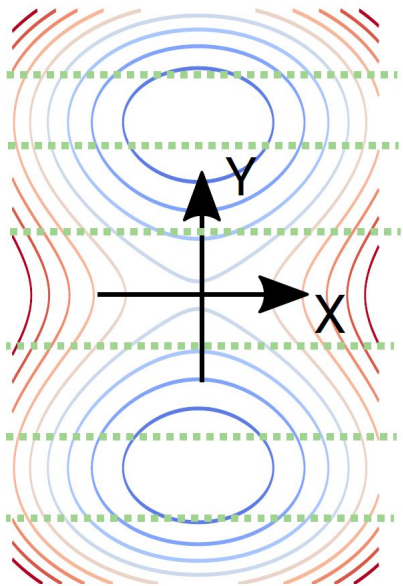
$$P(\mathbf{s}) \propto Q^{\frac{T}{T + \Delta T}}$$

# CV and Orthogonal Space Degeneracy

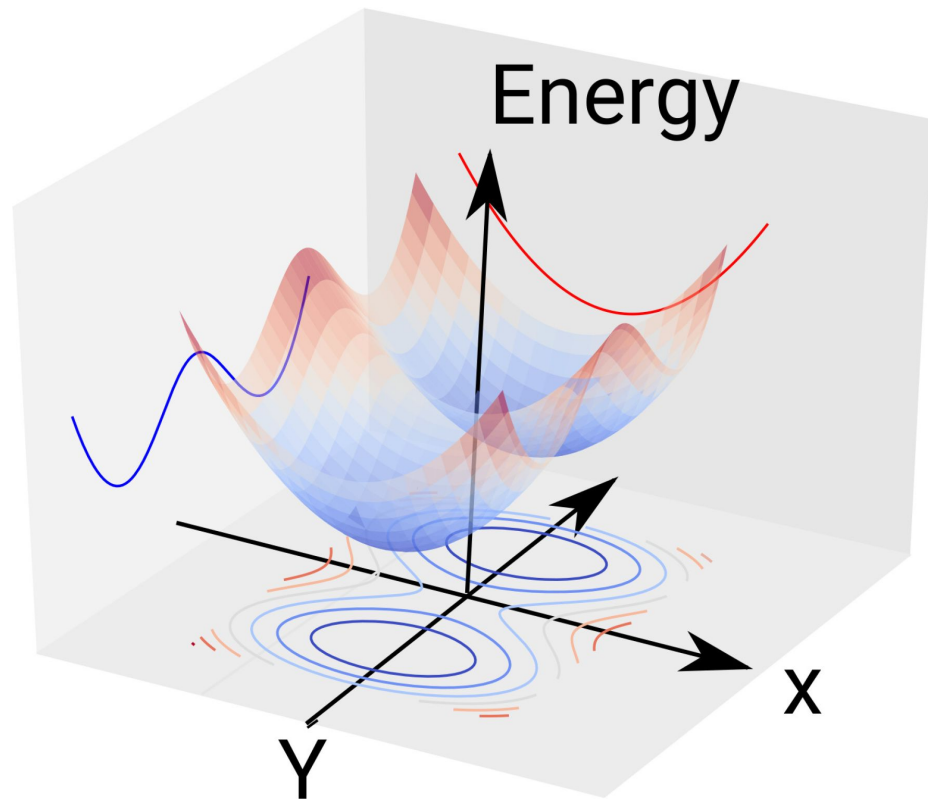
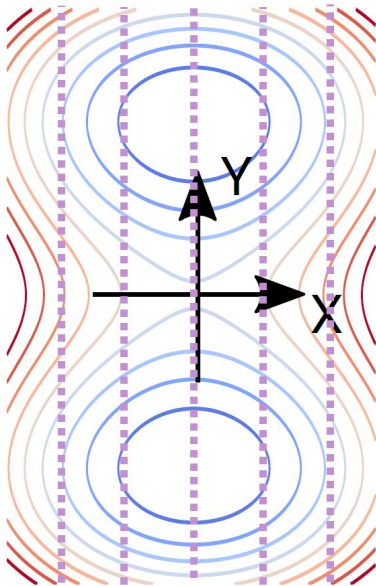


# CV and Orthogonal Space Degeneracy

CV: Y  
Orthogonal Space: X



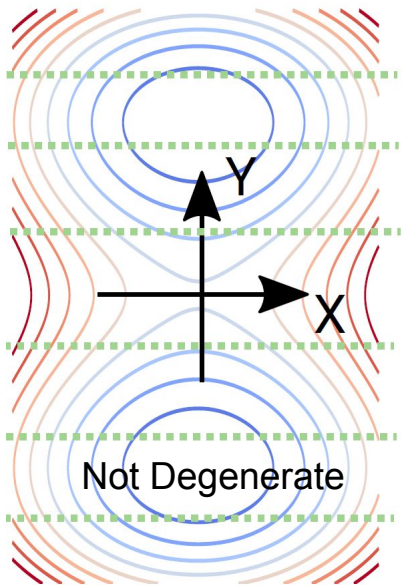
CV: X  
Orthogonal Space: Y



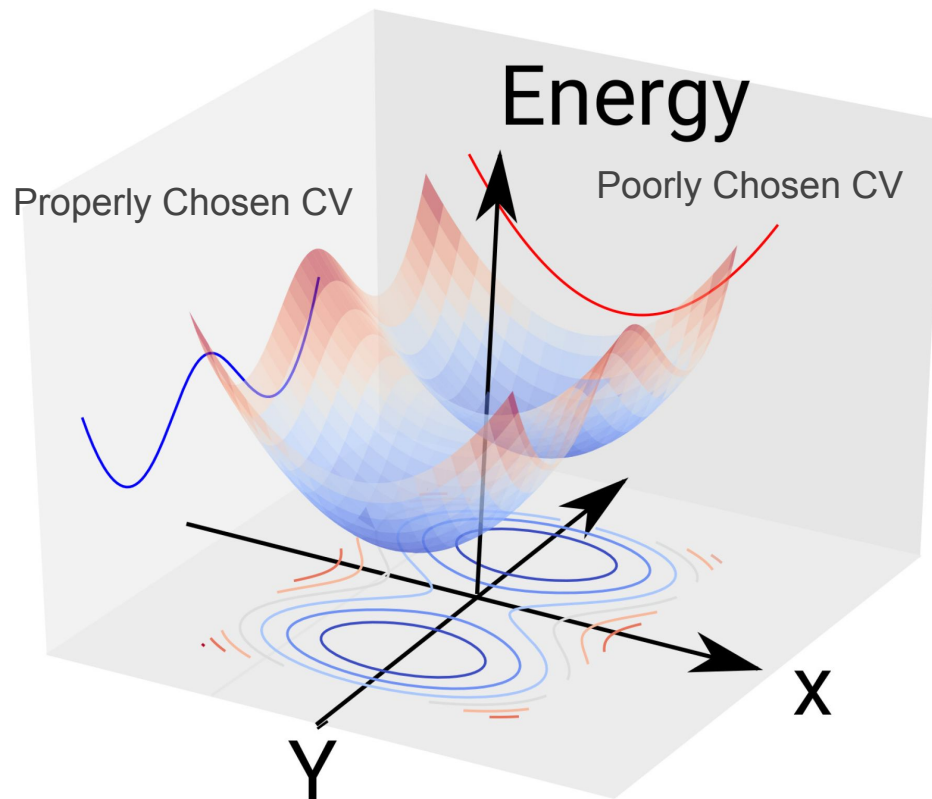
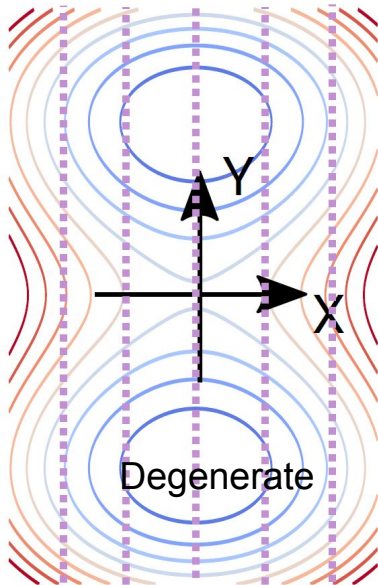


# CV and Orthogonal Space Degeneracy

CV: Y  
Orthogonal Space: X

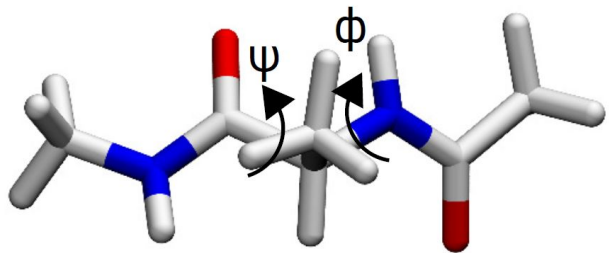


CV: X  
Orthogonal Space: Y





# Degeneracy and Sampling Efficiency: Case Study of Alanine Dipeptide

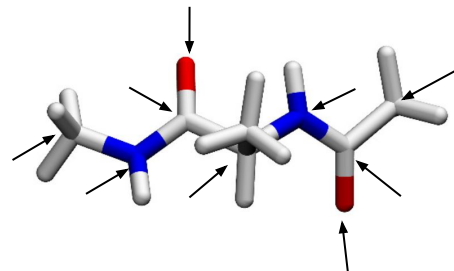


Ramachandran Dihedral Angle is believed as a set of appropriate CVs for alanine dipeptide.

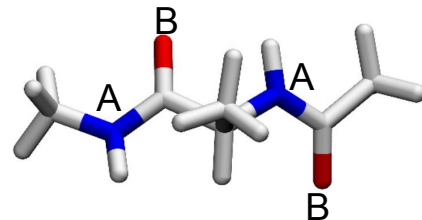
M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, *J. Chem. Phys.* **134**, 124116 (2011).

What about other CVs, e.g. radius and gyration (Rg) and number of hydrogen bond (NH)?

$$\text{Rg: } \sqrt{\frac{1}{N_b} \sum_{i=1}^{N_b} \|\mathbf{x}_i - \bar{\mathbf{x}}\|^2}$$



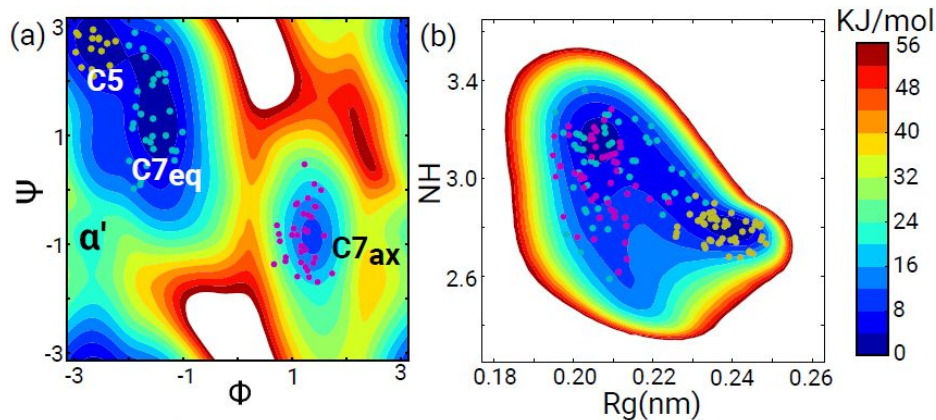
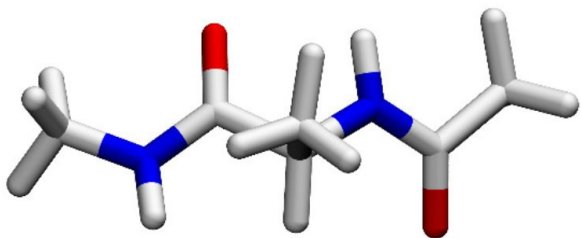
$$\text{NH: } \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}}$$



G. Bussi, F. L. Gervasio, A. Laio, M. Parrinello, *J. Am. Chem. Soc.* **128**, 13435 (2006)

J. B. Abrams and M. E. Tuckerman, *J. Phys. Chem. B* **112**, 15742 (2008).

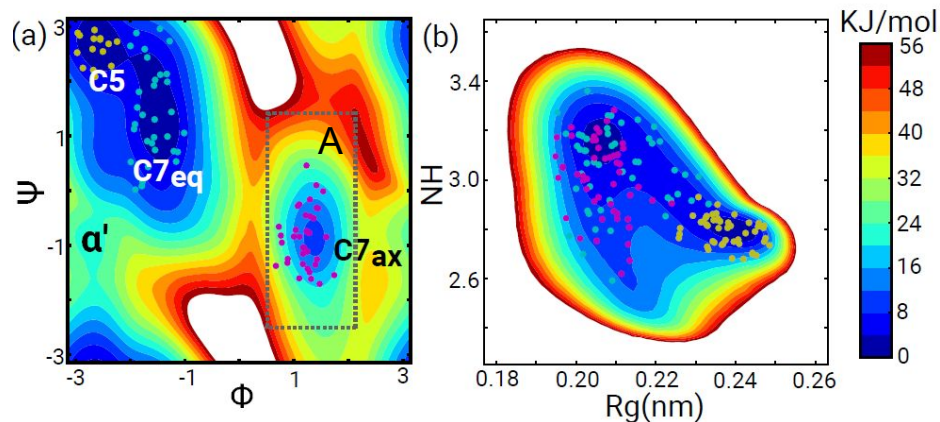
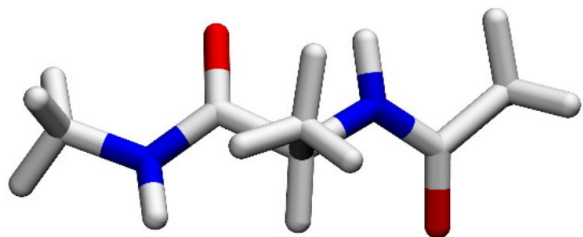
# Degeneracy and Sampling Efficiency: Case Study of Alanine Dipeptide



# Degeneracy and Sampling Efficiency: Case Study of Alanine Dipeptide

Correlation Function:

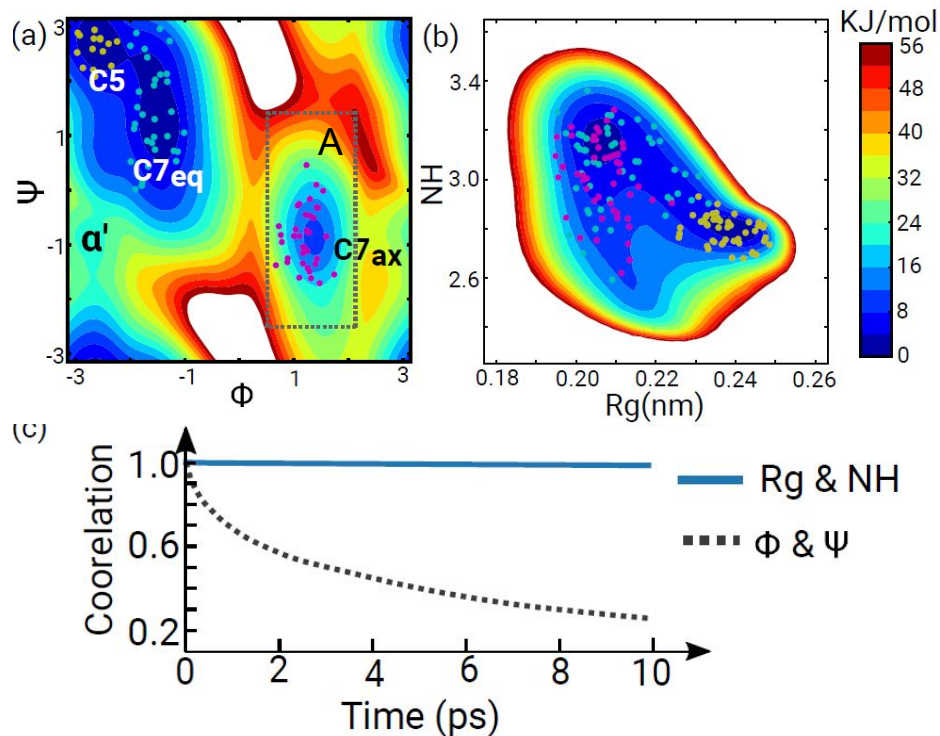
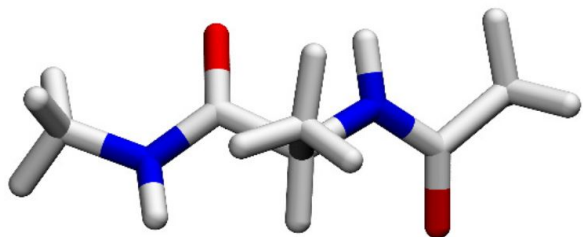
$$\frac{\langle \mathbb{1}_A(\mathbf{x}(0)) \mathbb{1}_A(\mathbf{x}(t)) \rangle}{\langle \mathbb{1}_A^2(\mathbf{x}(0)) \rangle}$$



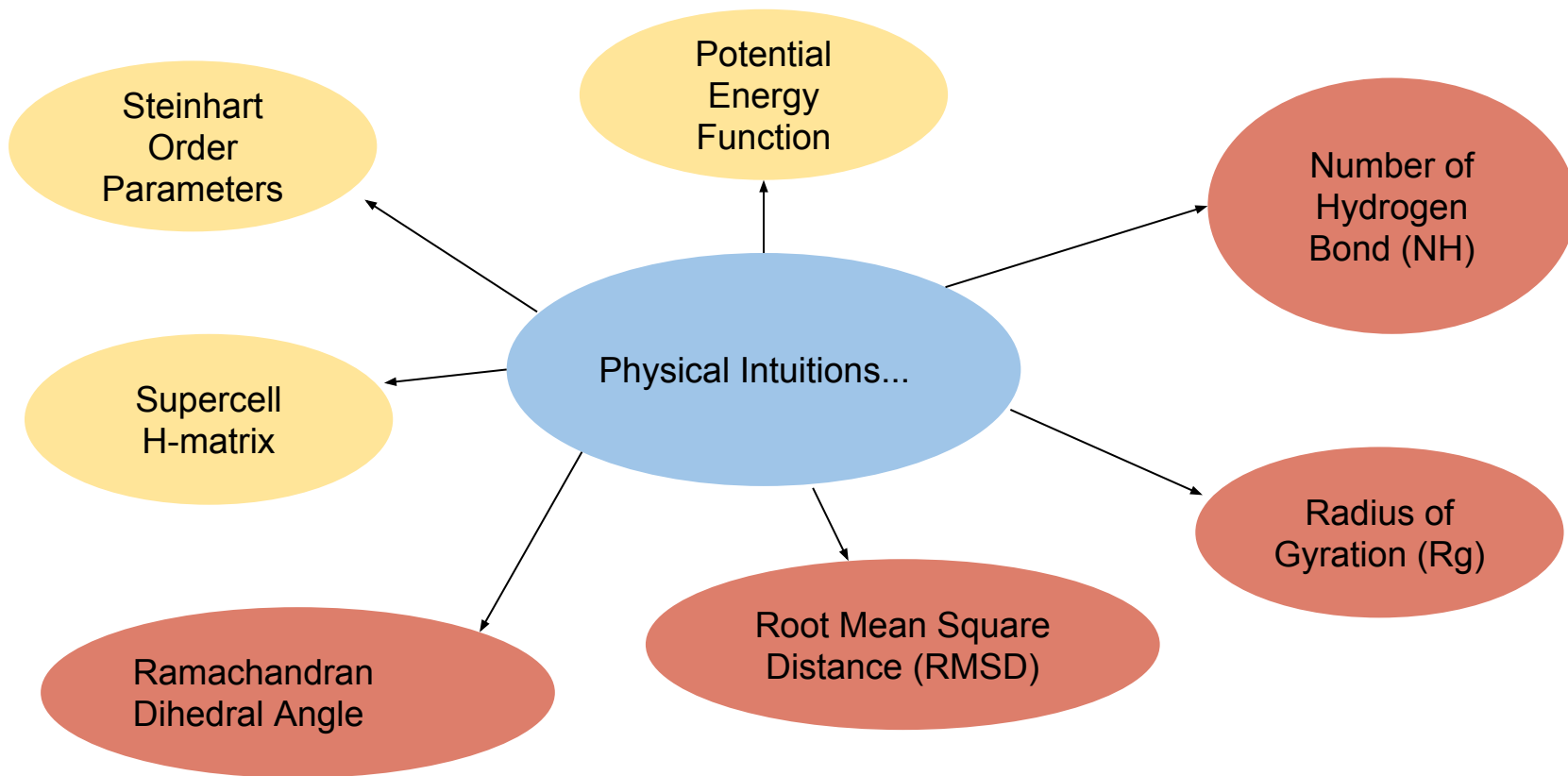
# Degeneracy and Sampling Efficiency: Case Study of Alanine Dipeptide

Correlation Function:

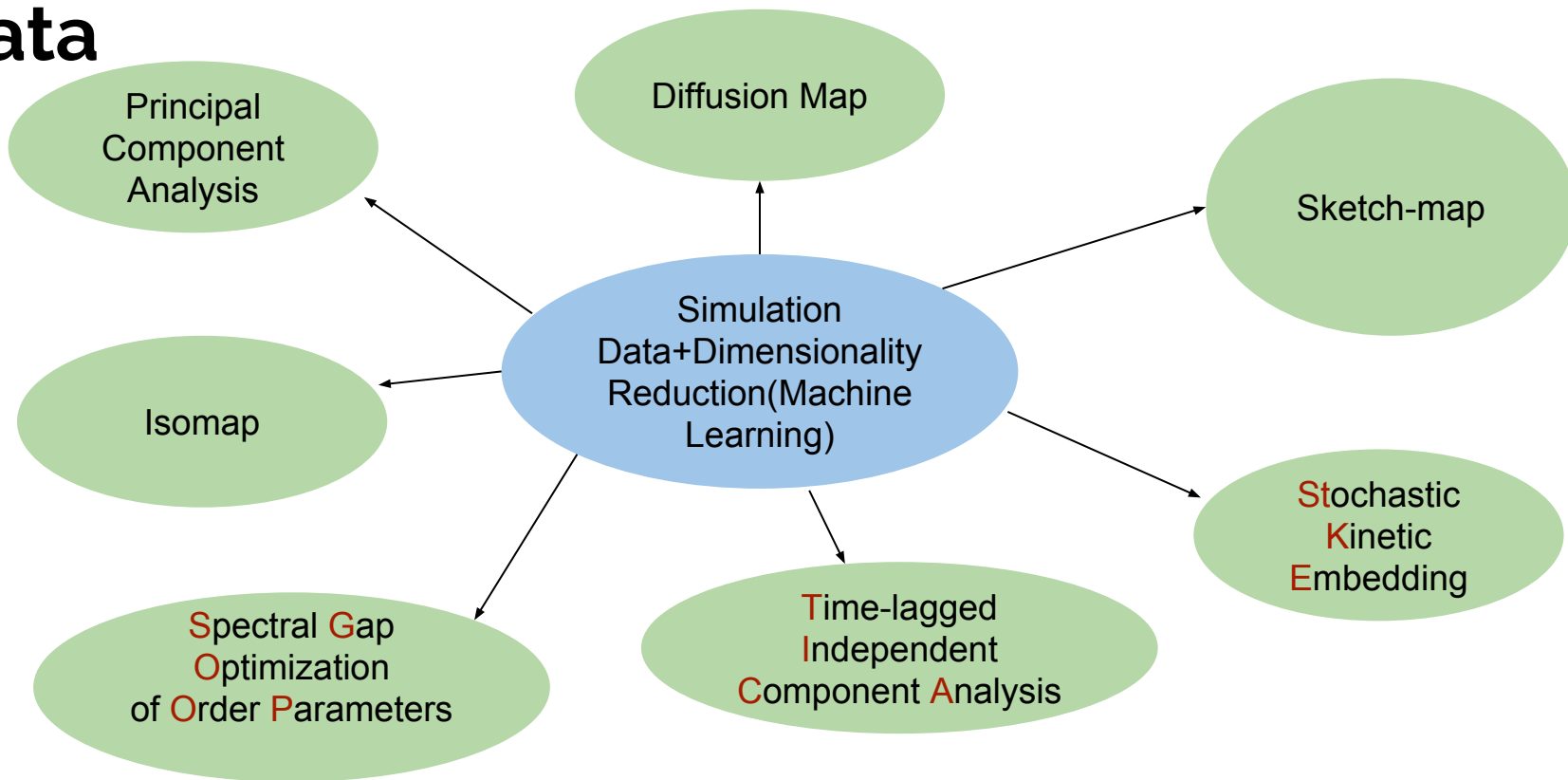
$$\frac{\langle \mathbb{1}_A(\mathbf{x}(0)) \mathbb{1}_A(\mathbf{x}(t)) \rangle}{\langle \mathbb{1}_A^2(\mathbf{x}(0)) \rangle}$$



# Generating CVs: Empirical Choices



# Generating CVs: Training from Simulation Data



# Eigenfunctions of Infinitesimal Generator

Assume Configurations are generated from a diffusion process:

$$\text{Friction} \longrightarrow \mu d\mathbf{x} = -\nabla_{\mathbf{x}} U dt + \sqrt{2\mu kT} dW \longleftarrow \text{Wiener Process (Brownian Motion)}$$

Assume  $f(\mathbf{x})$  is a smooth test function and an ensemble of trajectories starting from  $\mathbf{x}$  at time 0, we can evaluate the expectation of  $f(\mathbf{x})$  at time  $t$ , i.e.

$$g(\mathbf{x}, t) = \mathbb{E}(f(\mathbf{x}(t)) | \mathbf{x}(0) = \mathbf{x})$$

The evolution of  $g(\mathbf{x}, t)$  is described by backward Kolmogorov equation:

$$\mu \frac{\partial g}{\partial t} = -\nabla_{\mathbf{x}} U \cdot \nabla_{\mathbf{x}} g + kT \nabla_{\mathbf{x}}^2 g = \mathcal{L}g$$

Infinitesimal Generator



# Eigenfunctions of Infinitesimal Generator

The operator  $\mathcal{L}$  is elliptical with following eigenvalues and eigenfunctions:

$$\mathcal{L}\psi_i = -\lambda_i\psi_i \quad 0 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots$$

$g(\mathbf{x}, t)$  can be expressed as:

$$g(\mathbf{x}, t) = \mathbb{E}(f(\mathbf{x})) + \sum_{i \neq 0} C_i e^{-\lambda_i t} \psi_i(\mathbf{x})$$

It is easy to check in the long time limit

$$\lim_{t \uparrow \infty} g(\mathbf{x}, t) = \mathbb{E}(f(\mathbf{x}))$$

# Eigenfunctions of Infinitesimal Generator

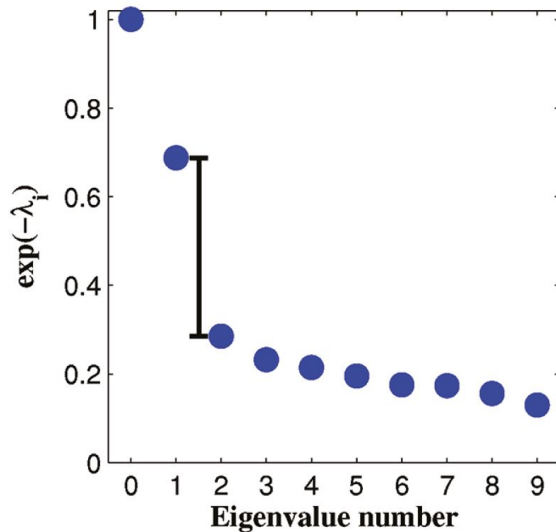
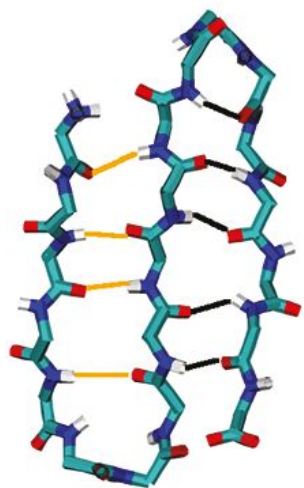
$$g(\mathbf{x}, t) = \mathbb{E}(f(\mathbf{x})) + \sum_{i \neq 0} C_i e^{-\lambda_i t} \psi_i(\mathbf{x})$$

What can we learn?

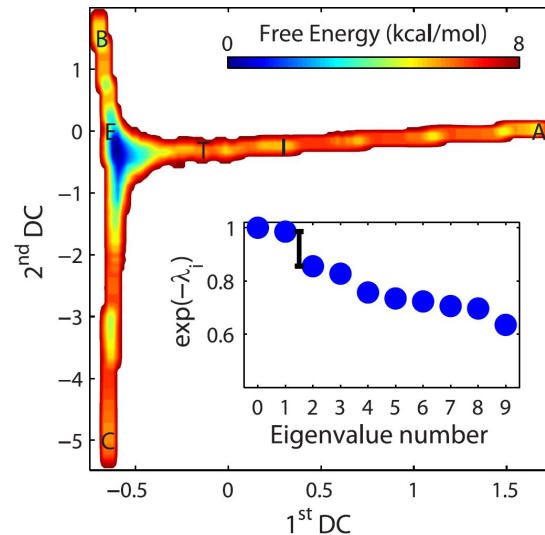
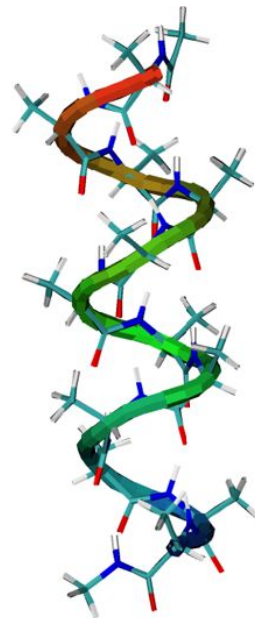
- $\lambda_i$  controls the equilibrium speed, smaller  $\lambda_i$  means slower equilibration.
- $\psi_i$  with small  $\lambda_i$  are natural “slow motions”.
- Assume there exists a spectral gap, i.e.  $\lambda_{i+1} \gg \lambda_i$ , the first  $i$   $\psi_i$  can be selected as CVs
- $\psi_i$  are usually highly nonlinear functions for real systems with complex interactions.

How to generate  $\psi_i$  from simulation data?

# Examples of Spectral Gap Existing in Chemical Systems



Beta3S



ALA12

# Diffusion Map

Assume configurations (samples) from MD simulation:  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ , we can define a Markov chain among samples with transition matrix  $M$ :

$$K_{ij} = \frac{\exp\{-|\mathbf{x}_i - \mathbf{x}_j|^2/2\sigma^2\}}{\sqrt{p(\mathbf{x}_i)p(\mathbf{x}_j)}} \quad M_{ij} = \frac{K_{ij}}{\sum_j K_{ij}}$$

where  $p(\mathbf{x}) = \sum_i \exp\{-|\mathbf{x} - \mathbf{x}_i|^2/2\sigma^2\}$  is the kernel density estimation of the Boltzmann distribution.

Values of  $\psi_i$  at each  $\mathbf{x}_i$  can be evaluated as the  $i$ 'th **right** eigenvector of  $M$ .

R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, *Multiscale Modeling and Simulation* **7**, 842 (2008).

M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, *J. Chem. Phys.* **134**, 124116 (2011).

# Time-lagged Independent Component Analysis

It has been proved that  $\mathbb{E}[\psi_i(\mathbf{x}(t))\psi_i(\mathbf{x}(t + \tau))] = e^{-\lambda_i\tau}$ . If  $\tilde{\psi}_i$  is an approximation of  $\psi_i$ ,  $\mathbb{E}[\tilde{\psi}_i(\mathbf{x}(t))\tilde{\psi}_i(\mathbf{x}(t + \tau))] \leq e^{-\lambda_i\tau}$ .

F. Noé and F. Nüske, *Multiscale Modeling & Simulation* **11**, 635 (2013).

G. Pálrez-Hernández, F. Paul, T. Giorgino, G. D. Fabritiis, and F. Noál, *J. Chem. Phys.* **139**, 015102 (2013).

M. M. Sultan and V. S. Pande, *J. Chem. Theory Comput.* 10.1021/acs.jctc.7b00182 (2017)

# Time-lagged Independent Component Analysis

It has been proved that  $\mathbb{E}[\psi_i(\mathbf{x}(t))\psi_i(\mathbf{x}(t + \tau))] = e^{-\lambda_i\tau}$ . If  $\tilde{\psi}_i$  is an approximation of  $\psi_i$ ,  $\mathbb{E}[\tilde{\psi}_i(\mathbf{x}(t))\tilde{\psi}_i(\mathbf{x}(t + \tau))] \leq e^{-\lambda_i\tau}$ .

$$\tilde{\psi}_i(\mathbf{x}) = \sum_k b_{ik} \chi_k(\mathbf{x})$$



Variational Principle

$$\mathbf{C}^\chi(\tau) \mathbf{b}_i = \varepsilon_i(\tau) \mathbf{b}_i, \text{ where } C_{ij}^\chi(\tau) = \mathbb{E}[\chi_i(\mathbf{x}(t))\chi_j(\mathbf{x}(t + \tau))]$$



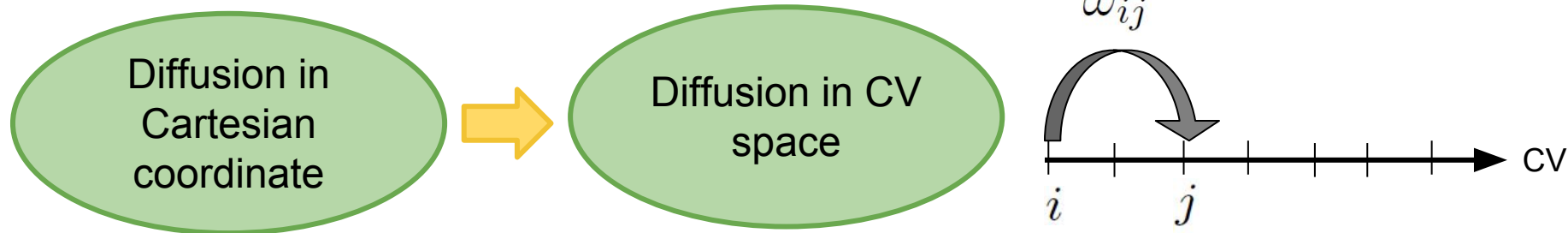
Selected Eigenvectors (CVs)

F. Noé and F. Nüske, *Multiscale Modeling & Simulation* **11**, 635 (2013).

G. Pálrez-Hernández, F. Paul, T. Giorgino, G. D. Fabritiis, and F. Noál, *J. Chem. Phys.* **139**, 015102 (2013).

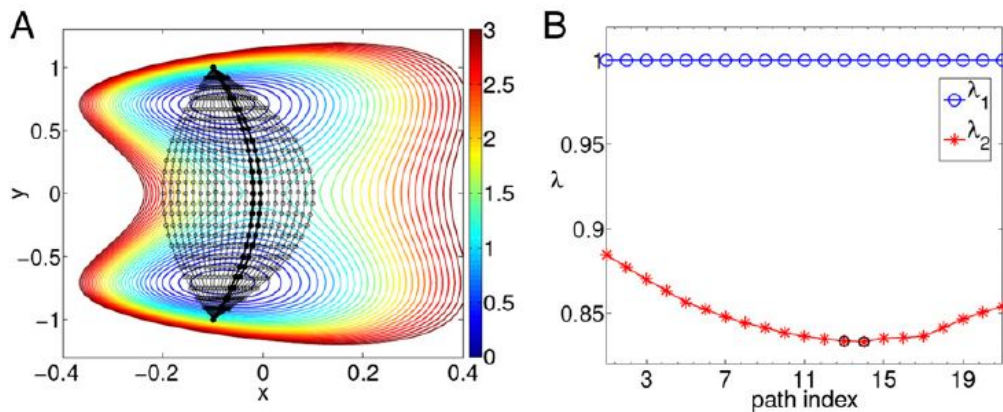
M. M. Sultan and V. S. Pande, *J. Chem. Theory Comput.* 10.1021/acs.jctc.7b00182 (2017)

# Spectral Gap Optimization of Order Parameters



$$\tilde{\psi}_i(\mathbf{x}) = \sum_k b_{ik} \chi_k(\mathbf{x})$$

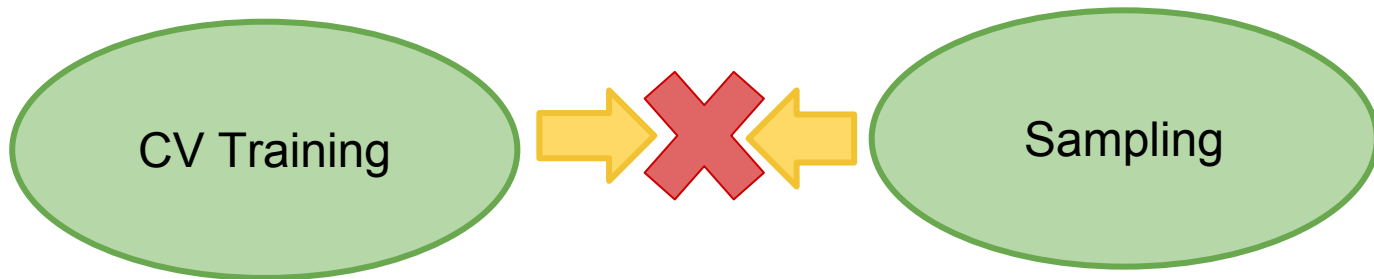
Optimize CV form so that  $\omega$  has maximum spectral gap.



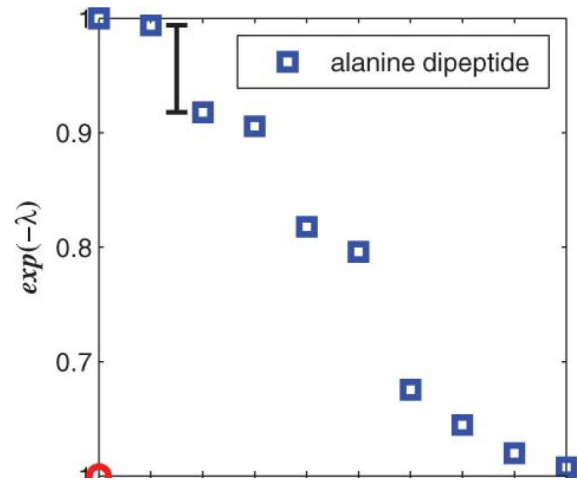


# Problems:

- Training requires complete sampling.

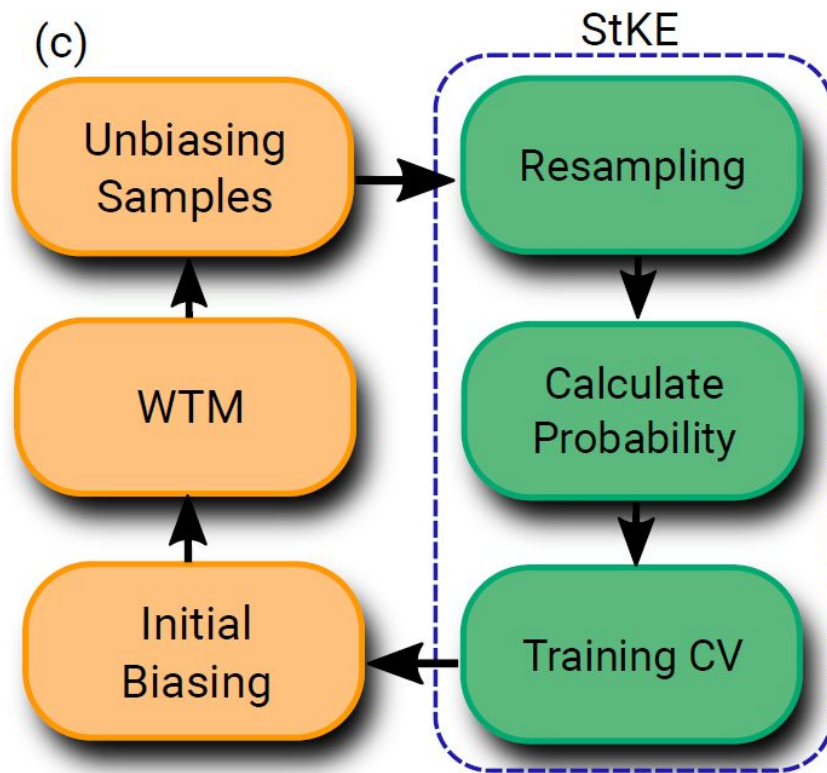
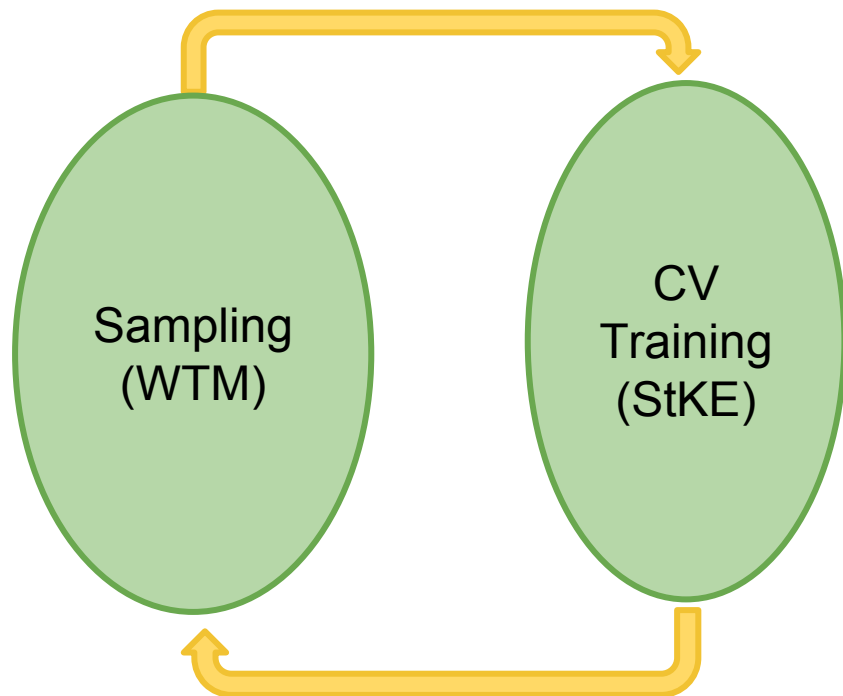


- Low Rank Approximation

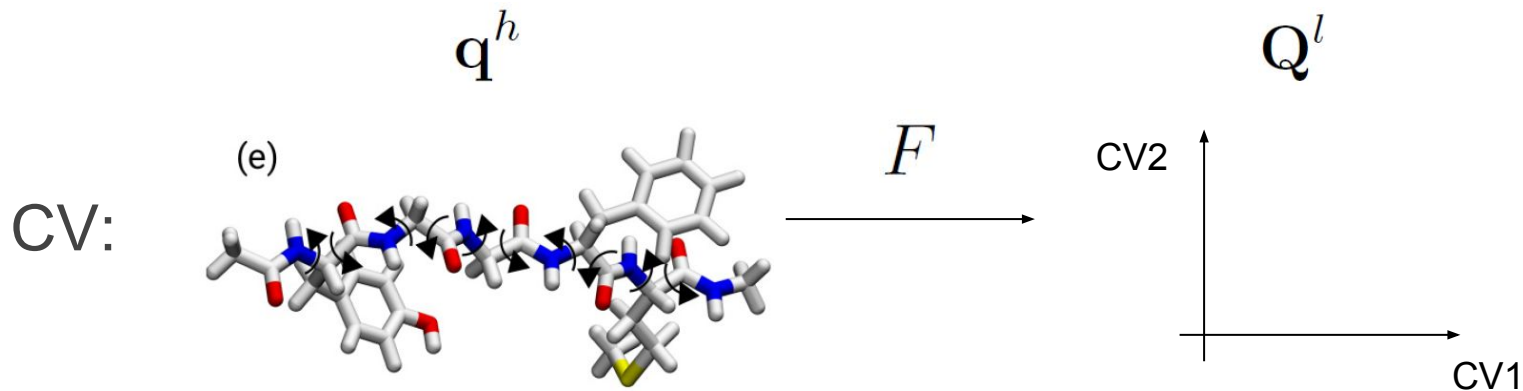


M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi,  
*J. Chem. Phys.* **134**, 124116 (2011).

# Active Enhanced Sampling



# Stochastic Kinetic Embedding



Sample:  $\{s_1^h, \dots, s_N^h\} \longrightarrow \{S_1^l, \dots, S_N^l\}$

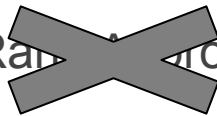
$M_{ij}^{high}$  (  $M_{ij}^{low}$  ) : Diffusion map transition matrix in  $q^h$  ( $Q^l$ ) space

# Stochastic Kinetic Embedding

Low Rank Approximation

# Stochastic Kinetic Embedding

Low Rank Approximation

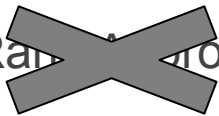


Kullback-Leibler divergence (t-SNE)

L. van der Maaten and G. Hinton, *J. Mach. Learn. Res.* **9**, 2579 (2008).

# Stochastic Kinetic Embedding

Low Rank Approximation



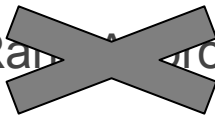
Kullback-Leibler divergence (t-SNE)

L. van der Maaten and G. Hinton, *J. Mach. Learn. Res.* **9**, 2579 (2008).

$$C = \sum_i \left( \sum_j M_{ij}^{high} \log \frac{M_{ij}^{high}}{M_{ij}^{low}} \right)$$

# Stochastic Kinetic Embedding


Low Rank Approximation




Kullback-Leibler divergence (t-SNE)

L. van der Maaten and G. Hinton, *J. Mach. Learn. Res.* **9**, 2579 (2008).

$$C = \sum_i \left( \sum_j M_{ij}^{high} \log \frac{M_{ij}^{high}}{M_{ij}^{low}} \right)$$


$$M_{ij}^{low}(\mathbf{S}_1^l, \dots, \mathbf{S}_N^l) = M_{ij}^{low}(F(\mathbf{s}_1^h; W), \dots, F(\mathbf{s}_N^h; W))$$

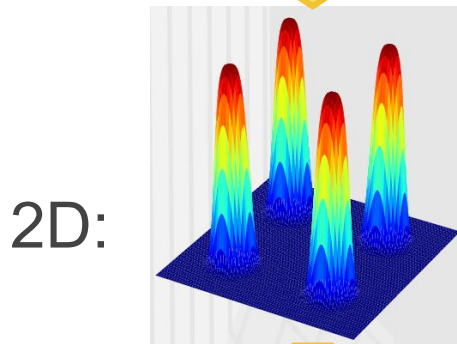
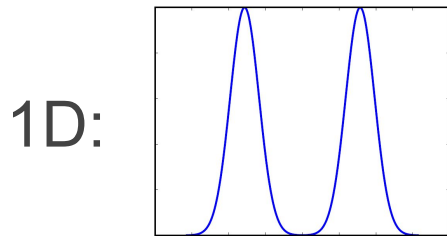
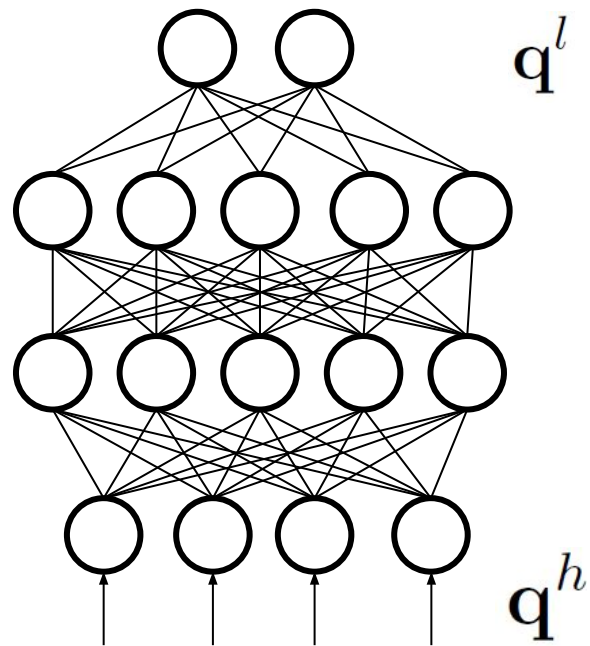


Parametrized  $\mathbf{S}^l = F(\mathbf{s}^h; W)$



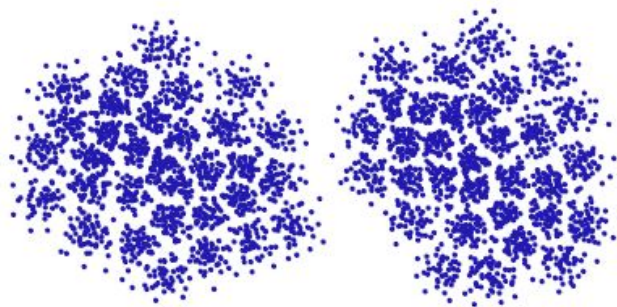
# Stochastic Kinetic Embedding

$F$  : Multilayer Perceptron

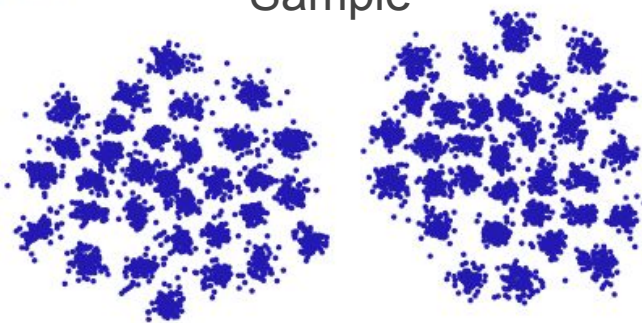


6D

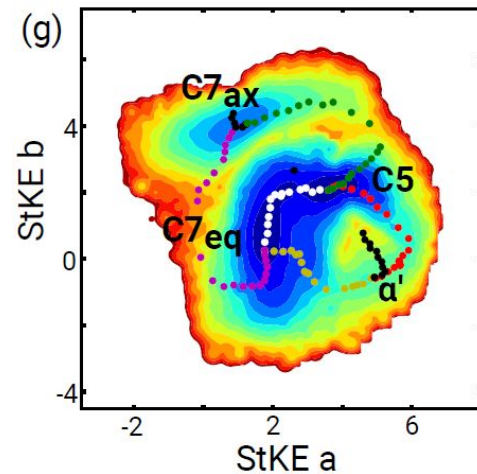
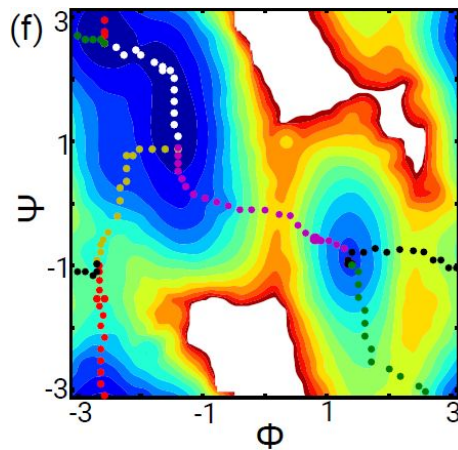
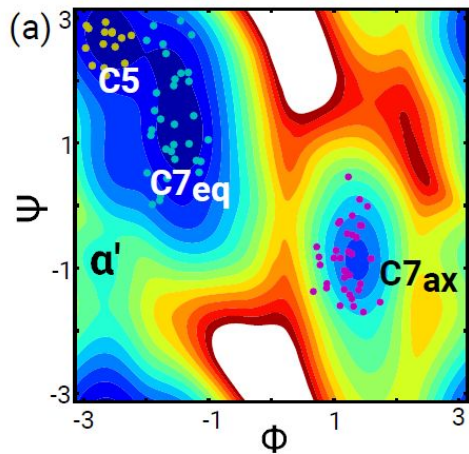
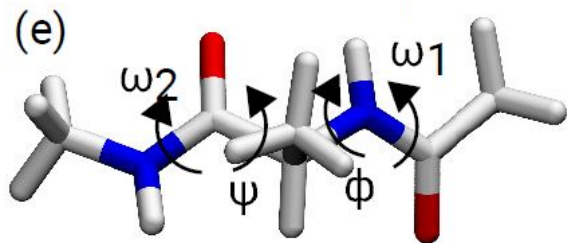
(a) Resample



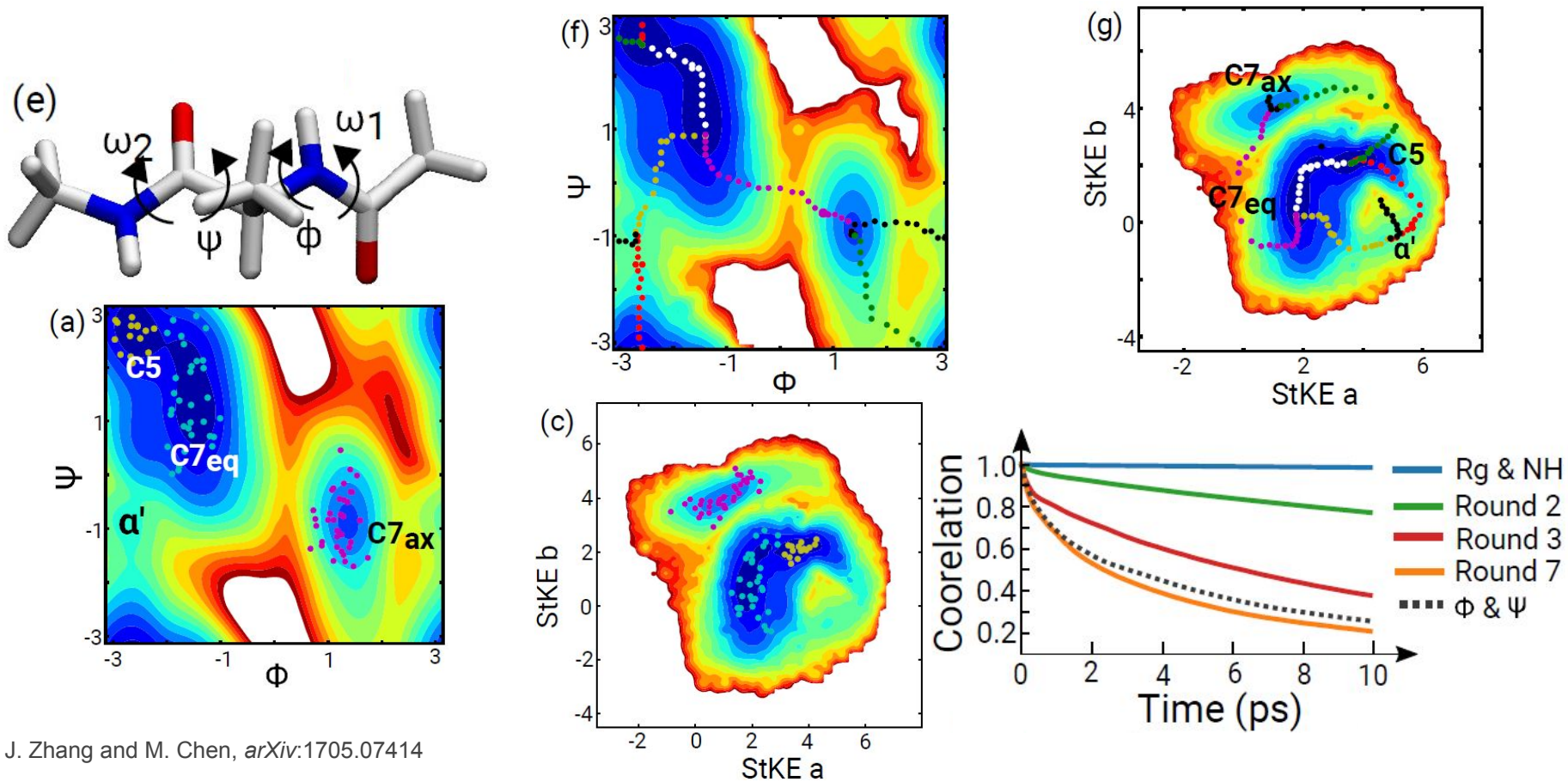
(b) Sample



# AES Example: Alanine Dipeptide

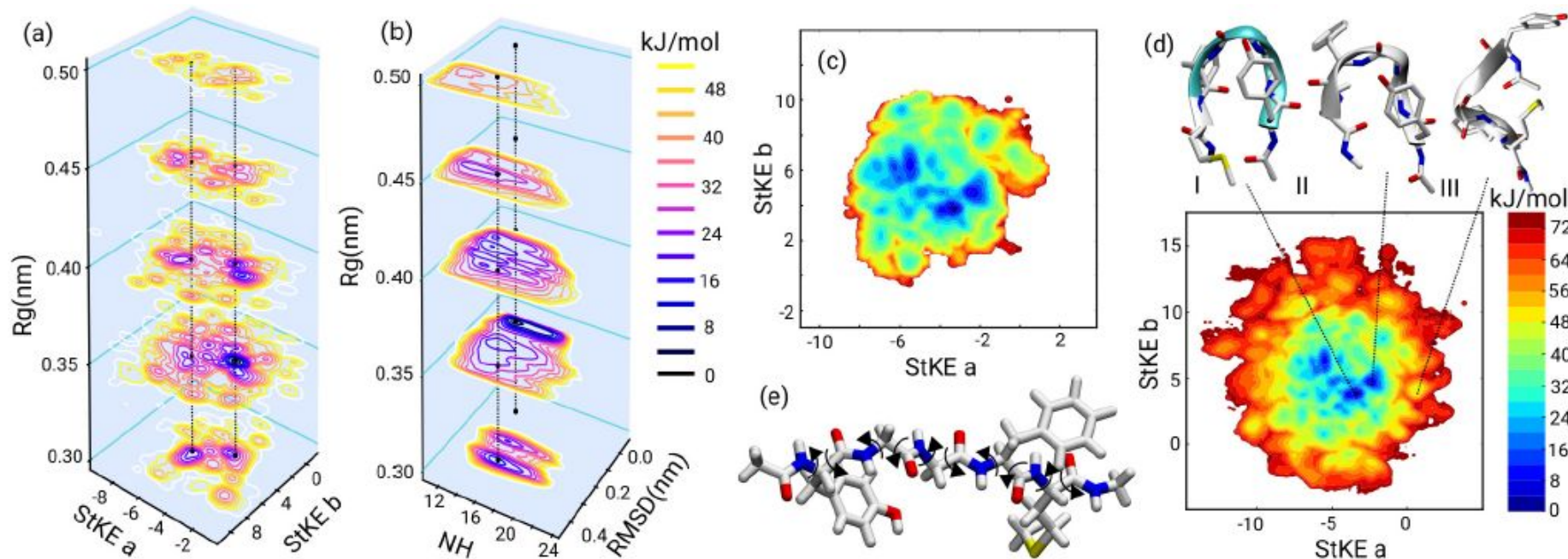


# AES Example: Alanine Dipeptide



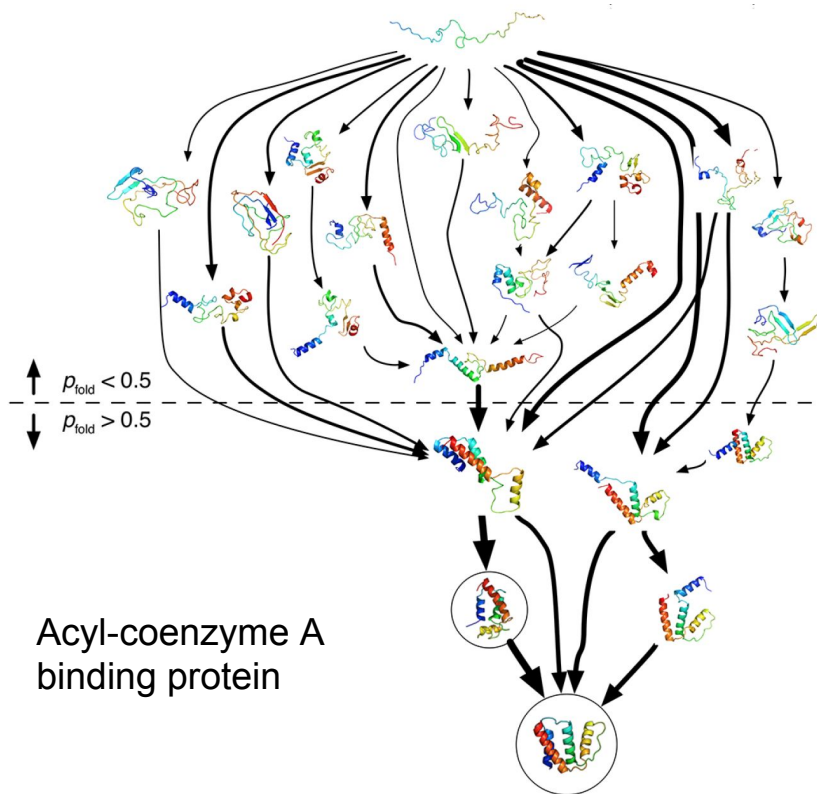


# Met-enkephalin: A Non-trivial Test



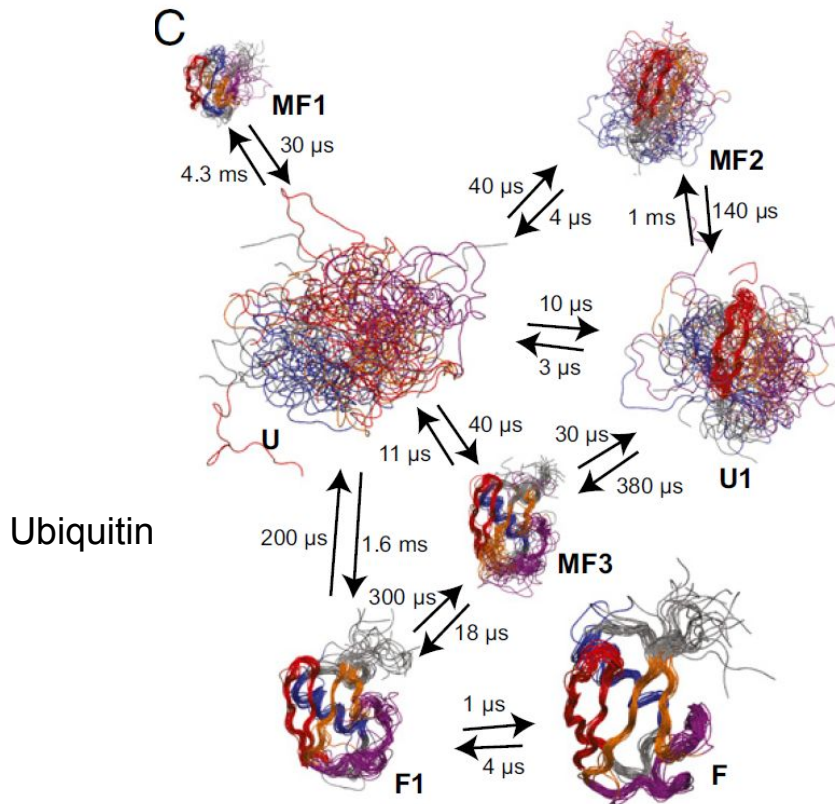
Tyr-Gly-Gly-Phe-Met

# Sampling Partially Folded/Unfolded States



Acyl-coenzyme A  
binding protein

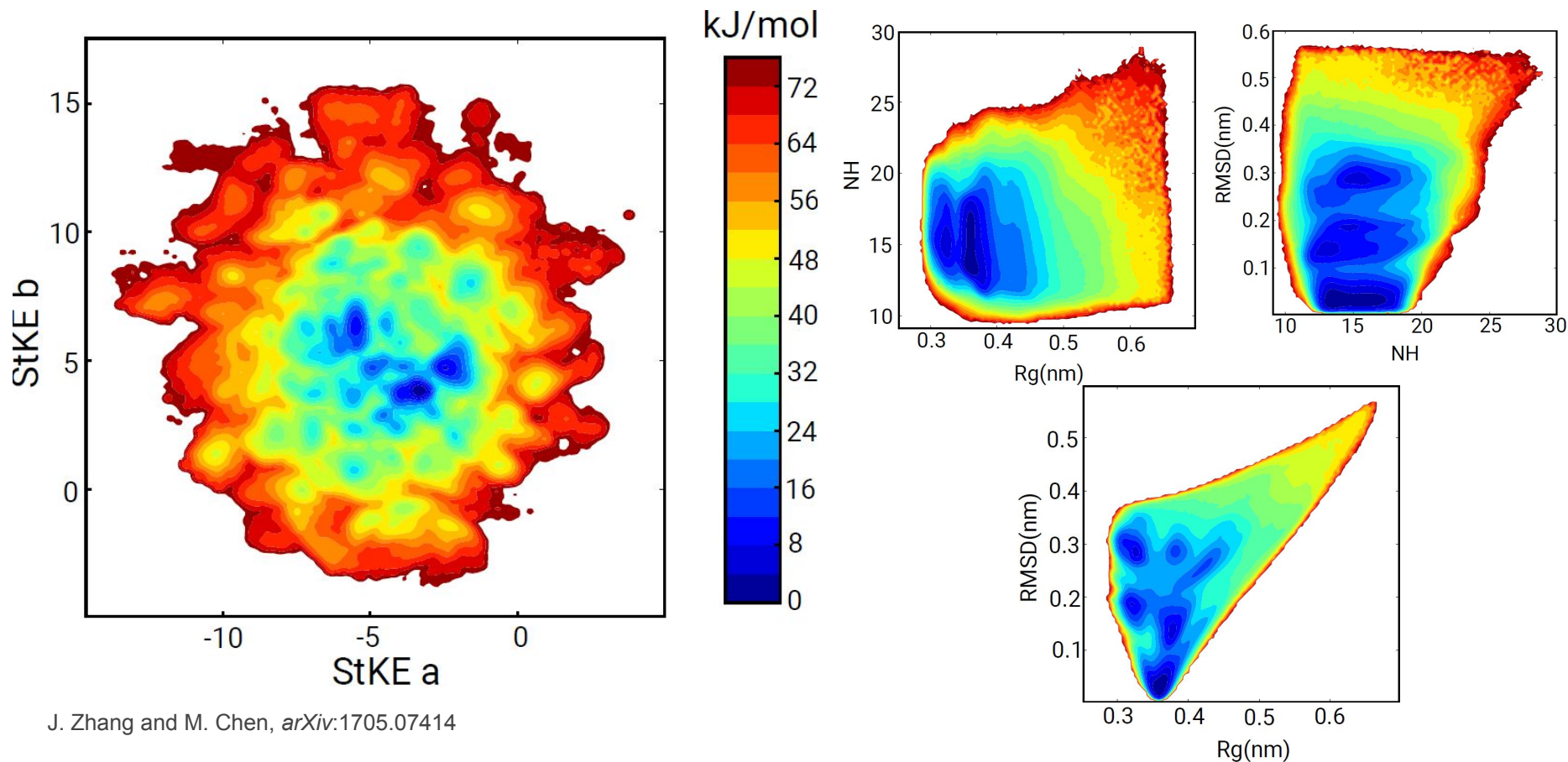
V. A. Voelz, et al., *J. Am. Chem. Soc.* **134**, 12565 (2012).



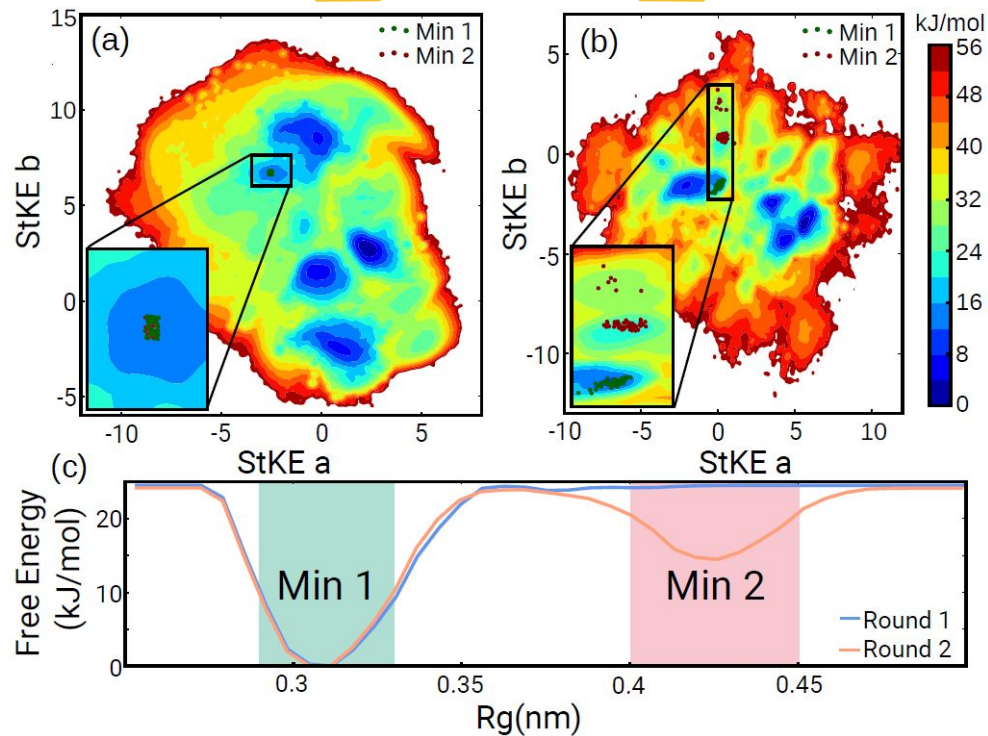
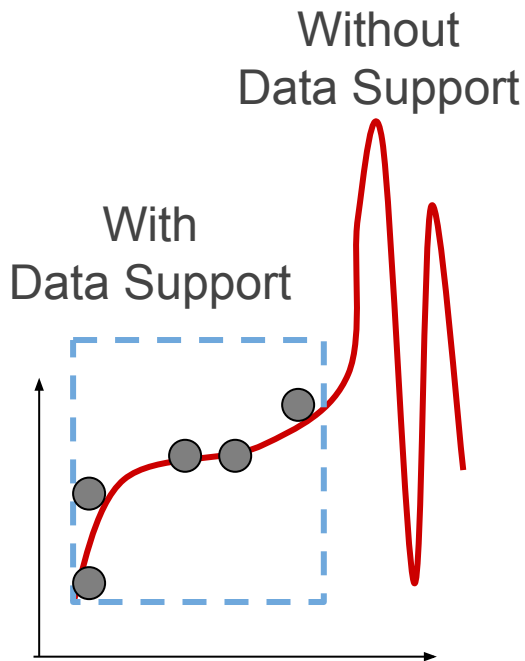
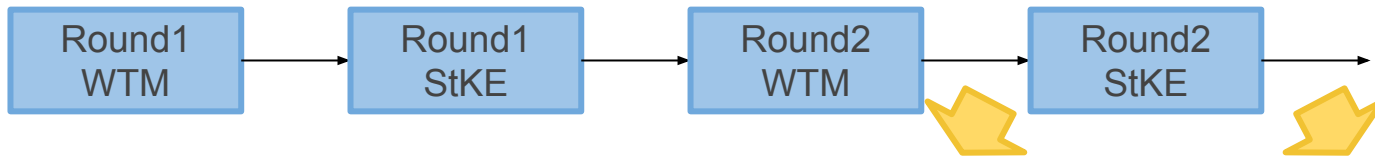
Ubiquitin

S. Piana, K. Lindorff-Larsen, and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5915 (2013).

# Met-enkephalin: A Non-trivial Test

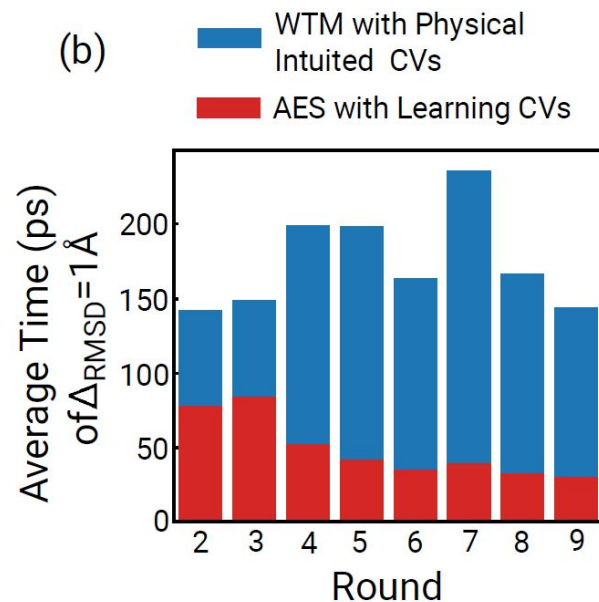
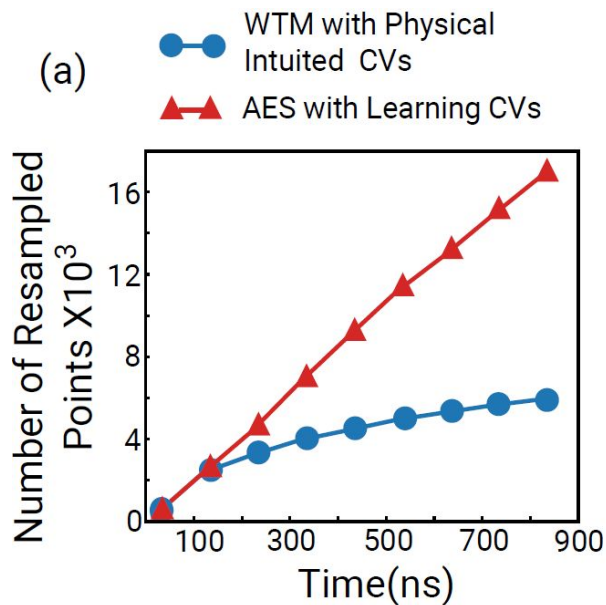
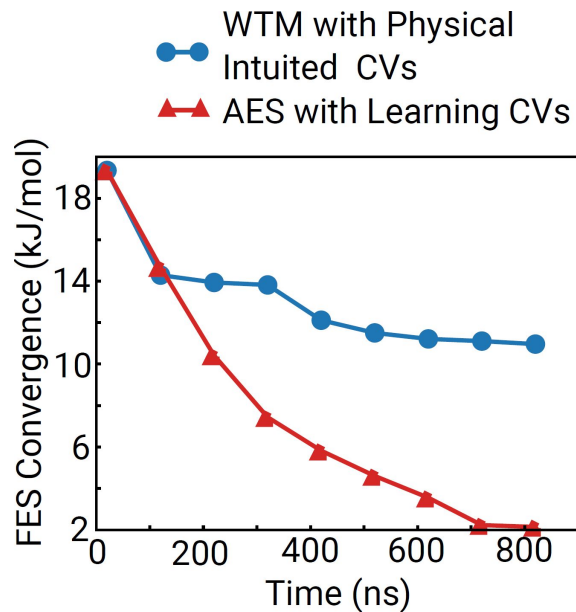


# Extrapolating with Physical CV





# AES: Sample Completeness and Sample Quality



# Conclusion

- CV selection is critical to enhanced sample efficiency.
- CV selection is a dimensionality reduction problem.
- Kinetic information is important for CV selection.
- “Catch 22” problem between sampling and CV training can be solved in a interactive way in AES.
- AES is able to remove degeneracy on the fly.
- AES can achieve better sampling compared to WTM with physical intuited CVs

# Acknowledgment

Prof. Mark E. Tuckerman

Prof. John Zhang

Prof. Yingkai Zhang

Stuffs in NYU Shanghai

Prof. Eran Rabani (Current Advisor)

Jing Zhang (Collaborator)