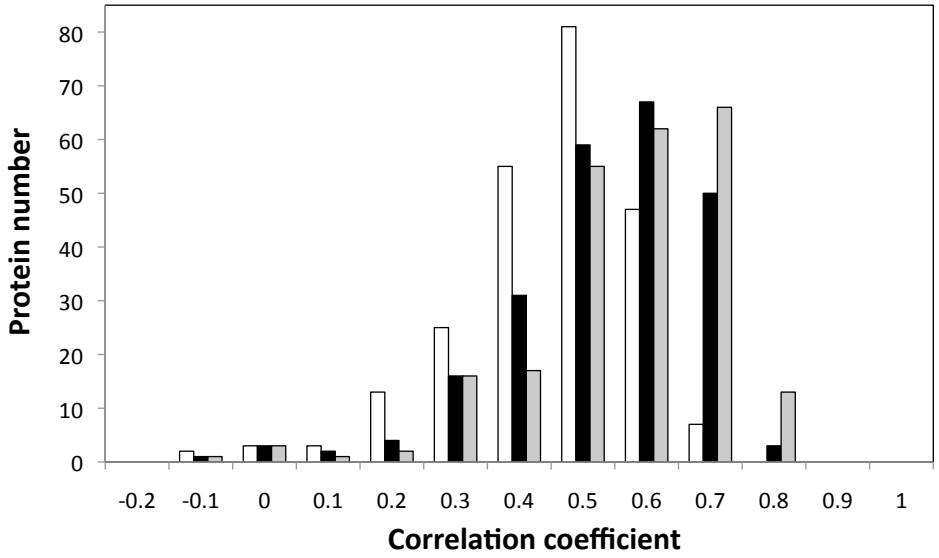
7-16

LIFE IS FULL OF SURPRISES
SCHULZ
© PNTS

Assignment 5 (or Final project)



Correlation distributions of different GNM methods with conservation scores

You may get good correlations between between GNM-BF and ConSURF scores. But it is highly likely that you will get very bad correlations.

Now, instead of presenting the results as they are, you may need to think of why they are so - not any reasons but those that are supported by evidence. And then draw your conclusions from your calculations.

From GNM to Contact number

Flexibility and packing in proteins

Halle showed that

$$\sigma_k = \int C^{-1} d\rho_k \rho_k^T \rho_k \exp(-\beta W(\rho_k))$$

$$\beta W(\rho_k) \sim \lambda_k \rho_k^T \rho_k$$

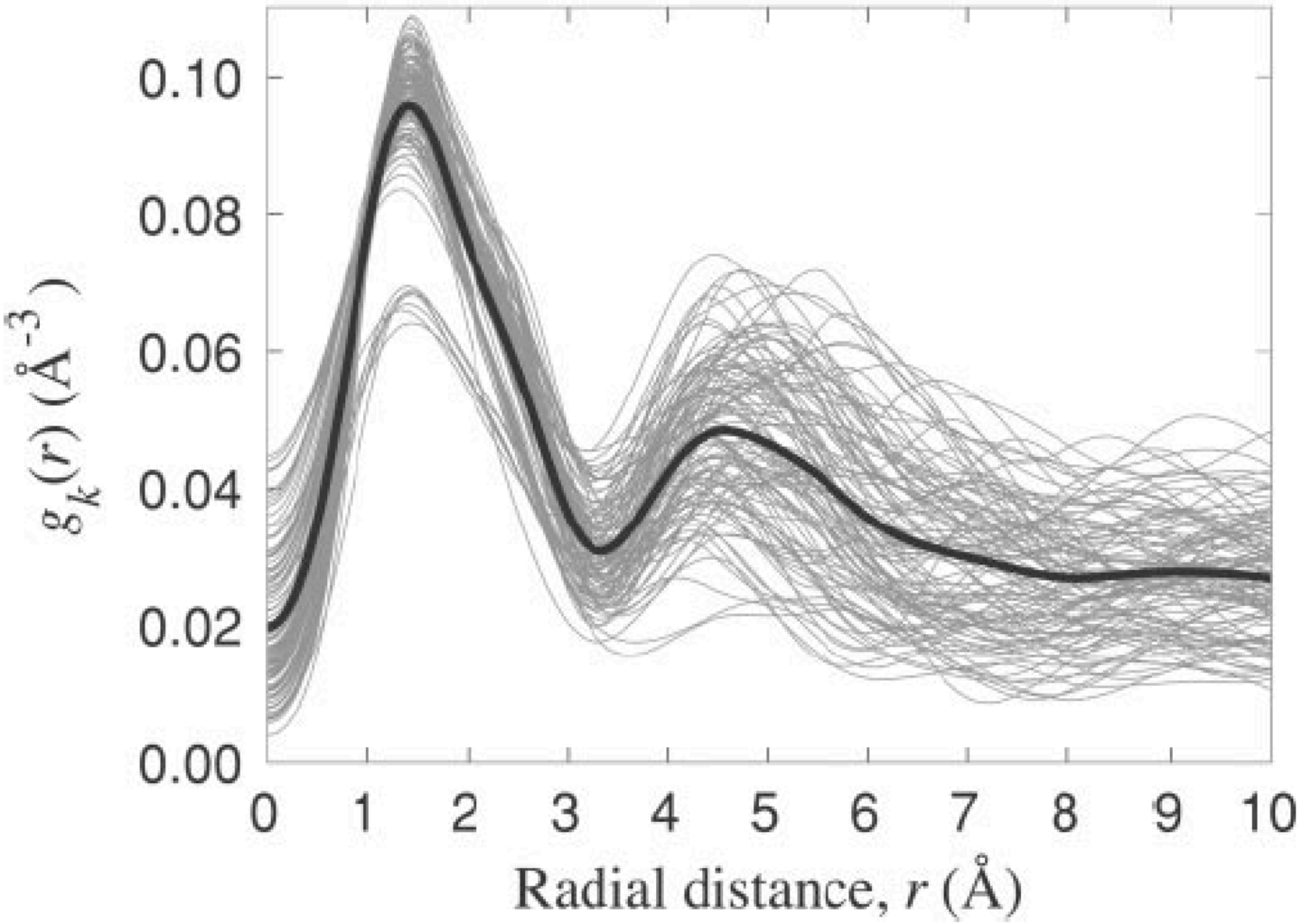
$$\sigma_k = \frac{3}{2\lambda_k} = \left(\frac{3}{2\lambda_k}\right)\left(\frac{1}{n_k}\right)$$

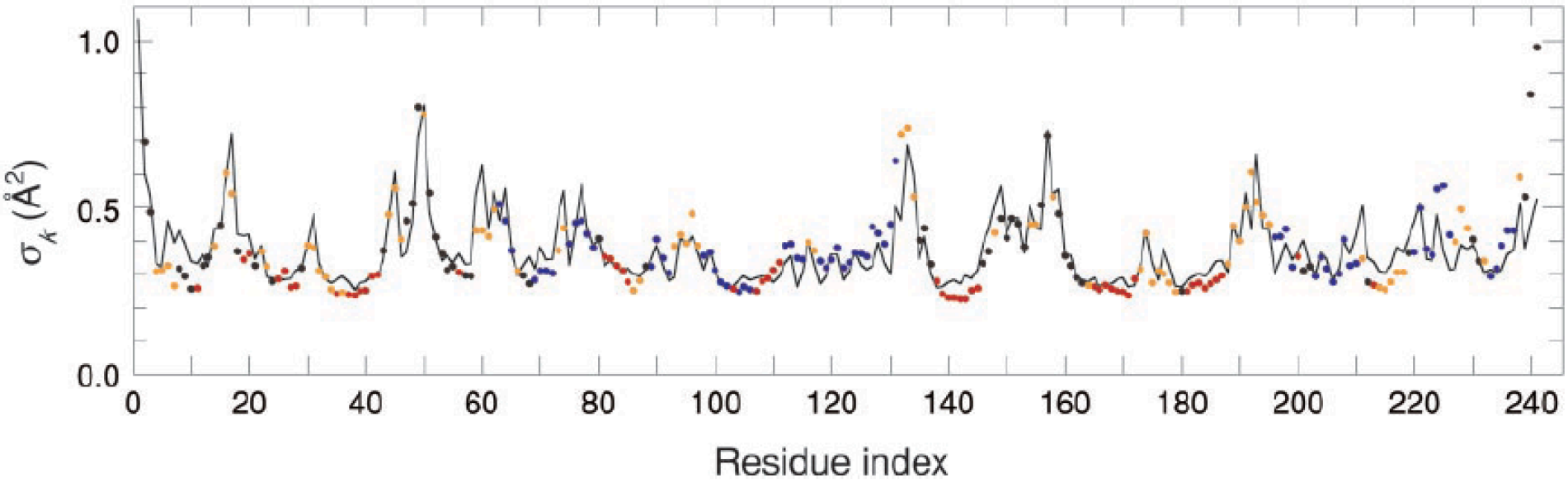
PNAS 2002;99;1274-1279

Ignoring off-diagonal elements of the Hessian matrix

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \cdots \\ h_{21} & h_{22} & h_{23} & \cdots \\ h_{31} & h_{32} & h_{33} & \cdots \\ \vdots & & & \end{pmatrix} \sim \begin{pmatrix} h_{11} & 0 & 0 & \cdots \\ 0 & h_{22} & 0 & \cdots \\ 0 & 0 & h_{33} & \cdots \\ \vdots & & & \end{pmatrix}$$

$$C = H^{-1} = \begin{pmatrix} h_{11}^{-1} & 0 & 0 & \cdots \\ 0 & h_{22}^{-1} & 0 & \cdots \\ 0 & 0 & h_{33}^{-1} & \cdots \\ \vdots & & & \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1} & 0 & 0 & \cdots \\ 0 & \frac{1}{n_2} & 0 & \cdots \\ 0 & 0 & \frac{1}{n_3} & \cdots \\ \vdots & & & \end{pmatrix}$$


0.10
0.08
0.06
0.04
0.02
0.00
$g_k(r)$ (Å^{-3})
0
1
2
3
4
5
6
7
8
9
10
Radial distance, r (Å)


1.0
0.5
0.0
σ_k (Å^2)
0
20
40
60
80
100
120
140
160
180
200
220
240
Residue index

Comparison between GNM and Contact Number (CN)

Table 1. Indicators for model predictions of Cα AMSDs for full protein set

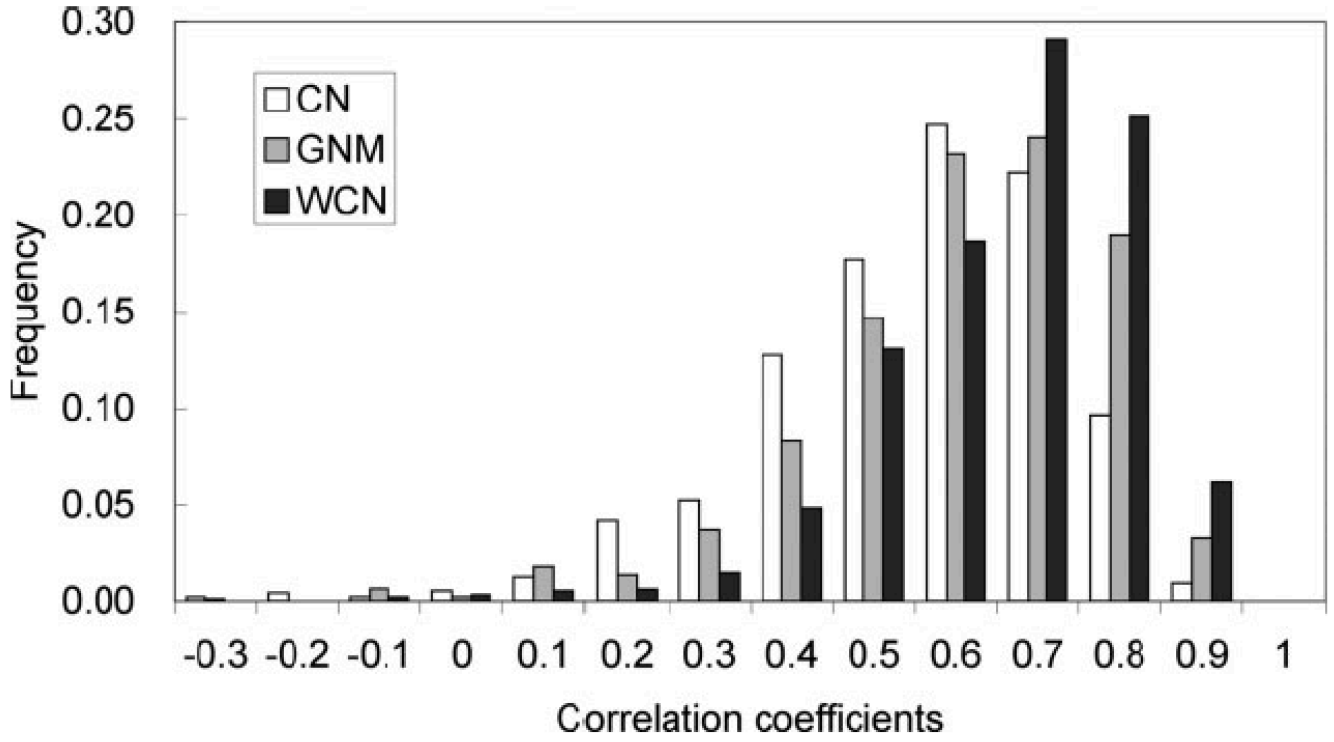
	Model	Density*	⟨Δ⟩†	Range of Δ	⟨ρ⟩†	Range of ρ
a	LDM	all/ref/fix	0.89 ± 0.27	0.63–2.19	0.62 ± 0.09	0.41–0.80
b	LDM	all/ref/scd	0.86 ± 0.26	0.62–2.09	0.64 ± 0.09	0.43–0.81
c	LDM	all/xtl/fix	0.75 ± 0.12	0.52–1.21	0.67 ± 0.09	0.45–0.83
d	LDM	all/xtl/scd	0.72 ± 0.11	0.53–1.13	0.70 ± 0.09	0.49–0.85
e	P-GNM	Cα/ref/fix	1.08 ± 0.42	0.65–3.06	0.58 ± 0.17	0.05–0.84
f	LDM	Cα/ref/fix	1.02 ± 0.32	0.74–2.58	0.51 ± 0.11	0.20–0.70
g	LDM	Cα/ref/scd	0.97 ± 0.29	0.68–2.32	0.58 ± 0.08	0.42–0.75

Comparison between GNM, CN and WCN

Table I

The Performance Breakdown of the WCN Model, the CN Model and the GNM for the Structures Classified According to the SCOP Classes

	WCN		CN		GNM	
SCOP classes	$\bar{c}$	$p_{0.5}$	$\bar{c}$	$p_{0.5}$	$\bar{c}$	$p_{0.5}$
All-α proteins	0.59	0.73	0.47	0.43	0.54	0.68
All-β proteins	0.64	0.82	0.51	0.58	0.58	0.73
α/β proteins	0.62	0.82	0.49	0.51	0.57	0.75
α + β proteins	0.60	0.77	0.49	0.51	0.54	0.65



CN and WCN

Contact Number

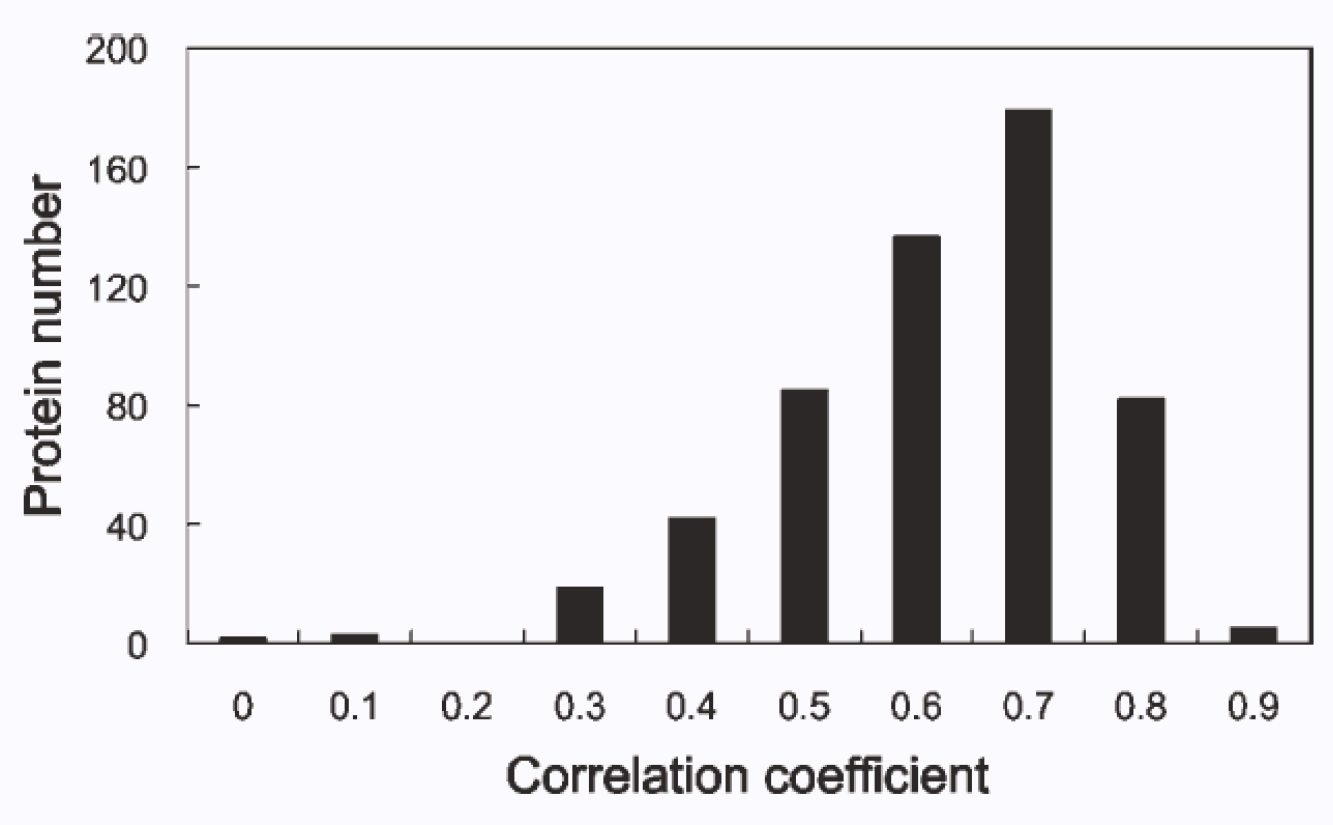
$$C_k = \sum_i H(r_{ik} - r_{cutoff})$$

Weighted Contact Number

$$W_k = \sum_i \frac{1}{r_{ki}^2}$$

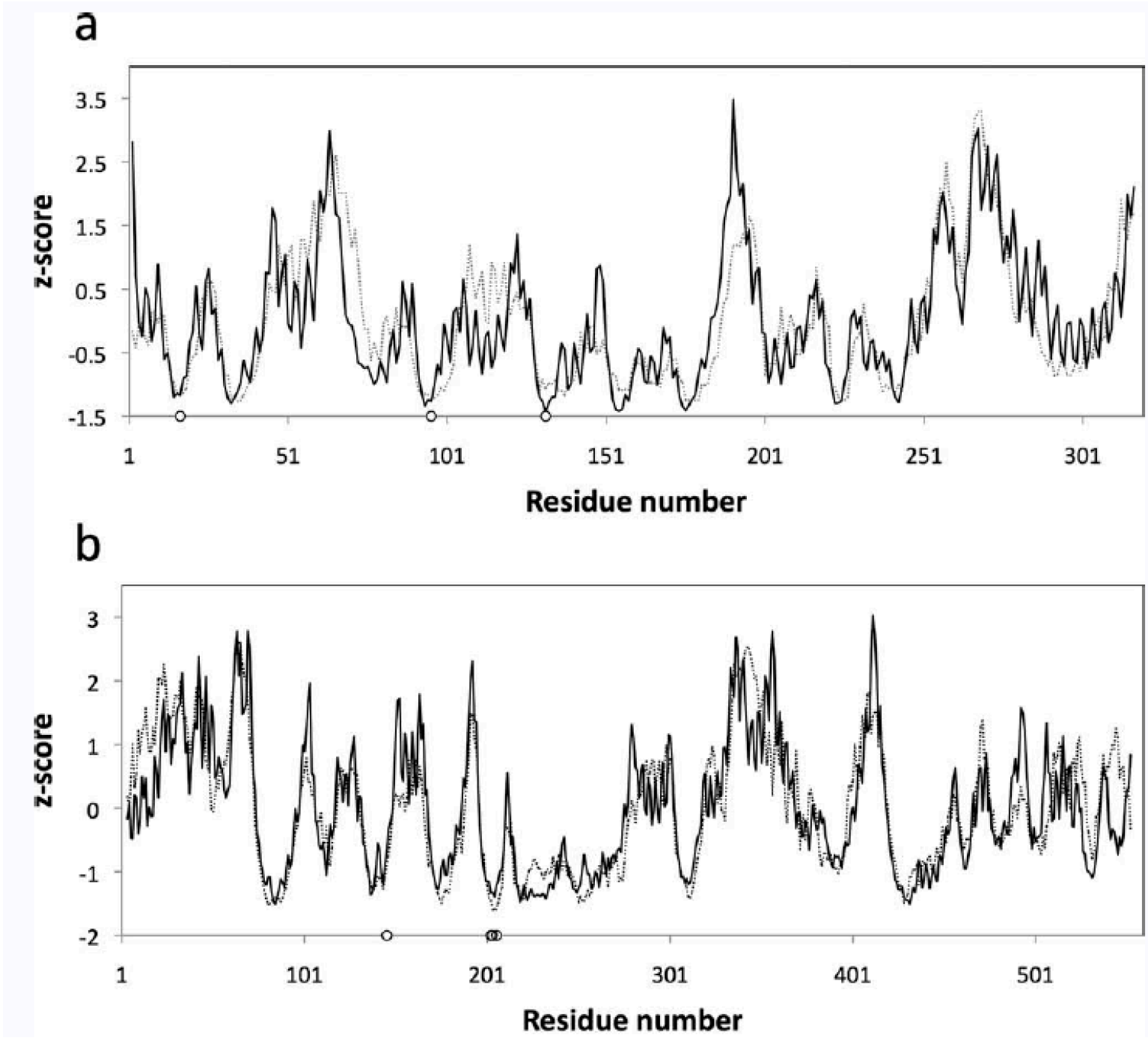
The link between contact number and conservation

The distribution of the correlations between WCN and conservation



Average correlation coefficient for 554 structures is 0.57

WCN vs. Conservation



Solid: WCN
Dotted: Consv

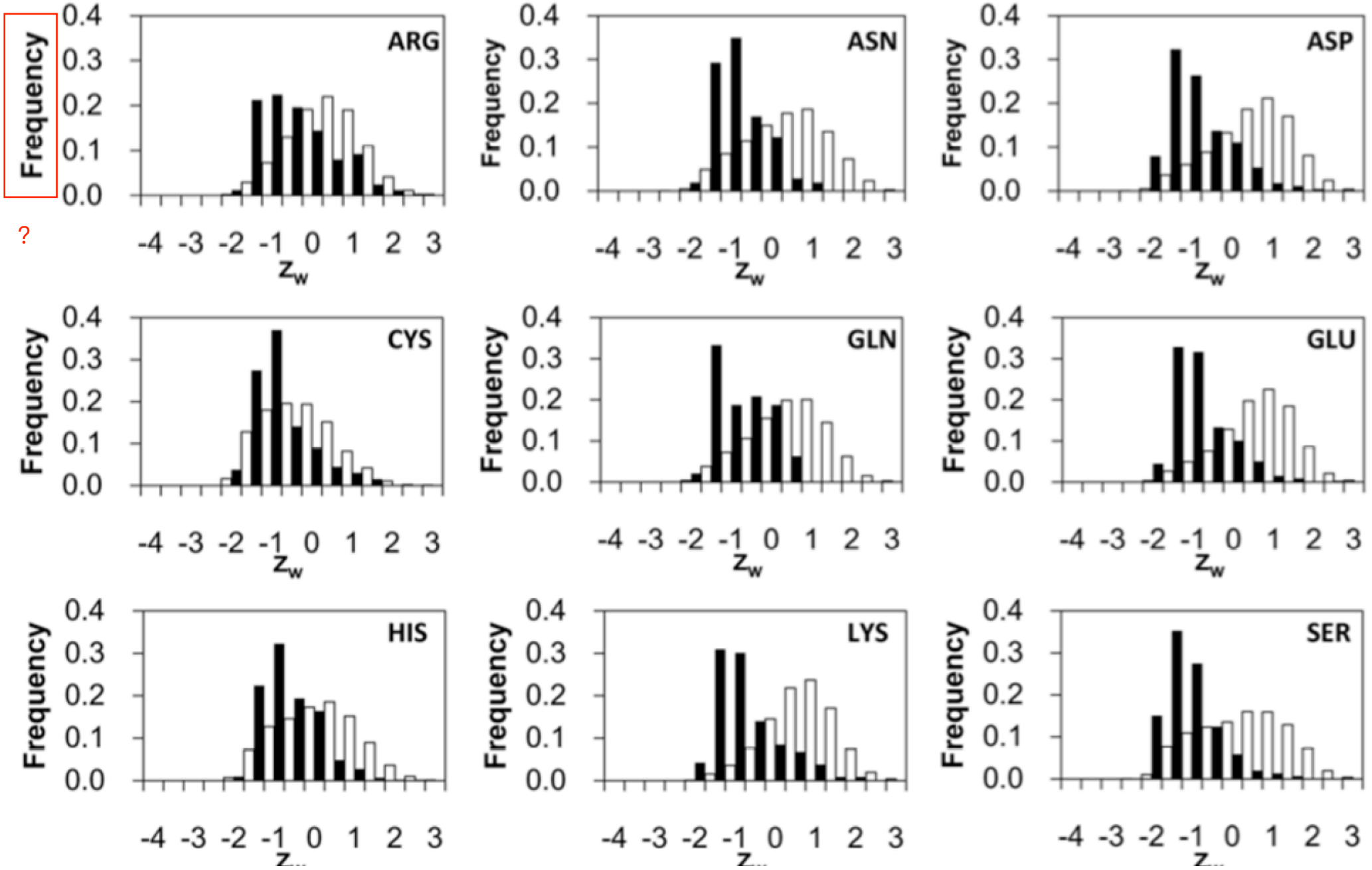
Solid: WCN
Dotted: Consv

1. The BF is correlated with GNM-B factors and WCN and CN (published)

2. The WCN and CN is associated with conservation (published)

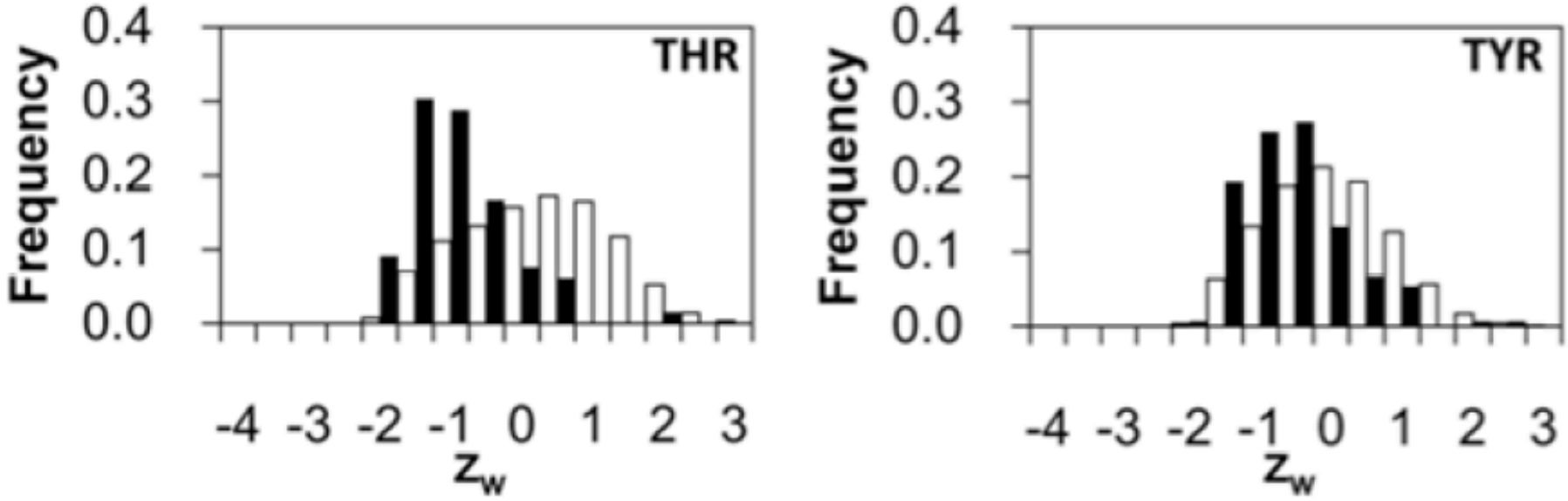
3. **Is BF associated with conservation – your final project?**

The distribution of WCN of catalytic amino acids



?

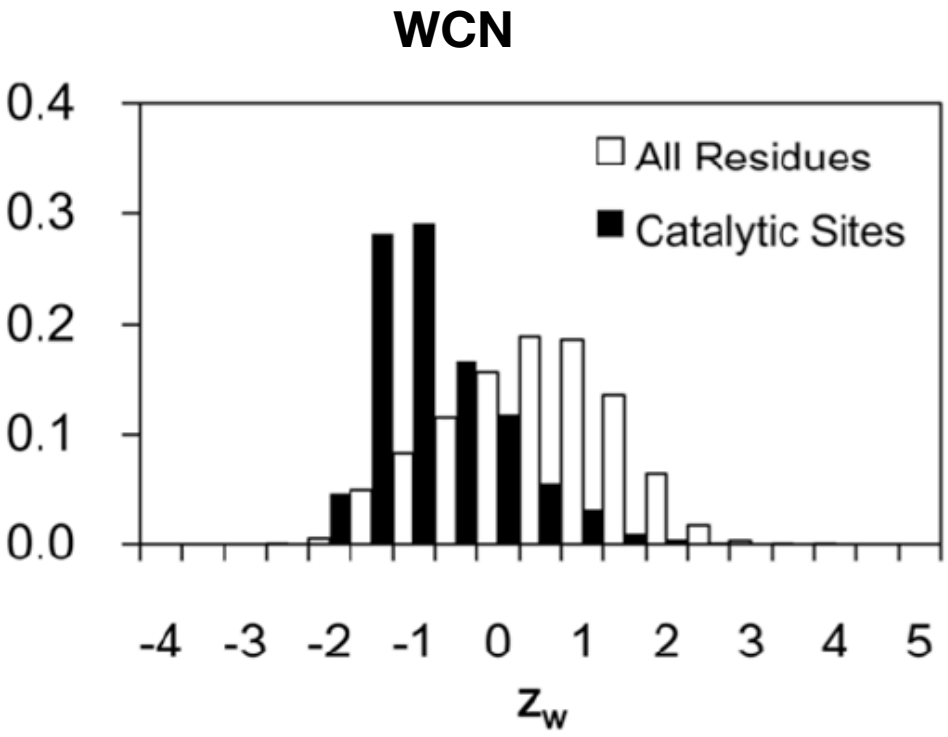
The distribution of WCN of an amino acid type

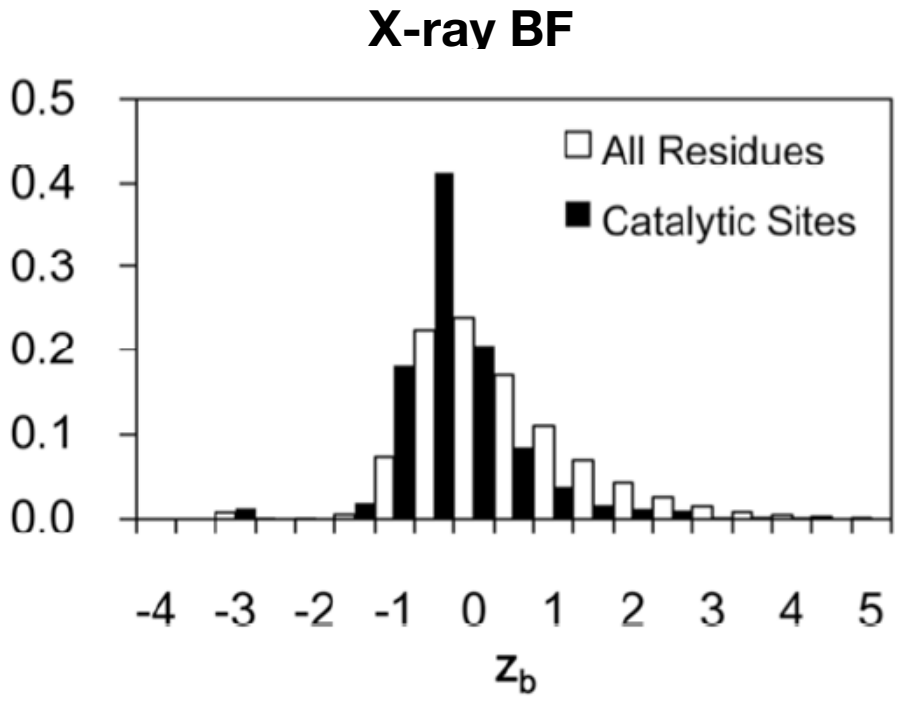


Catalytic residues

among the most considered residues

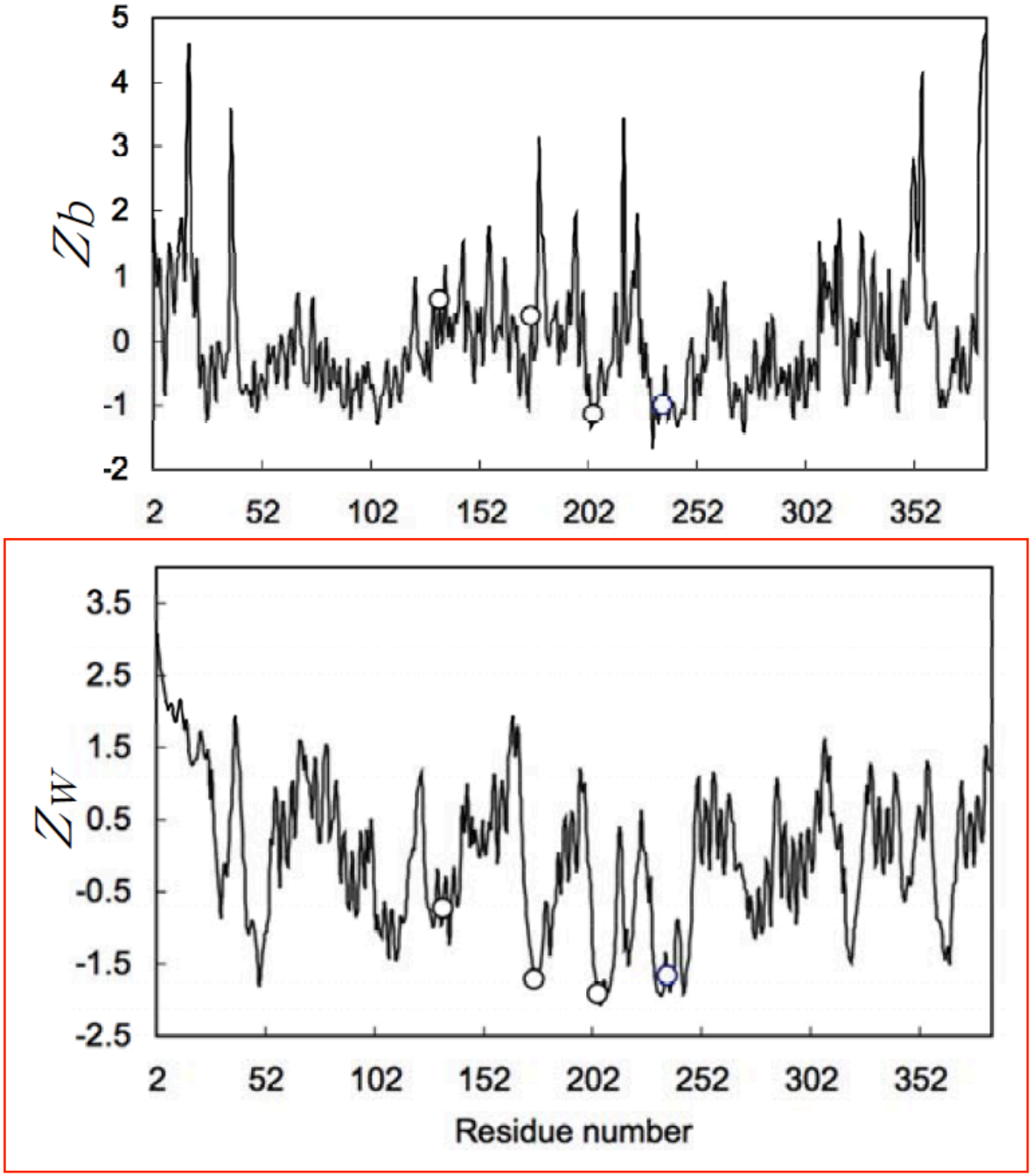
Comparison between WCN and X-ray BF distributions of catalytic residues vs. other residues



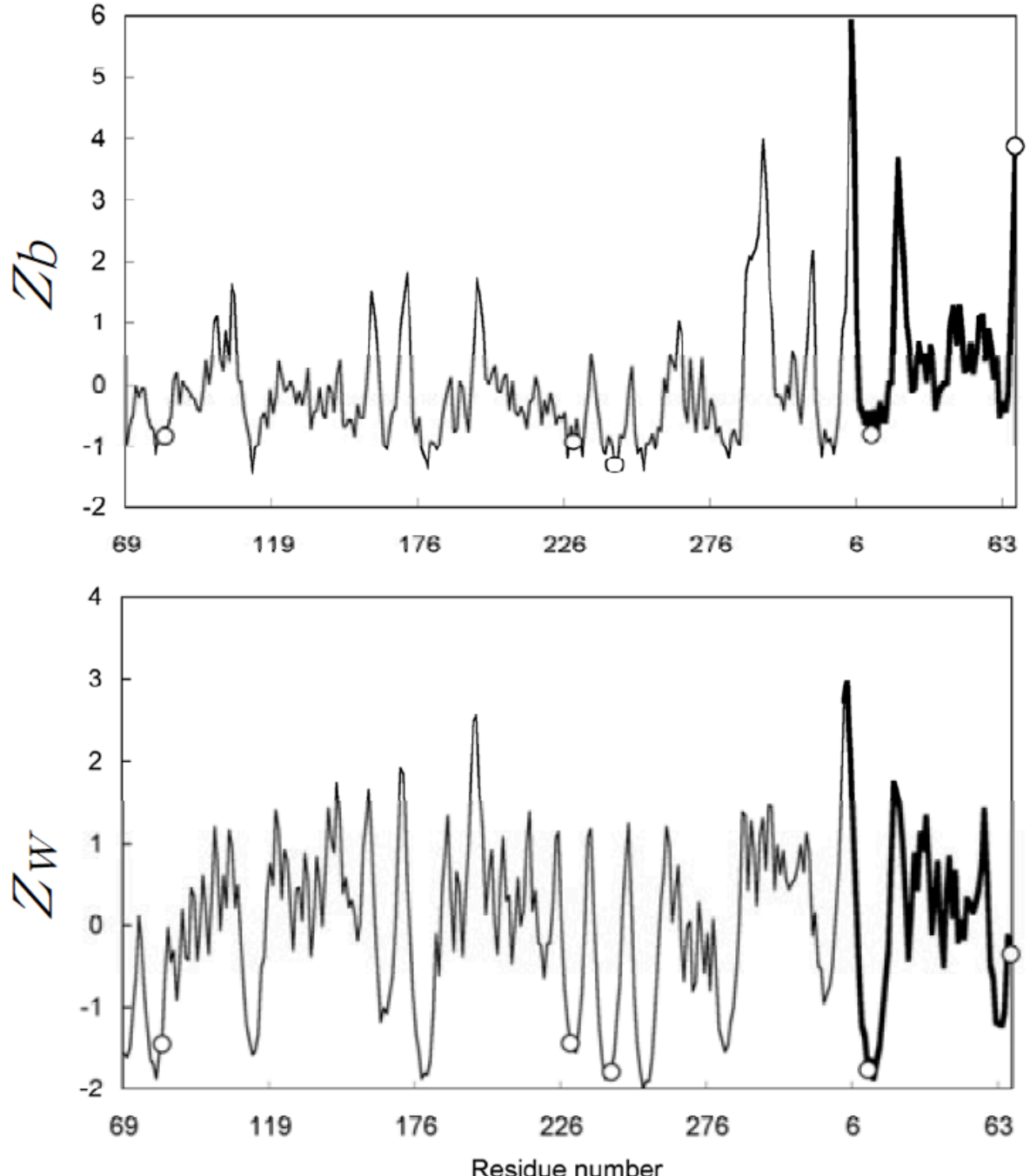


Not promising

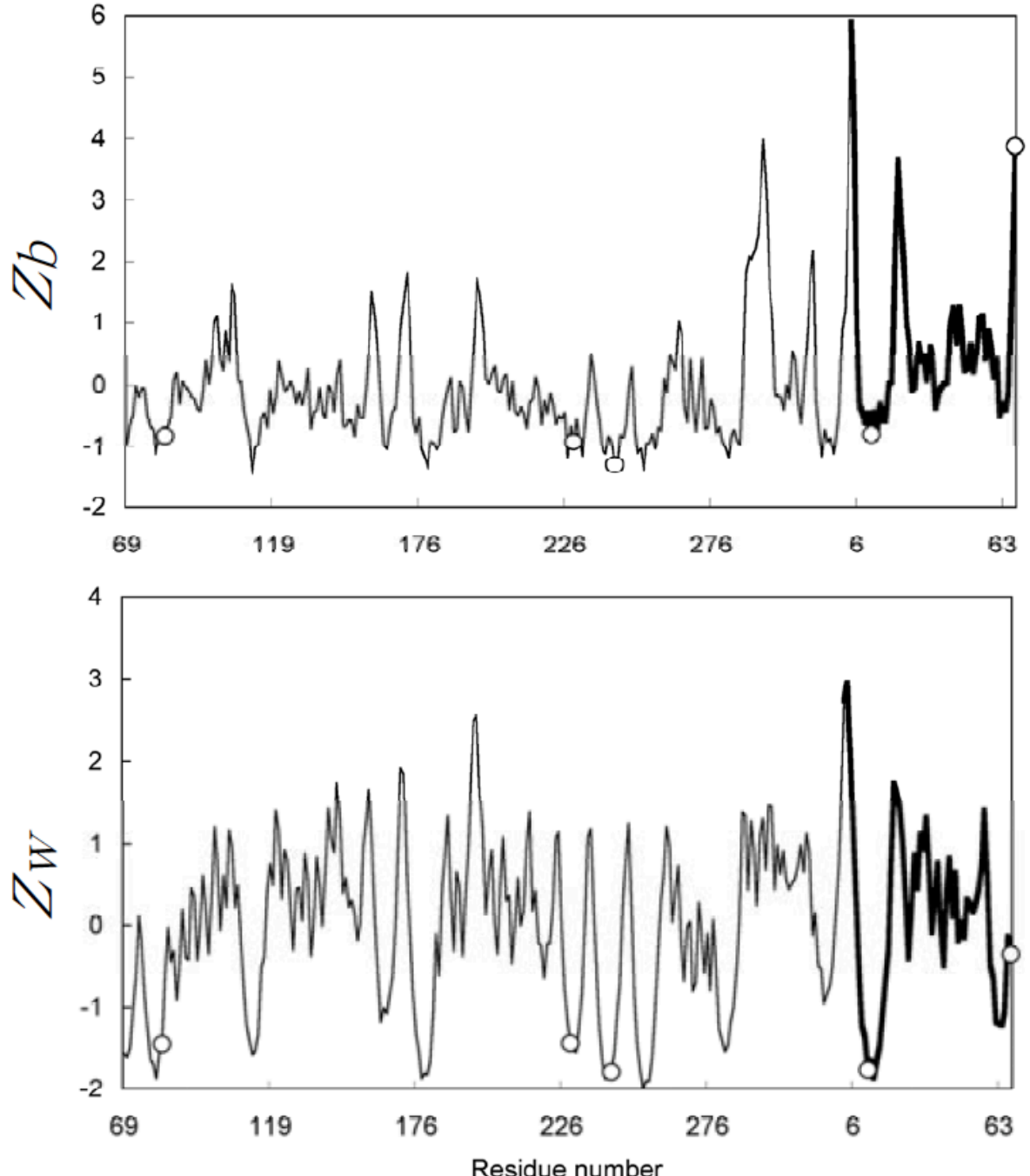
prediction of catalytic residues



prediction of catalytic residues



prediction of catalytic residues

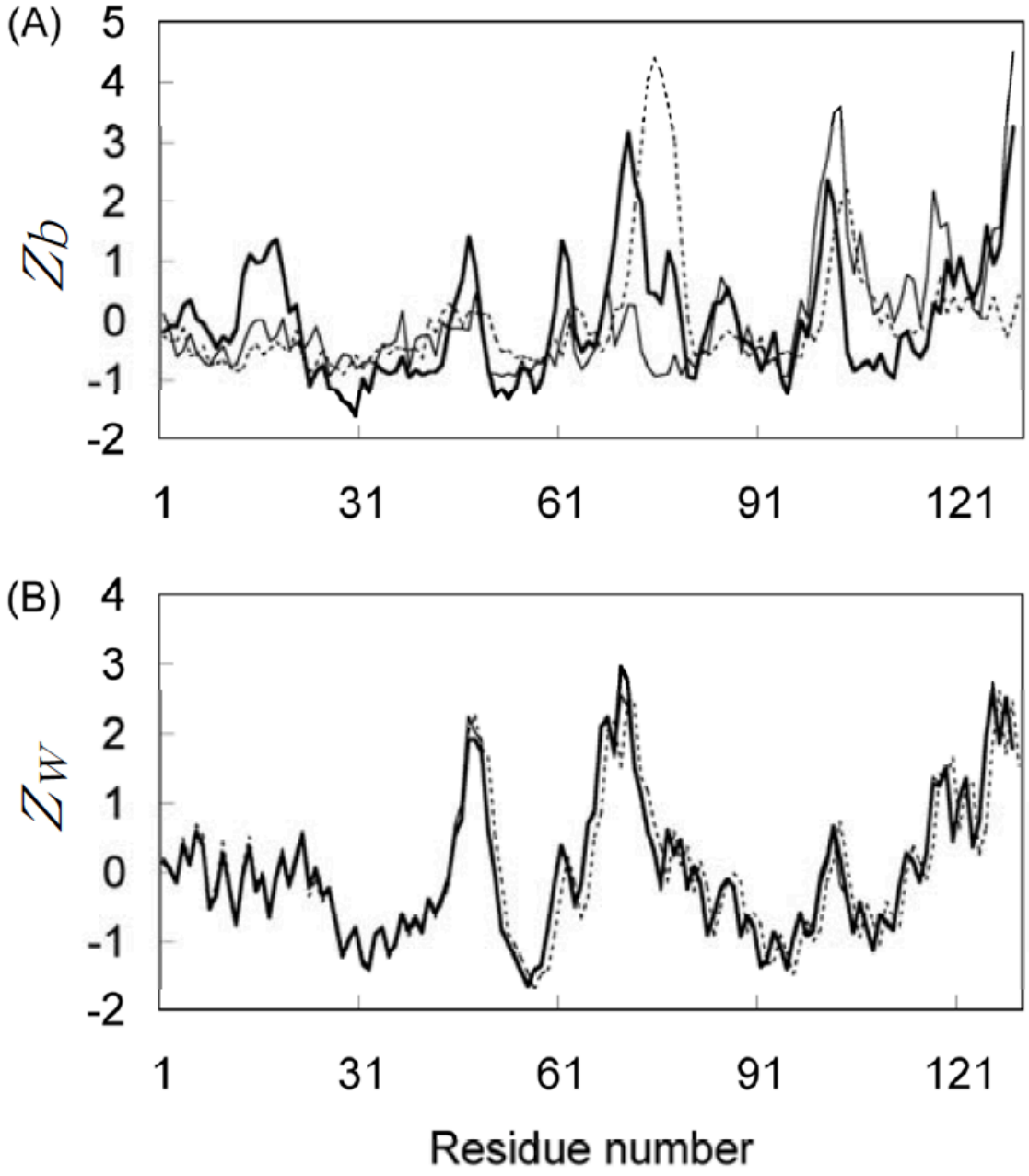


Why BF is not good at predicting catalytic residues?

If **B factors** appear to be correlated well with **WCN** and CN (published), and **WCN** is shown to be correlated well with **conservation scores** and is a good predictor of catalytic residues, then why **B-factors are such a poor predictor of catalytic residues**?

One of the reasons is that the values of X-ray B-factors are subject to the changes of many factors such as crystallization conditions, refinement methods, temperatures…

The x-rat B-factor profiles are different even for the same protein



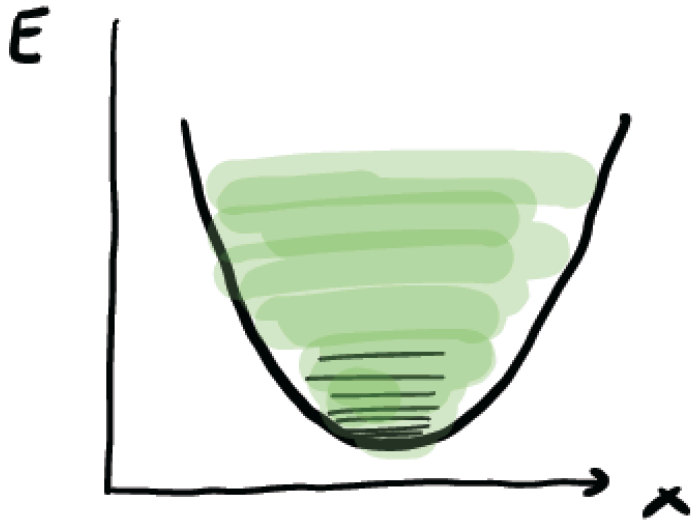
1. The **GNM-BF** is correlated with GNM-B factors and WCN and CN (published)

2. The WCN and CN is associated with conservation (published)

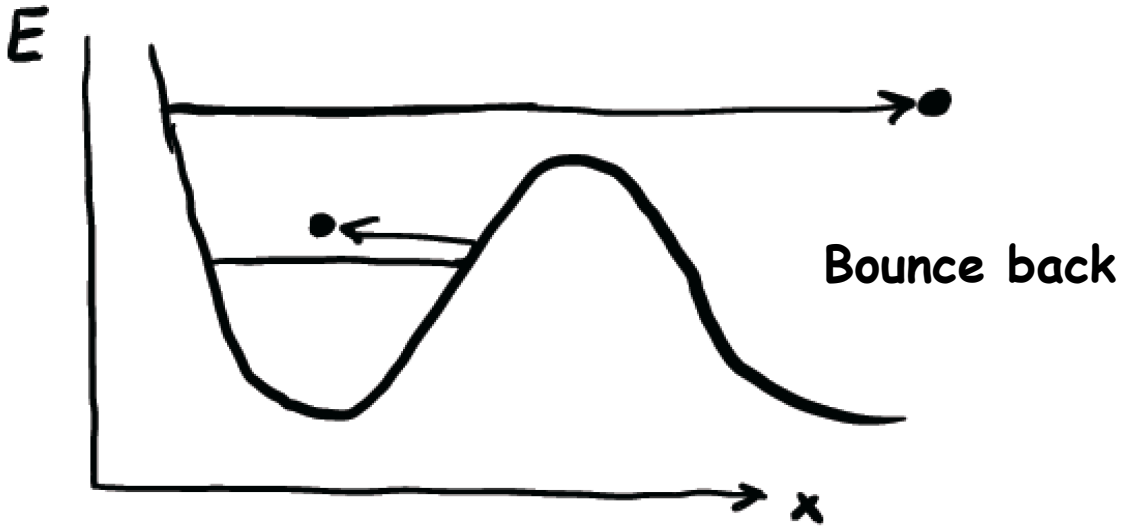
3. **Is GNM-BF associated with conservation – your final project?**

QM/MM simulation

Classical particles

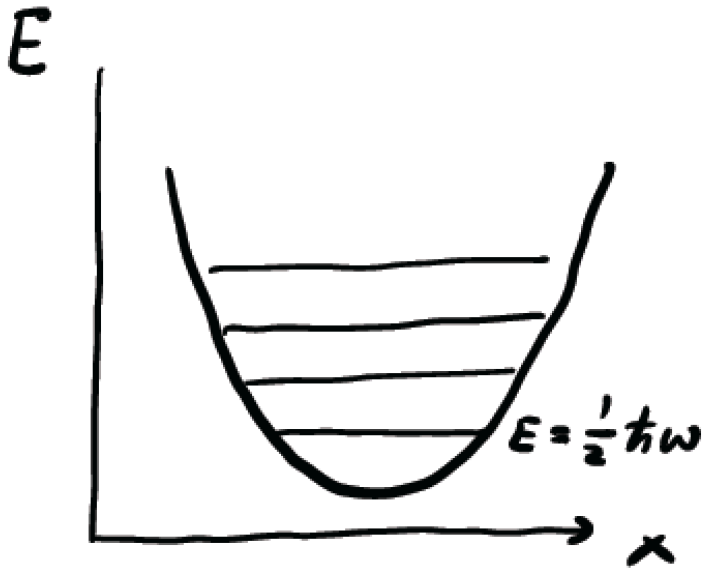


Continuous energy

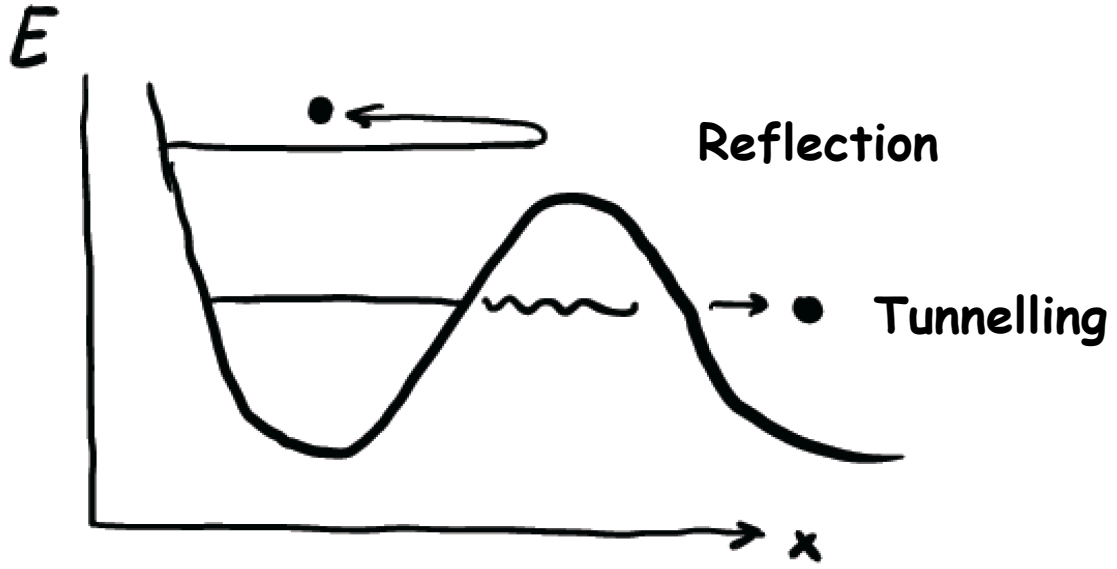


Bounce back

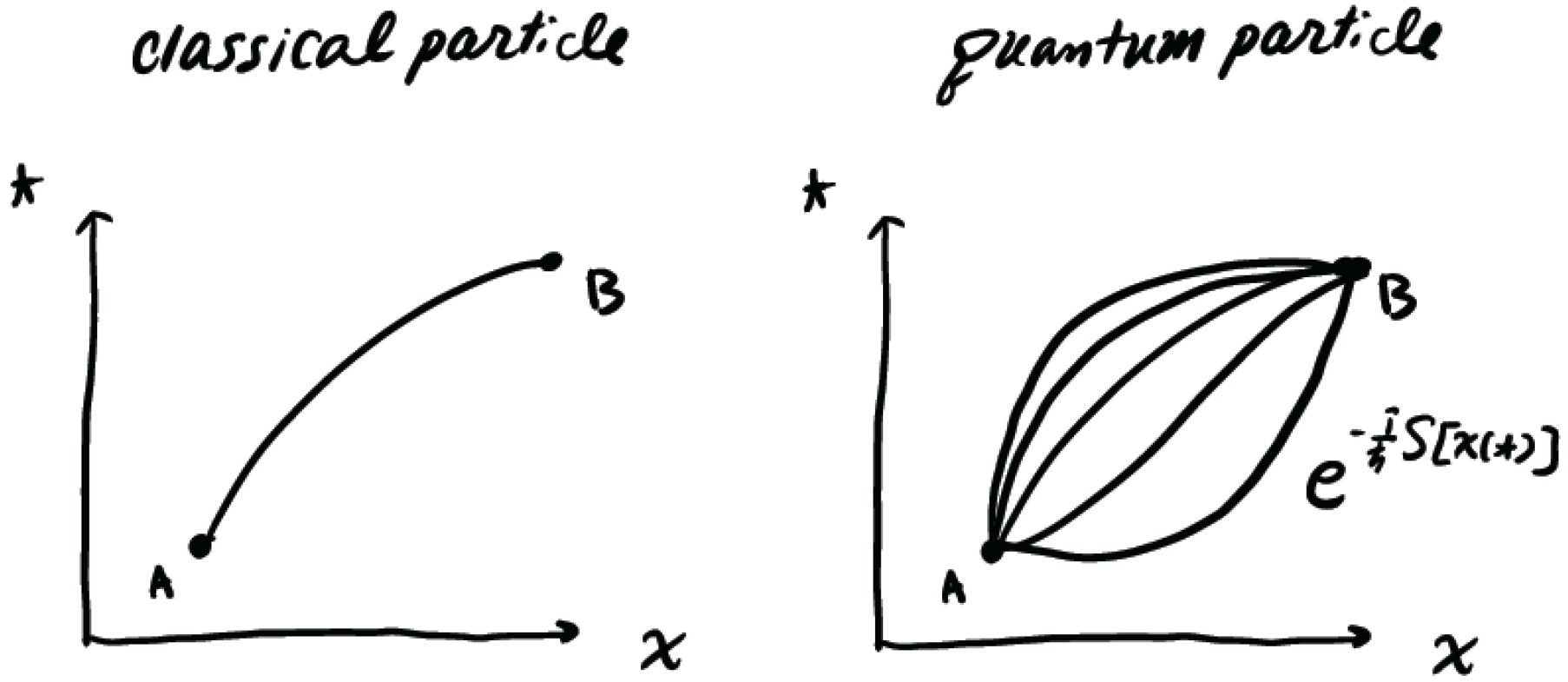
Quantum particles



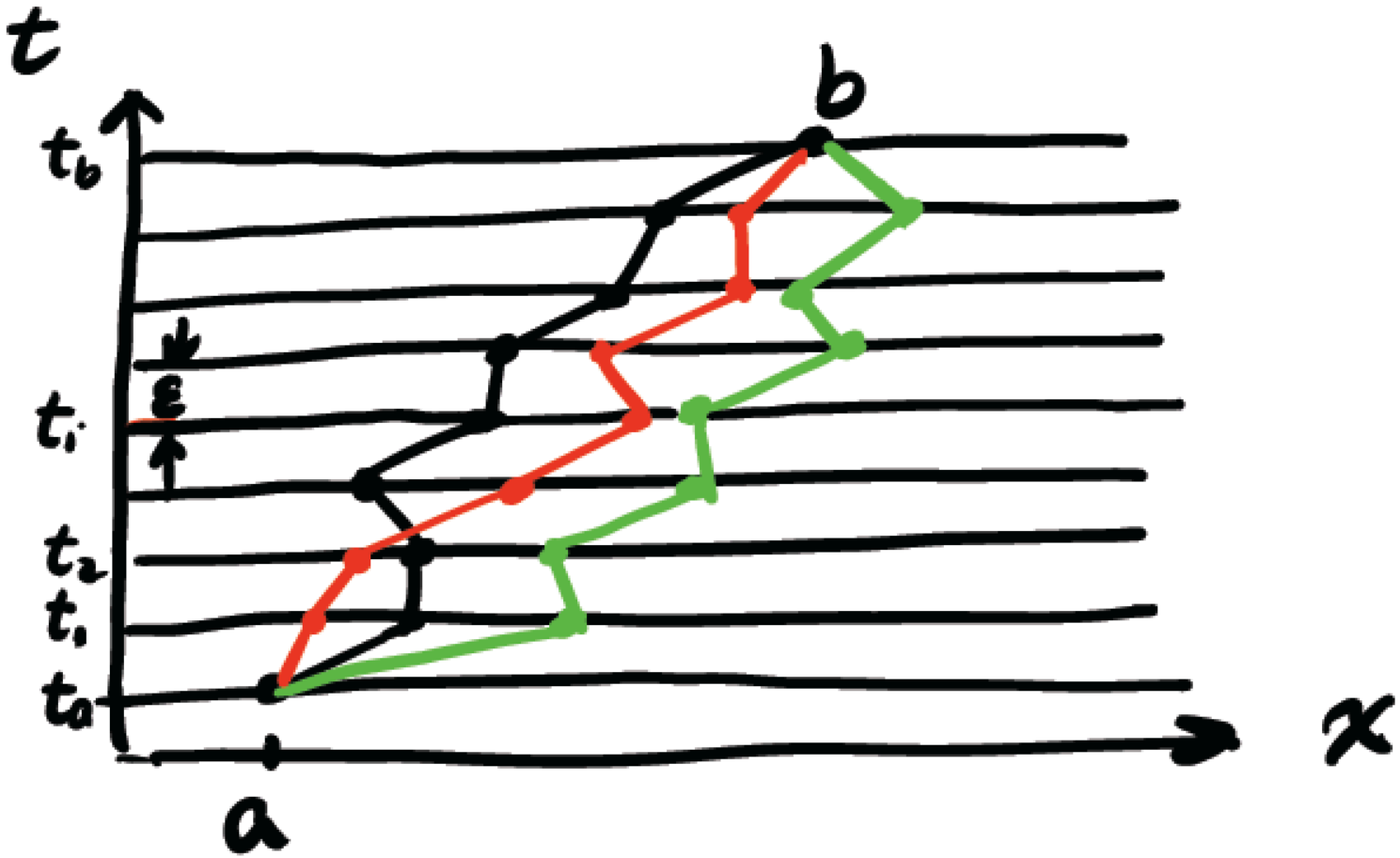
Zero-point energy



Feynman Path integral approach



Feynman Path integral approach



Double slit experiment

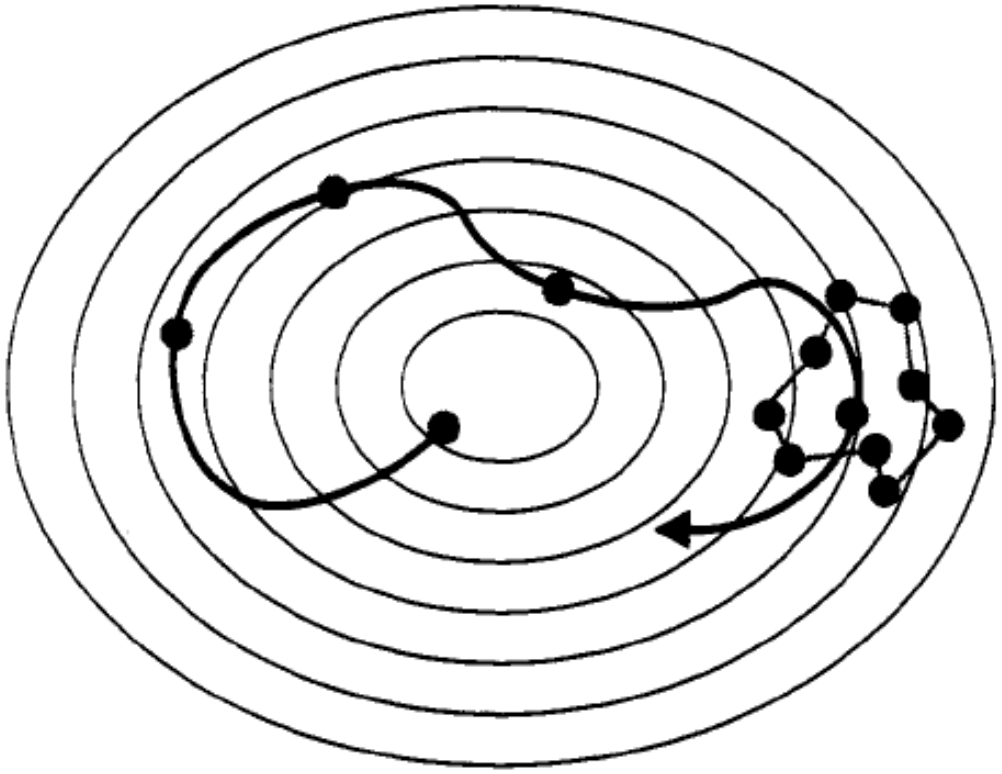
Double Slit Experiment explained! by Jim Al-Khalili

https://www.youtube.com/watch?v=A9tKncAdlHQ&frags=pl%2Cwn

Double Slit Experiment explained! by Dr. Quantum

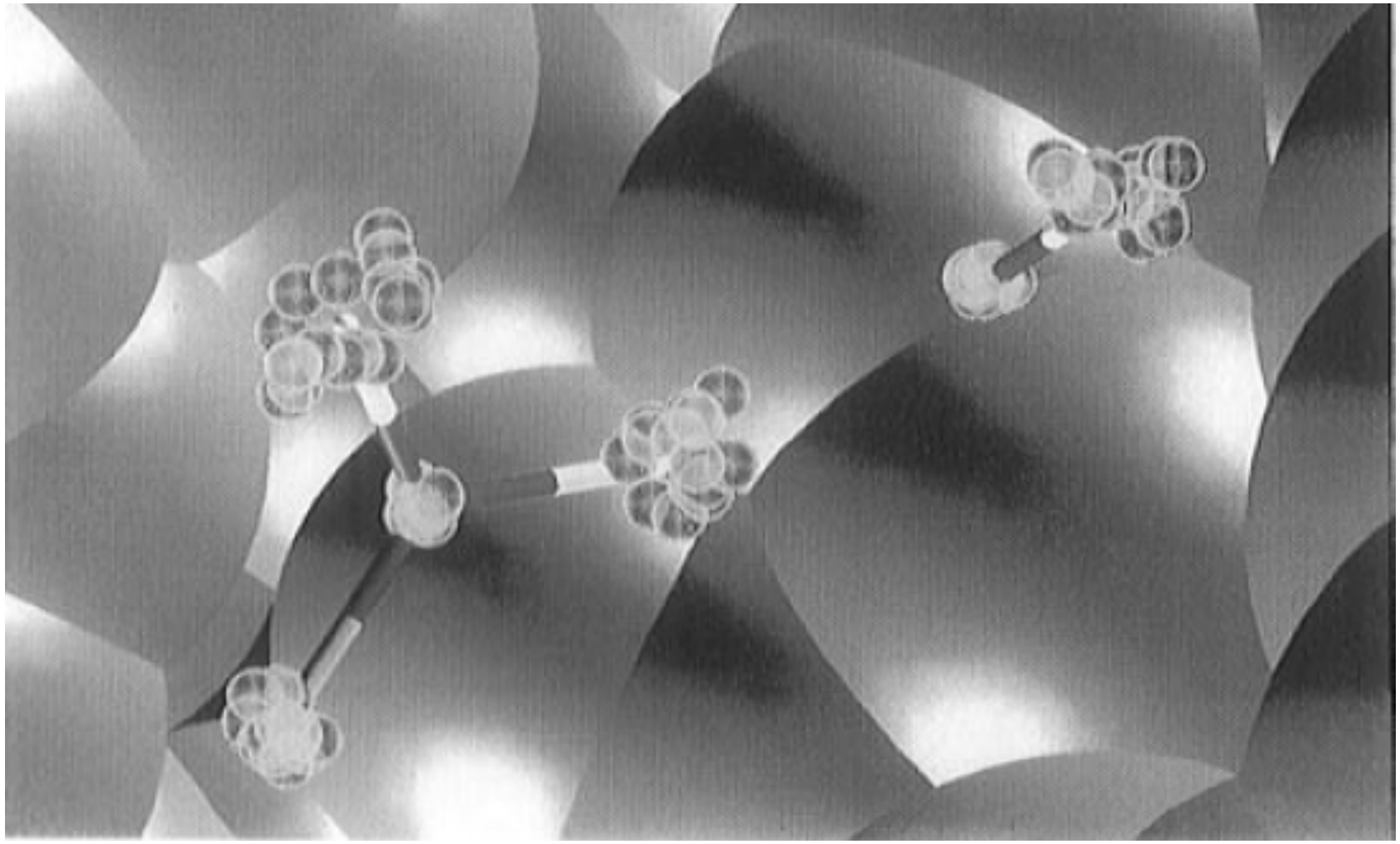
https://www.youtube.com/watch?v=DfPeprQ7oGc&frags=pl%2Cwn

Feynman Path integral

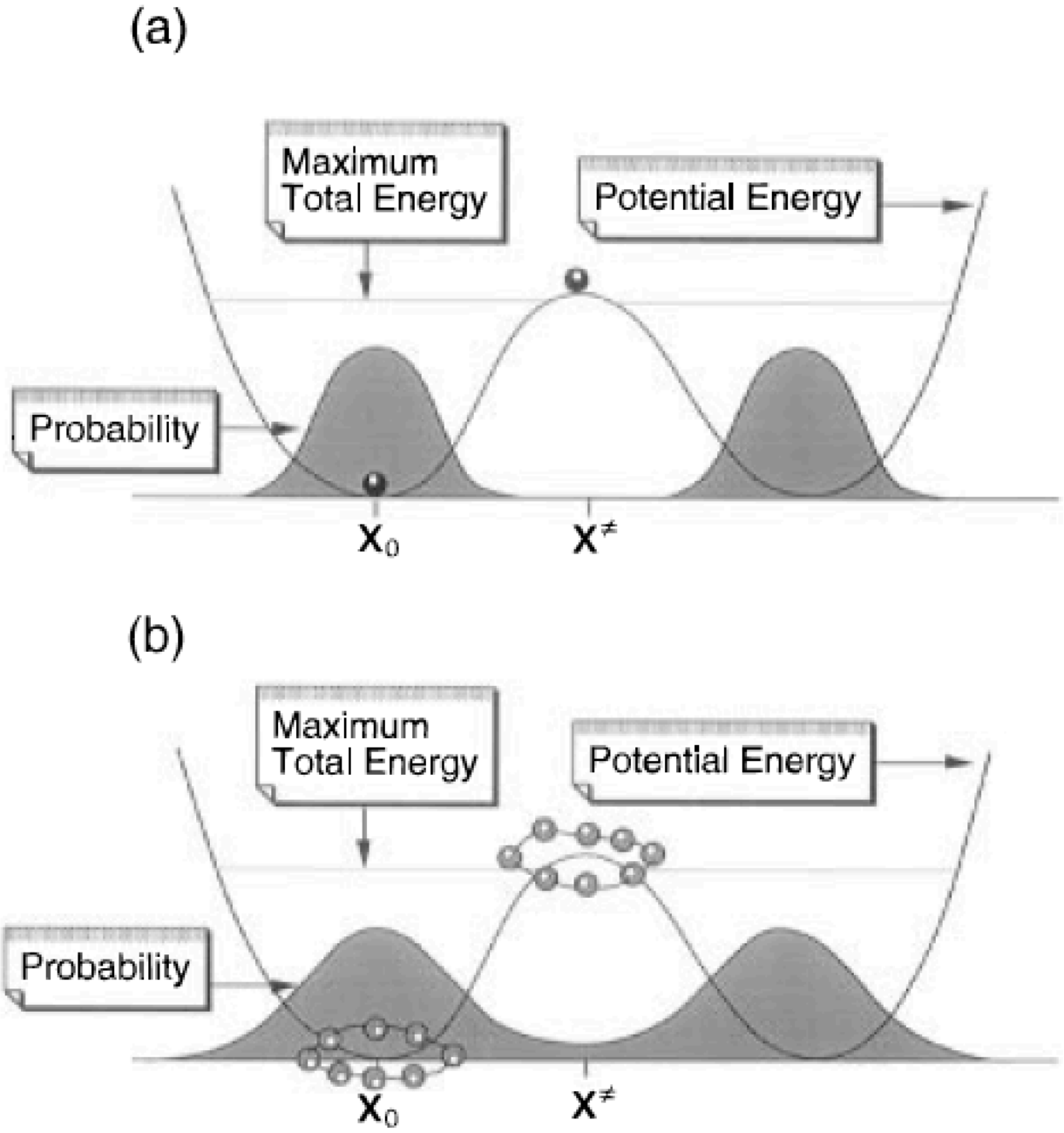


$$U_q = \frac{1}{p}\sum_{k=1}^{p}\frac{1}{2}M\Omega^2\Delta x_k^2 + U(x_k)$$

QM/MM simulation using Feynman path integral



J. Am. Chem. Soc. 1996, 118, 11745-11751


(a)
Maximum
Total Energy
Potential Energy
Probability
X_0
$X^{\neq}$
(b)
Maximum
Total Energy
Potential Energy
Probability
X_0
$X^{\neq}$