

# Advanced Sampling Techniques for Proteins

Anıl Kurut Şabanoğlu

13.03.2017

Monday Lunch Talks

Copenhagen

# My ultimate goal

## **To understand**

- *How* do proteins behave
- *Why* do proteins behave as they do
- What is the *reason* behind their *success* and *failure*
- What are the *consequences* of their failure

# What are Proteins?

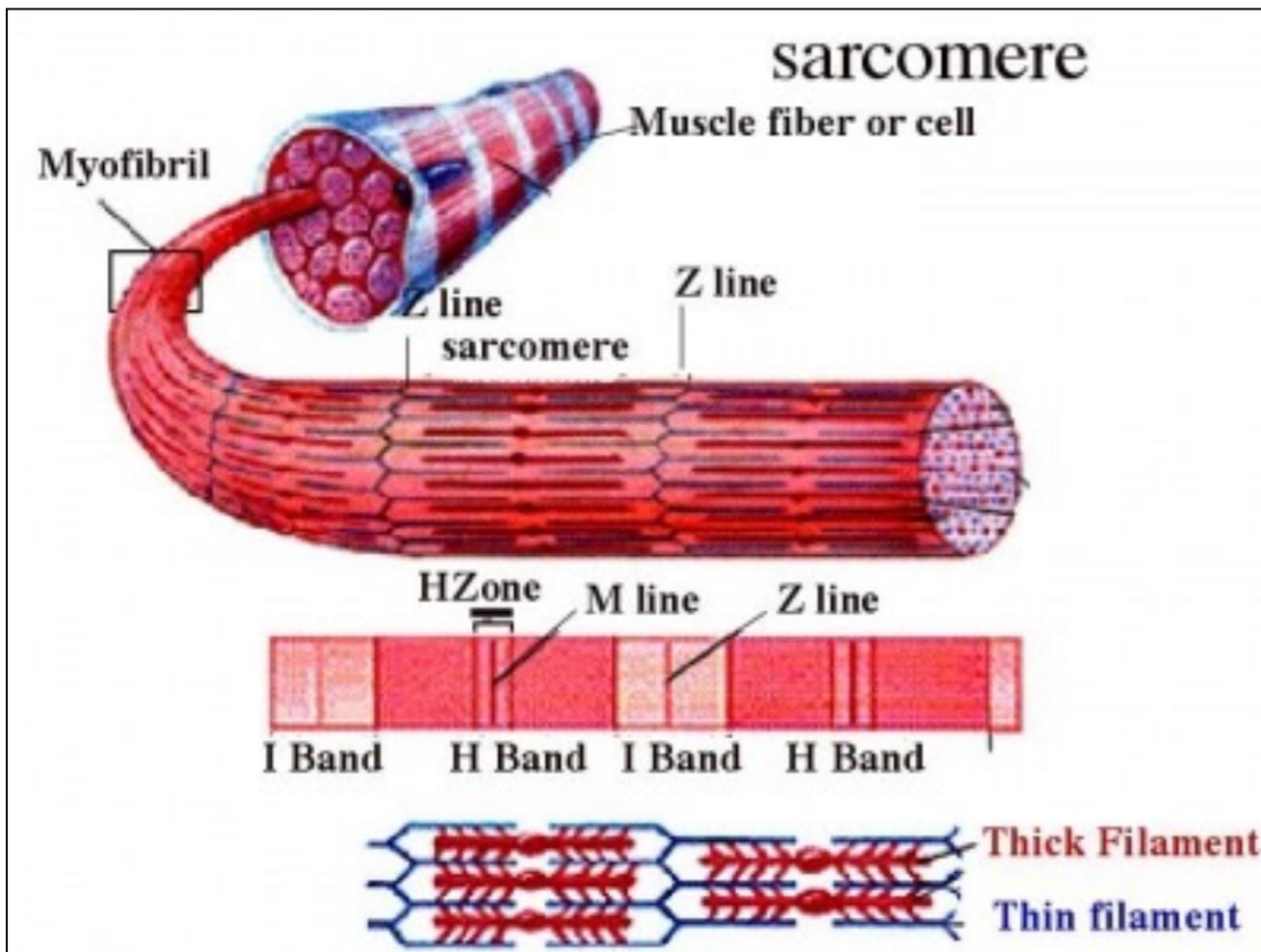
- Machineries of our body!
- Are responsible for almost all the processes
  - Body movement, transportation, structural support, communication, defence, synthesis and assembly of new machineries ... etc

# What are proteins?

- Muscle building v.s. protein:
- Muscle structure: Filaments of two proteins: Actin and Myosin

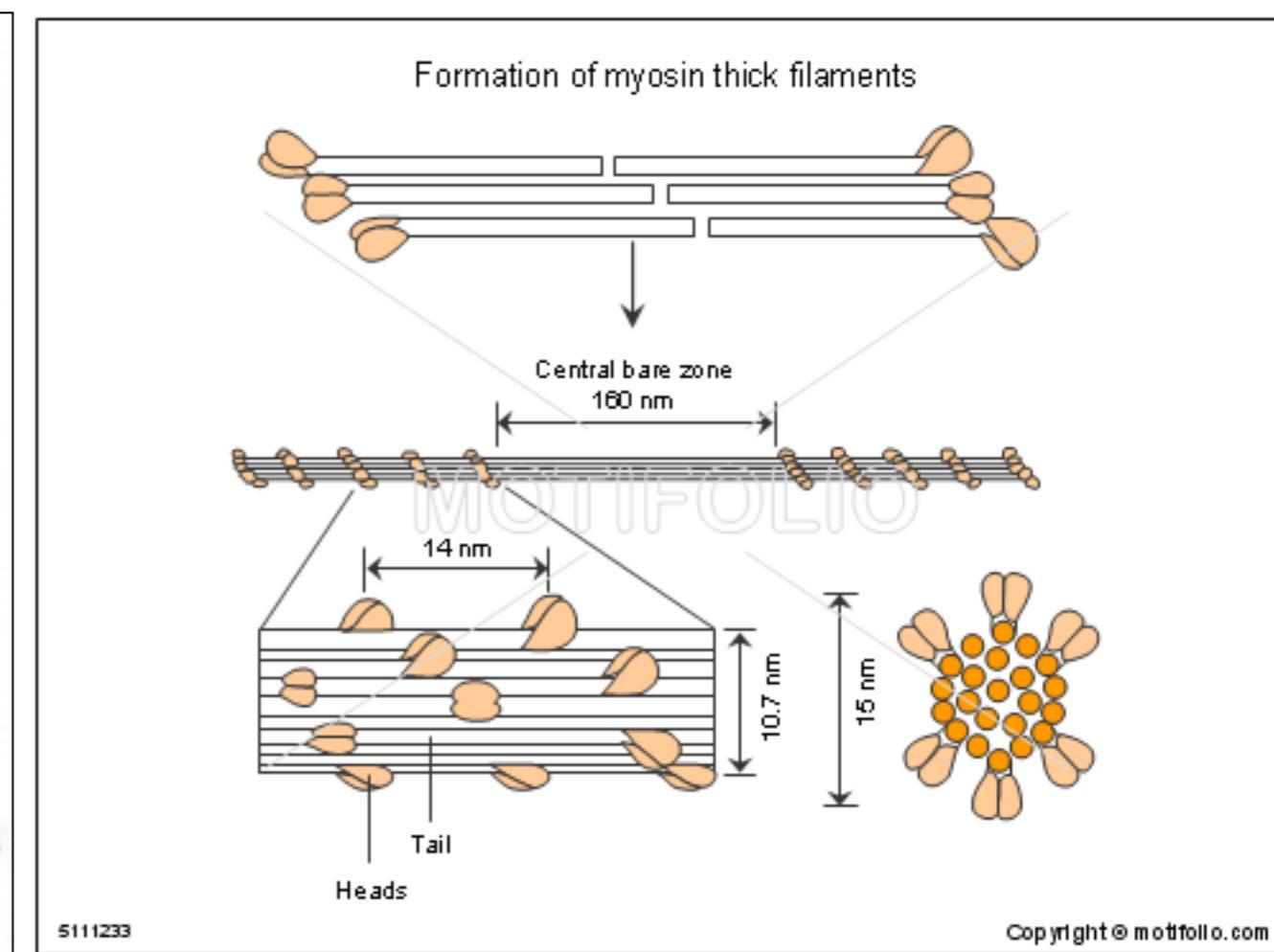
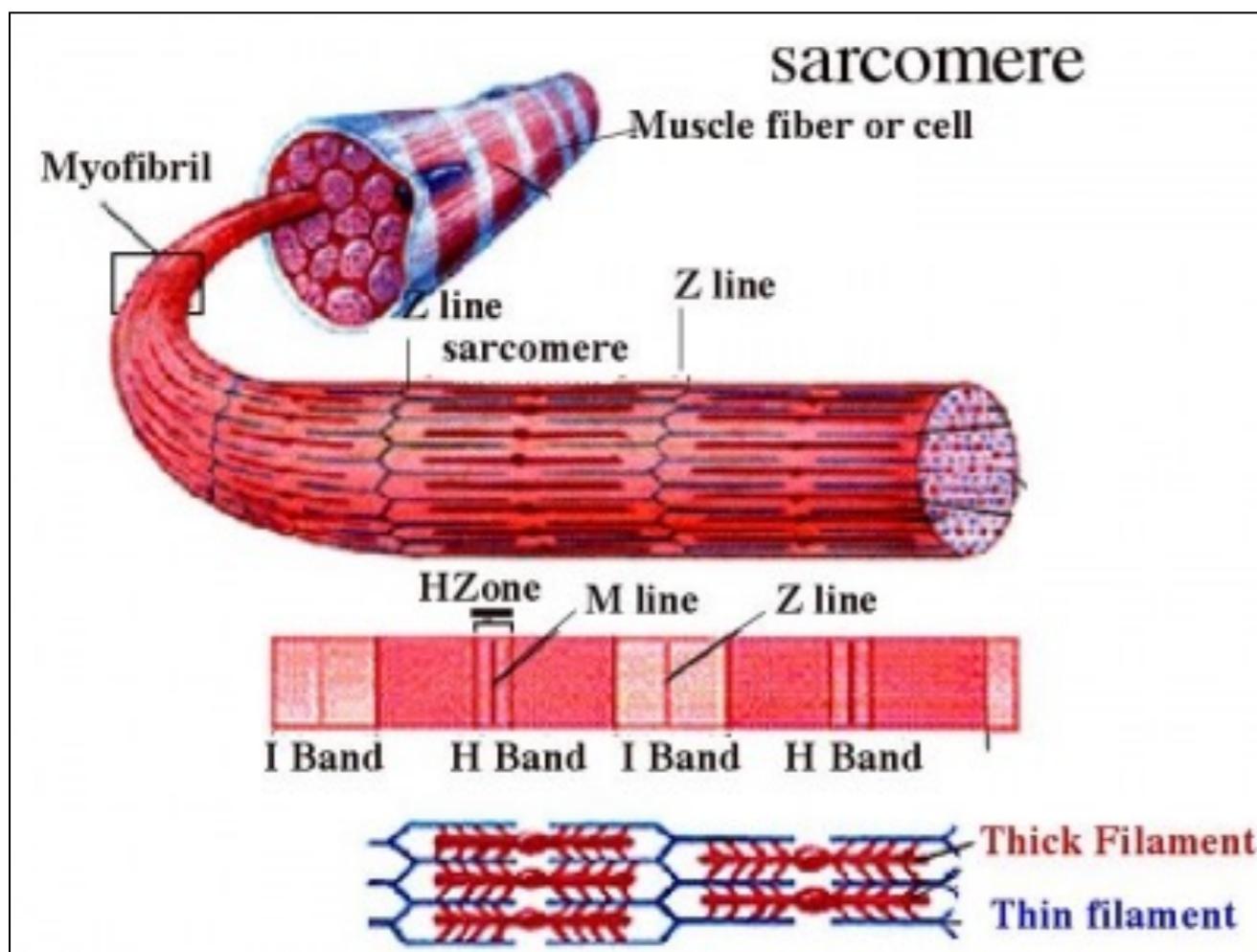
# What are proteins?

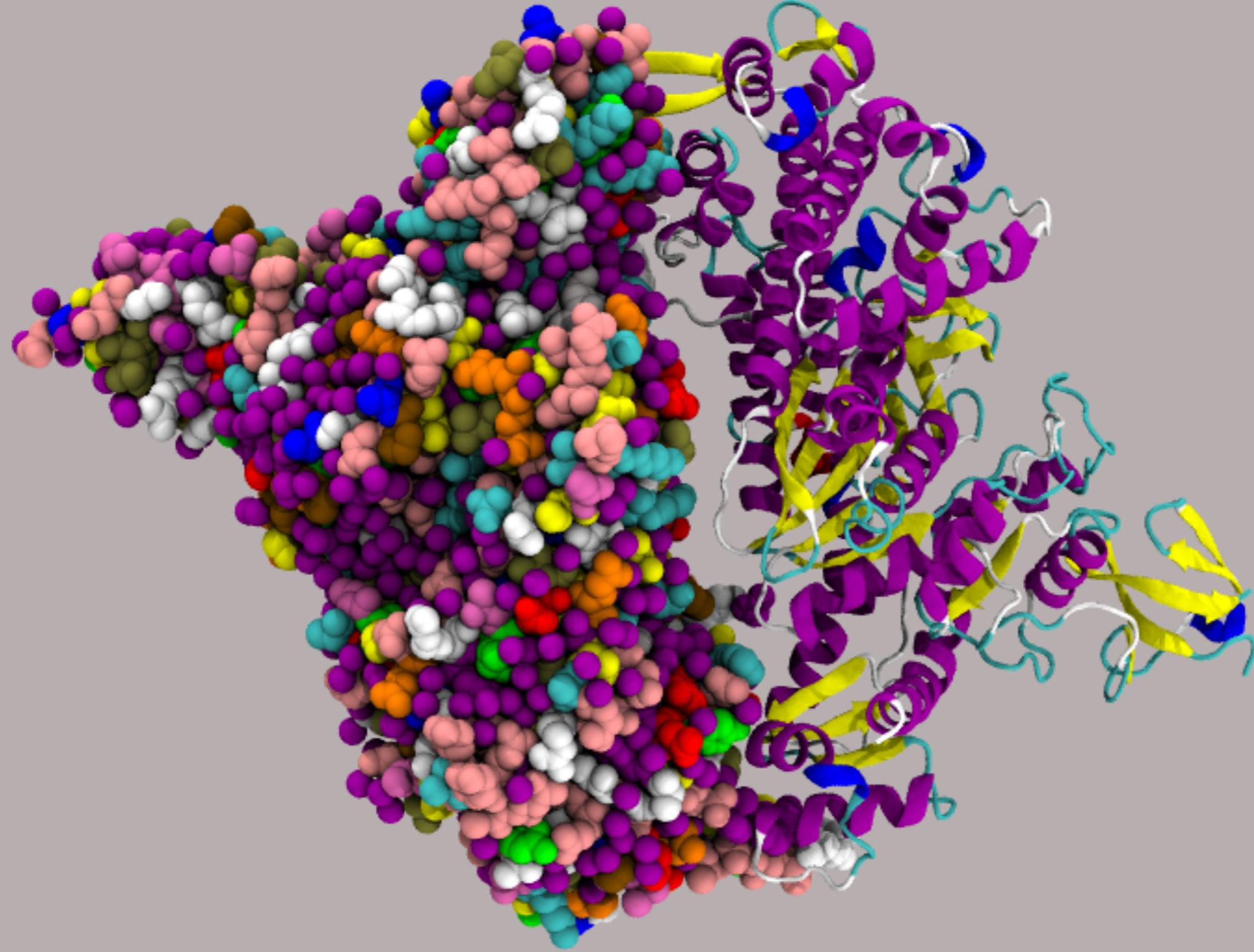
- Muscle building v.s. protein:
- Muscle structure: Filaments of two proteins: Actin and Myosin



# What are proteins?

- Muscle building v.s. protein:
- Muscle structure: Filaments of two proteins: Actin and Myosin





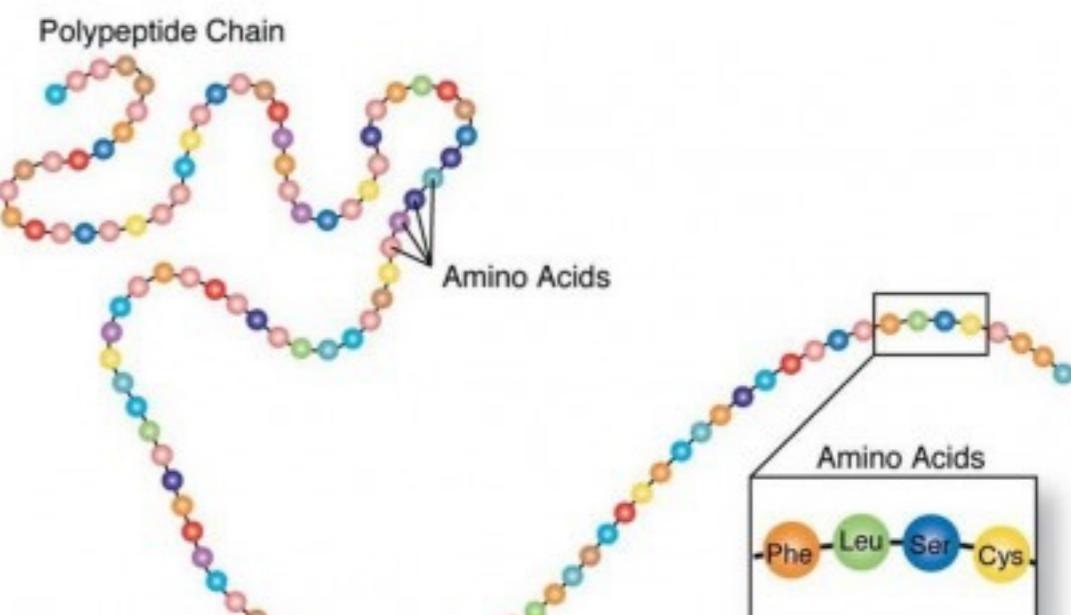
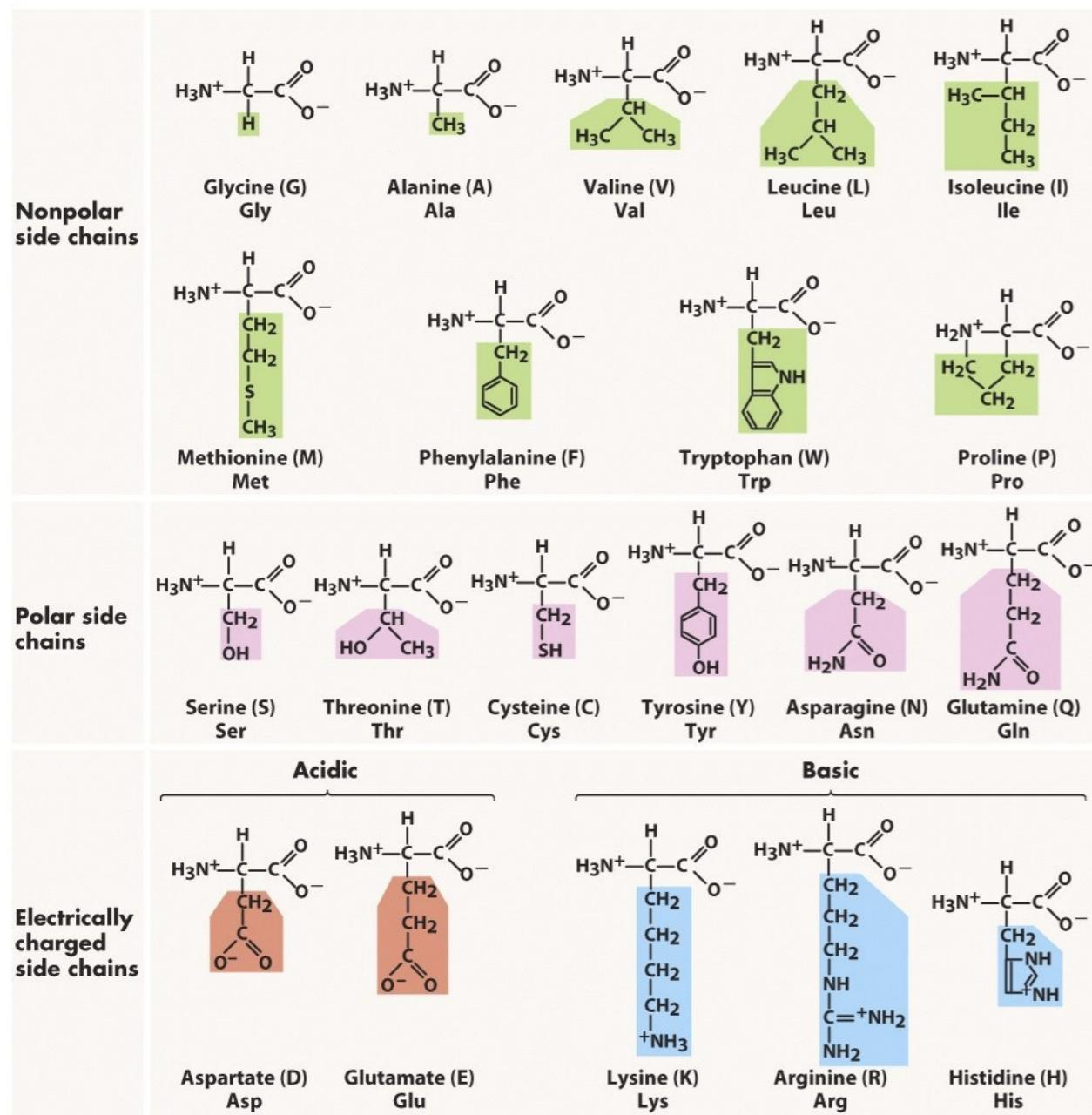
Myofibri



in



# Building Blocks of Proteins

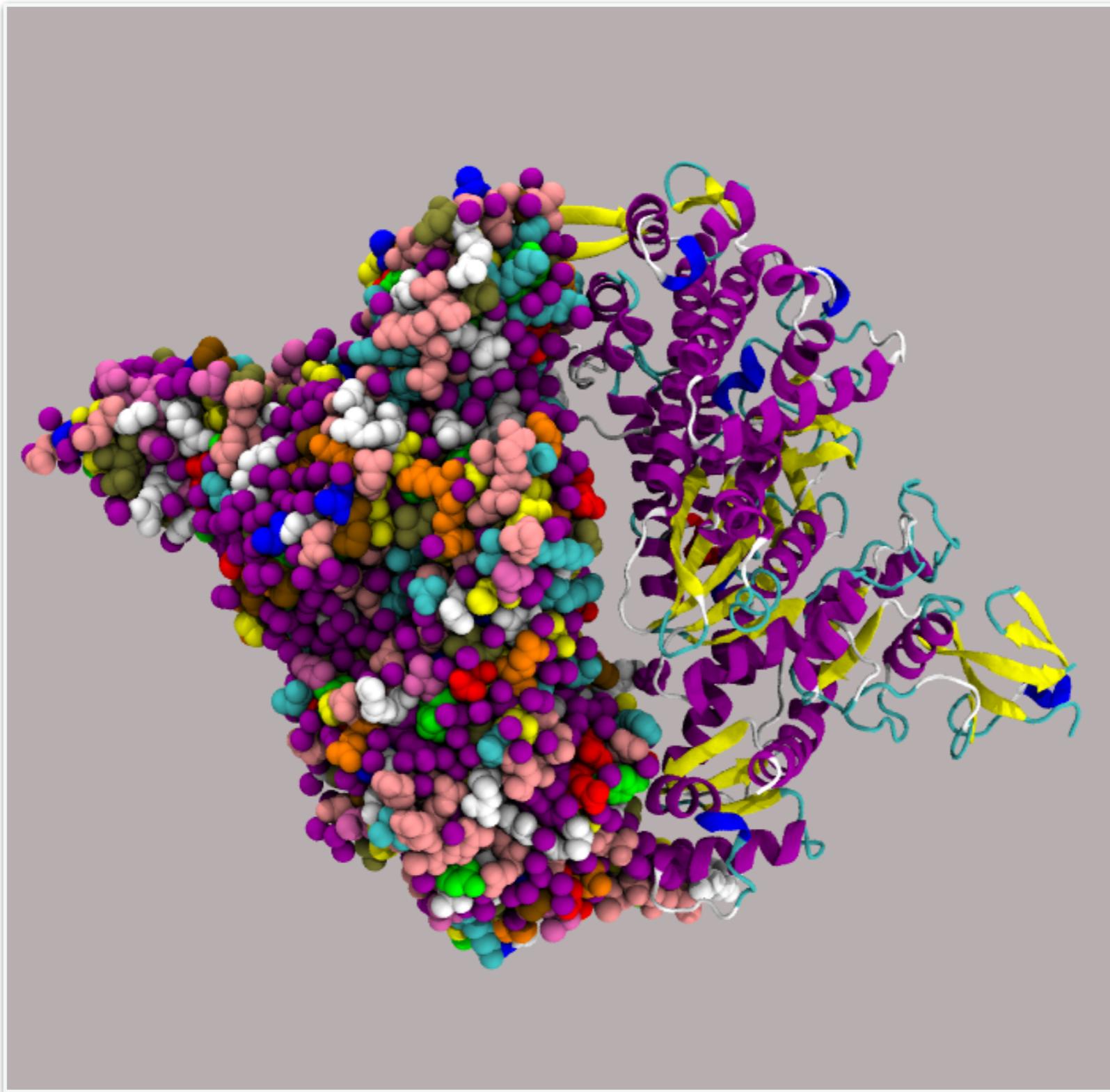


Amino Acids			
Ala: Alanine	Gln: Glutamine	Leu: Leucine	Ser: Serine
Arg: Arginine	Glu: Glutamic acid	Lys: Lysine	Thr: Threonine
Asn: Asparagine	Gly: Glycine	Met: Methionine	Trp: Tryptophane
Asp: Aspartic acid	His: Histidine	Phe: Phenylalanine	Tyr: Tyrosine
Cys: Cysteine	Ile: Isoleucine	Pro: Proline	Val: Valine

Figure 3-5 Biological Science, 2/e

© 2005 Pearson Prentice Hall, Inc.

# Muscle protein: Myosin



# Muscle protein: Myosin

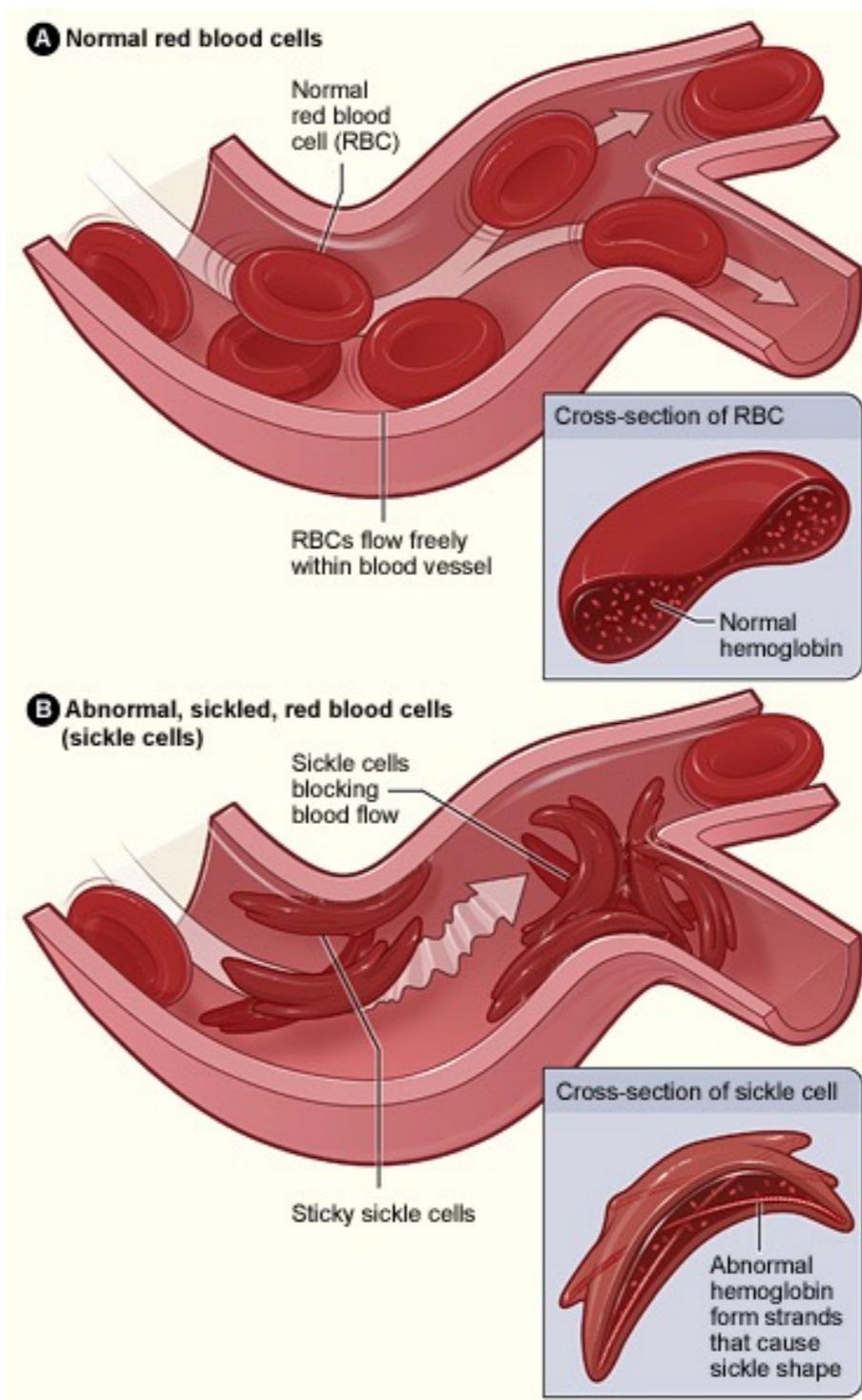
*MDNFFTEGTRVWLRENGQHFPSTVNCAEGIVVFRTDYG  
QVFTYKQSTITHQKVTA  
MHPTEEGVDDMASLTELHGGSIMYNLFQRYKRNQIYTYIGSILASVNPYQPIAGLYEPATMEQ  
YSRRHLGELPPHIFAIANE  
CYRCLWKRHDNQCILISGESGAGKTESTKLILKFLSVISQQSLELSLKEKTSCVERAILESS  
PIMEAFGNAKTVYNNNNSSRF  
GKFVQLNICQKGNIQGGRI  
VDYLLEKNRVVRQNPGERNYHIFYALLAGLEHEEREELYLS  
STPENYHYLNQSGC  
VEDKTISDQESFREVITAMDVMQFK  
EEVREVSRLLAGILHLGNIEFITAGGAQVSFKTALGRSAEL  
LGLDPTQLTDALTQRSMFLRGEEILTPLNVQQAVDSRDS  
LAMALYACCFEWVIKKINSRIKG  
NEDFKSIGILDIFGFENFE  
VNHF  
EQFNINYANEKLQEYFNKHIFSLEQLEY  
SREGLV  
WEDIDWIDNGECLDLIEKKLGLLALINEESHFPQATDSTLL  
EKLHSQHANNHFYVKPRVA  
VNNFGVKHYAGEVQYDVRI  
LEKNRDTFRDDLNL  
LRESRFDFIYDLFEHVSSRNNQDTLKCGSKHRRPTVSSQFKDSLHSLMATLSSNPFFVRCIKN  
MQKMPDQFDQAVVLNQLRYSGMLET  
VRIRKAGYAVRRPF  
QDFYKRYKVL  
MRNLALPEDVRGKCTSLLQLYDASNSEW  
QLGKT  
KVFLRESLEQKLEKRREEE : 741 residue*

# What can go wrong?

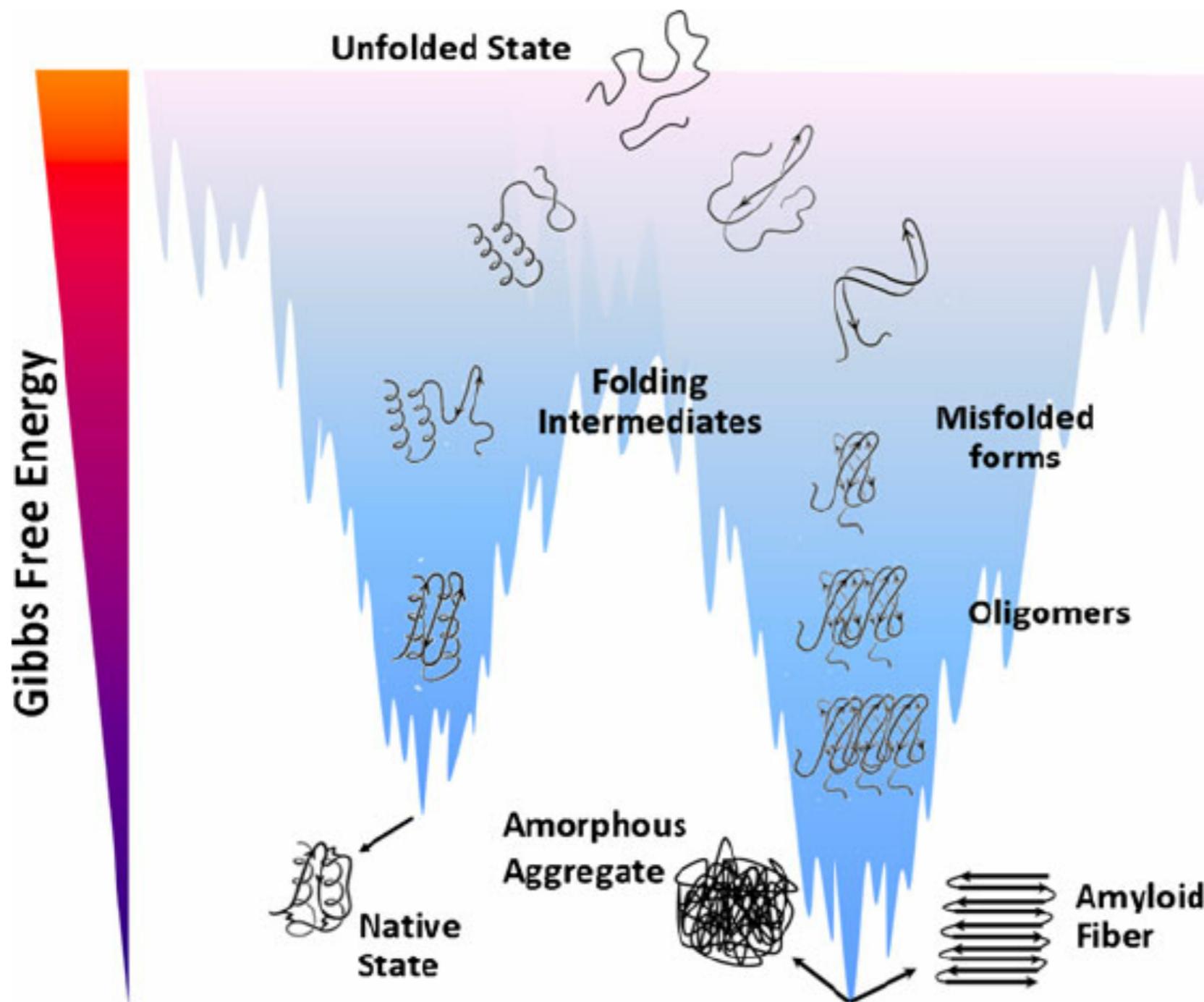
- Mutation in one of the residue:
- Misfolding: Become malfunctional, destructive structures
- Sickle-cell anaemia:

# What can go wrong?

- Mutation in one of the residue:
- Misfolding: Become malfunctional, destructive structures
- Sickle-cell anaemia:

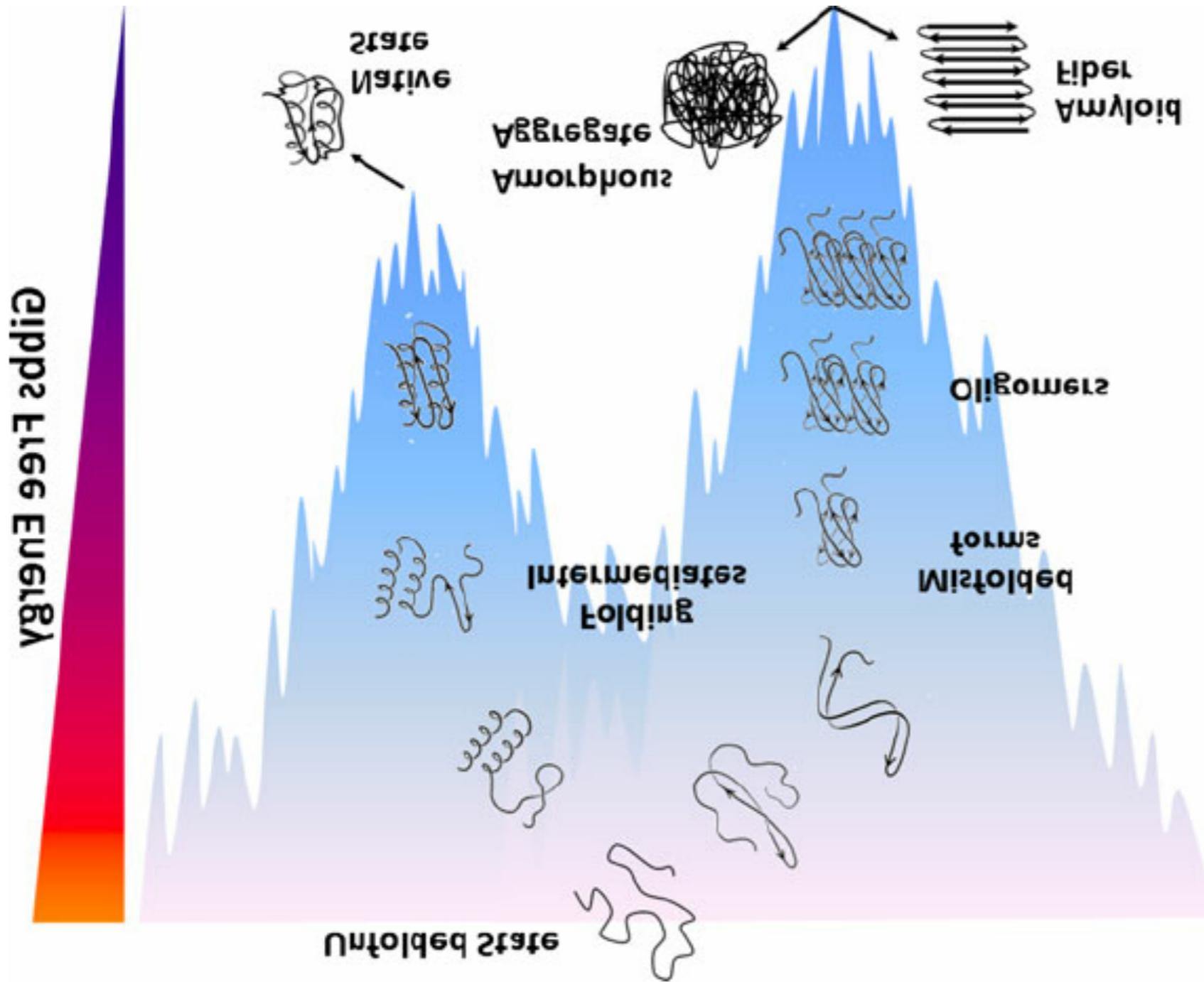


# Distribution of all possible states



$$P_i = \frac{\exp(-G_i/k_B T)}{\sum P_j \text{ all } j}$$

# Distribution of all possible states



$$P_i = \frac{\exp(-G_i/k_B T)}{\sum P_j \text{ all } j}$$



Input:

1) Color of a pixel in the picture:

**Sequence of the protein**

2) Average color in the picture

**Experimental measurements are averaged over all states**

3) Interaction model of the pixels

**Physics**



**My task is: Predict distribution of colors in the picture**



Input:

1) Color of a pixel in the picture:

Sequence of the protein

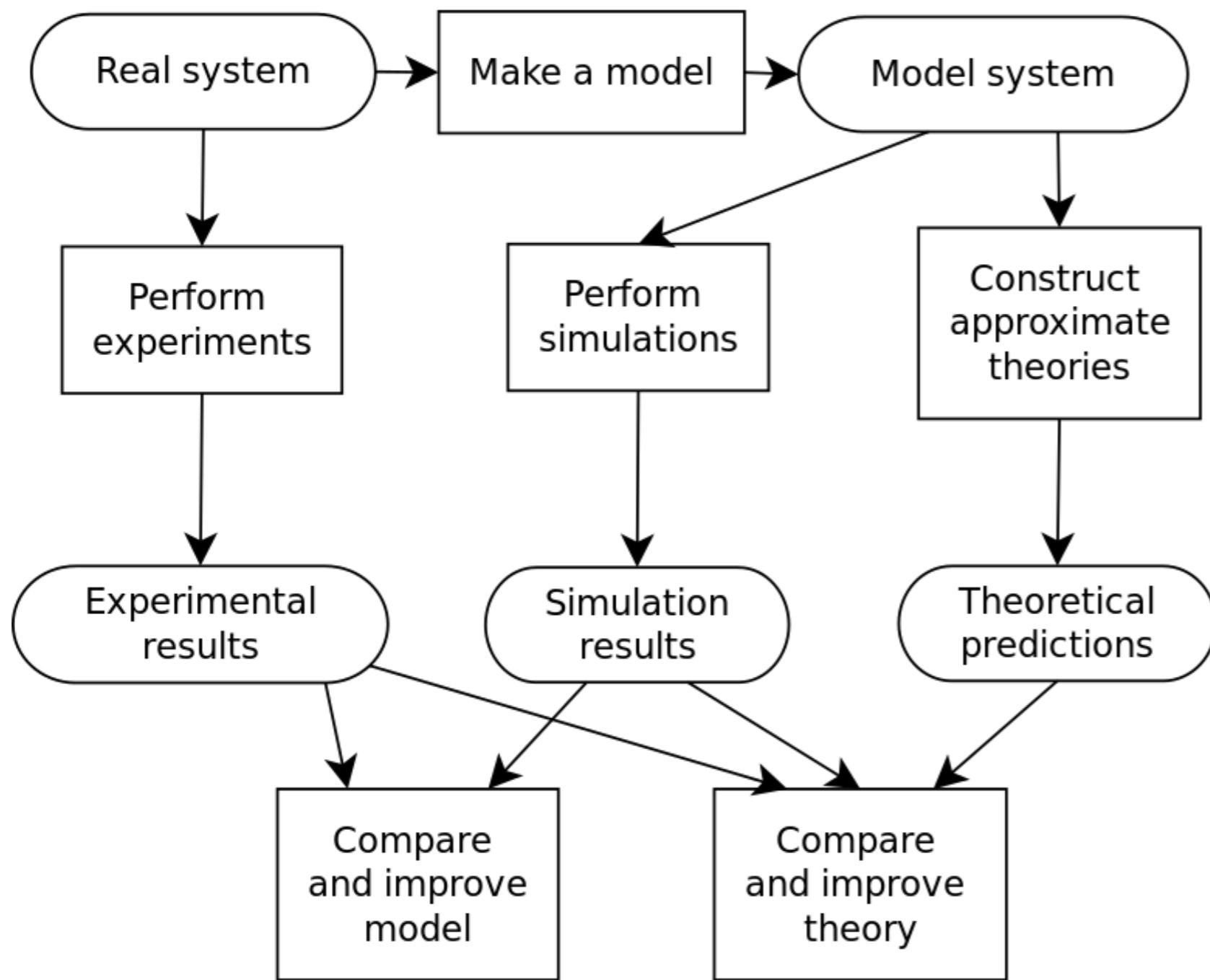
2) Average color in the picture

Experimental measurements are averaged over all states

3) Interaction model of the pixels

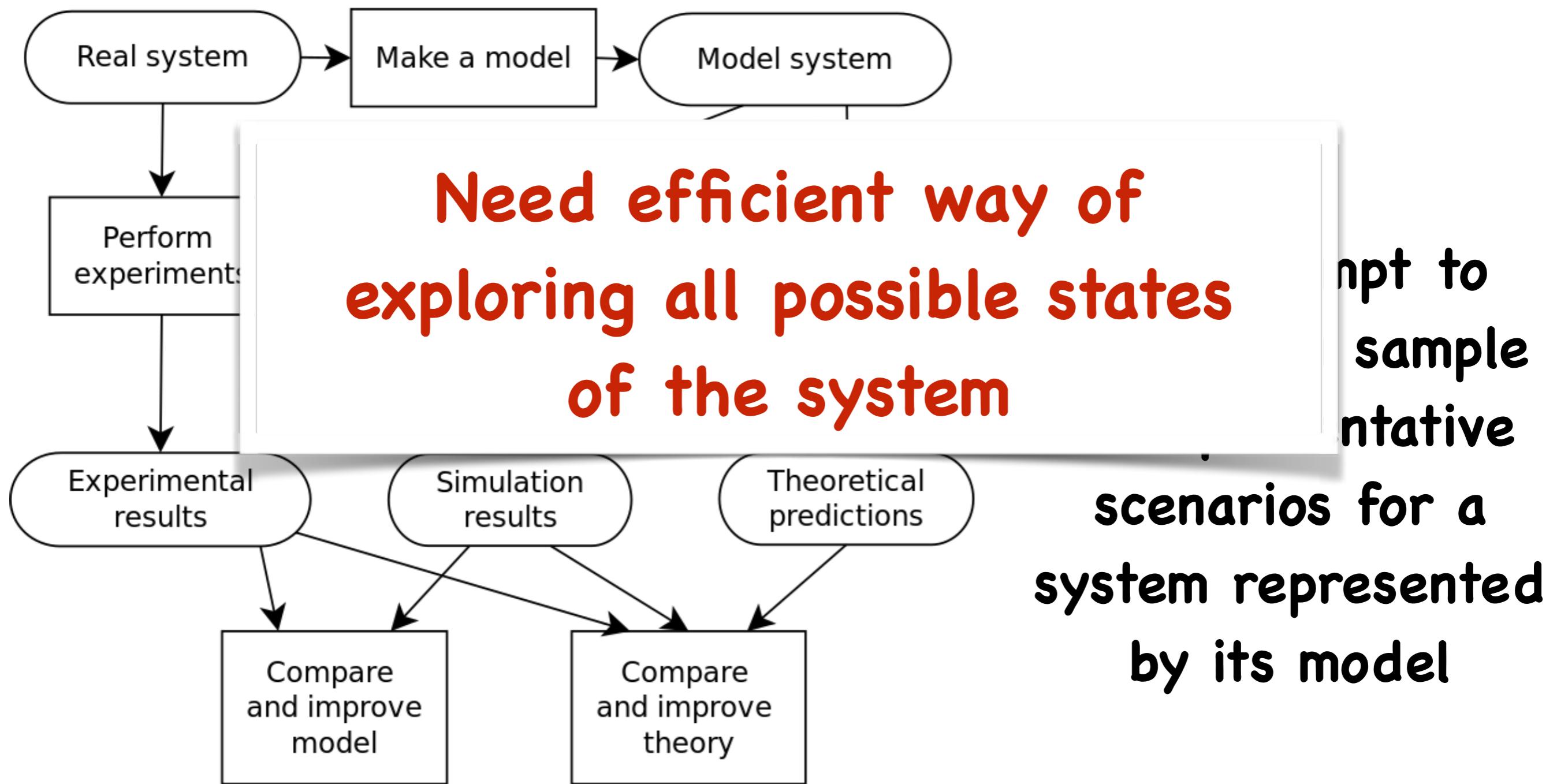
Physics

# Molecular simulations



- **The attempt to generate a sample of representative scenarios for a system represented by its model**

# Molecular simulations

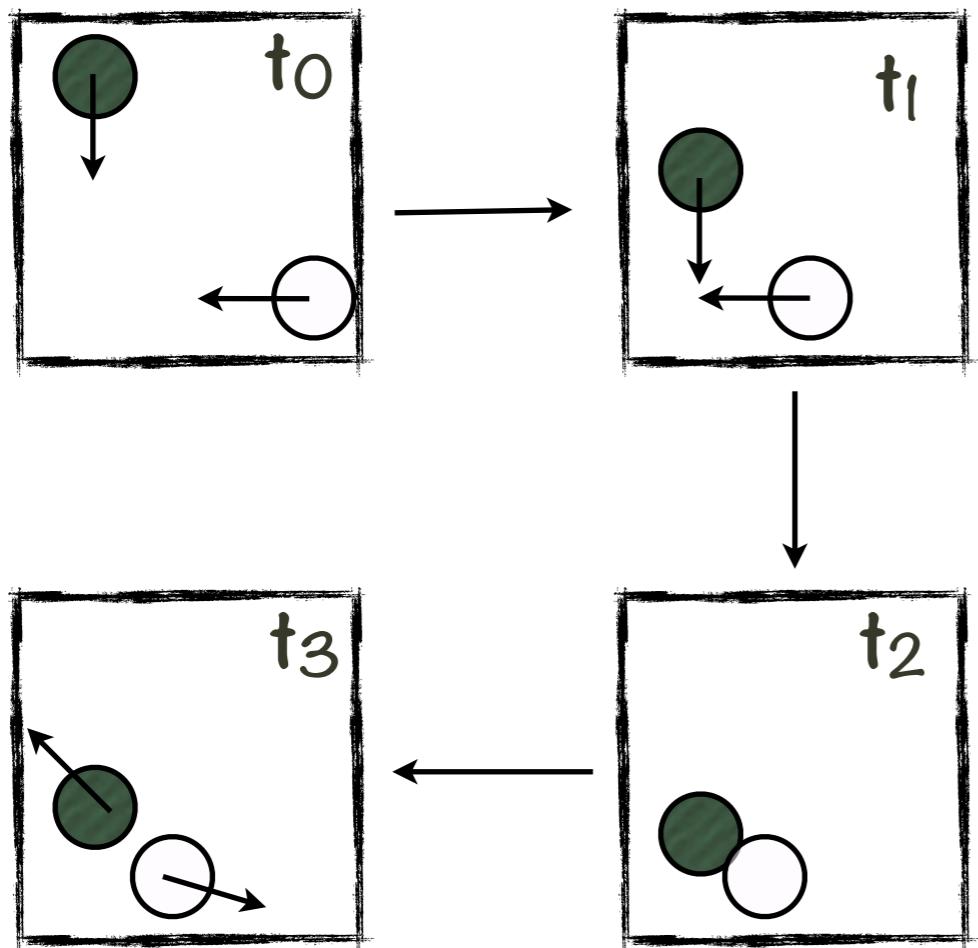


# Conventional Methods

## Molecular Dynamics

(MD):

Time averaging

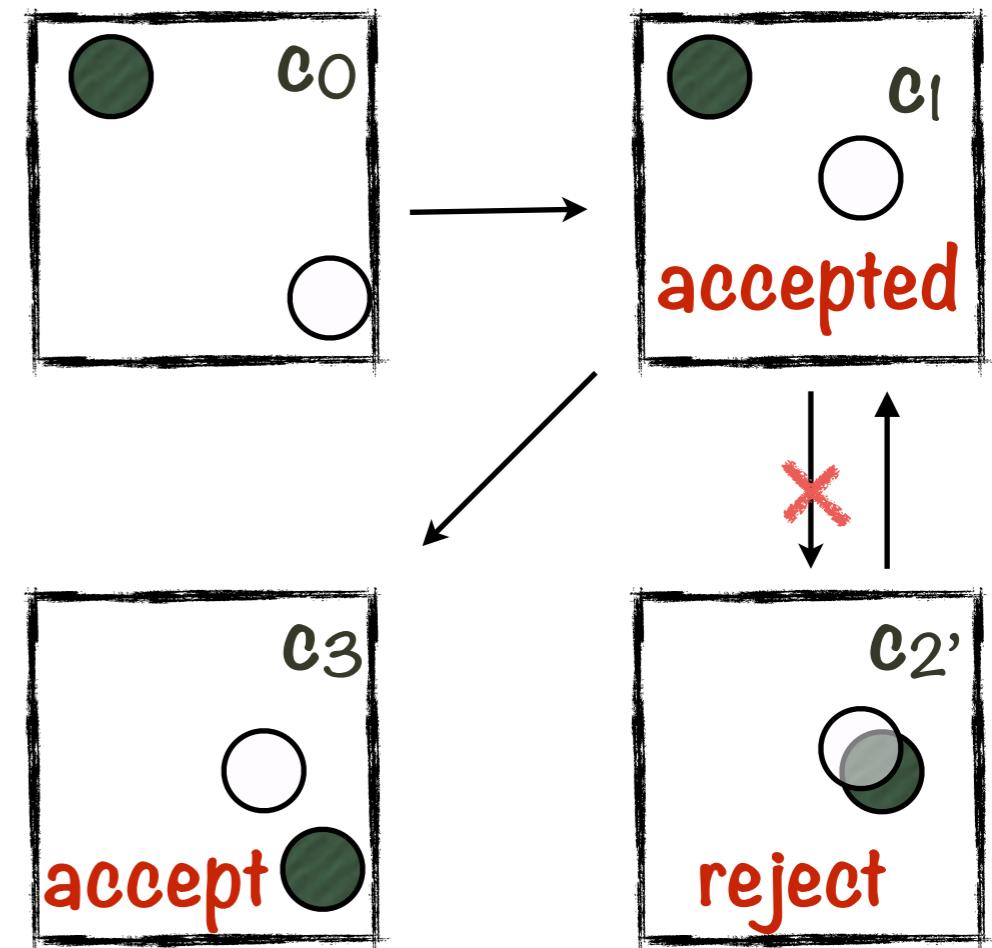


Integrating  
Newton Equations of Motions

## Monte Carlo

(MC):

Ensemble averaging



Random Displacement &  
Thermal averaging:  
weighted by

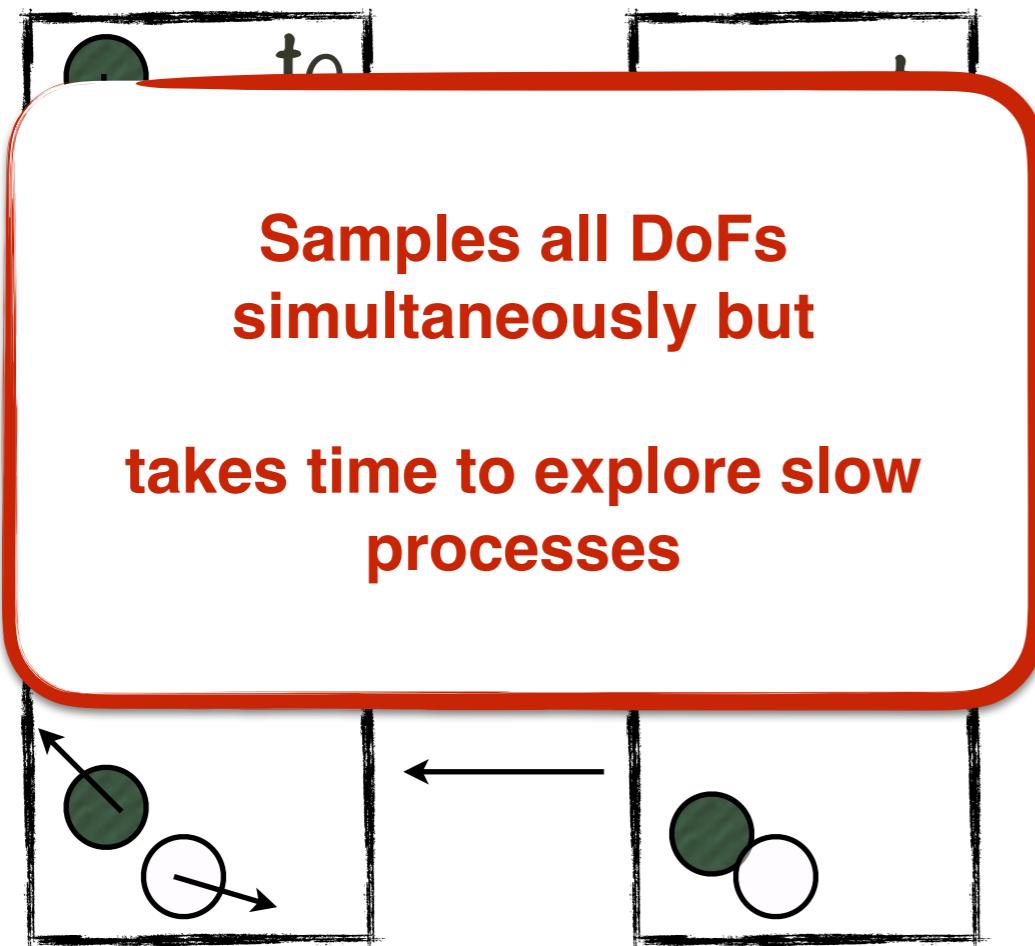
$$e^{-\frac{\Delta U(r)}{kT}}$$

# Conventional Methods

## Molecular Dynamics

(MD):

Time averaging



Integrating  
Newton Equations of Motions

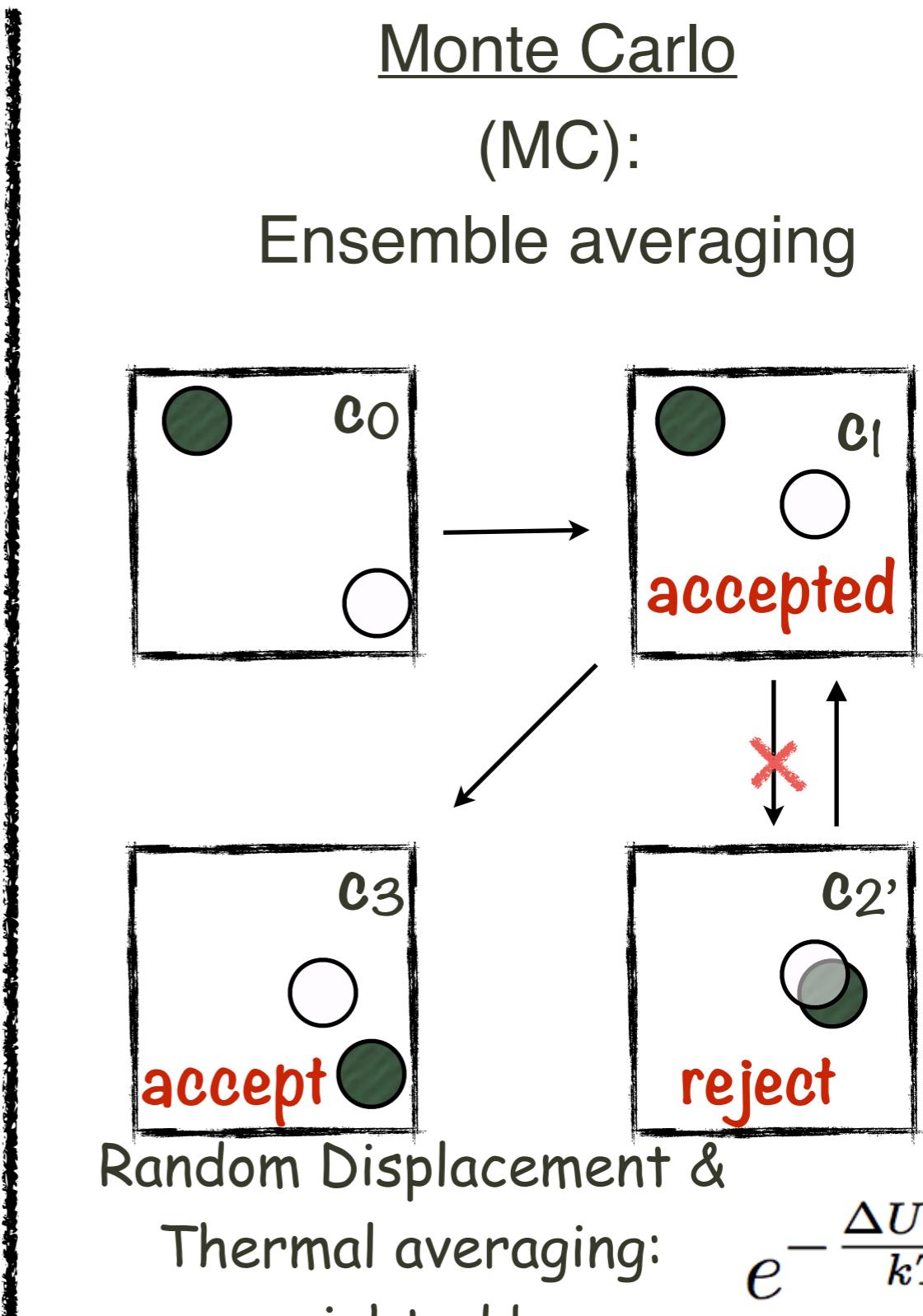
Samples all DoFs  
simultaneously but

takes time to explore slow  
processes

## Monte Carlo

(MC):

Ensemble averaging



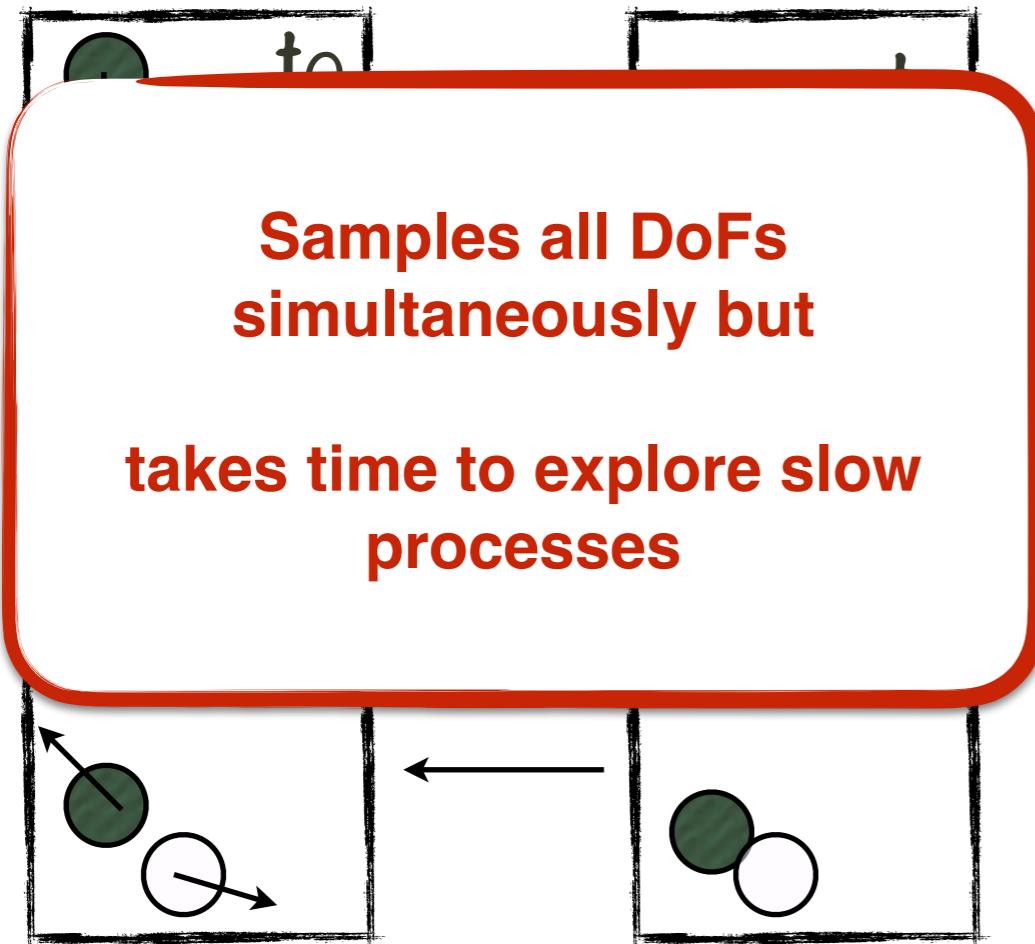
Random Displacement &  
Thermal averaging:  
weighted by  $e^{-\frac{\Delta U(r)}{kT}}$

# Conventional Methods

## Molecular Dynamics

(MD):

Time averaging

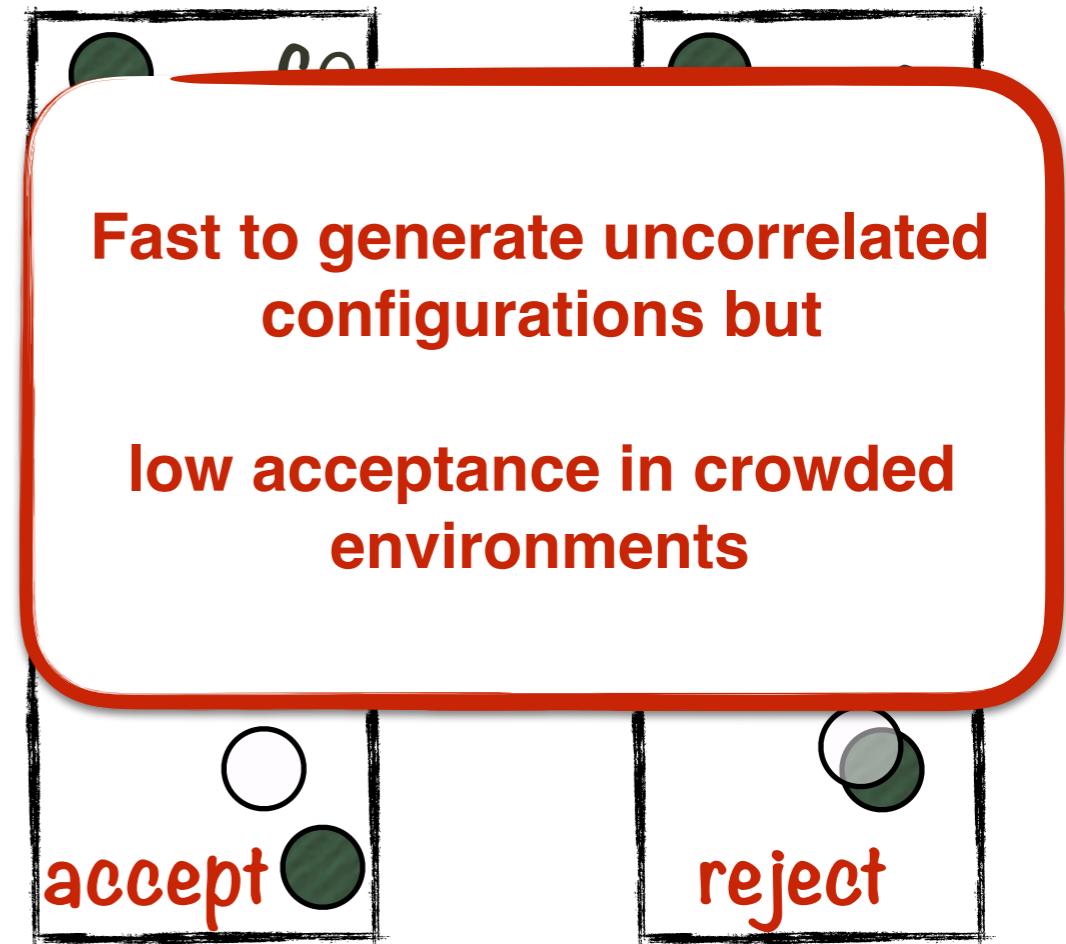


Integrating  
Newton Equations of Motions

## Monte Carlo

(MC):

Ensemble averaging



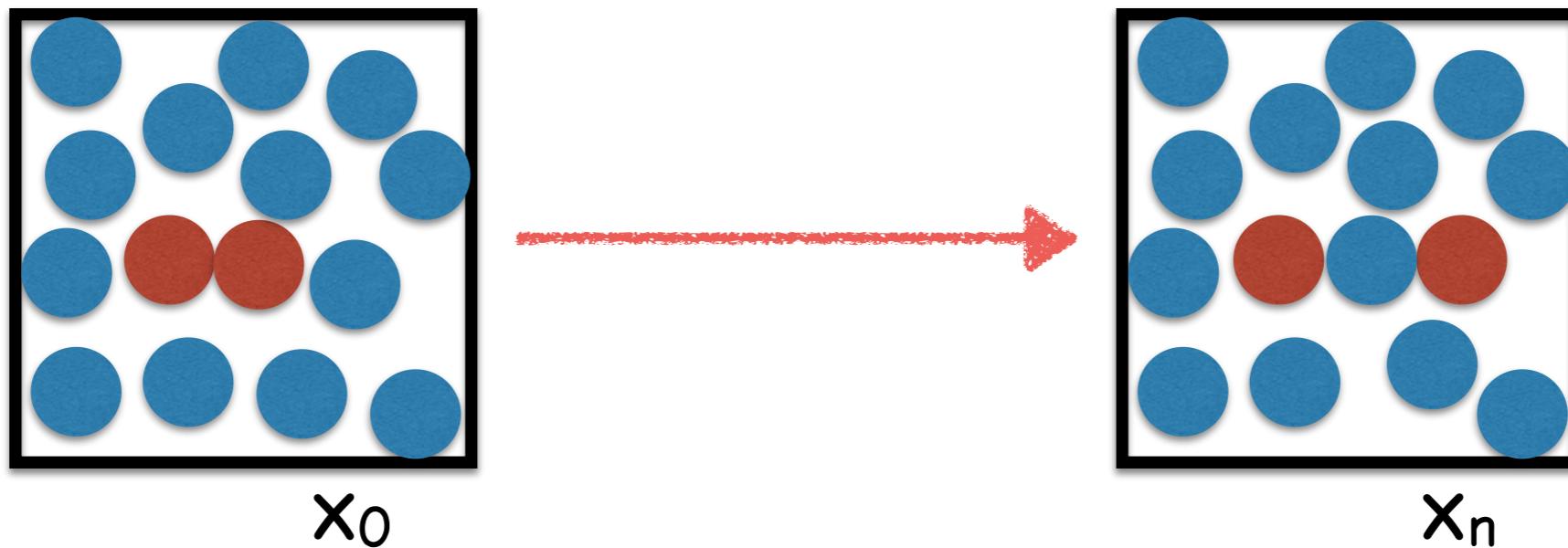
Random Displacement &  
Thermal averaging:  
 $e^{-\frac{\Delta U(r)}{kT}}$   
weighted by

# NCMC

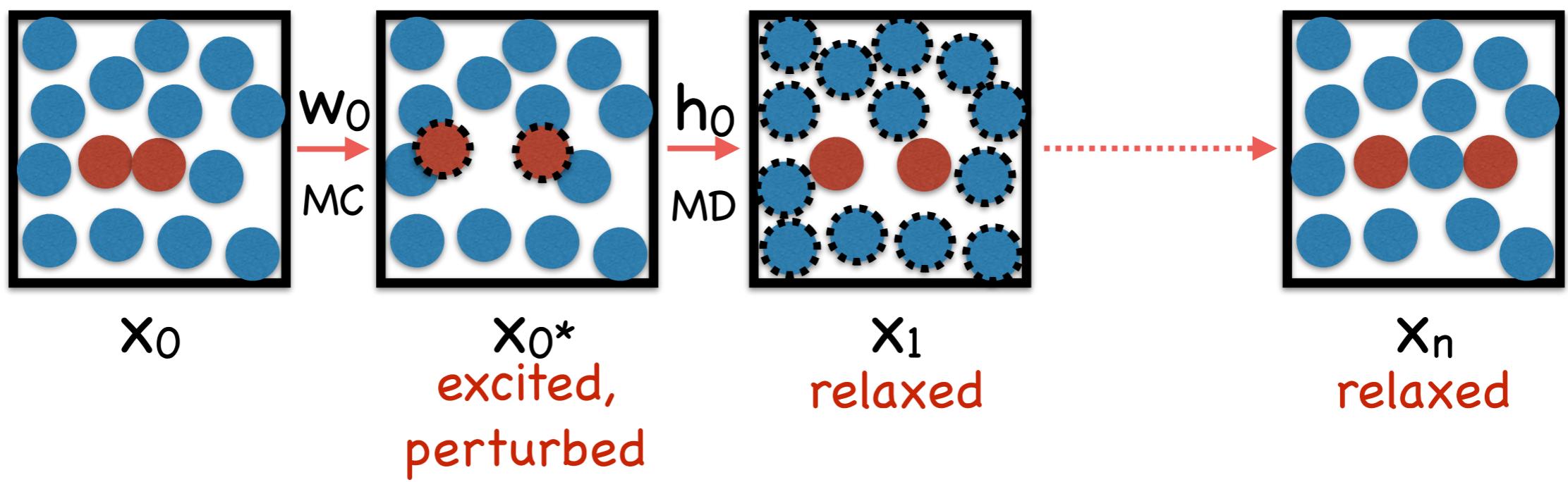
Combines best part of MD and MC

- Drive *slow* DoF's using MC moves:
  - Proposes an uncorrelated configuration for slow DoF's
- Relax rest of the DoF's using MD
  - Naturally evolves DoF's coupled with the driven DoF's

# NCMC



- Divide the move into smaller  $n$  steps:  $\text{NCMC}^n$



# NCMC



excited,  
perturbed

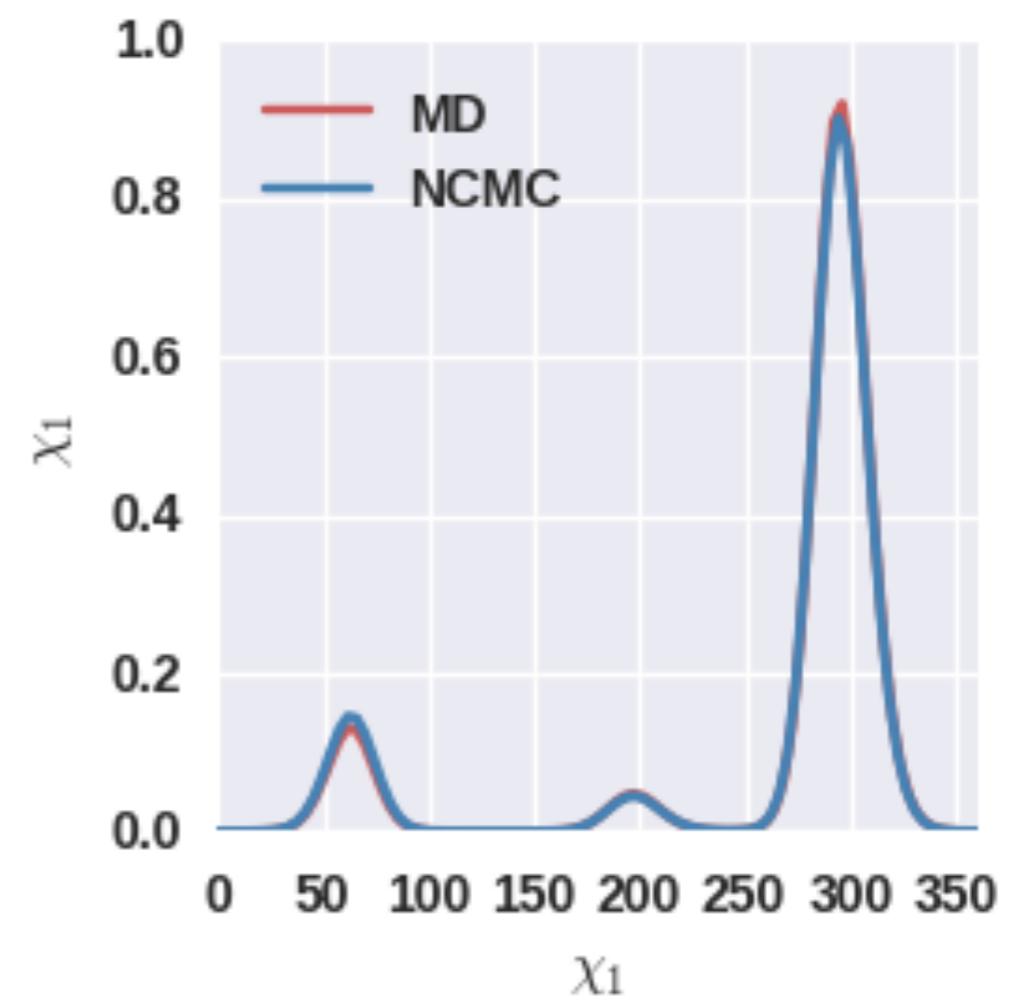
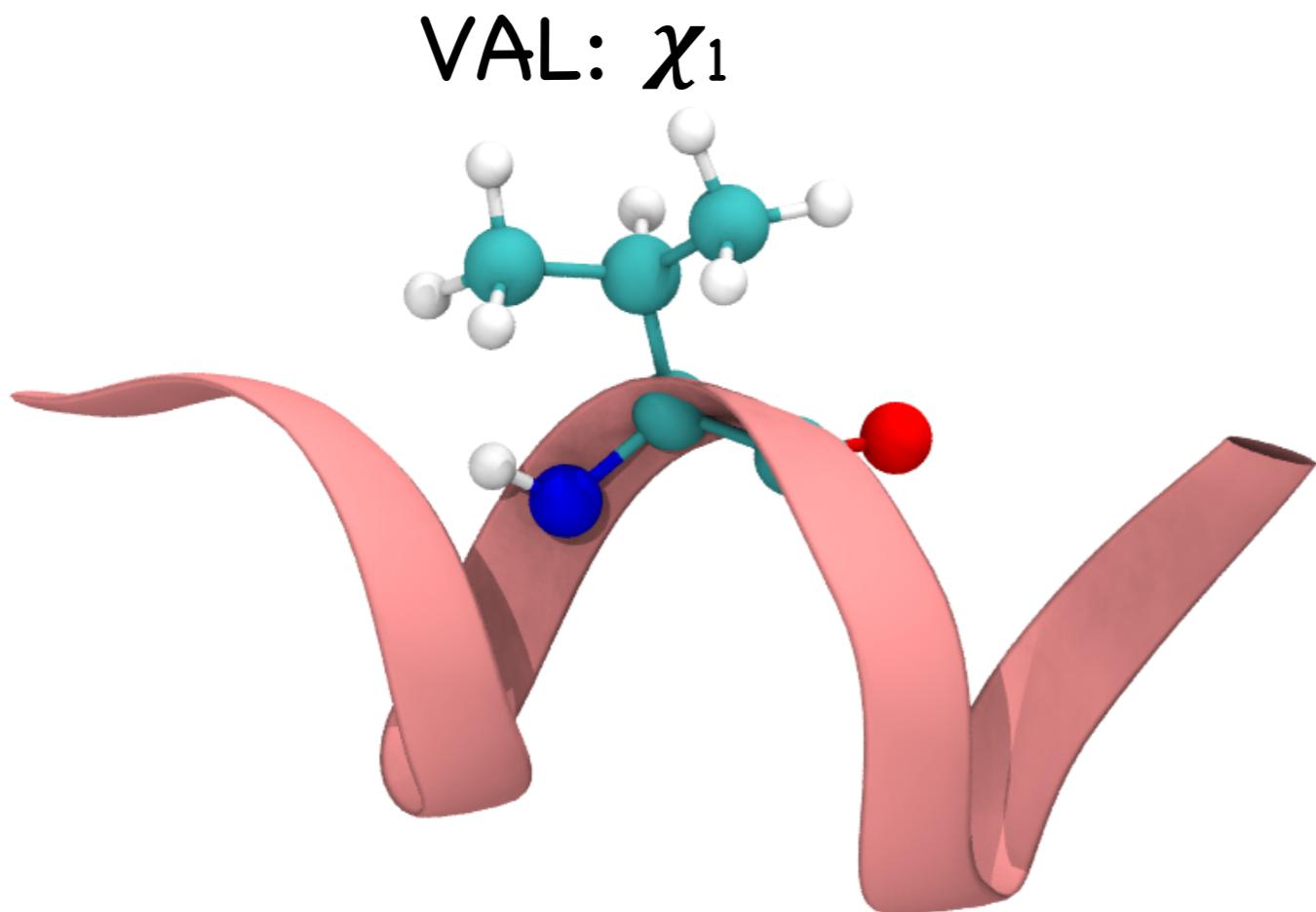
relaxed

relaxed

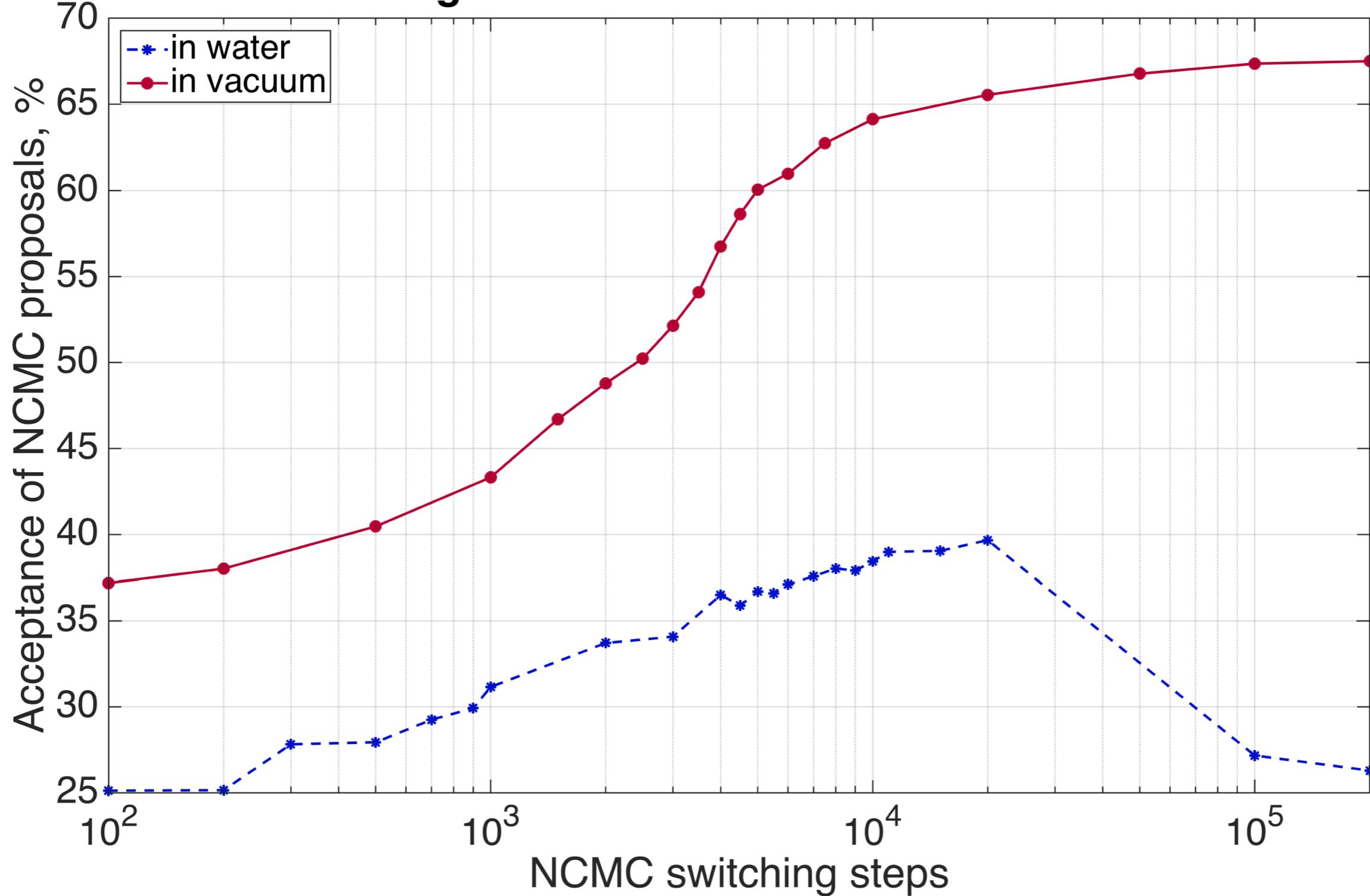
# System

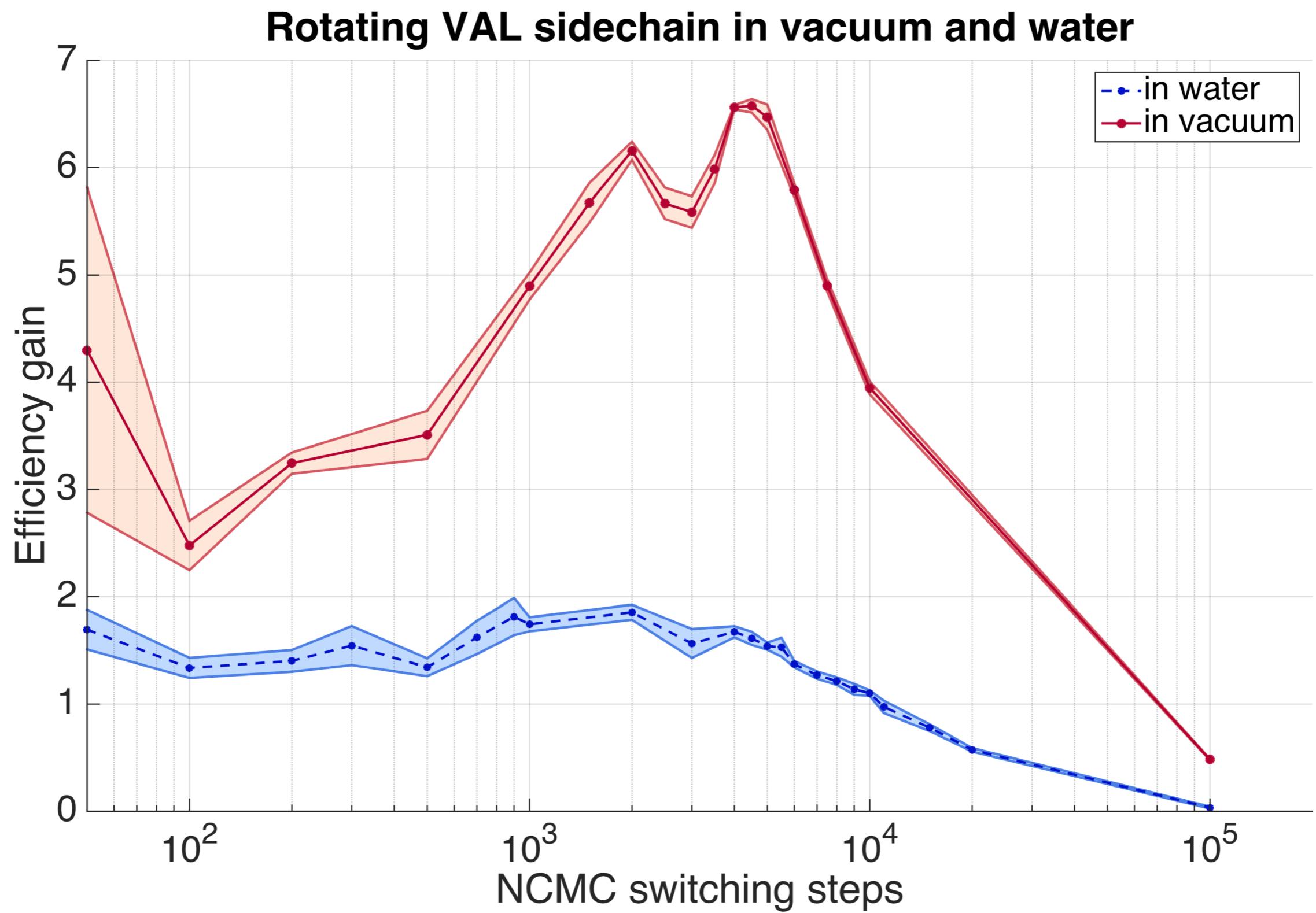
- Sampling side chain rotamer distribution both in *vacuum* and in *solvent*
- Peptide : capped AAAA**V**AAAAA,
- Backbone is restrained in a helical conformation

**small size & slow in MD**



## Rotating VAL sidechain in vacuum and water





# Conclusion

- Simulation: Imitation of real systems using its model
  - Distribution of all possible states
  - We need tricks to explore state-space efficiently
- NCMC: Combines the best properties of conventional methods.
- ~2 times faster in water and ~7 times faster in water.

# Acknowledgement

- Supervisor: Wouter Boomsma
- Kresten Lindorf Larsen

# Acknowledgement

- Sup
- Kres



Thank you