CS 279 Assignment 2 Kickstart

10/16/23

Software Setup

- PyMOL 2.5
- Python 3.9
- CS 279 Conda environment:
  - Making a conda environment: `conda create -n cs279 python=3.9`
  - Activating a conda environment: `conda activate cs279`

Force Fields and Free Energy Calculation

$$U = \sum_{\text{bonds}} k_b \left(b - b_0\right)^2$$ **Bond lengths ("Stretch")**

$$+ \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2$$ **Bond angles ("Bend")**

$$+ \sum_{\text{torsions}} \sum_{n} k_{\phi,n} \left[1 + \cos\left(n\phi - \phi_n\right)\right]$$ **Torsional/dihedral angles**

Bonded terms

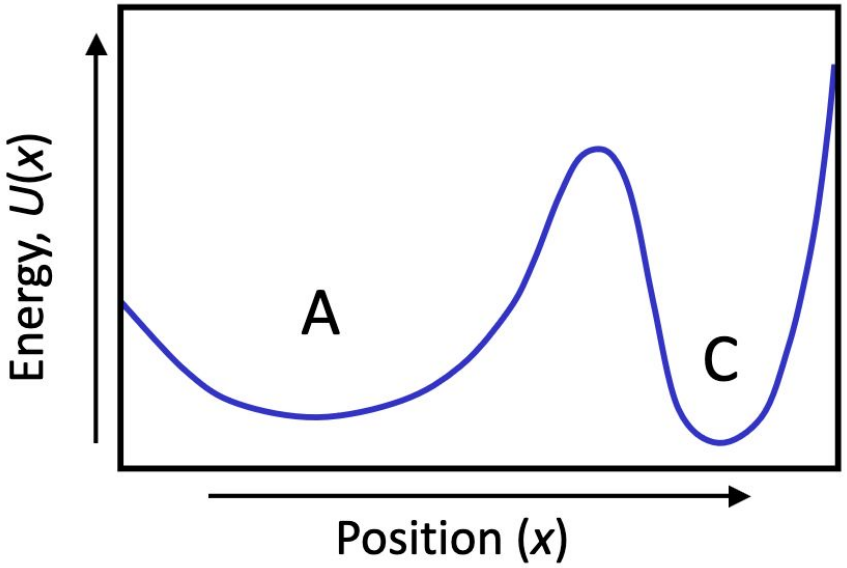
$$+ \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}}$$ **Electrostatics**

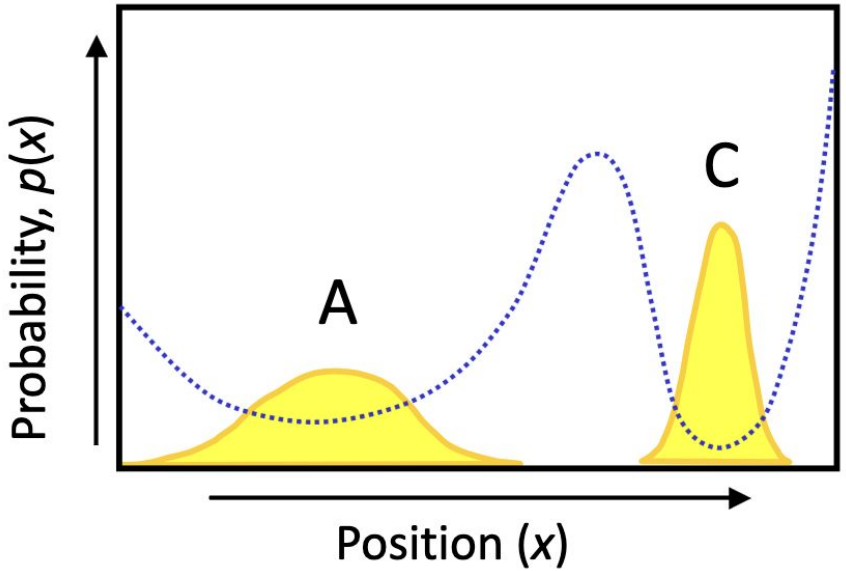
$$+ \sum_i \sum_{j>i} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$ **Van der Waals**

Non-bonded terms

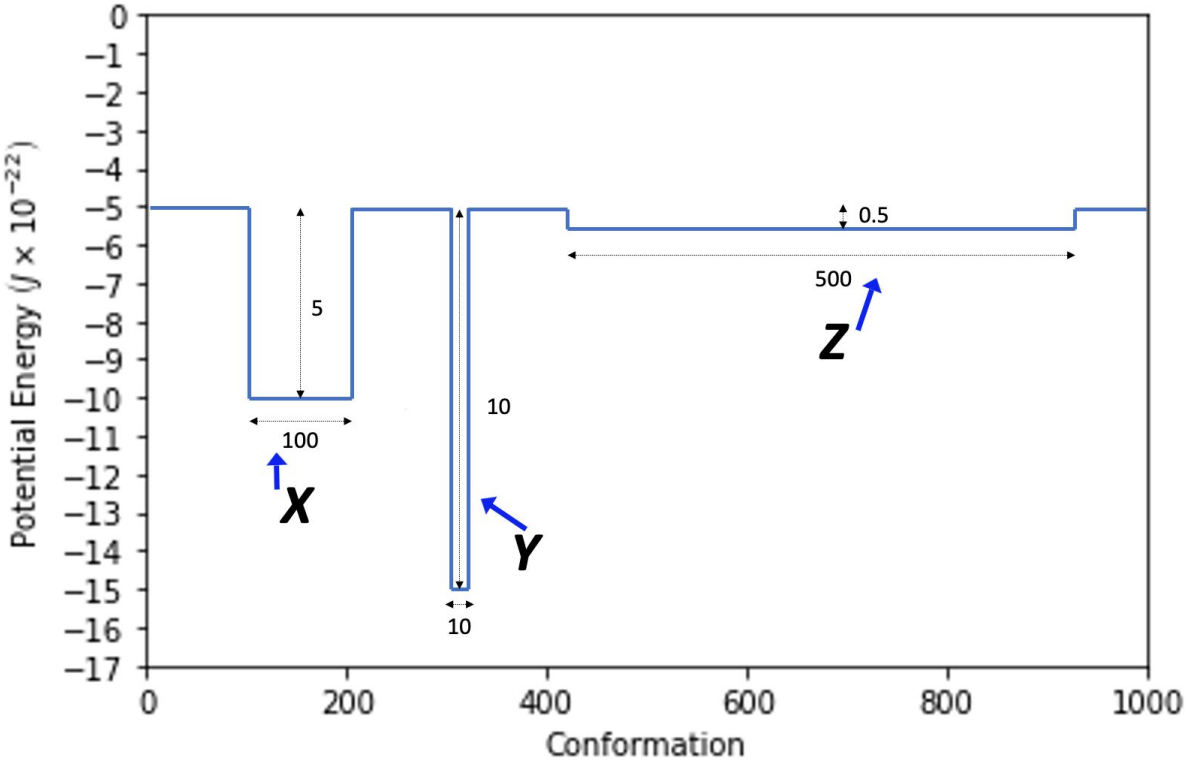
Force Fields and Free Energy Calculation

$$P(A) = \int_{x \in A} P(x) \propto \int_{x \in A} \exp\left({-U(x)} \Big/ {k_B T} \right) dx$$

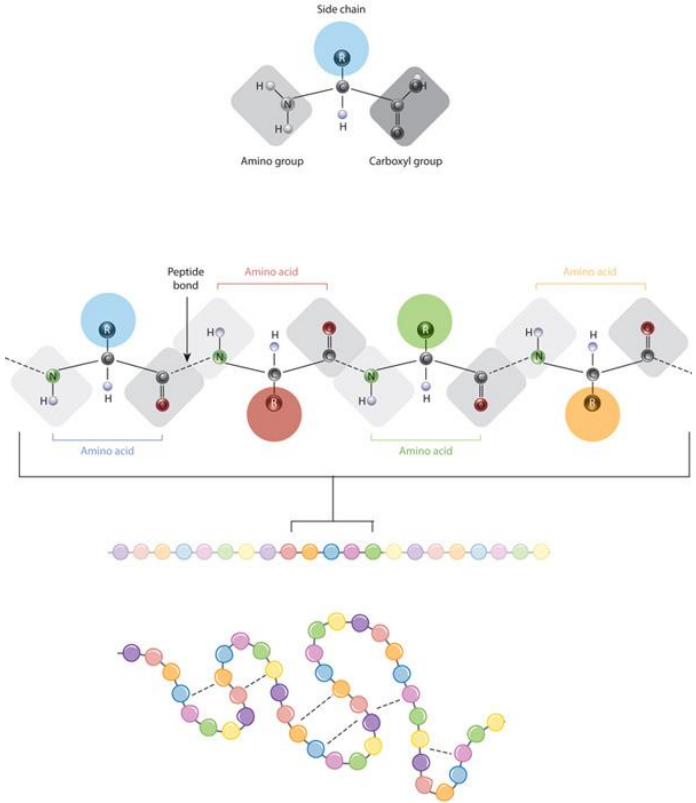




Force Fields and Free Energy Calculation



Deriving a knowledge-based energy function




```
train_structures = [ [
        ('ASP', 'CB', 1, (0, 0, 0)),
        ('LYS', 'CB', 10, (0, 0, 1))
    ]]
```

Deriving a knowledge-based energy function

$$G = \sum_{i=1}^{N} \sum_{j=i+1}^{N} g(d_{ij}, A_i, A_j)$$

scorer

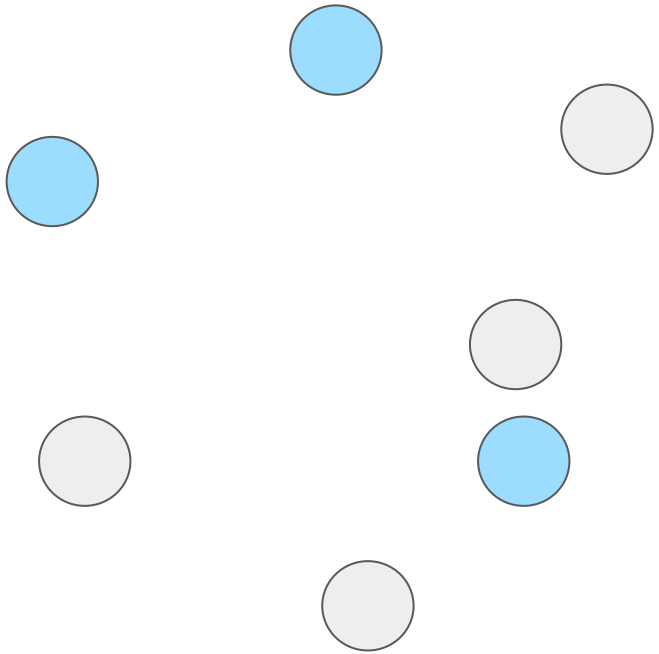
distributions_to_energy

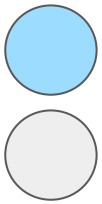
observed_distribution

$$g(d_{ij}, A_i, A_j) = -\log \frac{N_{obs}(d_{ij}, A_i, A_j)}{N_{exp}(d_{ij}, A_i, A_j)}$$

expected_distribution

Deriving a knowledge-based energy function





Aspartate

Arginine

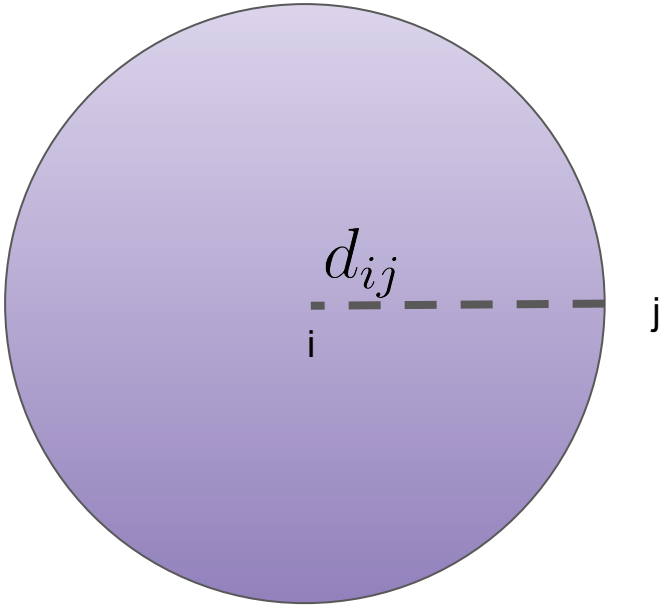
y = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

Deriving a knowledge-based energy function

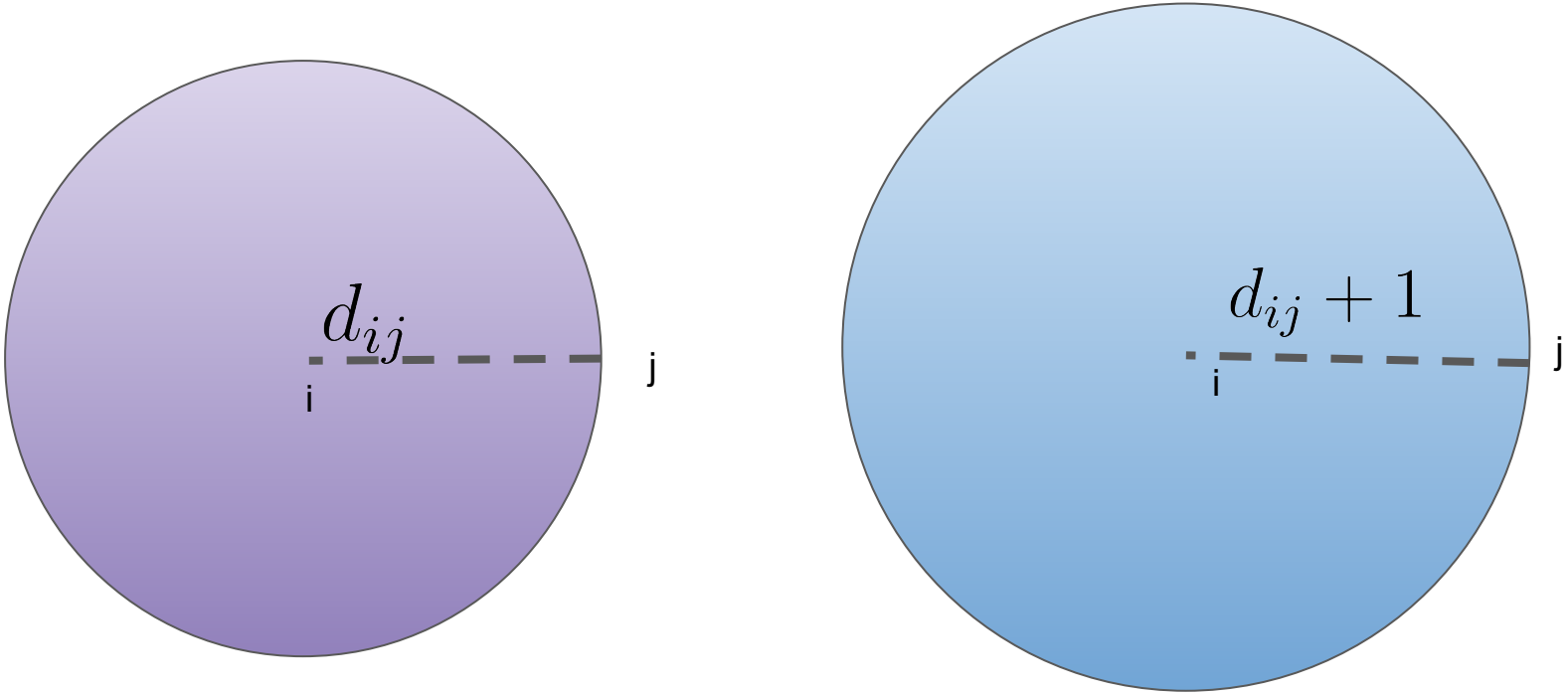
Tips for observed distance:

- Look at how the get_coords, get_distances, and distribution functions are implemented.
- Look at how the observed_distribution tests are implemented.

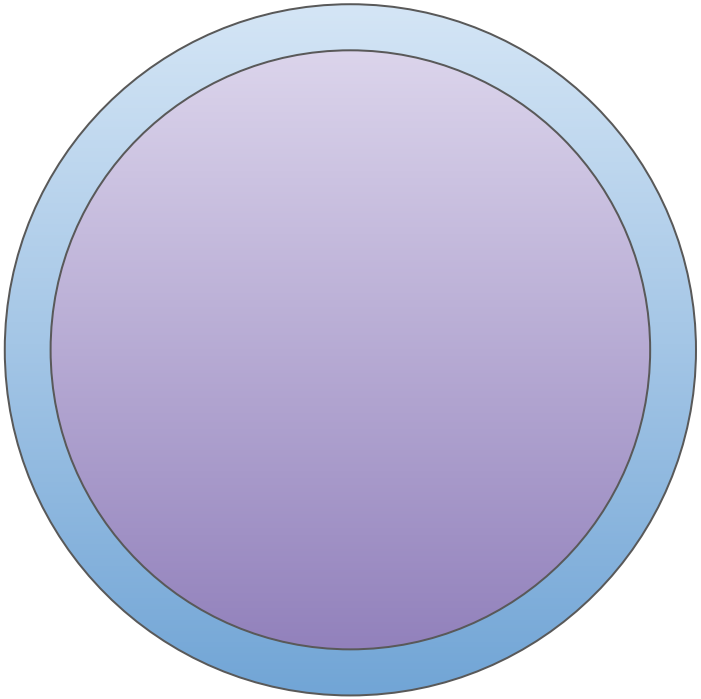
Deriving a knowledge-based energy function



Deriving a knowledge-based energy function



Deriving a knowledge-based energy function



Deriving a knowledge-based energy function

$$G = \sum_{i=1}^{N} \sum_{j=i+1}^{N} g(d_{ij}, A_i, A_j)$$

scorer

distributions_to_energy

observed_distribution

$$g(d_{ij}, A_i, A_j) = -\log \frac{N_{obs}(d_{ij}, A_i, A_j)}{N_{exp}(d_{ij}, A_i, A_j)}$$

expected_distribution

Protein structure prediction with AlphaFold