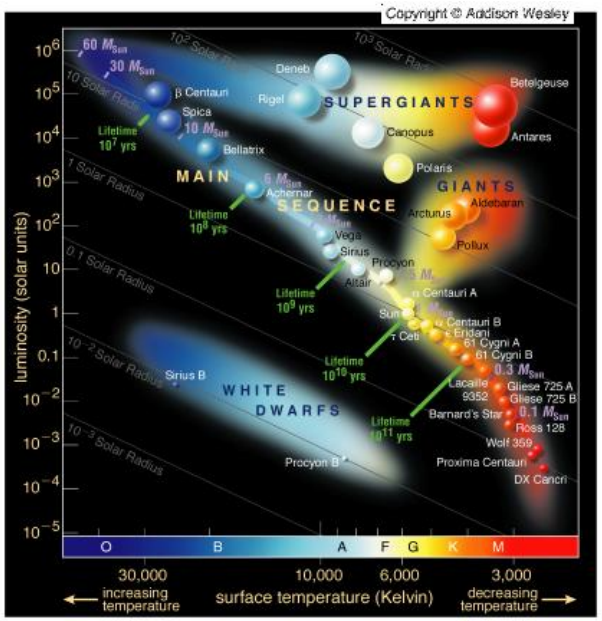
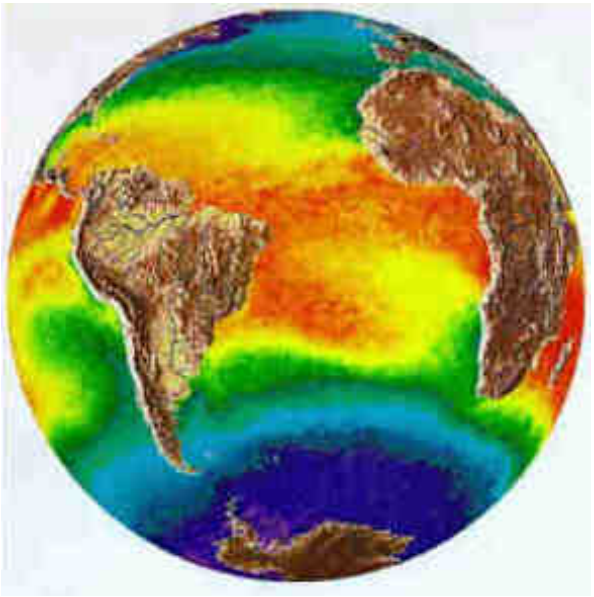
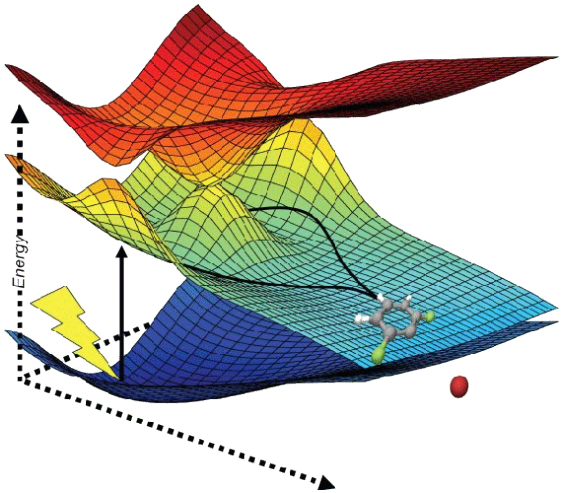
Brief Review: Dimensionality Reduction via Sketch Maps and Diffusion Maps to Analyze Molecular Dynamics Data

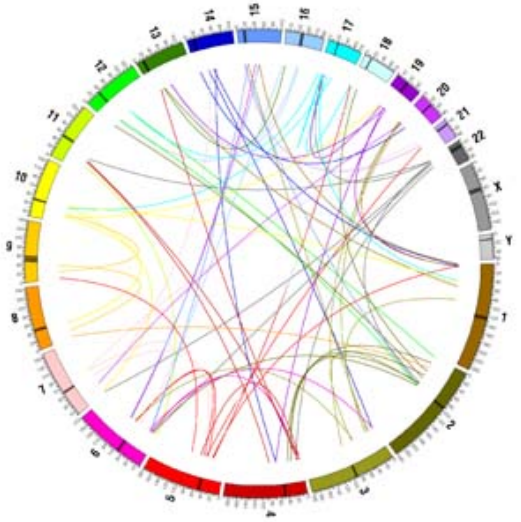
Bryan R. Goldsmith
10/15/2015

Dimensionality reduction amounts to finding low-dimensional structure in high-dimensional data





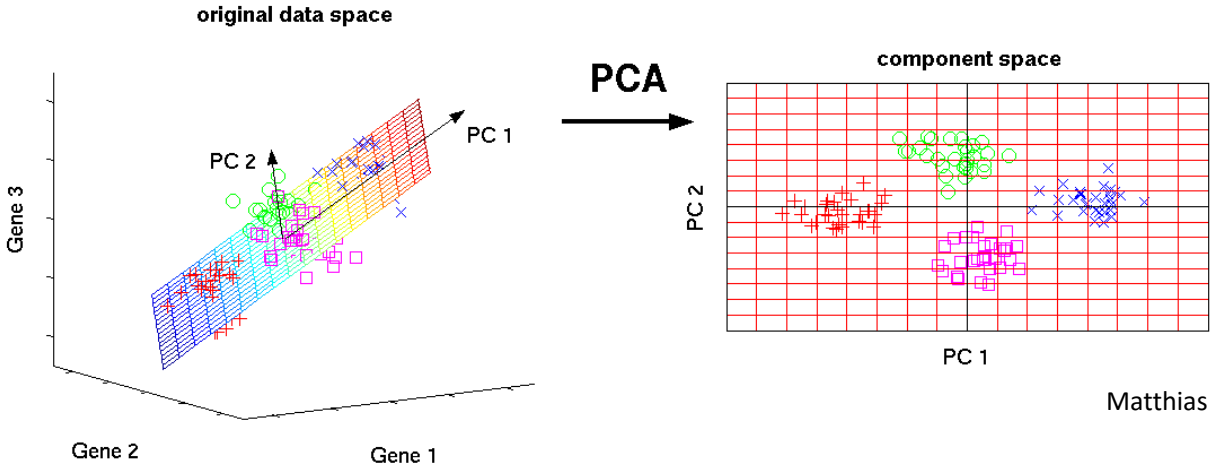




Common dimensionality reduction methods

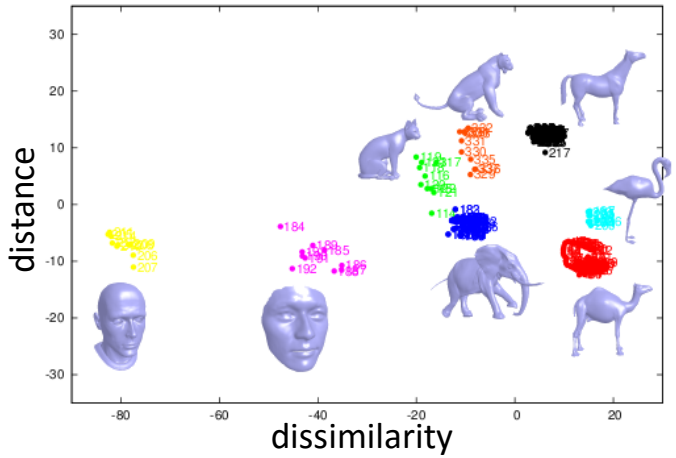
Principal component analysis (PCA)

Finds a low-dimensional embedding of the data that best preserves their variance as measured in high-dimensional input space



Matthias Scholz, Ph.D. thesis

Multidimensional scaling (MDS)



Classical MDS finds an embedding that preserves the inter-point distances

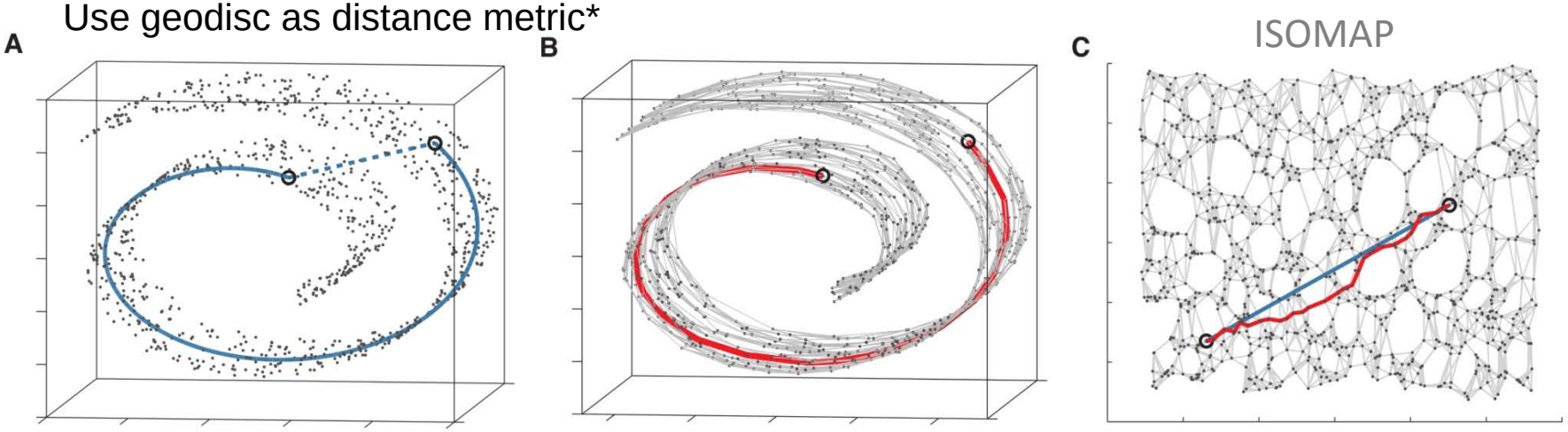
Equivalent to PCA when using Euclidean distances

PCA and MDS cannot detect nonlinear structures

Low-free-energy part of phase space has a complex structure
- Nonuniform dimensionality, that it is nonlinear, that is nonuniformly sampled.

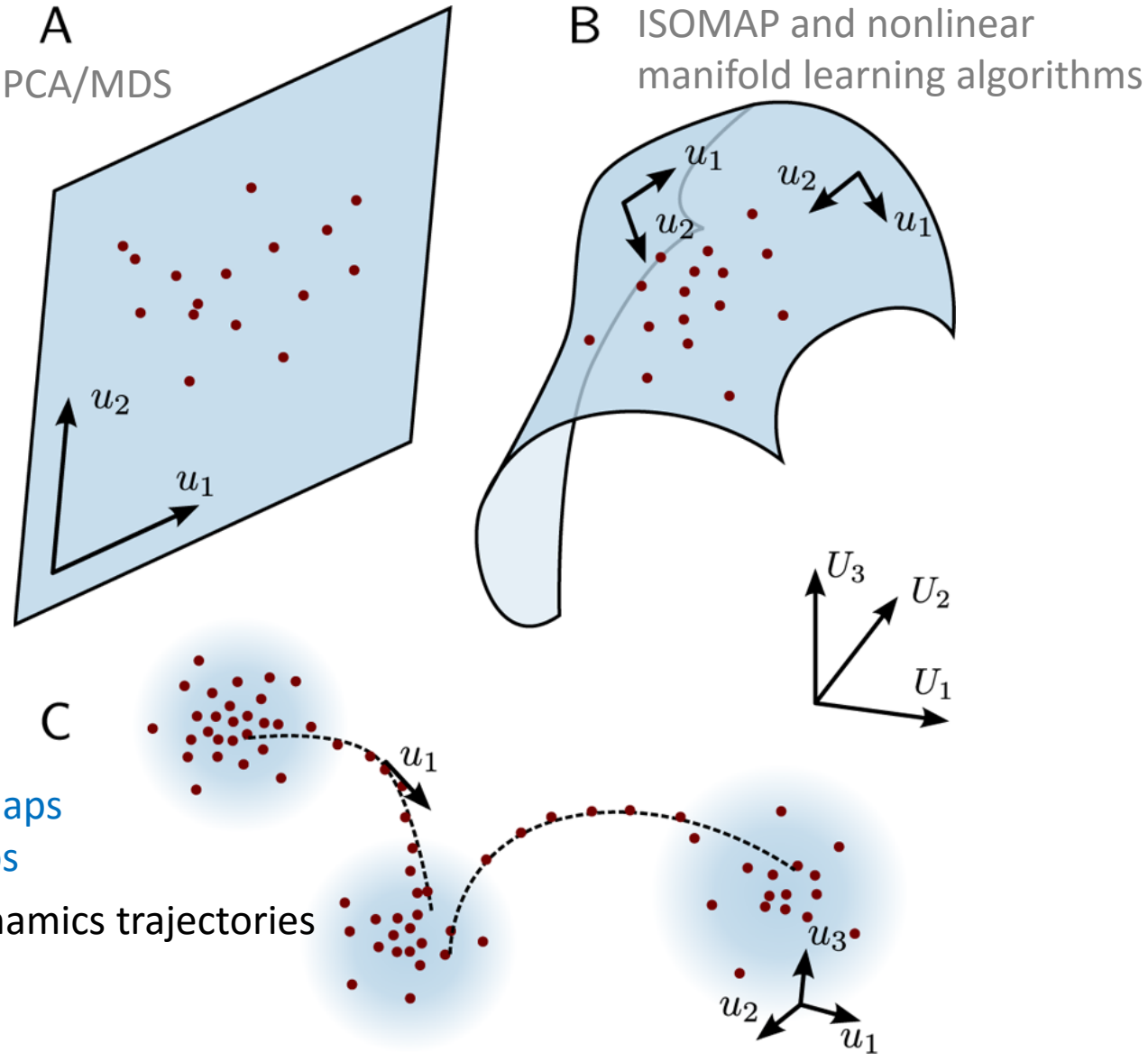
A Global Geometric Framework for Nonlinear Dimensionality Reduction
J. B. Tenenbaum, V. de Silva, J. C. Langford, Science, 290, 22 (2000). ~8600 citations

Use geodisc as distance metric*



*Assumes that the low-dimensional manifold is uniformly sampled and there are no “holes.”
This assumption is often violated in molecular dynamics simulations

Configurations of data that can be visualized using dimensionality reduction algorithms



Diffusion maps

Sketch maps

Molecular dynamics trajectories

Michele Ceriotti J. Chem. Theory Comput. 2013, 9, 1521−1532

Diffusion Maps

Extracts low-dimensionality structure of long *unbiased* molecular dynamics trajectories[1]

high friction, the Fokker–Planck equation governs the temporal evolution of the probability distribution $p(\boldsymbol{x}, t)$ at any configuration $\boldsymbol{x} \in \mathbb{R}^{3N}$ of the system,

$$\frac{\partial p}{\partial t} = -\sum_{i}^{3N} \frac{\partial}{\partial x_i}\left(\frac{1}{\beta}\frac{\partial}{\partial x_i} + \frac{\partial E}{\partial x_i}\right) p = -\mathbf{H}_{\mathrm{FP}}\, p,$$

General solution

$$p(\boldsymbol{x}, t) = \phi_0(\boldsymbol{x}) + \sum_{i=1}^{\infty} c_i \phi_i(\boldsymbol{x}) e^{-\lambda_i t},$$

For systems with one (or a few) slow process(es) dominating the dynamics (such as the crossing of a free-energy barrier), the eigenspectrum will present a gap; i.e., $\lambda_{k+1} \gg \lambda_k$ for some k, and the evolution of the probability distribution toward equilibrium may be approximated as the first k terms of the general solution,

$$p(\boldsymbol{x}, t) = \phi_0(\boldsymbol{x}) + \sum_{i=1}^{k} c_i \phi_i(\boldsymbol{x}) e^{-\lambda_i t},$$

Diffusion Map Method:

1) Define a weighted graph on the simulation data $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\varepsilon_i \varepsilon_j}\right),$

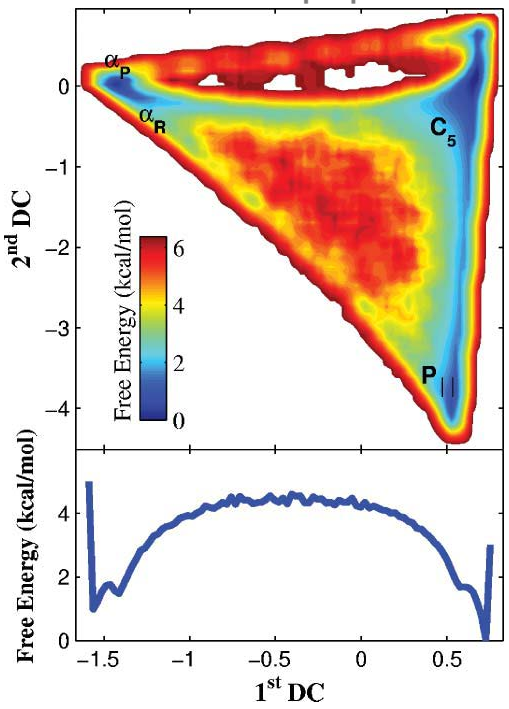
2) Determine the first few eigenvalues and eigenvectors via a random walk on the graph*

* This effectively approximates the eigenfunctions of the backward Fokker–Planck operator

[1] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker
PNAS, 102, 21, 2005, 7426 cited ~660

Example: Determination of reaction coordinates via locally scaled diffusion map

alanine dipeptide



The first diffusion coordinate is a good reaction coordinate for diffusive barriers, but lacks simple physical interpretation

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

A concern with diffusion maps

Not readily amenable to systems that require enhanced sampling

Currently the only means to assign physical meaning to diffusion map order parameters is to correlate them with combinations of physical variables[1]

- This could be accelerated with using likelihood maximization approaches[2]

[1] A. L. Fergusona, A. Z. Panagiotopoulosa, P. G. Debenedetti, and Ioannis G. Kevrekidis, *PNAS* 107, 31, 13597 (2010)
[2] B. Peters and B. L. Trout, J. Chem. Phys. **125**, 054108 (2006)

Sketch Map: obtaining a low-dimensional representation of phase space explored during enhanced molecular dynamics

attempts to uncover simple collective behavior in the conformation spaces of molecules in an automated fashion

Demonstrating the transferability and the descriptive power of sketch-map
M. Ceriotti, G. A. Tribello and M. Parrinello, J. Chem. Theory Comput. 9 1521 (2013)
Sketch-map is used to analyze the results of long parallel tempereing simulations of Lennard Jones clusters.

Locating binding poses in protein-ligand systems using reconnaissance metadynamics
P. Söderhjelm, G. A. Tribello, and M. Parrinello, Proc. Acad. Natl. Sci. U.S.A. 109 5170 (2012)
Sketch-map is used to analyze the results from a reconnaissance metadynamics simulation of benzamidine binding to a trypsin protein

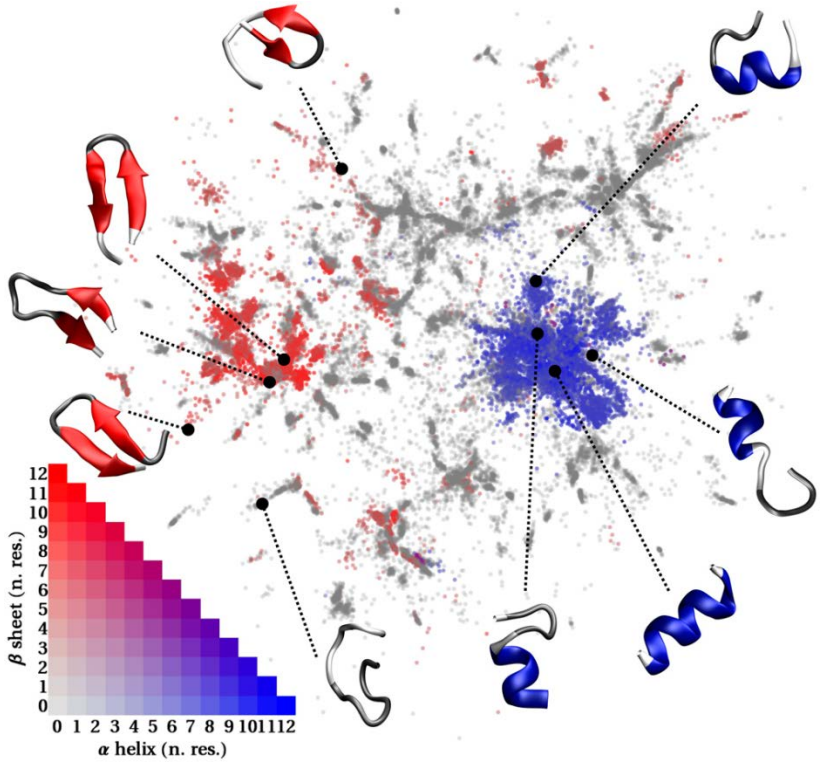
Using sketch-map coordinates to analyze and bias molecular dynamics simulations
G. A. Tribello, M. Ceriotti and M. Parrinello, Proc. Acad. Natl. Sci. U.S.A. 109 5196 (2012)
A paper showing how sketch-map coordinates can be used to bias molecular molecular dynamics simulations.

Simplfying the representation of complex free energy landscapes using sketch-map
M. Ceriotti, G. A. Tribello and M. Parrinello, Proc. Acad. Natl. Sci. U.S.A. 108 13023 (2011)
The original paper on the methodology.

The sketch map idea

Map of Germany sketched from memory.
Shows connections between important neighbors while distorting global layout

Sketch map scheme of Ceriotti *et al.* in mapping the conformation space of peptides



The social network (of protein conformations)
J. D. Chodera and V. S. Pande, *PNAS,*108, 12969 (2011)

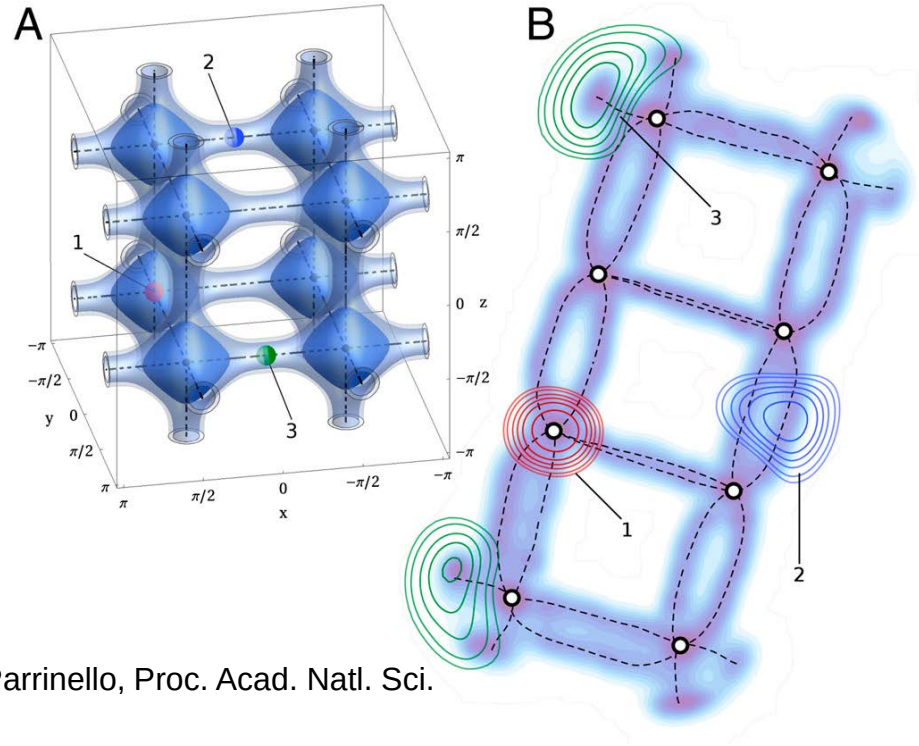
Sketch Map Goals:

Analyze data from accelerated sampling simulations on rough free-energy surfaces

Produce a mapping of phase space that can serve as a set of CVs for calculations that extract quantitative free energies

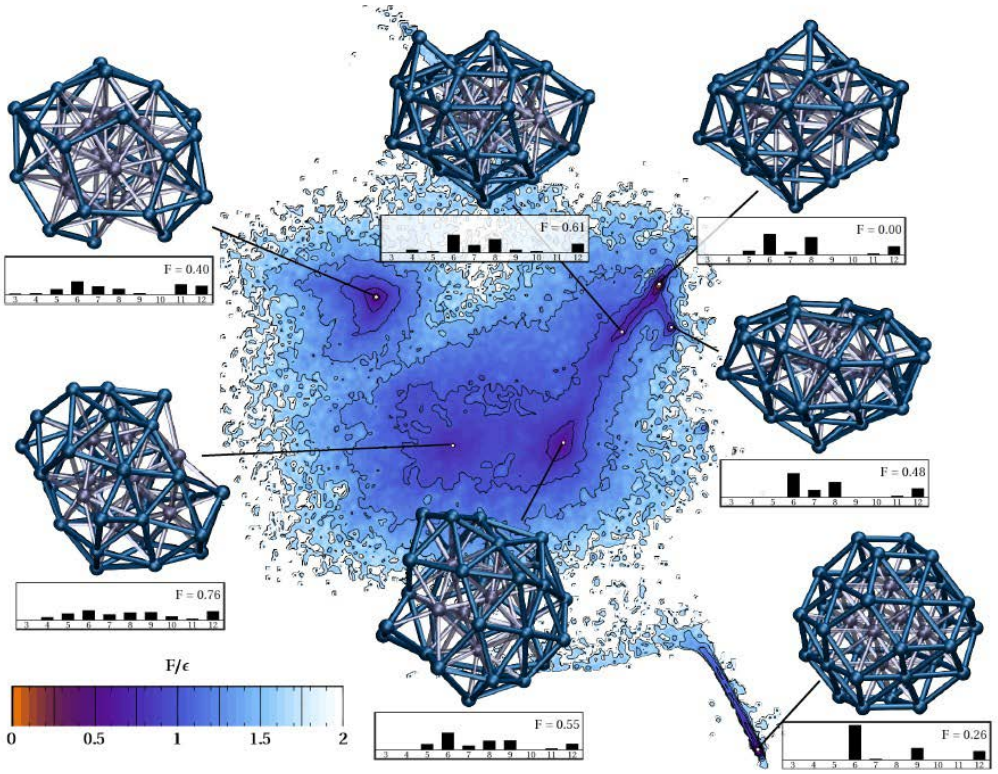
A projection that preserves the local connectivity relationship between conformational basins can be constructed

A complex free-energy surface that is periodic in three directions (A) and its sketch-map projection (B).

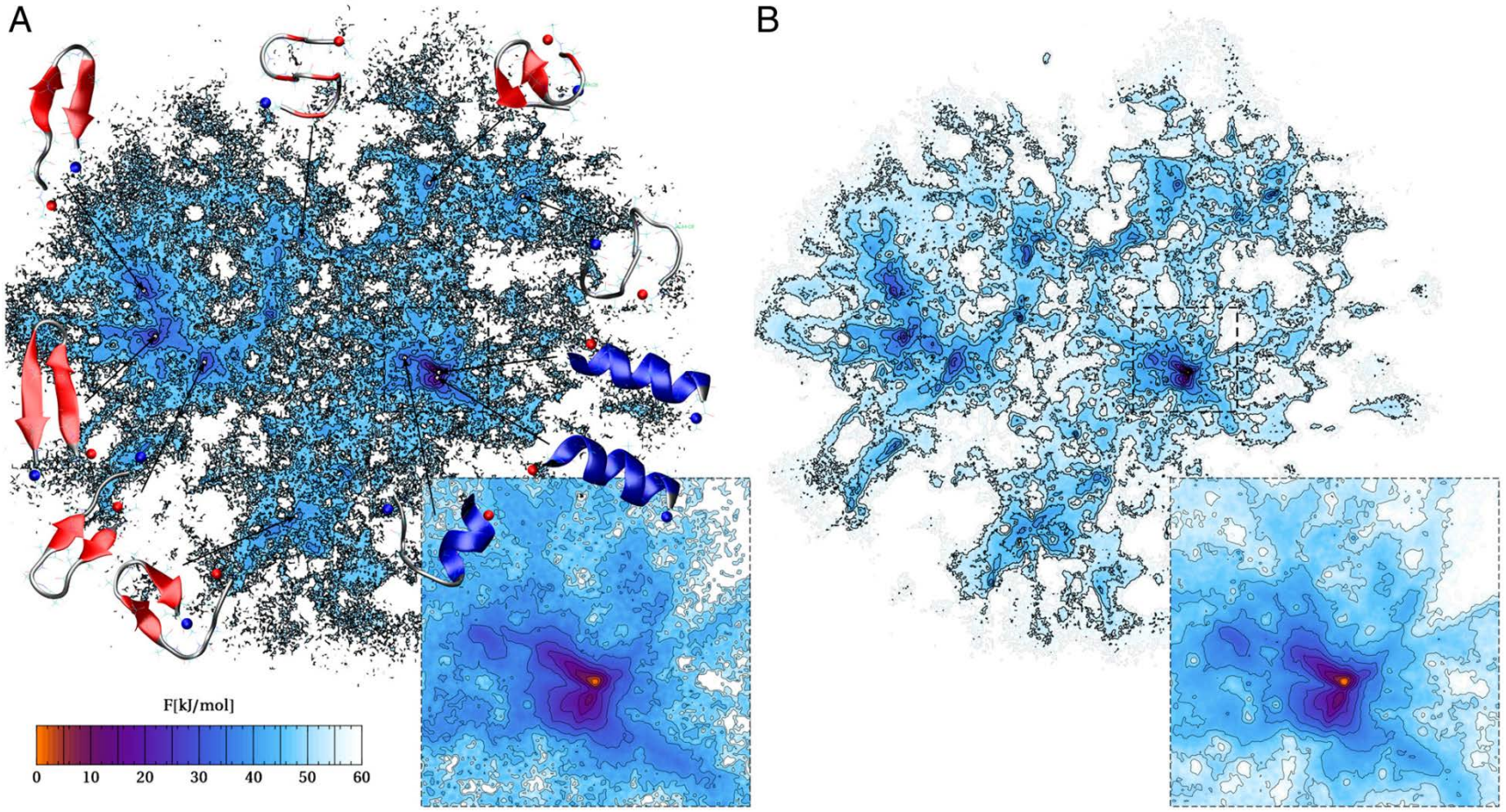


G. A. Tribello, M. Ceriotti and M. Parrinello, Proc. Acad. Natl. Sci. U.S.A. 109 5196 (2012)

Sketch Map for analyzing molecular dynamics trajectories



Sketch map for accelerating phase space exploration



The free-energy landscape for ala12 in implicit solvent calculated from parallel tempering (A) and field-overlapmetadynamics (B) simulations.

Note, Diffusion maps can also be used to accelerate dynamics: J. Phys. Chem. B 2013, 117, 12769−12776

Diffusion maps and Sketch maps are both promising approaches for dimensionality reduction of molecular dynamics data to find underlying structure and visualize data.

Diffusion map:

Con

Diffusion coordinates lack physical interpretation

Can be time consuming to construct

Only applied to simple systems so far that don't require accelerated sampling

Only works on Boltzmann-distributed data so far

Pro

Diffusion coordinates can be used as reaction coordinates, unlike sketch map coordinates (have dynamical information)

Could extract order parameters or physical reaction coordinates from diffusion coordinates

Can be used to accelerate dynamics

Sketch map:

Con

No unique mapping

No axis on figures

Lack physical information on transition between basins and neglects dynamics

Pro

Works on complicated energy landscapes

Works on accelerated MD data

Simple postprocessing of data to visualize structure in complex data

Can be used to accelerate dynamics