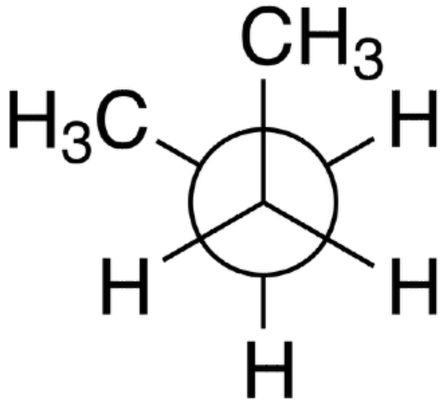
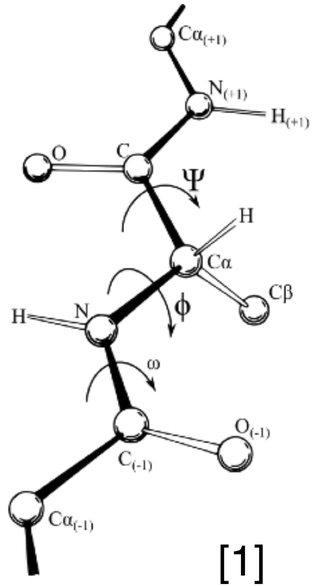
Dimensionality reduction

- Biomolecular simulations have 3N dimensions, but in practice, functionally relevant macromolecular motion can be described with a lot fewer.
- Cartesian coordinates (x, y, and z) of each atom are the most common description but molecular systems can be described by other coordinate systems
- You have probably already seen
  - the free energy of butane as a function of the dihedral angle
  - the Ramachandran diagram, where protein backbones are described by the ϕ and ψ angles



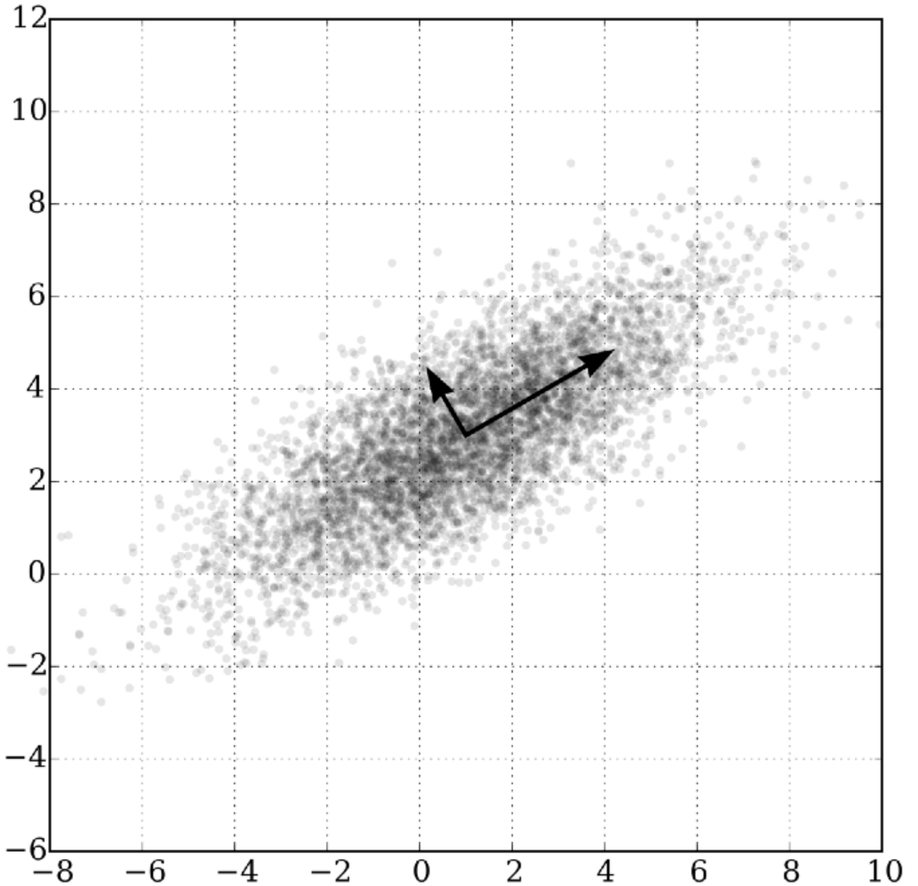


[1]

Principal component analysis (PCA)

- An *automated* way to do dimensionality reduction
- A linear transformation of coordinates in decreasing order of variance
  - First principal component has the largest variance
  - Second principal component has second largest variance
  - And so forth
- Dimensions can be reduced by keeping the highest-variance dimensions

- See https://en.wikipedia.org/wiki/Principal_component_analysis



[2]