



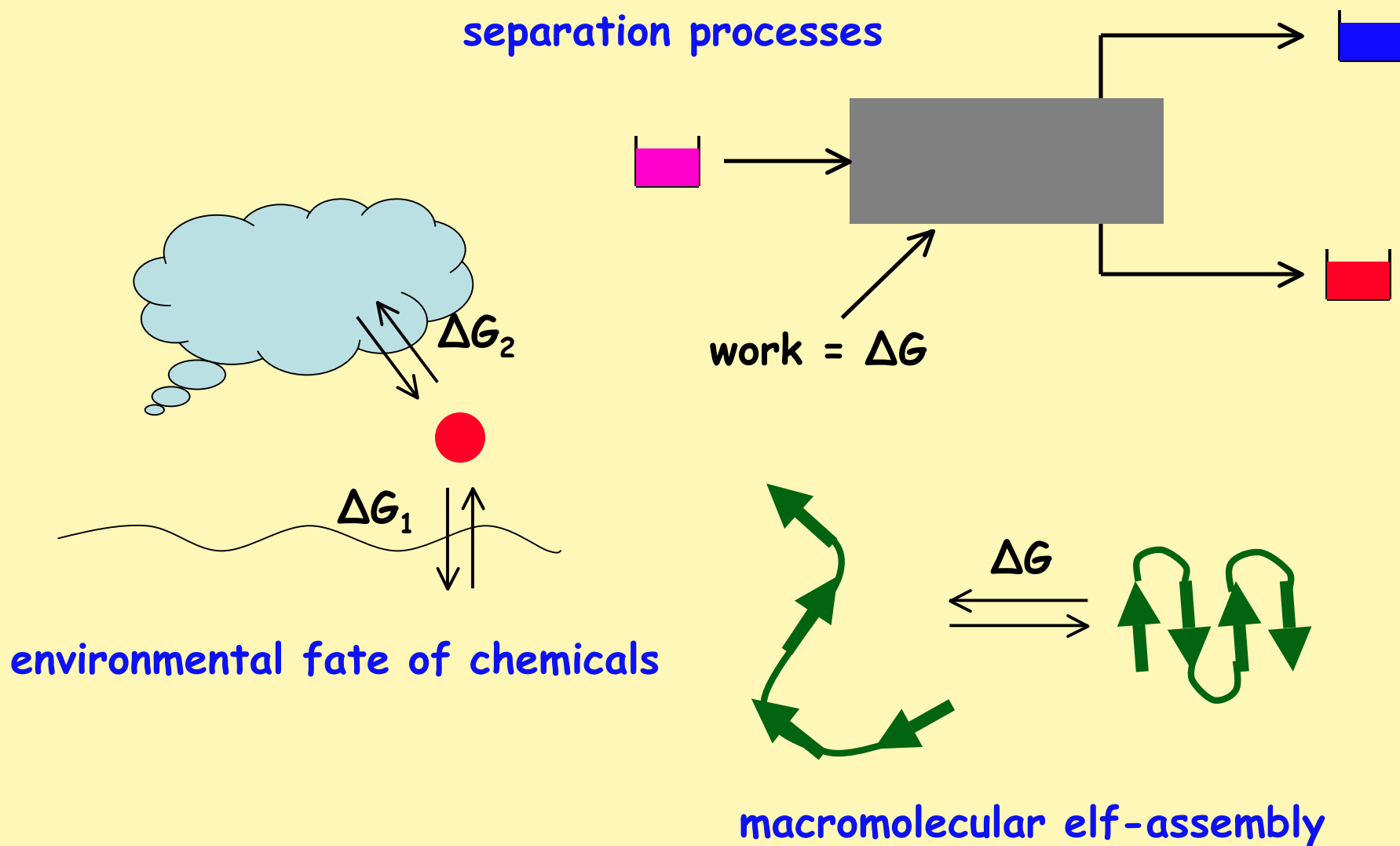
Self-Assembly in Aqueous Environments: Molecular Simulations-Based Calculations of Free Energies

Michael Paulaitis and Sowmi Utiramerur

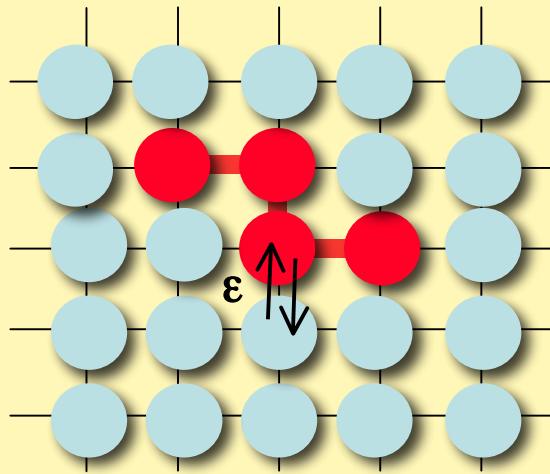
William G. Lowrie Department of Chemical & Biomolecular Engineering
Ohio State University

November 15, 2009

Significance of Free Energy Calculations



Chemical Engineering (Primitive) Free Energy Calculations



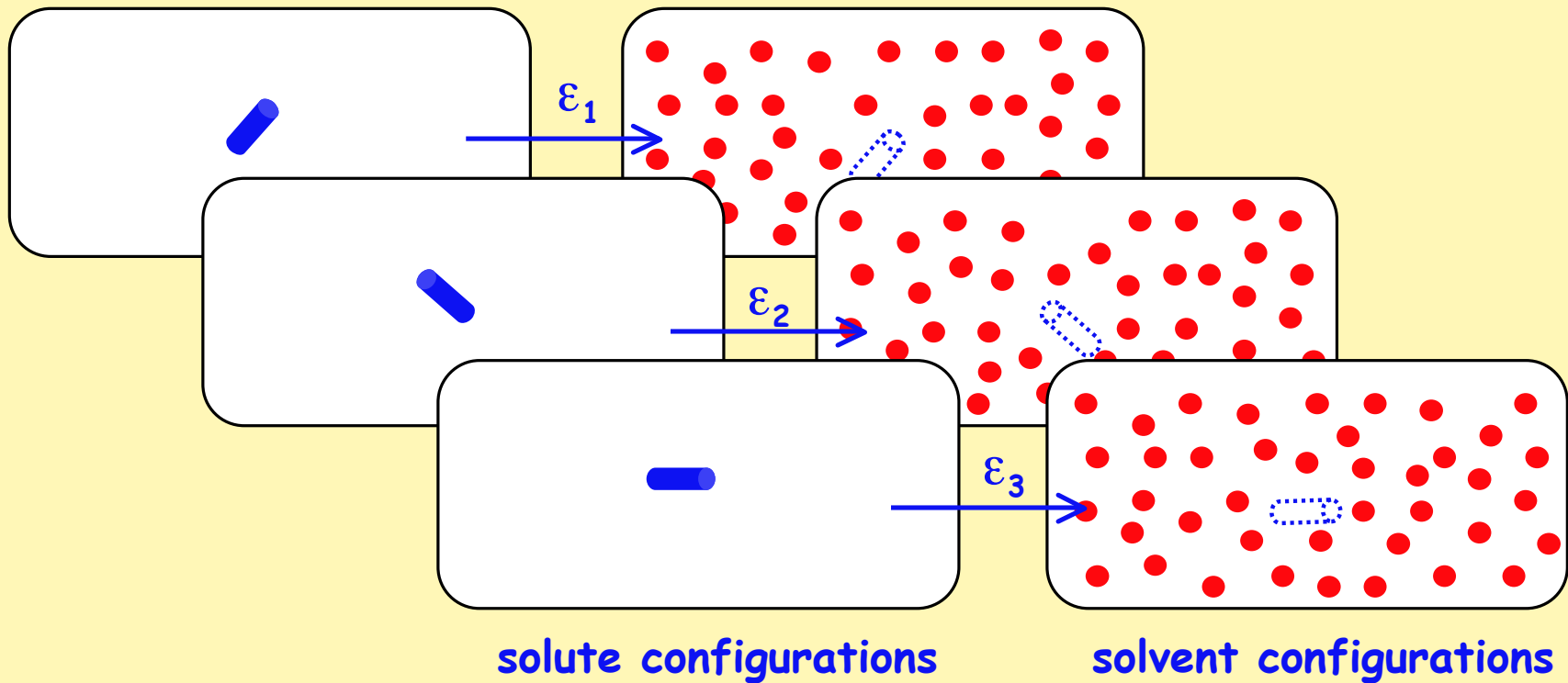
- lattice models (energetic interactions)
 - Flory-Huggins models (entropic interactions)
 - group contribution models (additive interactions)
- > supported by high-throughput measurements of γ^∞

Molecular Simulations for γ^∞ Determinations

(multiscale modeling approaches)

- Free energy perturbation methods
- Thermodynamic integration methods
- Nonequilibrium methods (Jarzynski's method)
- Potential Distribution Theorem methods

Potential Distribution Theorem



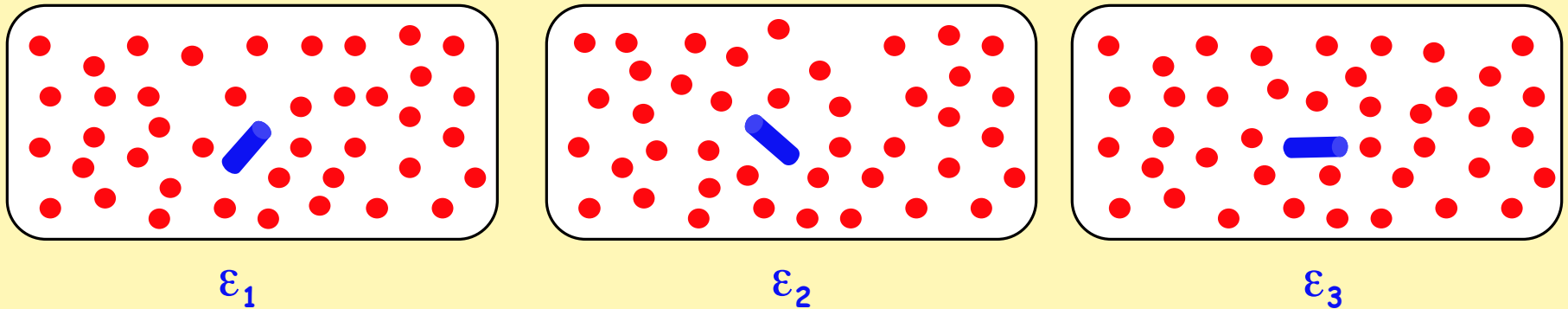
$$e^{-\beta\Delta G} \Leftrightarrow \sum_i e^{-\beta\epsilon_i} \Leftrightarrow \int e^{-\beta\epsilon} P^{(0)}(\epsilon) d\epsilon$$

Potential Distribution Theorem

$$e^{-\beta\Delta G} \Leftrightarrow \sum_i e^{-\beta\varepsilon_i} \Leftrightarrow \int e^{-\beta\varepsilon} P^{(0)}(\varepsilon) d\varepsilon$$

- free energy of solvation depends on energetic interactions of the solute with the surrounding environment.
- free energy of solvation depends on the configurations of the solute and the configurations of the surrounding molecular environment.
- self-assembly involves the selection of those solute configurations that have favorable interactions with the surrounding environment.

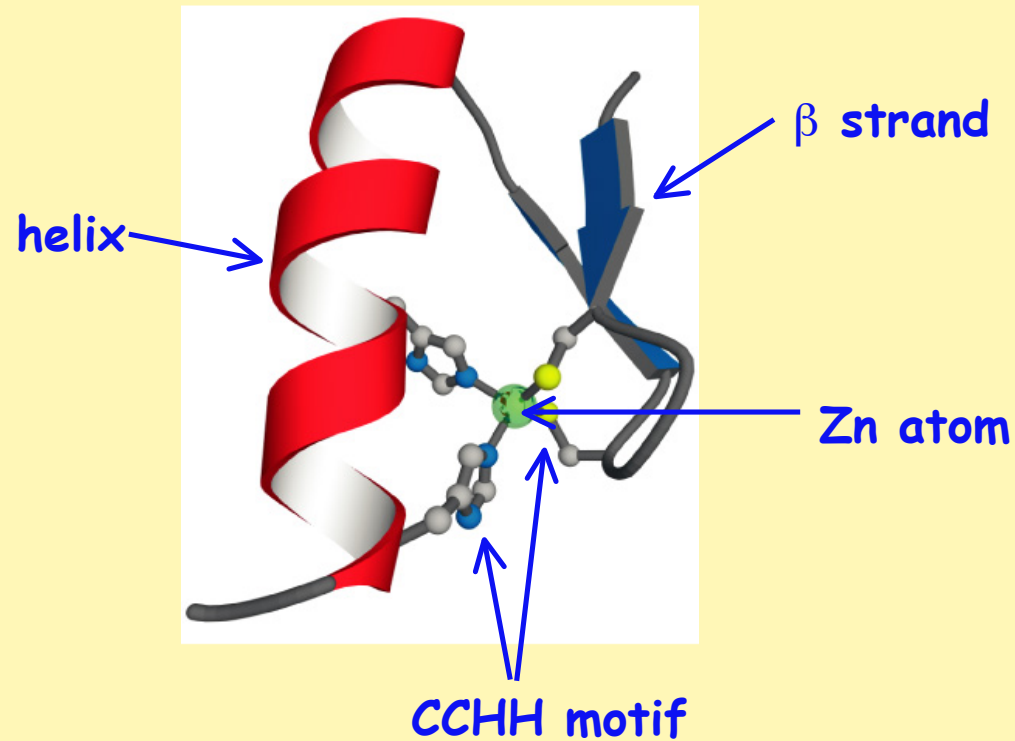
Inverse Potential Distribution Theorem



$$e^{\beta\Delta G} \Leftrightarrow \sum_i e^{\beta\epsilon_i} \Leftrightarrow \int e^{\beta\epsilon} P(\epsilon) d\epsilon$$

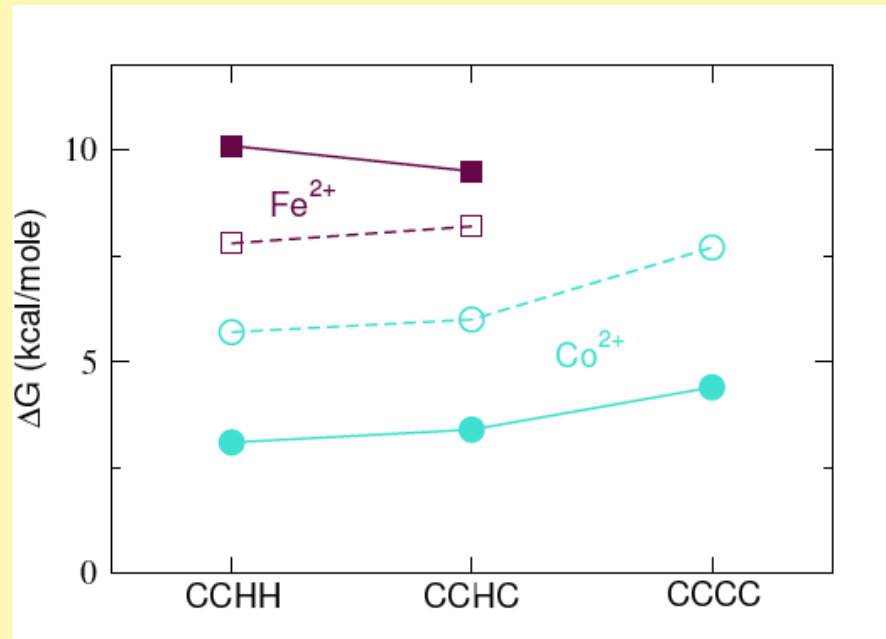
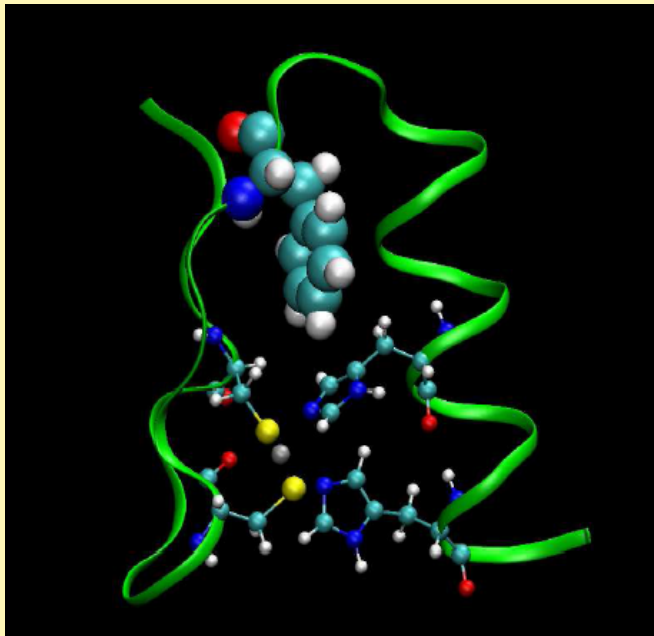
- ϵ is the interaction energy of the solute coupled its surrounding molecular environment.
- $P(\epsilon)$ characterizes how this interaction energy is distributed.

Zinc Finger Domains



- DNA binding protein domain.
- biomedical engineering interest for gene therapy.
- model for metal-mediated protein folding.
- role of zinc in the (mis)folding/aggregation of β -amyloid?

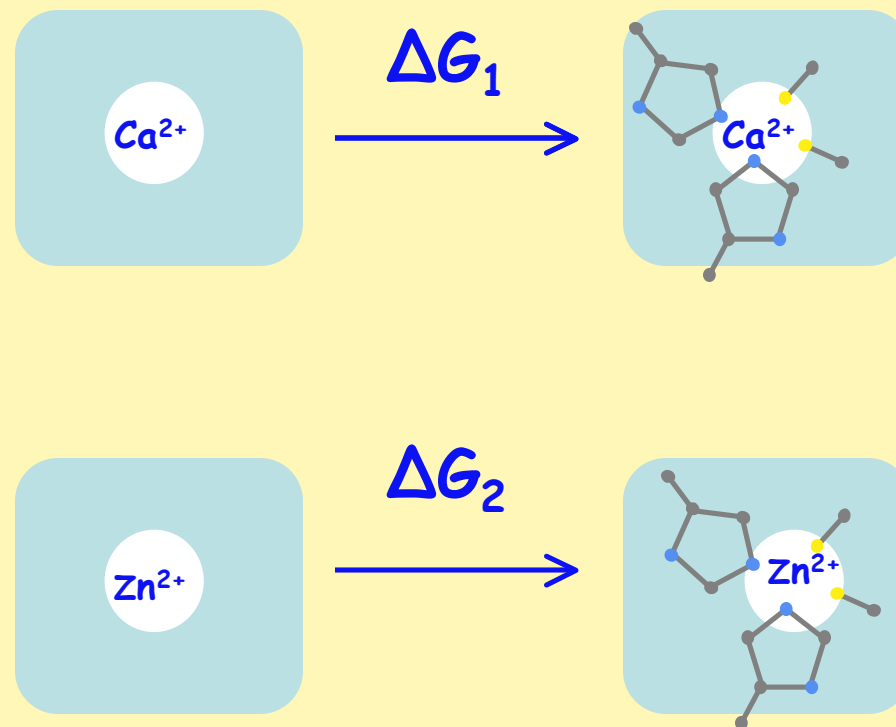
Zinc Finger Domains: Scientific/Technical Questions



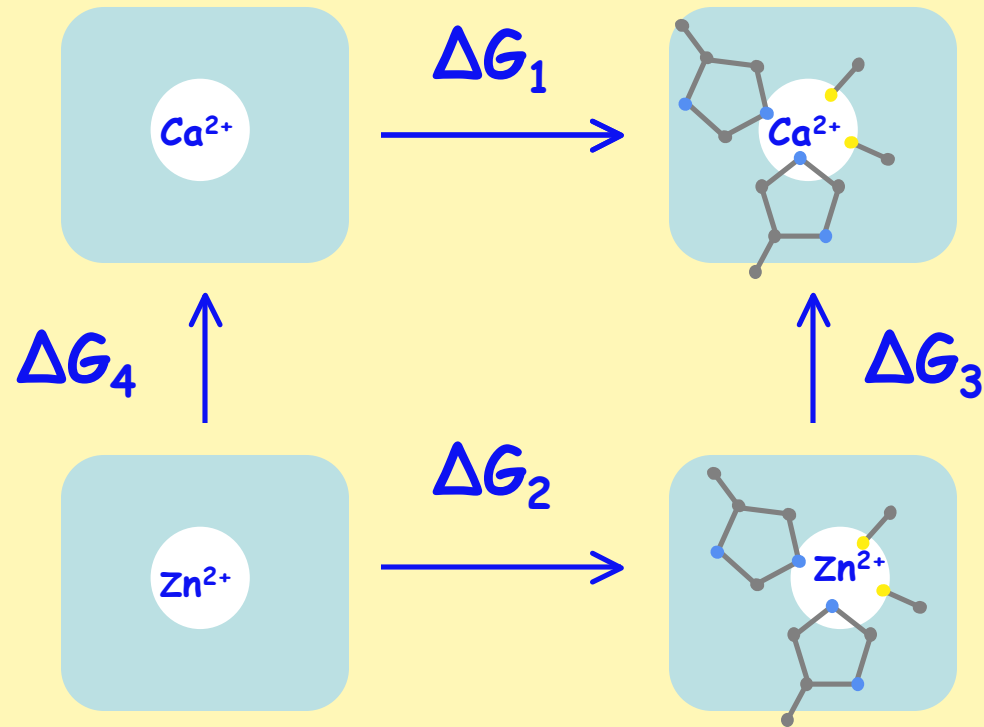
- stability of different divalent cations.
- effect of different binding motifs: CCHH, CCHC, CCCC.
- mutations in the secondary structure of the protein.

Problem Statement

Given molecular dynamics simulation data, compute the relative stability of Ca^{2+} to Zn^{2+} in the standard zinc finger protein (CCHH binding motif).



Thermodynamic Cycle: $\Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$



$$\Delta G = k_B T \cdot \ln \int_{-\infty}^{+\infty} e^{\beta \varepsilon} P(\varepsilon) d\varepsilon$$

Molecular Dynamics Simulation Details

- all simulations with NAMD2.6 (Ohio Supercomputing Center)
- periodic boundary conditions
- minimization performed for 2 ps, followed by 5 ns simulations
- timestep of 2 fs used in the simulations
- particle mesh Ewald summation method for electrostatics
- TIP3P water model (9261 waters of hydration)
- 15,853 total atoms in system (ion, protein, and water)

Files of simulation results

- pure water simulation: `water_in_water.txt`
- zinc ion in water simulation: `zn_in_water.txt`
- calcium ion in water simulation: `ca_in_water.txt`
- zinc ion in znfinger protein: `zn_in_znfinger.txt`
- calcium ion in znfinger protein: `zn_in_znfinger.txt`

NAMD Simulation Results

```
ENERGY: 1100    0.0000    0.0000    0.0000    0.0000    -490.6034    28.9520    0.0000
0.0000    0.0000    -461.6514    -0.0000    -461.6514    -461.6514    0.0000    -1343.4318    -18.1834
174527.1094    -1343.4318    -18.1834
```

```
ENERGY: 1110    0.0000    0.0000    0.0000    0.0000    -484.4644    29.9365    0.0000
0.0000    0.0000    -454.5279    -0.0000    -454.5279    -454.5279    0.0000    -1352.2676    -15.4463
170764.8017    -1352.2676    -15.4463
```

```
ENERGY: 1120    0.0000    0.0000    0.0000    0.0000    481.1824    26.6705    0.0000
0.0000    0.0000    -454.5119    -0.0000    -454.5119    -454.5119    0.0000    -1364.9075    -20.8647
171967.0738    -1364.9075    -20.8647
```

•
•
•
•

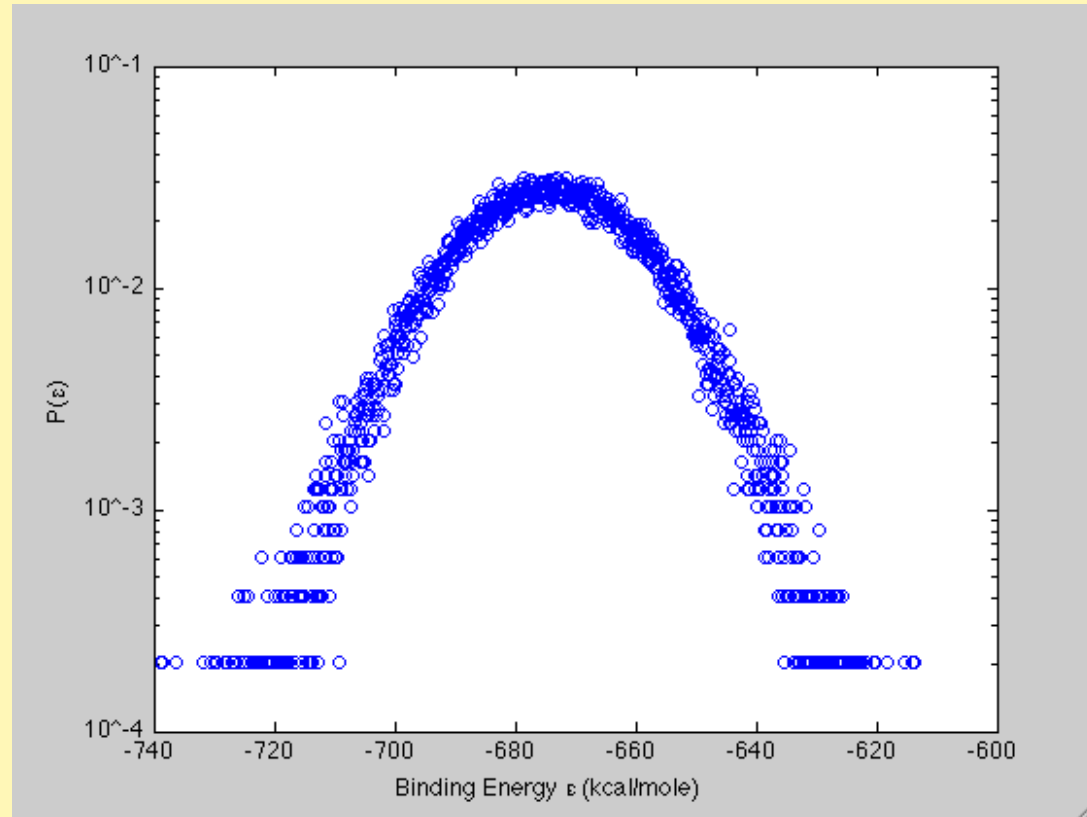
```
ENERGY: 501070    0.0000    0.0000    0.0000    0.0000    -674.3671    28.9084    0.0000
0.0000    0.0000    -645.4587    -0.0000    -645.4587    -645.4587    0.0000    -1466.2261    -53.9675
162590.8297    -1466.2261    -53.9675
```

```
ENERGY: 501080    0.0000    0.0000    0.0000    0.0000    -692.4820    30.8248    0.0000
0.0000    0.0000    -661.6571    -0.0000    -661.6571    -661.6571    0.0000    -1468.9976    -53.7013
162550.2491    -1468.9976    -53.7013
```

```
ENERGY: 501090    0.0000    0.0000    0.0000    0.0000    -699.7397    26.5131    0.0000
0.0000    0.0000    -673.2266    -0.0000    -673.2266    -673.2266    0.0000    -1488.2278    -66.0151
162484.5489    -1488.2278    -66.0151
```

Graduate student vs. undergraduate student projects.

NAMD Simulation Results: Zn^{2+} in Zinc Finger

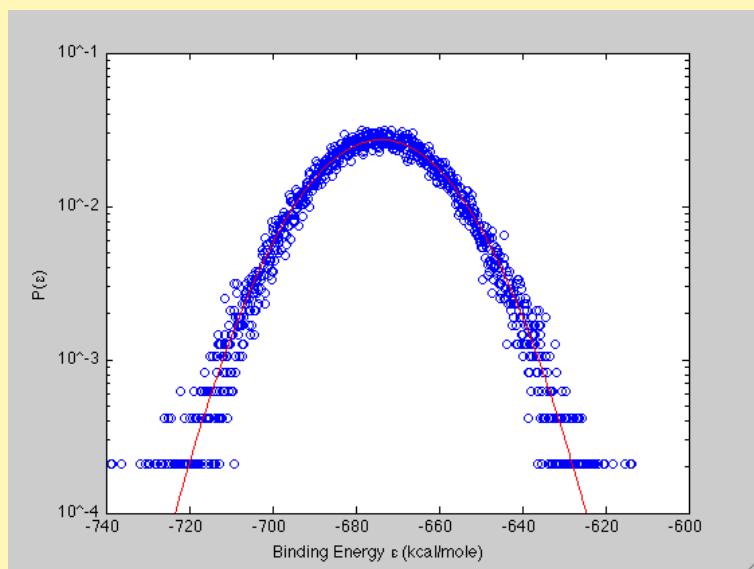


mean binding energy: $\langle \epsilon \rangle = -674.0$ kcal/mol

variance: $\sigma^2/2k_B T = +176.1$ kcal/mol

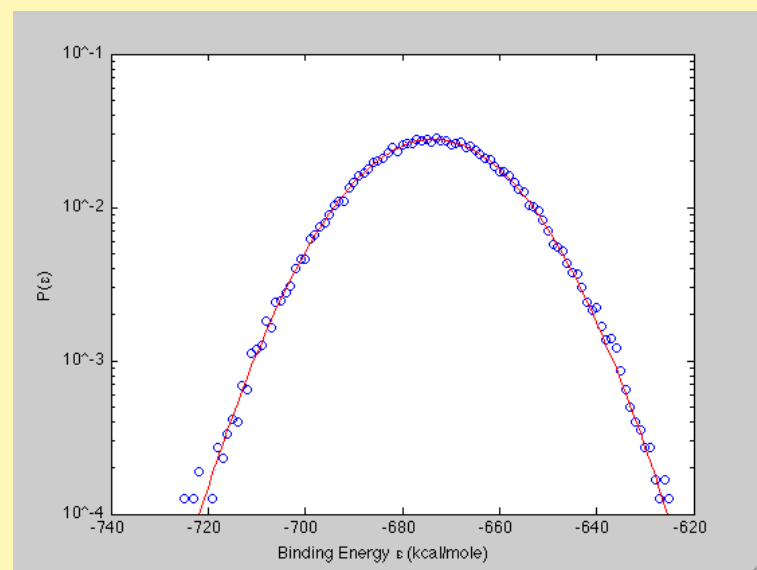
NAMD Simulation Results: Zn^{2+} in Zinc Finger

Effect of Energy Bin Size



$$\langle \epsilon \rangle = -674.0 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +176.0 \text{ kcal/mol}$$

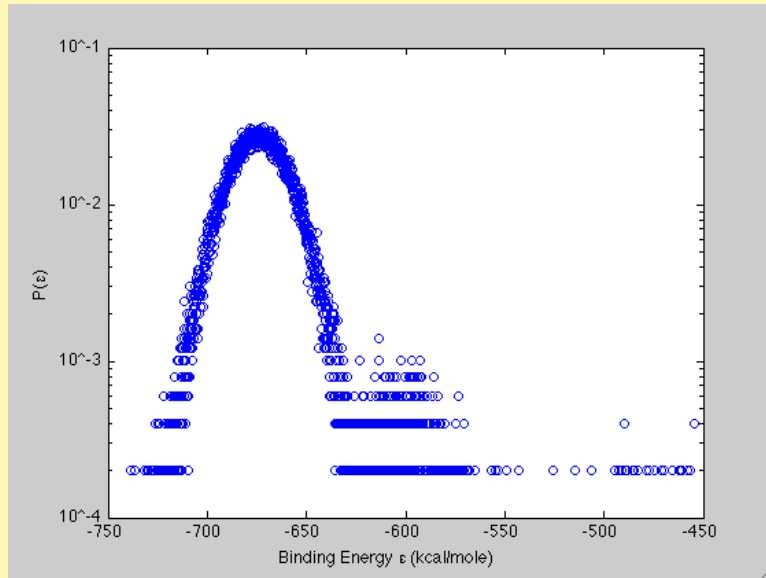
Increasing binding energy bin size



$$\langle \epsilon \rangle = -674.0 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +176.0 \text{ kcal/mol}$$

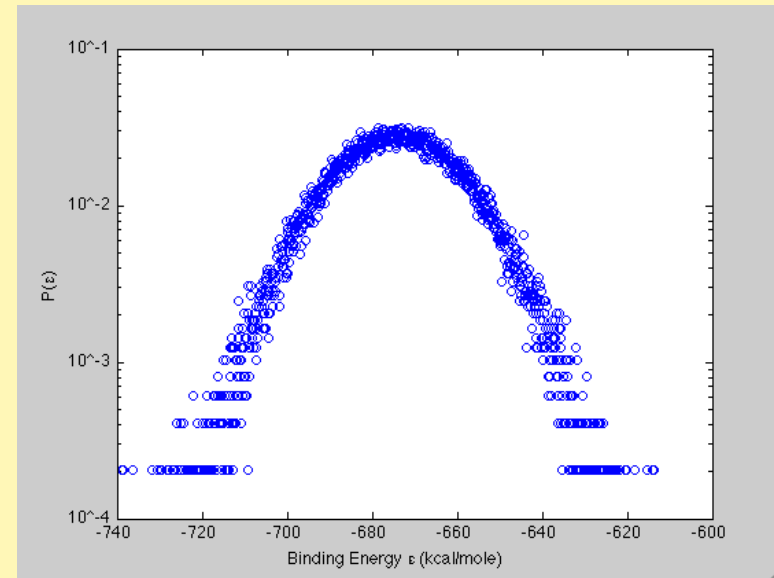
NAMD Simulation Results: Zn²⁺ in Zinc Finger

Effect of Proper Equilibration



$$\langle \epsilon \rangle = -672.9 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +258.1 \text{ kcal/mol}$$

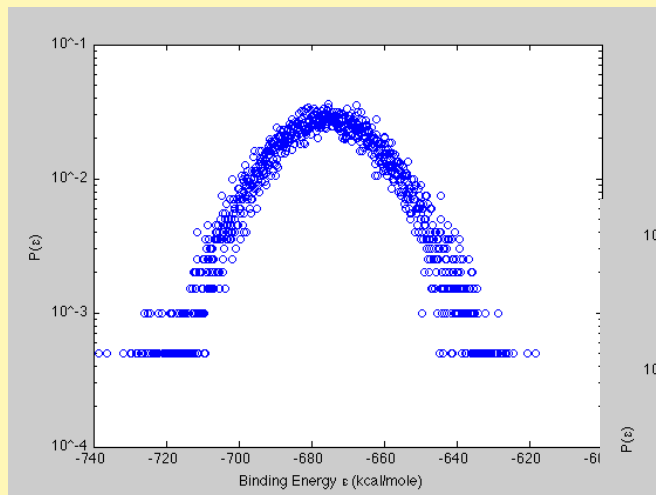
Eliminating initial configurations →



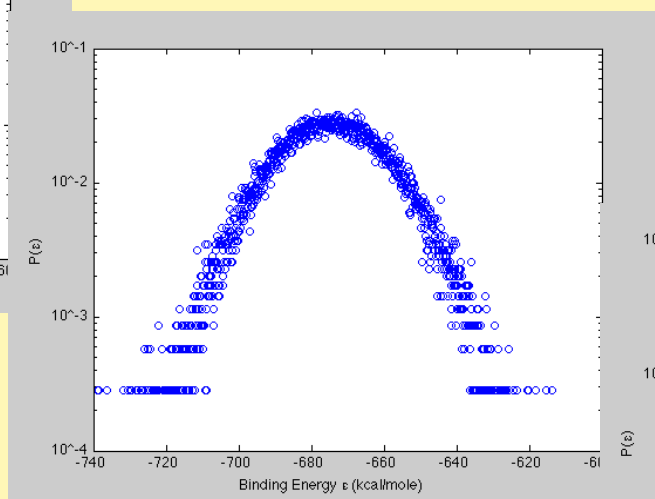
$$\langle \epsilon \rangle = -674.0 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +176.1 \text{ kcal/mol}$$

NAMD Simulation Results: Zn^{2+} in Zinc Finger

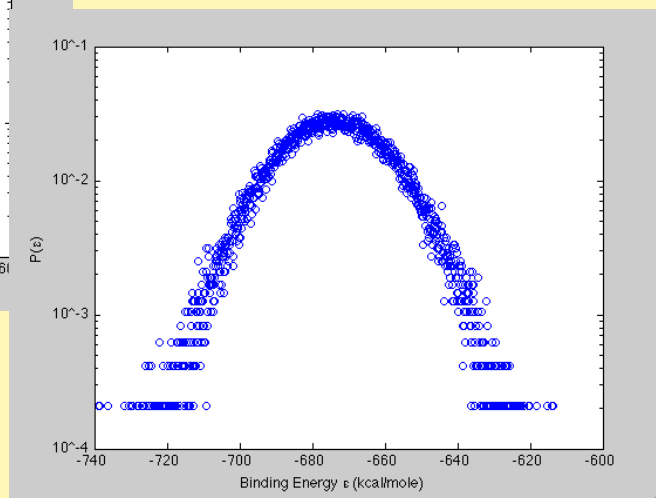
Is the simulation trajectory long enough?



$$\langle \epsilon \rangle = -675.6 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +176.8 \text{ kcal/mol}$$



$$\langle \epsilon \rangle = -674.4 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +176.4 \text{ kcal/mol}$$

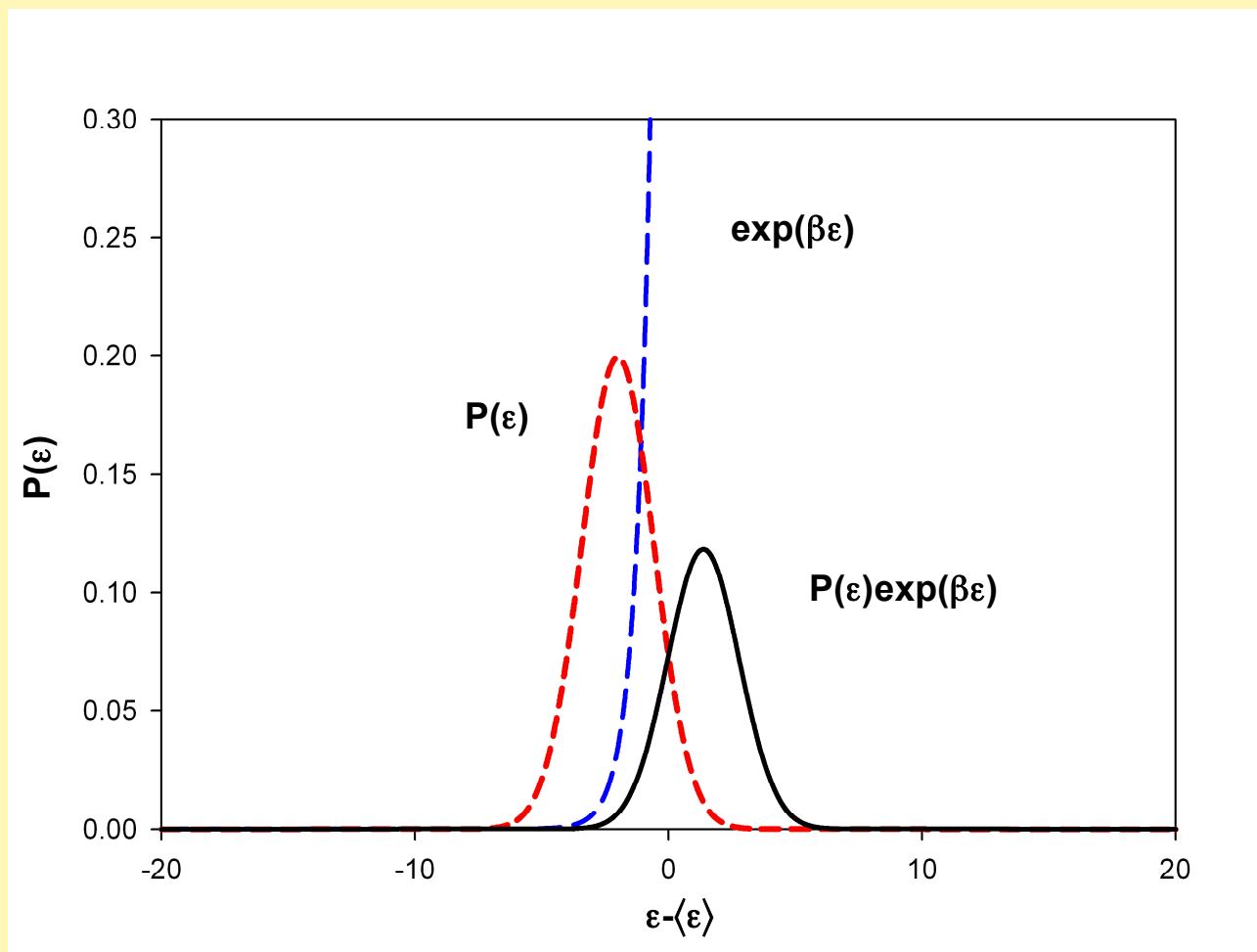


$$\langle \epsilon \rangle = -674.0 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +176.0 \text{ kcal/mol}$$

Increasing the length of trajectory 

NAMD Simulation Results: Zn^{2+} in Zinc Finger

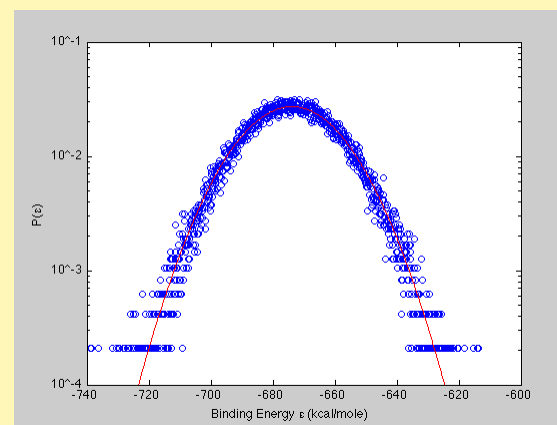
Adequate sampling of the high-energy tail?



NAMD Simulation Results: Zn^{2+} in Zinc Finger

Gaussian approximation

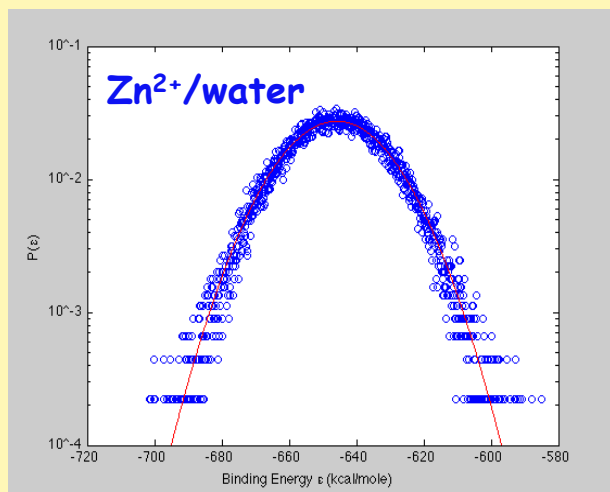
$$P(\varepsilon) = \exp\left[-\frac{(\varepsilon - \langle \varepsilon \rangle)^2}{2\sigma^2}\right]$$



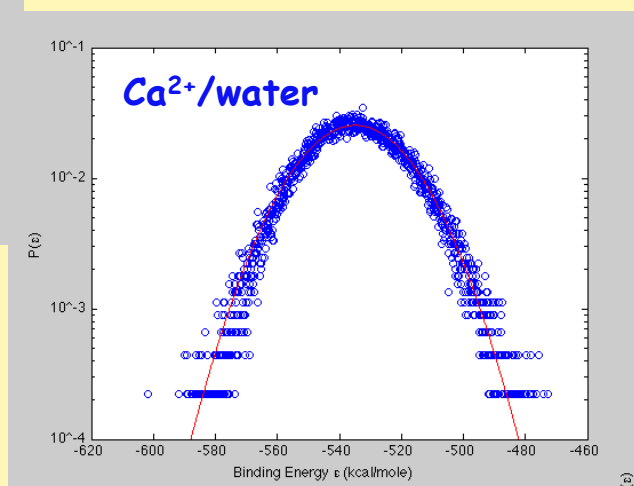
$$\Delta G = k_B T \cdot \ln \int_{-\infty}^{+\infty} e^{\beta \varepsilon} P(\varepsilon) d\varepsilon$$

$$\Delta G = \langle \varepsilon \rangle + \sigma^2 / 2k_B T$$

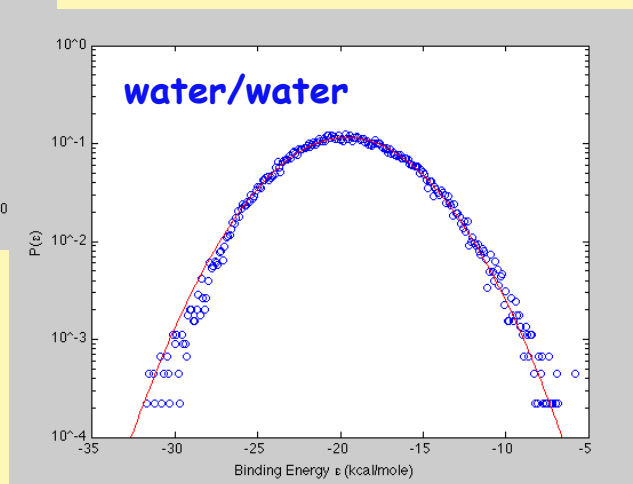
NAMD Simulation Results: non-Gaussian behavior



$$\langle \epsilon \rangle = -646.0 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +173.3 \text{ kcal/mol}$$



$$\langle \epsilon \rangle = -534.7 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +200.3 \text{ kcal/mol}$$

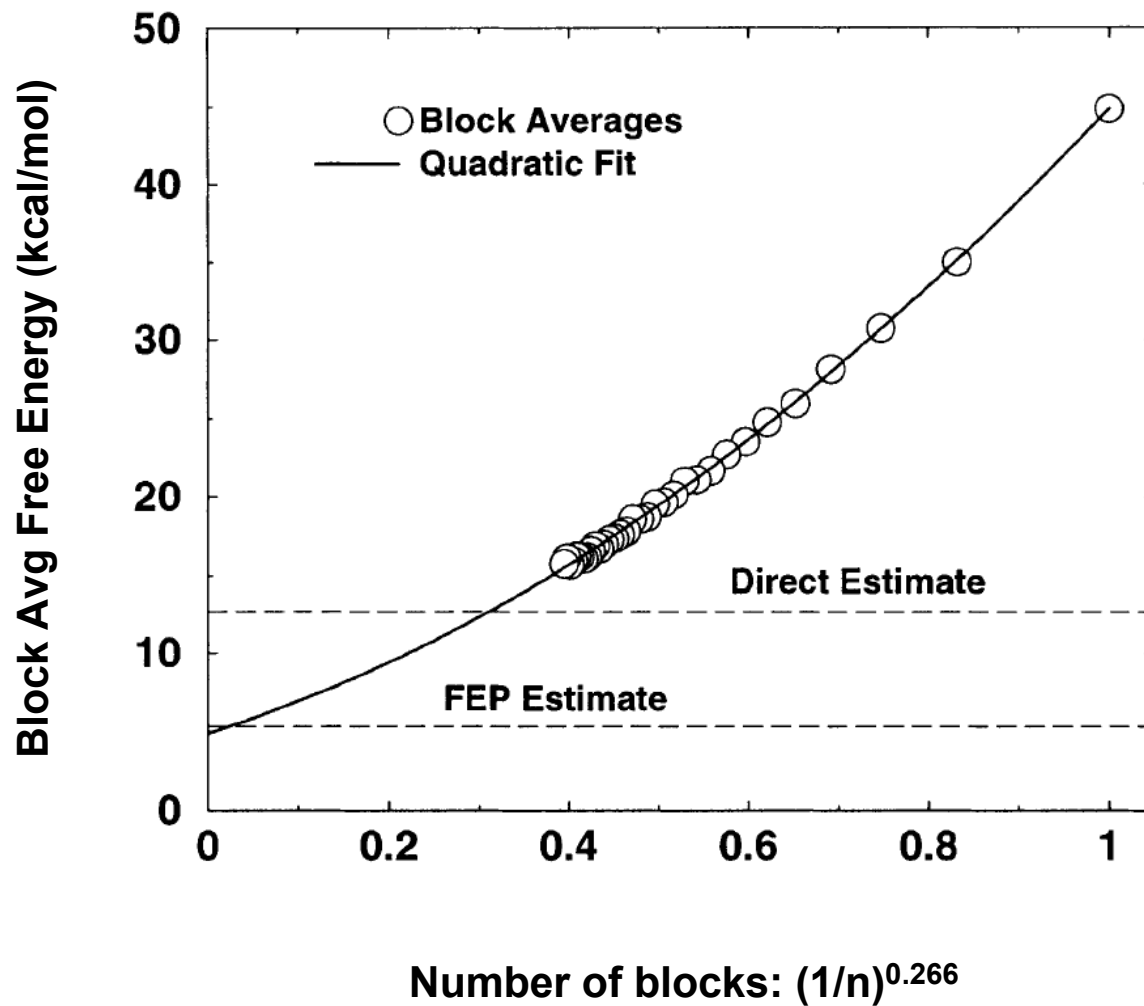


$$\langle \epsilon \rangle = -19.6 \text{ kcal/mol}$$
$$\sigma^2/2k_B T = +9.9 \text{ kcal/mol}$$

increasing skewness

NAMD Simulation Results

Uncertainties from Block Averaging



References

- JM Berg; HA Godwin, “Lessons from Zinc-Binding Peptides,” *Annu. Rev. Biophys. Biomol. Struct.* **26**:357 (1997).
- LR Pratt; D Asthagiri, “Potential Distribution Methods and Free Energy Models of Molecular Solution,” in *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, C Chipot and A Pohorille, editors, Springer Ser. in Chemical Physics **86**:323 (2007).
- TL Beck; ME Paulaitis; LR Pratt, *The Potential Distribution Theorem and Models of Molecular Solutions*, Cambridge Univ. Press (2006).
- D Asthagiri; LR Pratt; ME Paulaitis; SB Rempe, “Hydration Structure and Free Energy of Biomolecularly Specific Aqueous Dications, Including Zn²⁺ and First Transition Row Metals,” *J. Am. Chem. Soc.* **126**:1285 (2004).