Lecture 22

What is Monte Carlo (MC) method ?

The Monte Carlo method is a numerical method for statistical simulation which utilizes sequences of random numbers to perform the simulation.
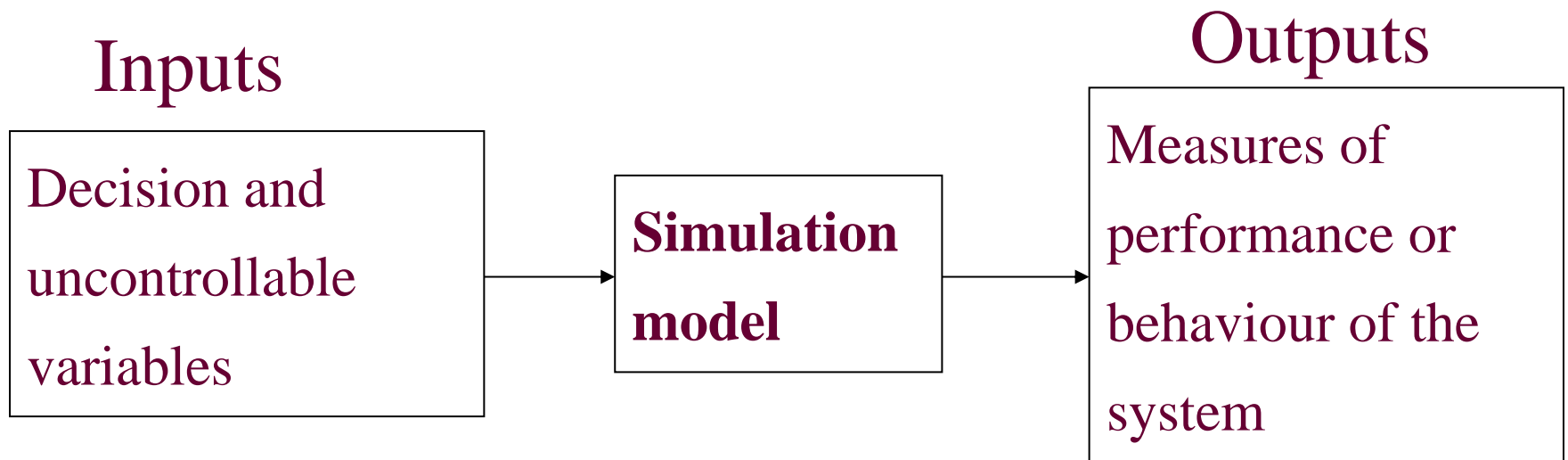
What the meaning of MC simulation?

MC simulation is a versatile tool to analyze and evaluate complex measurements

Constructing a *model* of a *system*.

Experimenting with the model to draw inferences of the system's behavior

A simulation model

Inputs

Decision and uncontrollable variables

Simulation model

Outputs

Measures of performance or behaviour of the system

A simulation model cont..

- Model inputs capture the environment of the problem

- The simulation model

  – Conceptual model: set of assumptions that define the system

  – Computer code: the implementation of the conceptual model

- Outputs describe the aspects of system behaviour that we are interested in

Random numbers

Uniform Random numbers or pseudo-random numbers (PRN)
are essentially independent random variables uniformly
Distributed over the unit interval (0,1).

The PRNs are good if they are uniformly distributed,
statistically independent and reproducible.
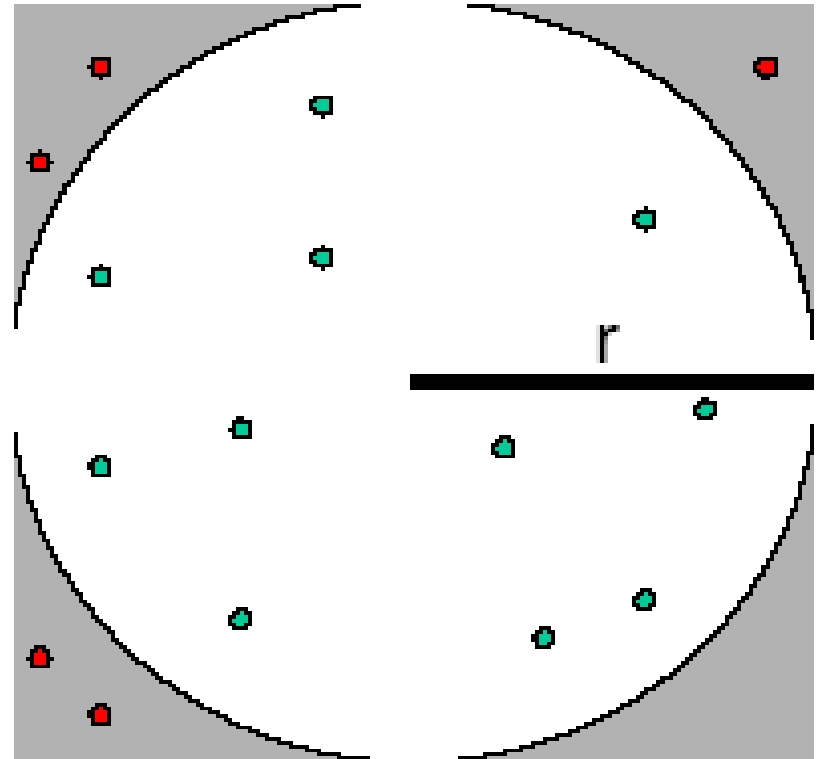
Classic Example

Find the value of π

Use the reject and accept method
Or hit and miss method

The area of square=(2r)²

The area of circle = r²π

$$\frac{area \cdot of \cdot square}{area \cdot of \cdot circle} = \frac{4r^2}{\pi r^2} = \frac{4}{\pi}$$
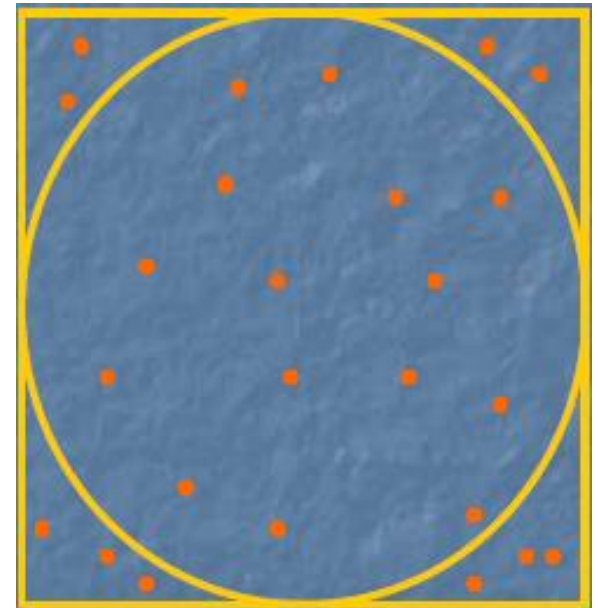
$$\pi = 4 * \frac{area \cdot of \cdot circle}{area \cdot of \cdot square}$$

r

Cont....

$$\frac{area.of.circle}{area.of.square} = \frac{\#.of.dots.inside.circle}{total.number.of.dots}$$

Hit and miss algorithm

- Generate two sequences of N of PRN :: $R_i,,R_j$

- $X_i = -1 + 2R_i$

- $Y_j = -1 + 2R_j$

- Start from s=zero

- If $(X^2 + Y^2 < 1)$ $s = s + 1$

- # of dots inside circle=s

- total number of dots=N

$$\pi = 4 * S / N$$

Random versus Pseudo-random

- Virtually all computers have "random number" generators
- Their operation is deterministic
- Sequences are predictable
- More accurately called "pseudo-random number" generators
- In this chapter "random" is shorthand for "pseudo-random"
- "RNG" means "random number generator"

Properties of an Ideal RNG

- Uniformly distributed
- Uncorrelated
- Never cycles
- Satisfies any statistical test for randomness
- Reproducible
- Machine-independent
- Changing "seed" value changes sequence
- Easily split into independent subsequences
- Fast
- Limited memory requirements

No RNG Is Ideal

- Finite precision arithmetic $\Rightarrow$ finite number of states $\Rightarrow$ cycles
  - Period = length of cycle
  - If period > number of values needed, effectively acyclic
- Reproducible $\Rightarrow$ correlations
- Often speed versus quality trade-offs

Linear Congruential RNGs

$$X_i = (a \times X_{i-1} + c) \bmod M$$

Modulus

Additive constant

Multiplier

Sequence depends on choice of seed, X_0

Period of Linear Congruential RNG

- Maximum period is M

- For 32-bit integers maximum period is 2^{32}, or about 4 billion

- This is too small for modern computers

- Use a generator with at least 48 bits of precision

Producing Floating-Point Numbers

- X_i, a, c, and M are all integers

- X_is range in value from 0 to M-1

- To produce floating-point numbers in range [0, 1), divide X_i by M

Defects of Linear Congruential RNGs

- Least significant bits correlated
  - Especially when M is a power of 2
- k-tuples of random numbers form a lattice
  - Points tend to lie on hyperplanes
  - Especially pronounced when k is large

# Lagged Fibonacci RNGs

$$X_i = X_{i-p} * X_{i-q}$$

1. p and q are lags, $0 < p < q$
2. $*$ is any binary arithmetic operation
   a. Addition modulo M
   b. Subtraction modulo M
   c. Multiplication modulo M
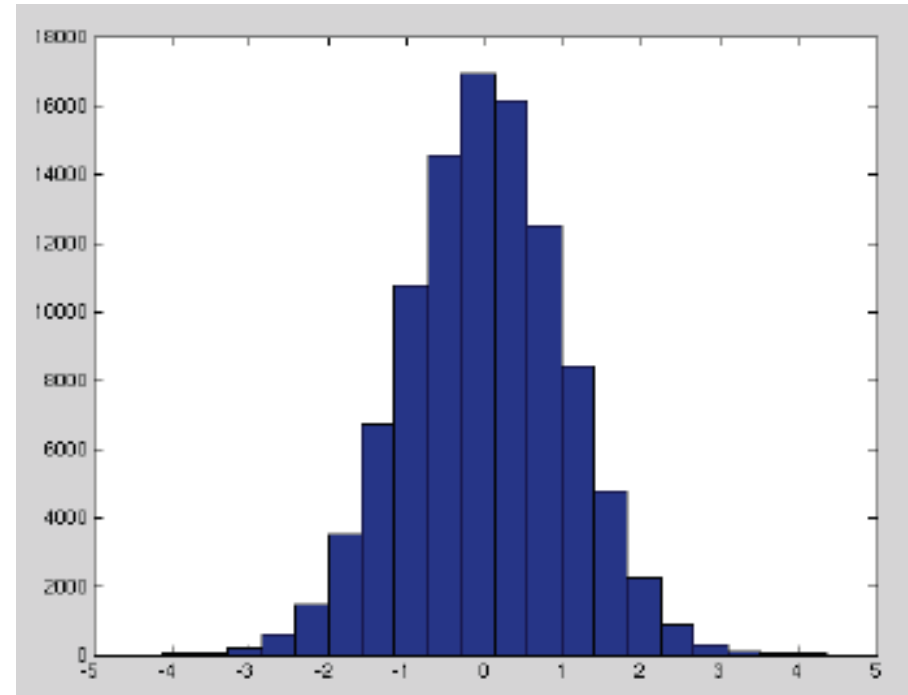   d. Bitwise exclusive or
3. M is usually a power of 2

Properties of Lagged Fibonacci RNGs

- Require p seed values

- Careful selection of seed values, p, and q can result in very long periods and good randomness

- For example, suppose M has b bits

- Maximum period for additive lagged Fibonacci RNG is $(2^p - 1)2^{b-1}$
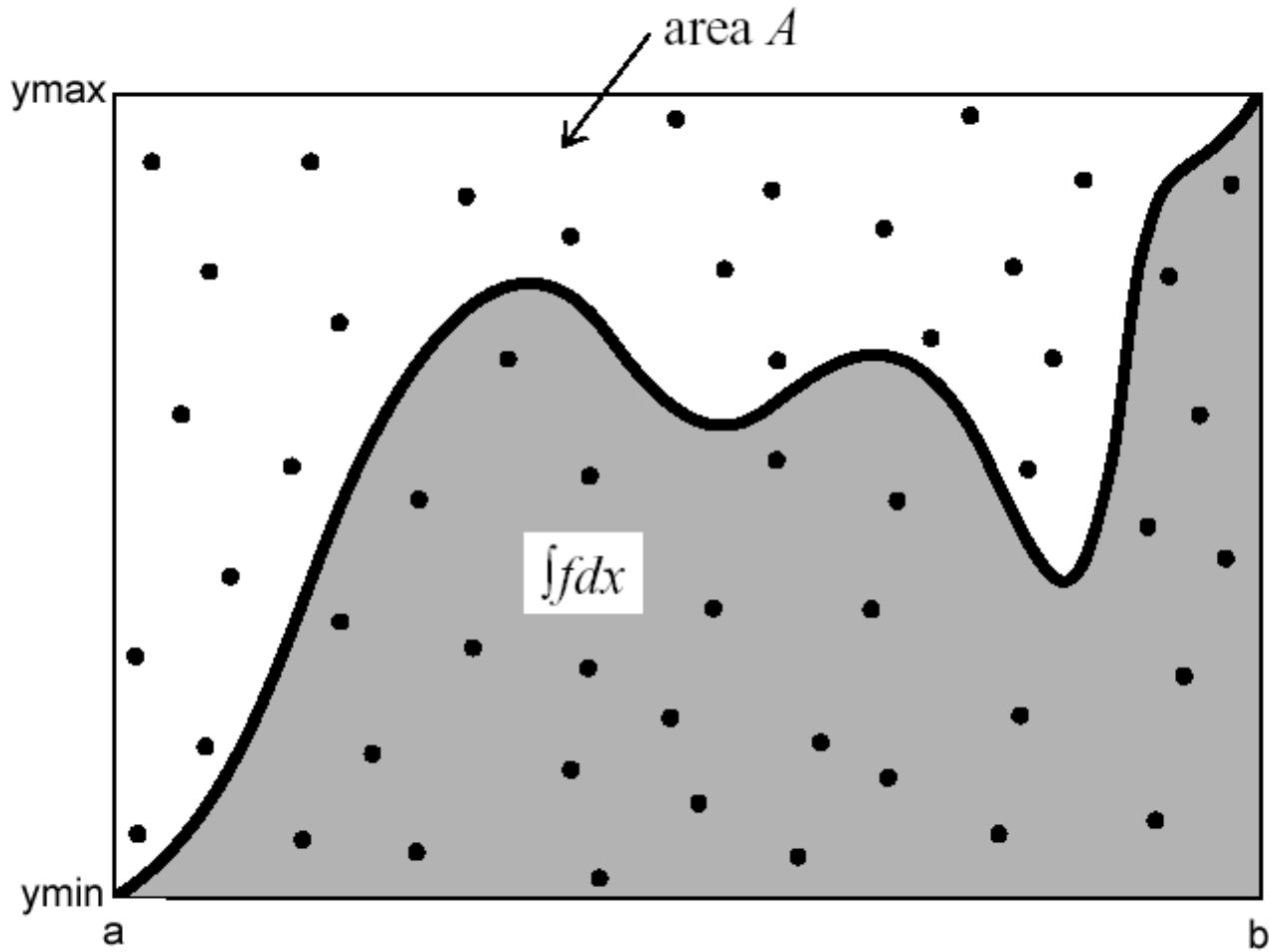
Types of distribution
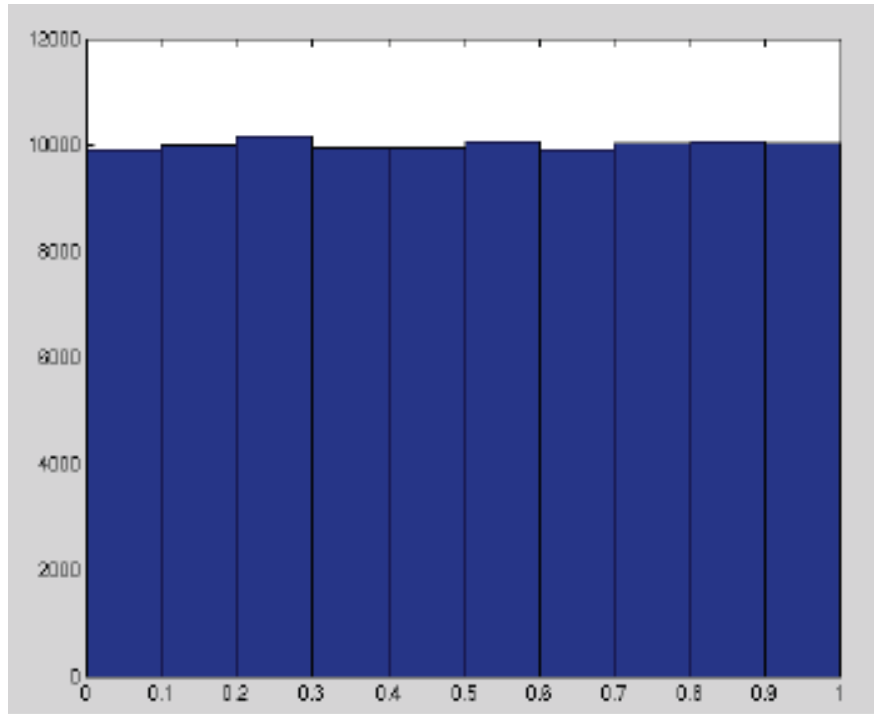
Uniform distribution

Gaussian or normal distribution
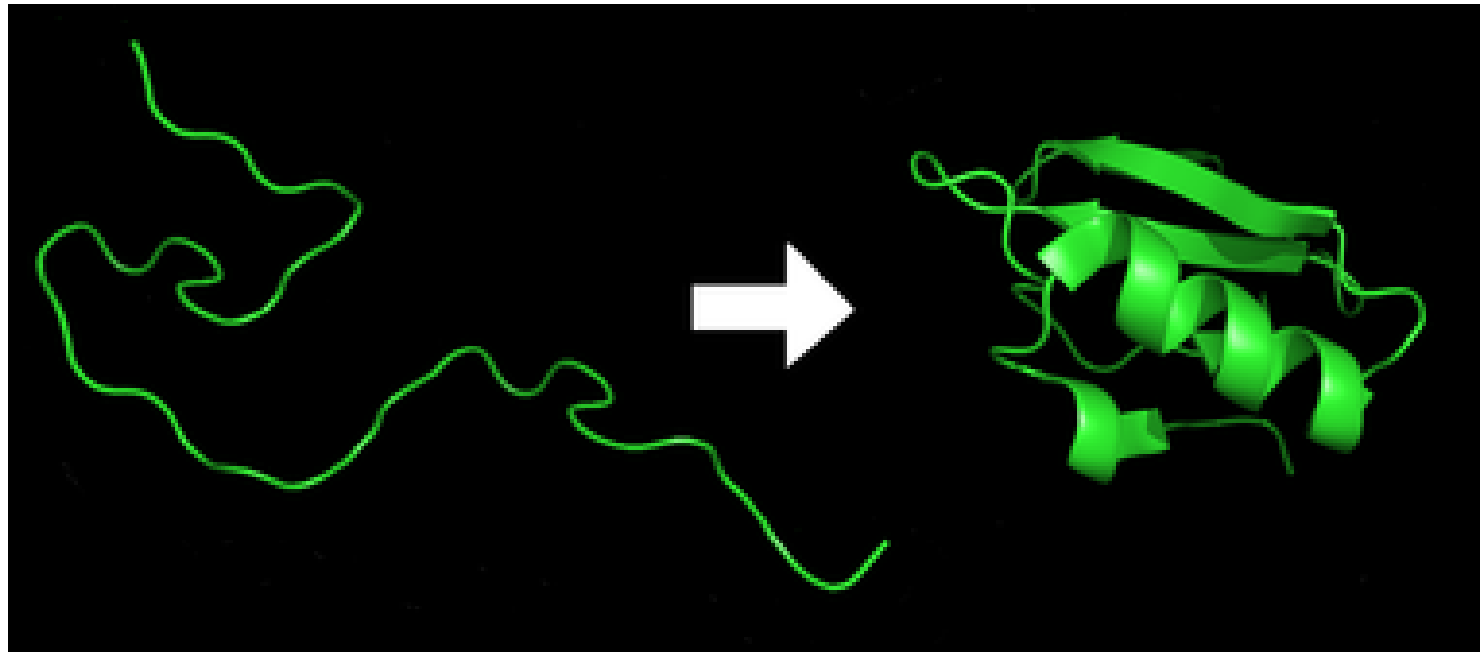
Monte Carlo Integration

# Lecture 23-

Protein Folding

Protein Folding

Protein folding is the process by which a protein structure assumes its functional shape or conformation. It is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from random coil.

Folding, unfolding, misfolding

- The correct three-dimensional structure is essential to function, although some parts of functional proteins may remain unfolded.

- Failure to fold into native structure generally produces inactive proteins, but in some instances misfolded proteins have modified or toxic functionality.

- Several neurodegenerative and other diseases are believed to result from the accumulation of amyloid fibrils formed by misfolded proteins.

- Many allergies are caused by incorrect folding of some proteins, for the immune system does not produce antibodies for certain protein structures

Folding, unfolding, misfolding

Unfolded

Properly folded protein

Toxic protein clump
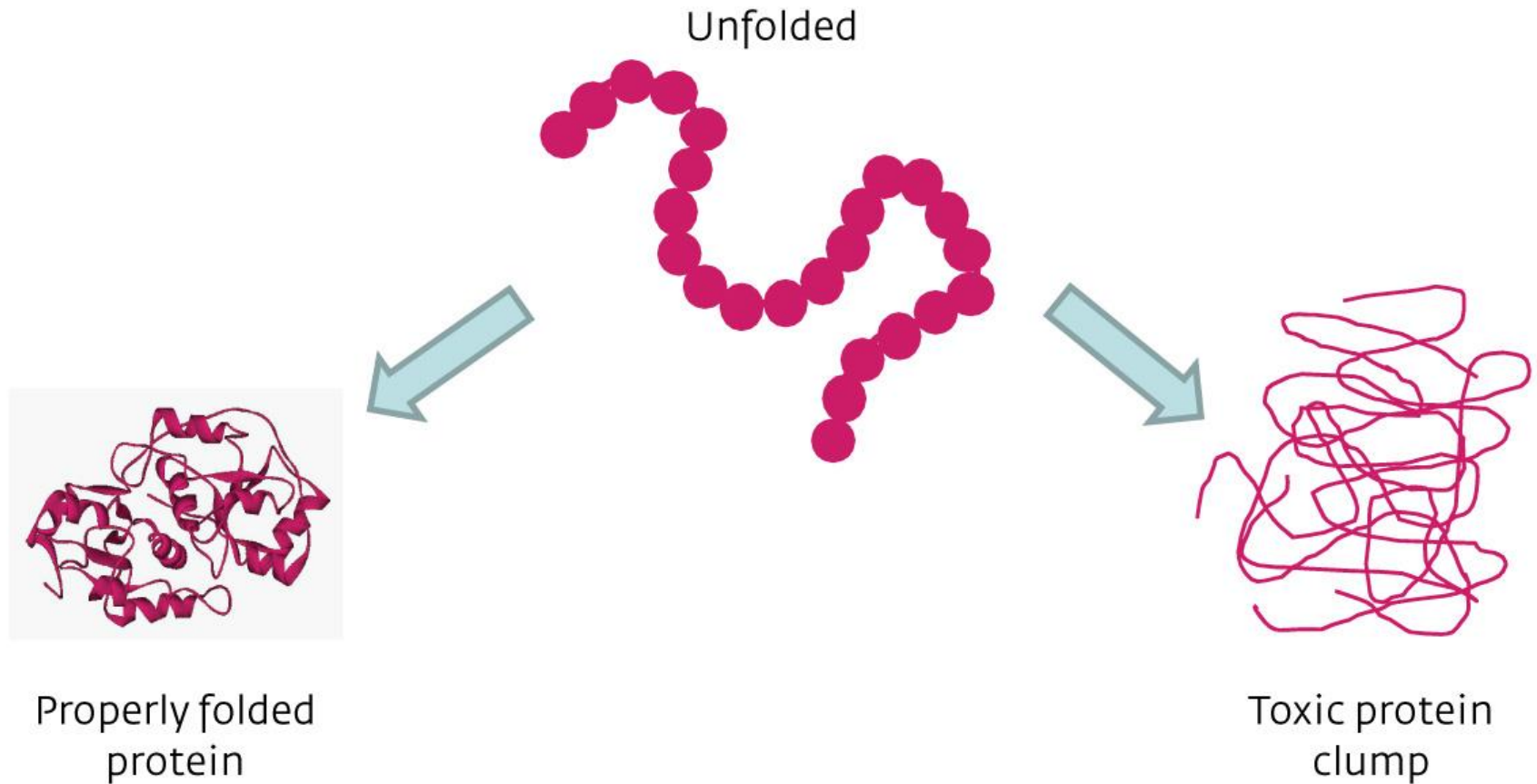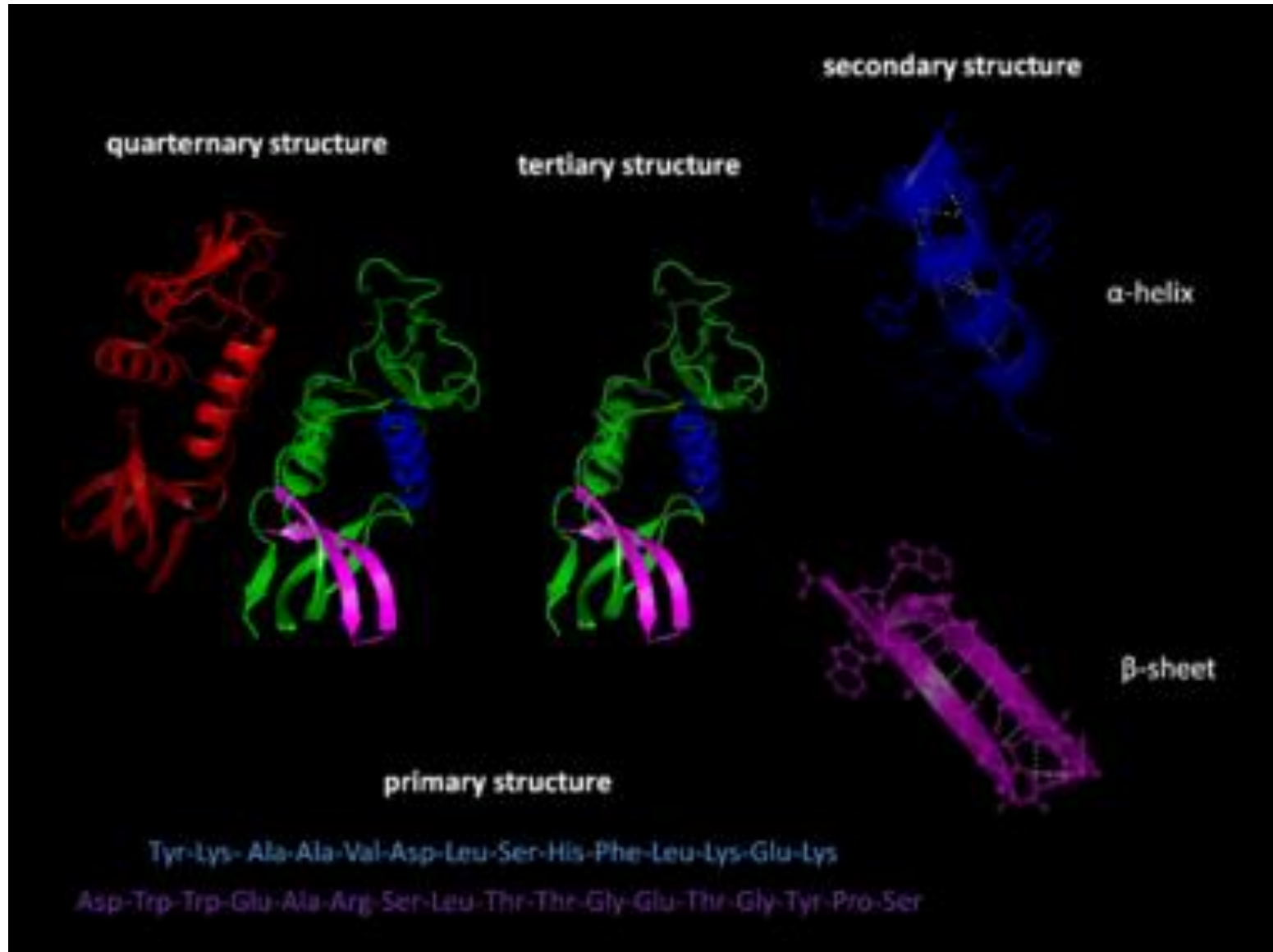
Folding driving forces
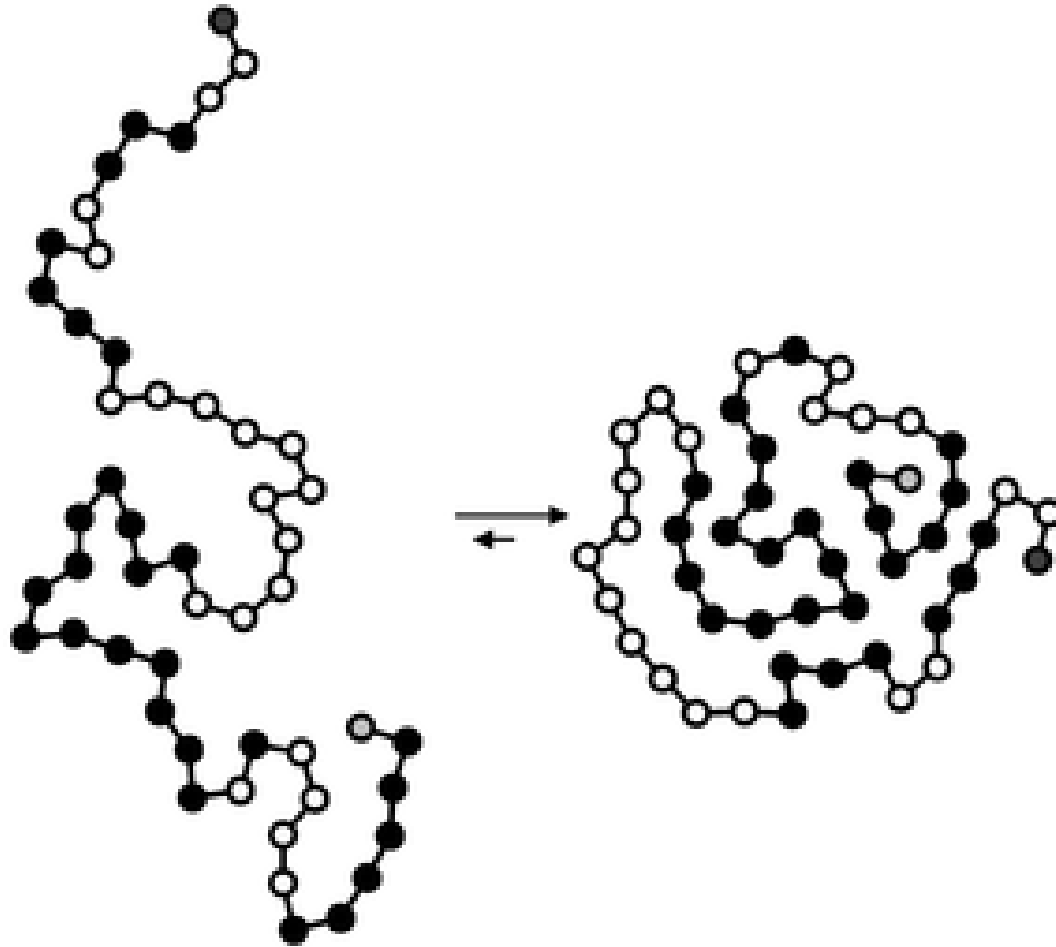
- Minimizing hydrophobic side-chains exposed to water,

- solvent (water or lipid bilayer),

- concentration of salts,

- pH,

- temperature,

- possible presence of cofactors, molecular chaperones.

- intramolecular hydrogen bonds

- van der Waals interaction

- Electrostatic interaction

- And many more …

Folding Simulation

- Video

Protein Structure

Protein Folding Models

Folding often begins co-translationally, so that the N-terminus of the protein begins to fold while the C-terminal portion of the protein is still being synthesized by the ribosome.

- The diffusion collision model, in which a nucleus is formed, then the secondary structure is formed, and finally these secondary structures are collided together and pack tightly together.

- The nucleation-condensation model, in which the secondary and tertiary structures of the protein are made at the same time.

Relationship between folding and amino acid sequence

Anfinsen's dogma

The native structure is determined only by the protein's amino acid sequence.
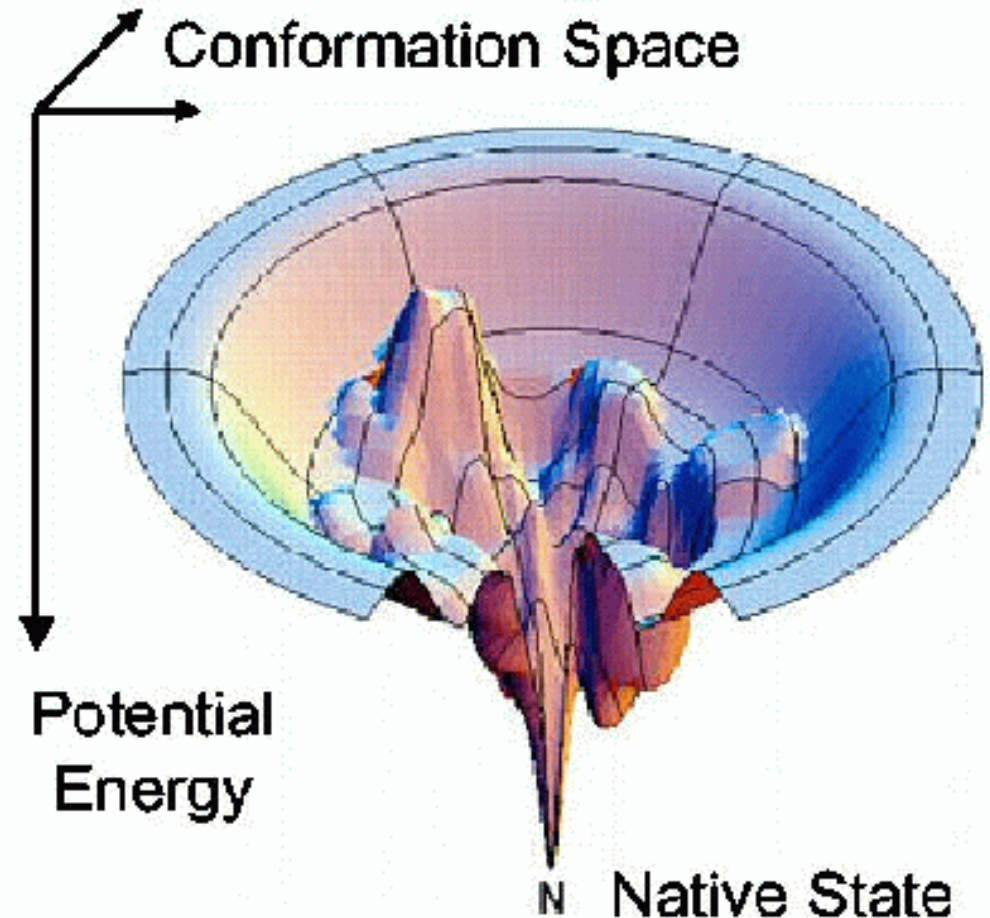
Tested the folding on ribonuclease A.

Limitations:
    Uniqueness
    Stability
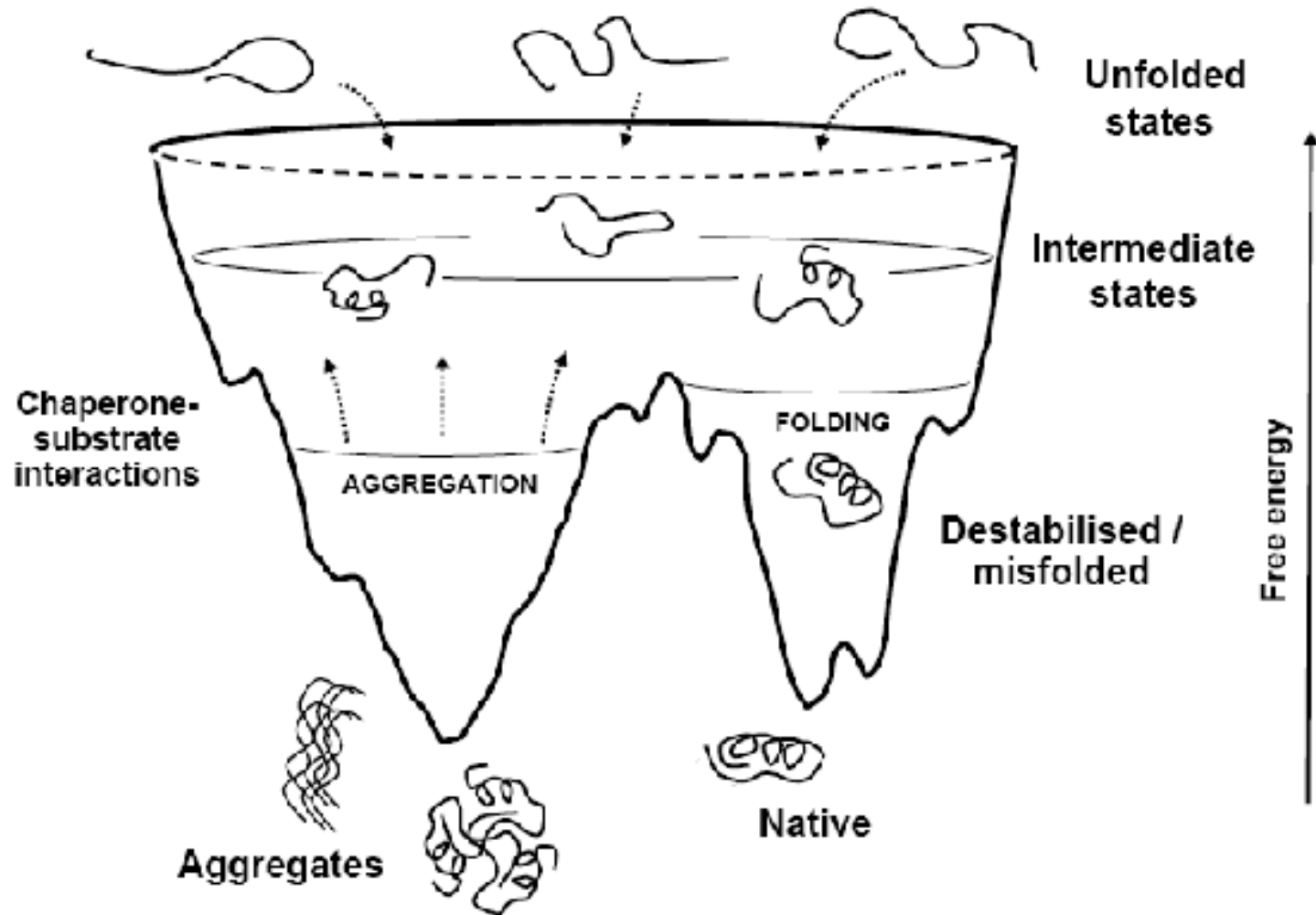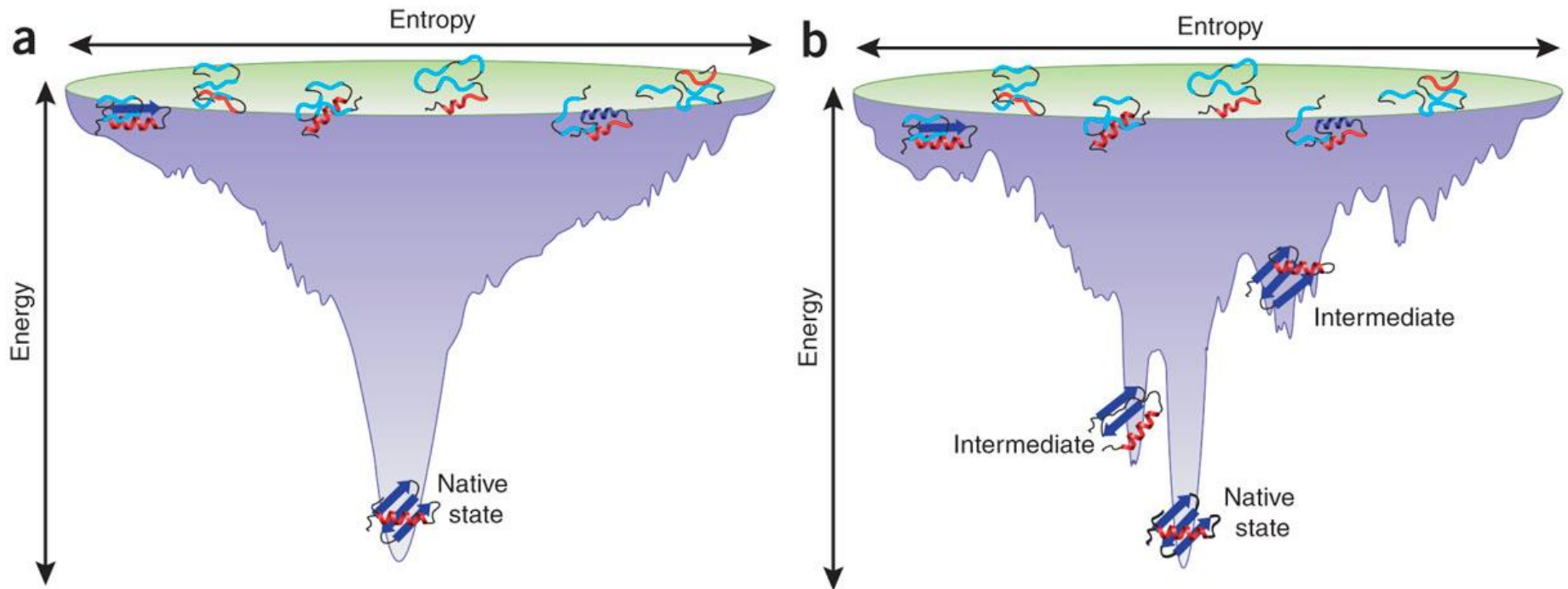   Kinetical accessibility

Energy landscape

In physics and biochemistry, an ***energy landscape*** is a mapping of all possible conformations of a molecular entity, or the spatial positions of interacting molecules in a system, and their corresponding energy levels, typically Gibbs free energy, on a two- or three-dimensional Cartesian coordinate system.
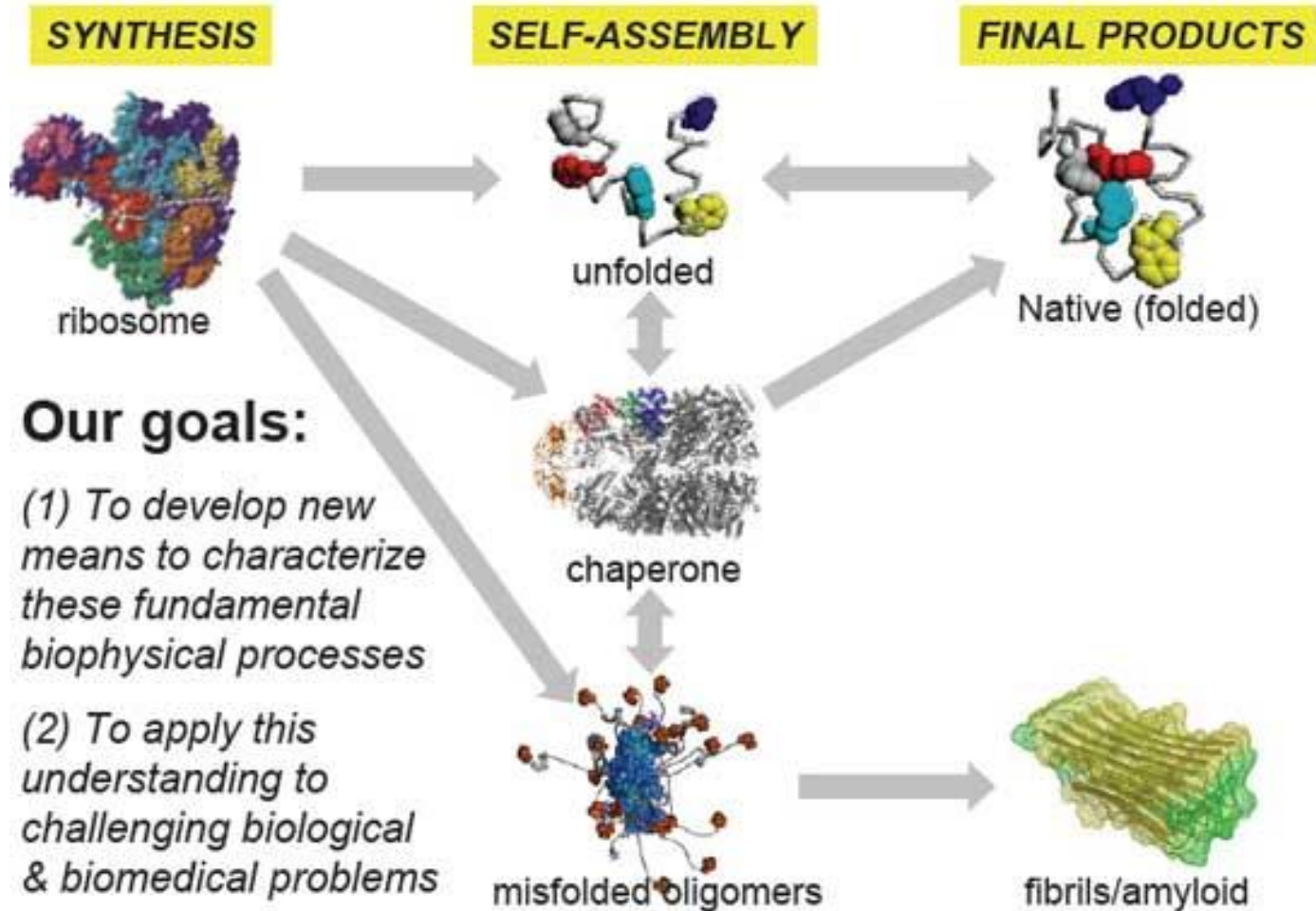

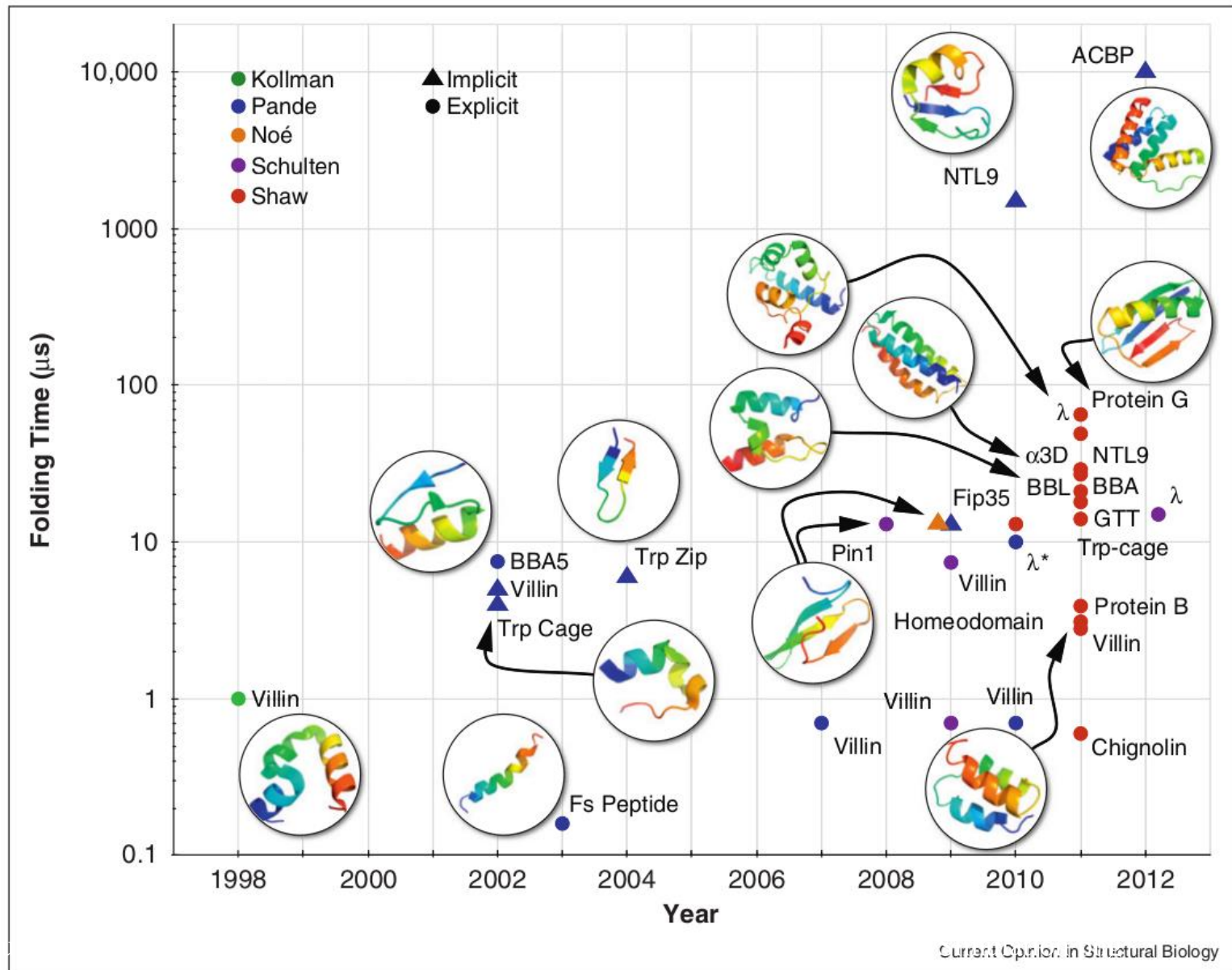Dill, K.A. and Chan, S. (1997) *Nature Structural. Biology*, 4:10–19

Energy landscape

Leandro, Paula & Gomes, Cláudio. (2008). Mini reviews in medicinal chemistry. 8. 901-11. 10.2174/138955708785132783.

Folding funnel

In vivo folding

Protein folding in the cell

SYNTHESIS · SELF-ASSEMBLY · FINAL PRODUCTS

ribosome · unfolded · Native (folded)

chaperone

Our goals:

(1) To develop new means to characterize these fundamental biophysical processes

(2) To apply this understanding to challenging biological & biomedical problems

misfolded oligomers · fibrils/amyloid

In-silico Folding

Current Opinion in Structural Biology

Post Translation Modifications

1. PCAF interacts with and acetylates Tat on lysine 28

2. K28-acetylated Tat interacts with P-TEFb

3. P-TEFb/Tat complex binds TAR RNA

4. p300 acetylates Tat on lysine 50

5. K50-acetylated Tat dissociates from TAR
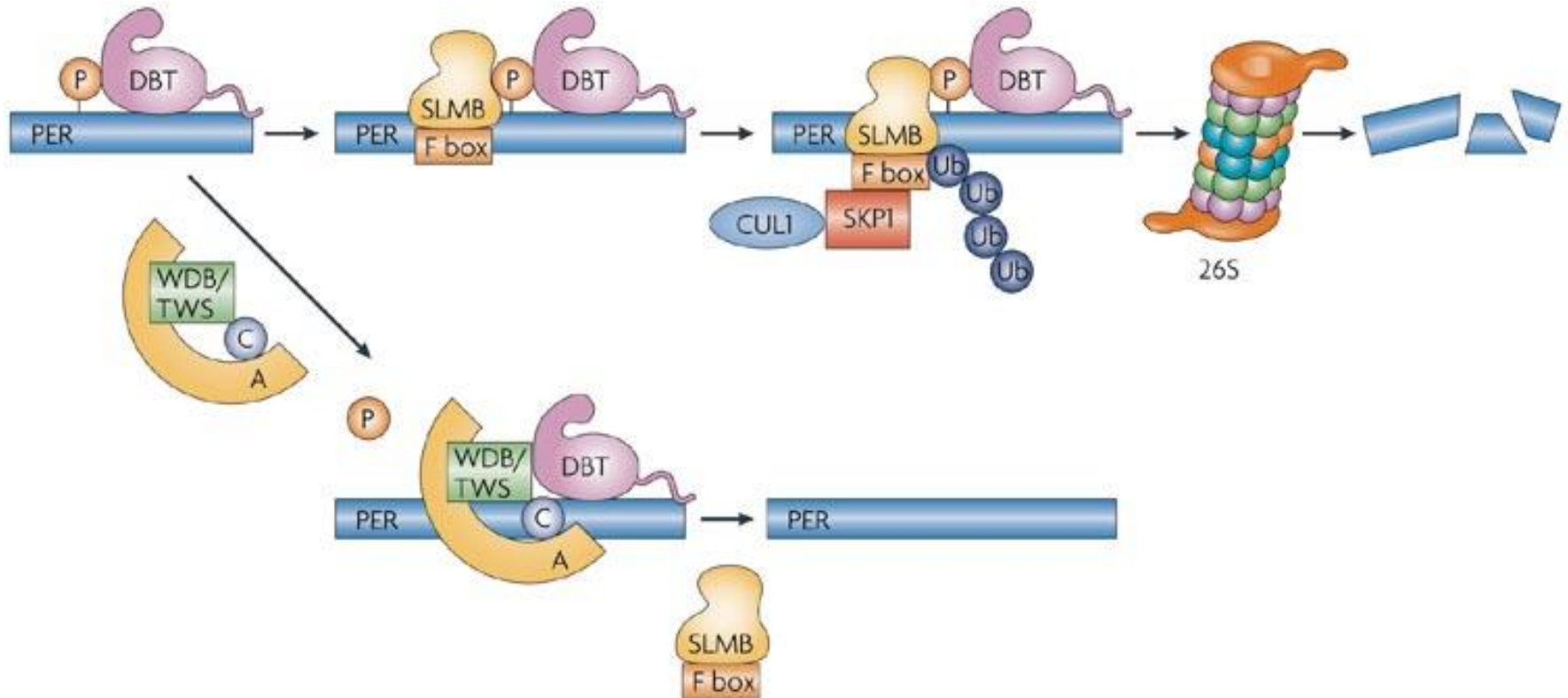
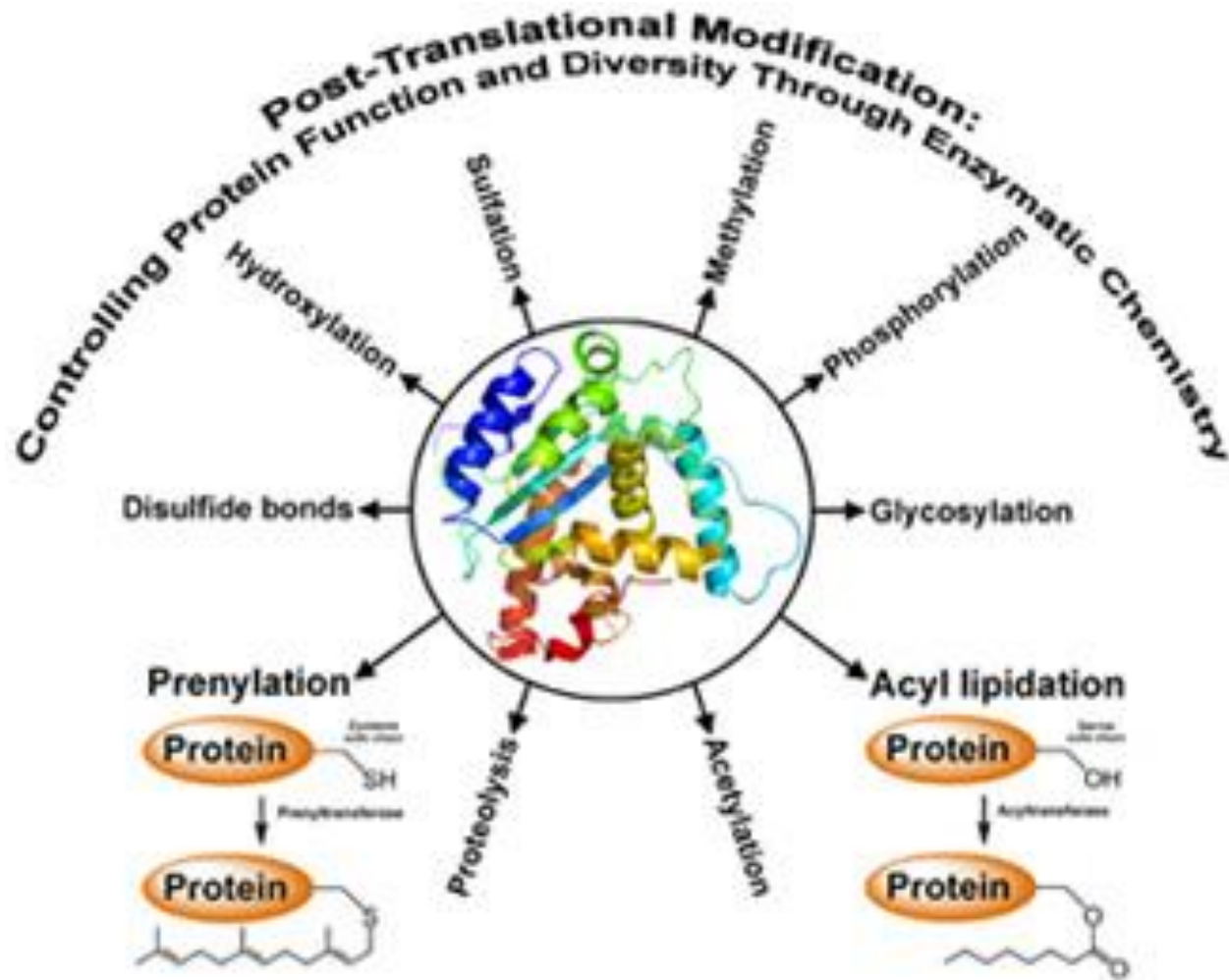K50-acetylated Tat interacts with PCAF

Fig.2: Regulation of the viral transactivator Tat transcriptional activity by post translational modifications
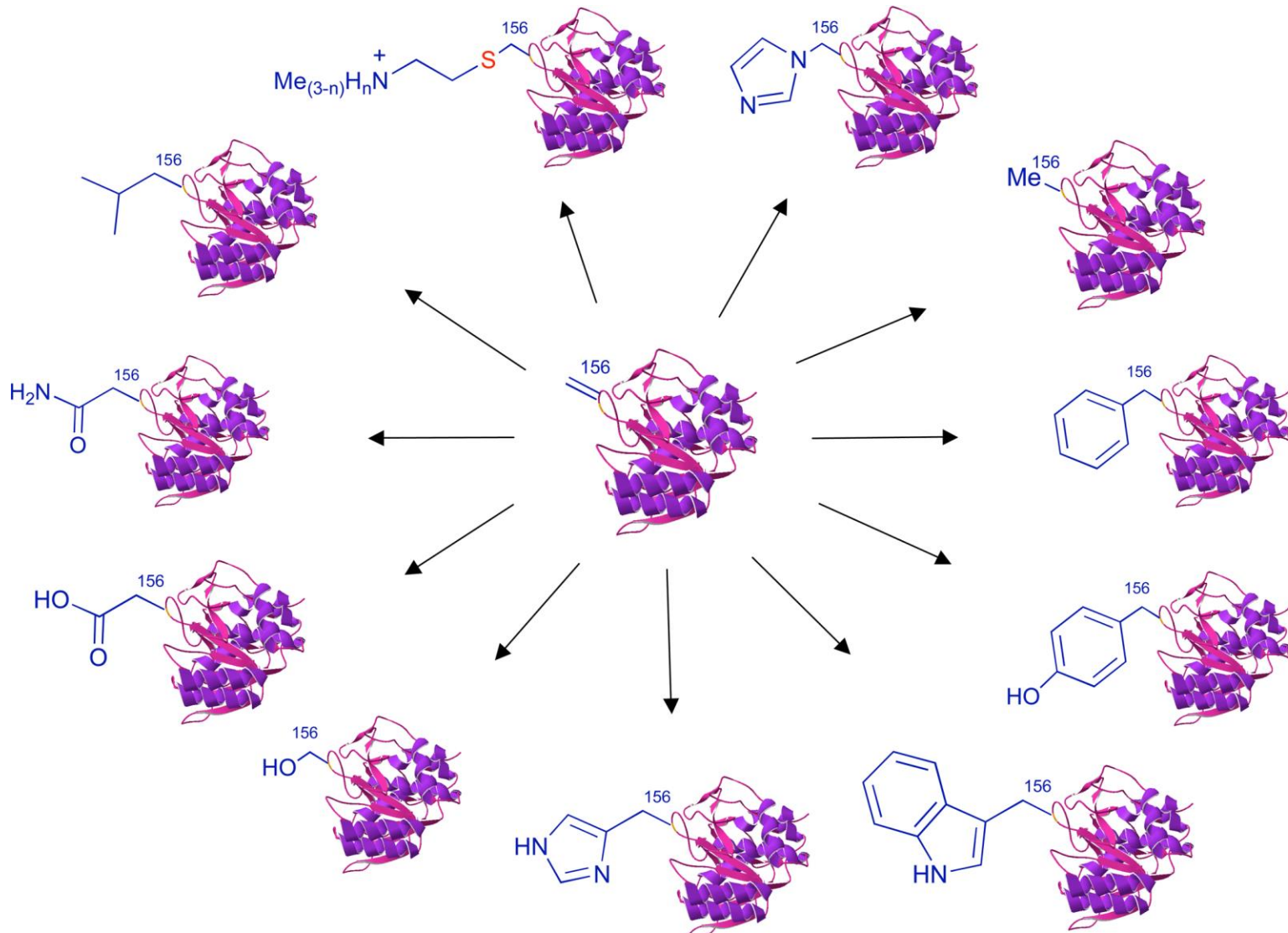
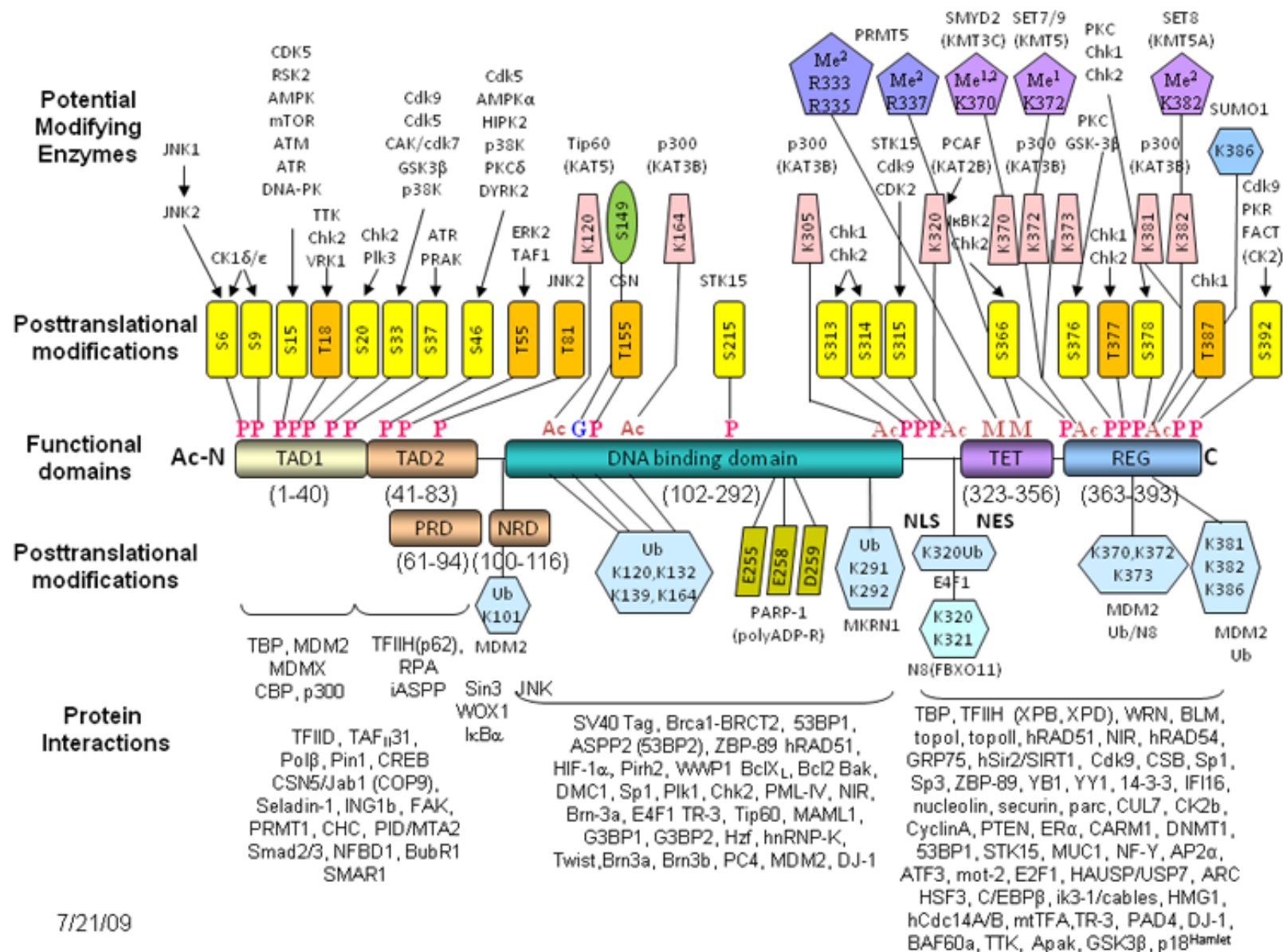# Post Translation Modifications

Post Translation Modifications

Post Translation Modifications

Koshland *et al*, *Proc Natl Acad Sci USA* **1966**, *56*, 1606

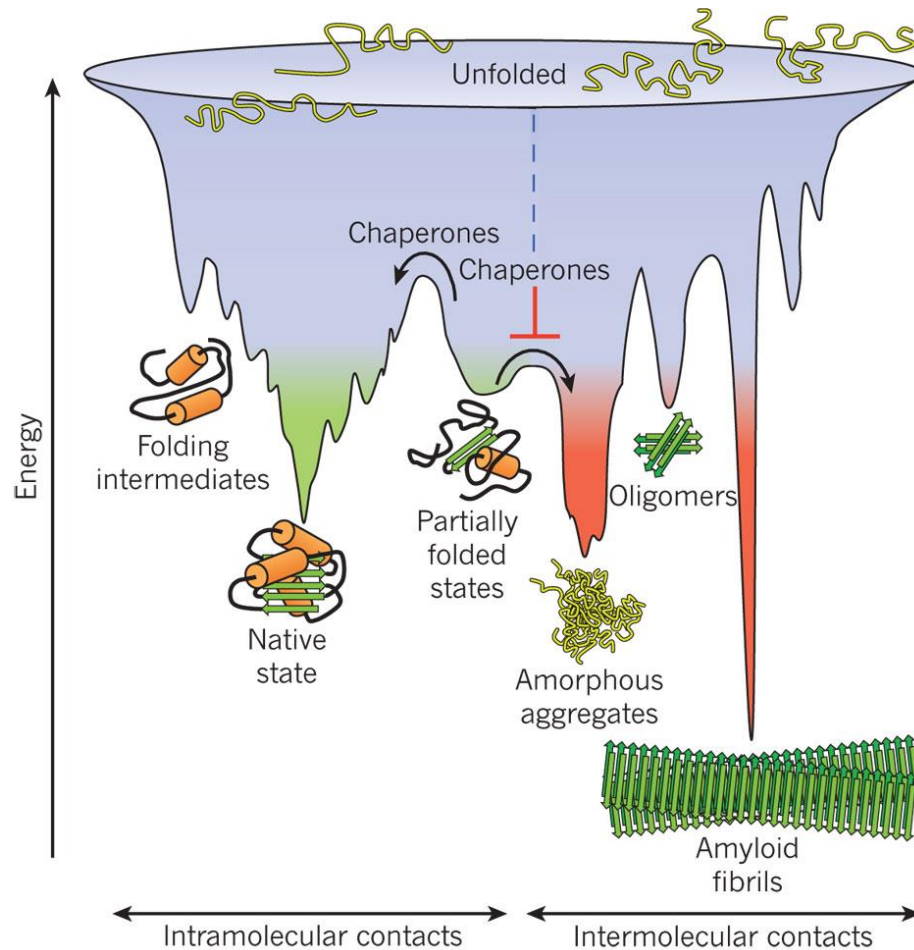# Human p53 Posttranslational Modifications

External factors on protein trajectories

Modification of the local minima by external factors can also induce modifications of the folding trajectory.

- Temperature,

- Electric, and/or magnetic fields,
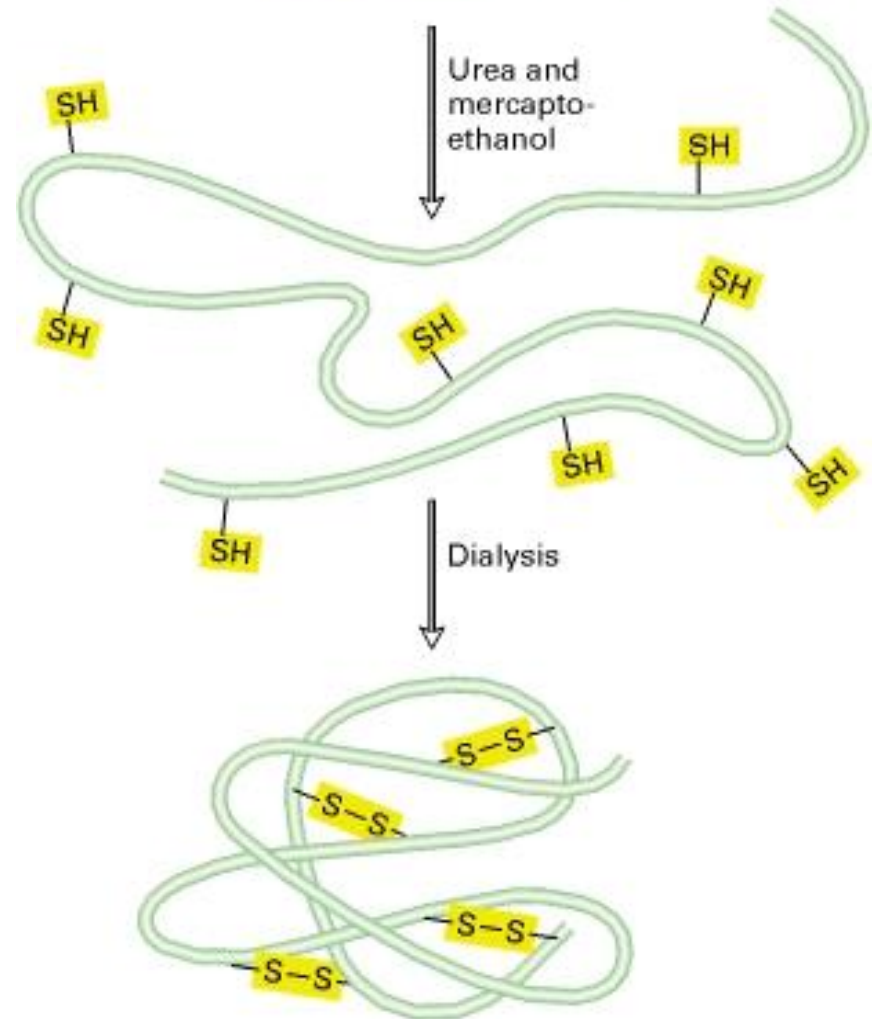
- Molecular crowding

- Space constraints.

Folding in funnel

In vitro denaturation and renaturation of proteins

Treatment with an 8 M urea solution containing mercaptoethanol ($HSCH_2CH_2OH$) completely denatures most proteins.

Disruption of the native state

Native state or biochemically functional forms may be disrupted for

- Thermal instability: Temperatures above or below the admissible range

- High concentrations of solutes

- Inadmissible pH

- Presence of chemical denaturants can do the same.

Denature, Refolding, Aggregates

- A fully denatured protein lacks both tertiary and secondary structure, and exists as a so-called random coil.

- Mostly denaturation is irreversible.

- Chaperones or heat shock proteins protect against the denaturing.

- In some situations some misfolded proteins are unfold, for a second chance to refold properly. This function is crucial to prevent the risk of precipitation into insoluble amorphous aggregates.

# Incorrect fold and neurodegenerative disease

Aggregated/Misfolded proteins accompany illnesses:

- Creutzfeldt-Jakob disease,

- Bovine spongiform encephalopathy (mad cow disease),

- Amyloid-related illnesses such as Alzheimer's disease

- Familial amyloid cardiomyopathy or polyneuropathy,

- Intracytoplasmic aggregation diseases such as Huntington's and Parkinson's disease.

- Antitrypsin-associated emphysema,

- Cystic fibrosis

- Lysosomal storage diseases,

Experimental techniques

- Protein nuclear magnetic resonance spectroscopy
- Circular dichroism
- Dual polarisation interferometry
- Vibrational circular dichroism of proteins
- Studies of folding with high time resolution
- Proteolysis
- Optical tweezers

Folding funnel
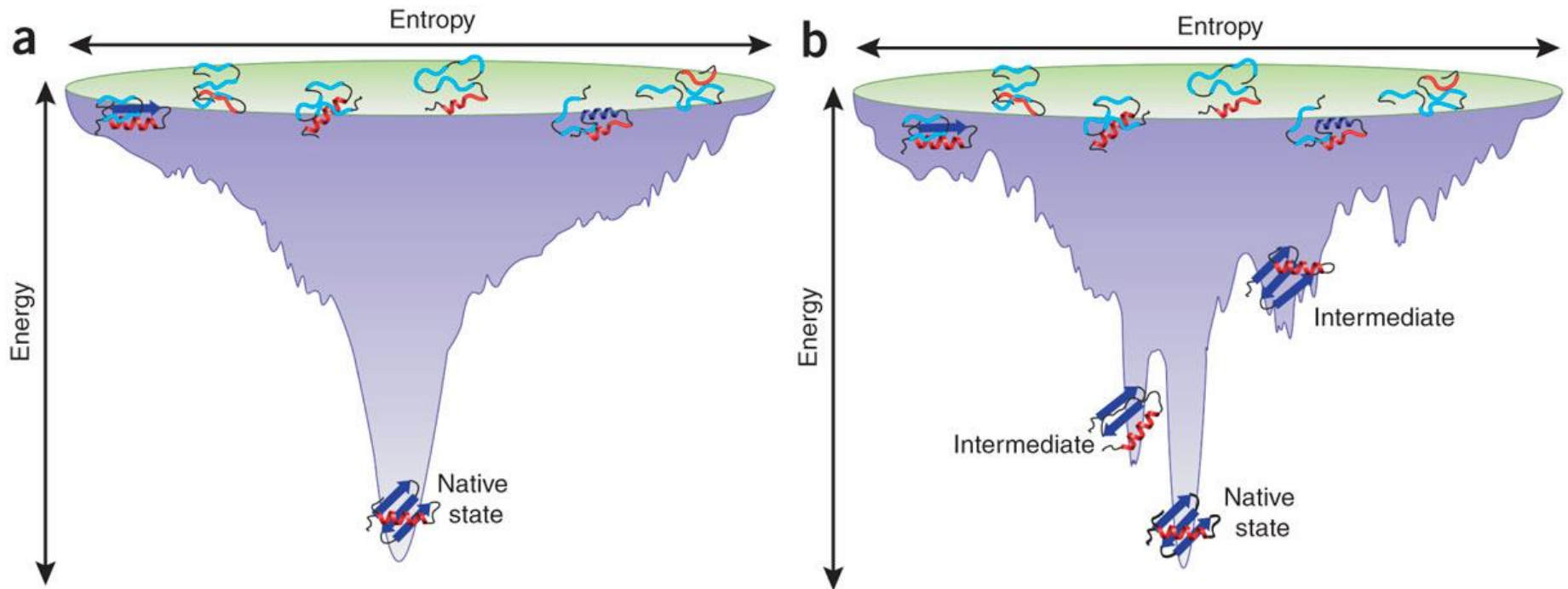
Levinthal Paradox

Levinthal proposed that a random conformational search does not occur, and the protein must, therefore, fold through a series of meta-stable intermediate states.