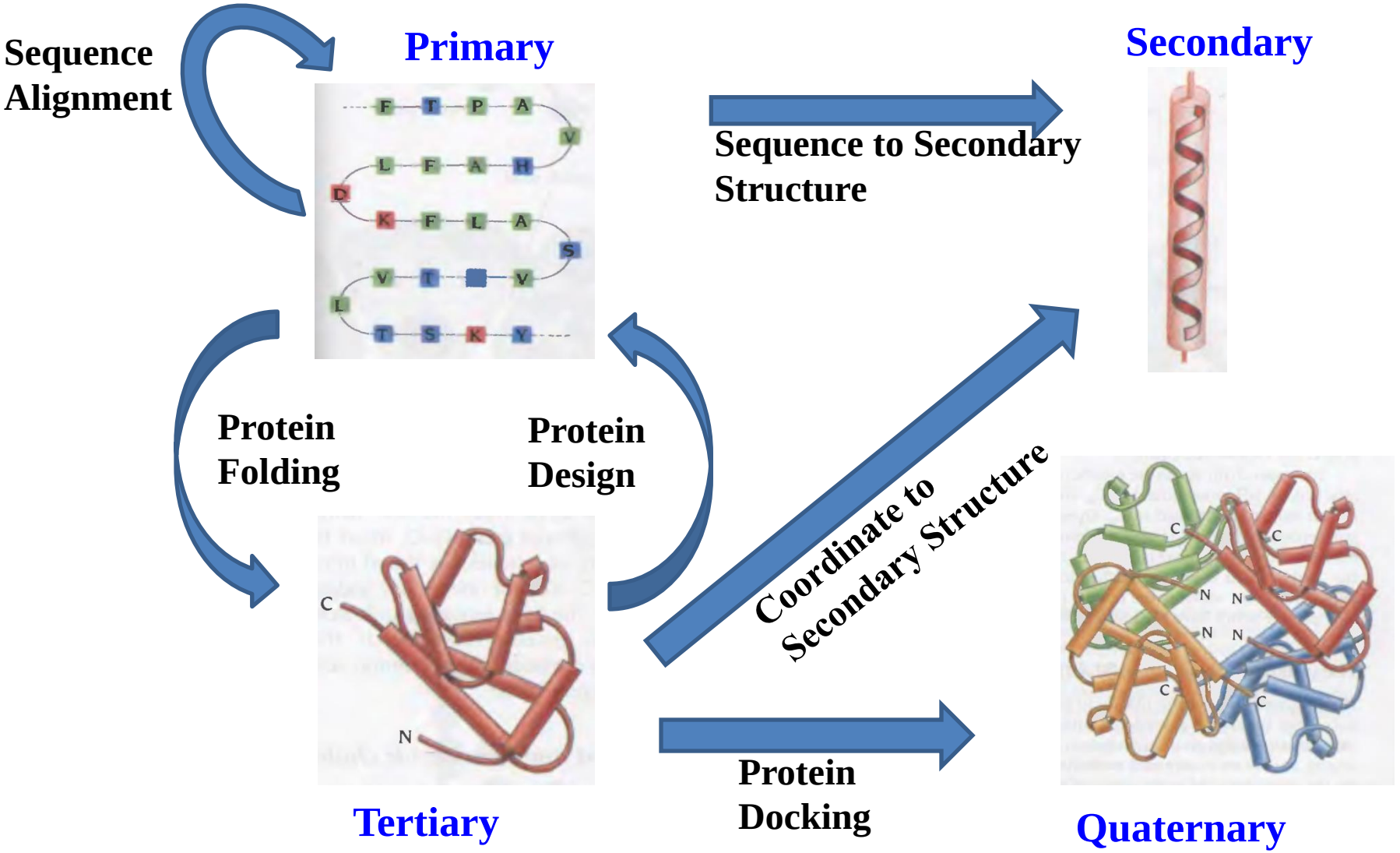
Lecture 17-18

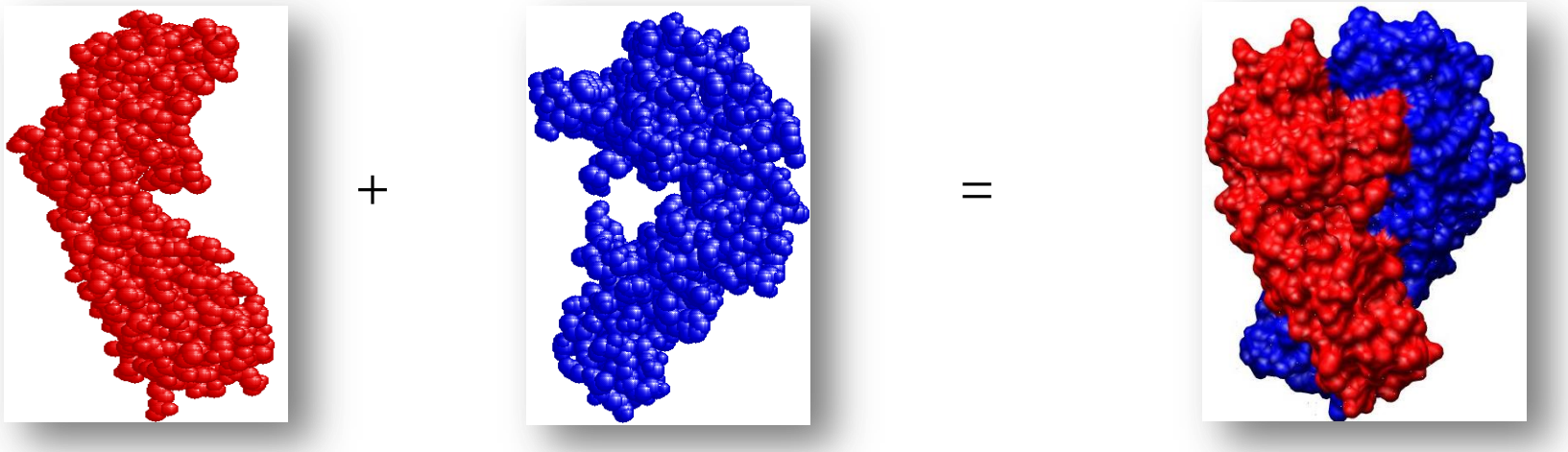
Protein-protein docking

Computational Methods in Proteins



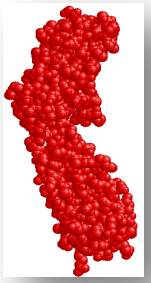
Protein-Protein Docking

Ab initio **Protein-protein docking** is the determination of the molecular structure of *complexes* formed by two or more proteins without the need for *experimental* measurement.

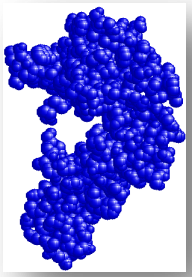


Docking Types

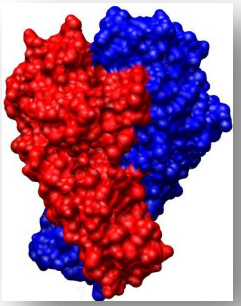
- Based on crystallization information
  - Bound docking
  - Unbound docking/predictive docking



(A)



(B)



(C)

- Based on protein flexibility
  - Rigid Body
  - Flexible Body

Rigid to Flexible

- Protein as a rigid body – a conceptual model
  - No bond stretching/bending
  - Less number of degrees of freedom for the atoms
  - Manageable possibilities
  - Not a practical scenario.

- Protein as a flexible body - a reality
  - Real life situation
  - Possibilities are enormous
  - Mathematical modeling of flexible body is difficult.

Protein Flexibility

Bad News	Good News
Dimensionality of the problem	Whole protein is not flexible ☺
If whole protein is flexible	☹
It varies from case to case	Predict the flexible regions a prior ☺
If flexibility occurs at some distance place not at the contact area	☹
If conformational changes occur after the binding/docking	☹

Protein flexibility is largely infeasible.
Ligand flexibility is possible.

Algorithms for Protein Docking

- Shape Complementarity

  (Connolly 1986, Cazals *et al* 2003)
- Fourier Technique

  (Katchalski-Katzir *et al*, 1992 (FTDock))
- Shape Polynomial and Shape Descriptor

  (Taubin *et al* 1989, Goldman *et al* 2000, Kazhdan *et al* 2003)
- Docking using Biochemical Information

  (Gabb *et al* 1997, Ritchie and Kemp 2000, Chen and Weng 2002, Fernandez-Recio *et al* 2002, Dominguez *et al* 2003 (HADDOCK), Murakami and Jones 2006 (SHARP))
- Meta-heuristic Approach

  (SA - Goodsell and Olson 1990, GA - Poirrette *et al* 1997, Gardiner *et al* 2001)
- Machine Learning Approach

  (Neuvirth *et al* 2004, Bradford and Westhead 2005, Qi *et al* 2006)

Docking Search Strategies

- **Pseudo Random**
  - Simulated Annealing / Monte Carlo
  - Genetic Algorithms
- **Directed Search**
  - Geometric Hashing
  - Spherical Harmonic Surface Triangles
- **Brute-Force Search**
  - Explicit Grid Correlations
  - Fast Fourier Transform (FFT) Correlations
  - Spherical Polar Fourier Correlations

Scoring in docking methods

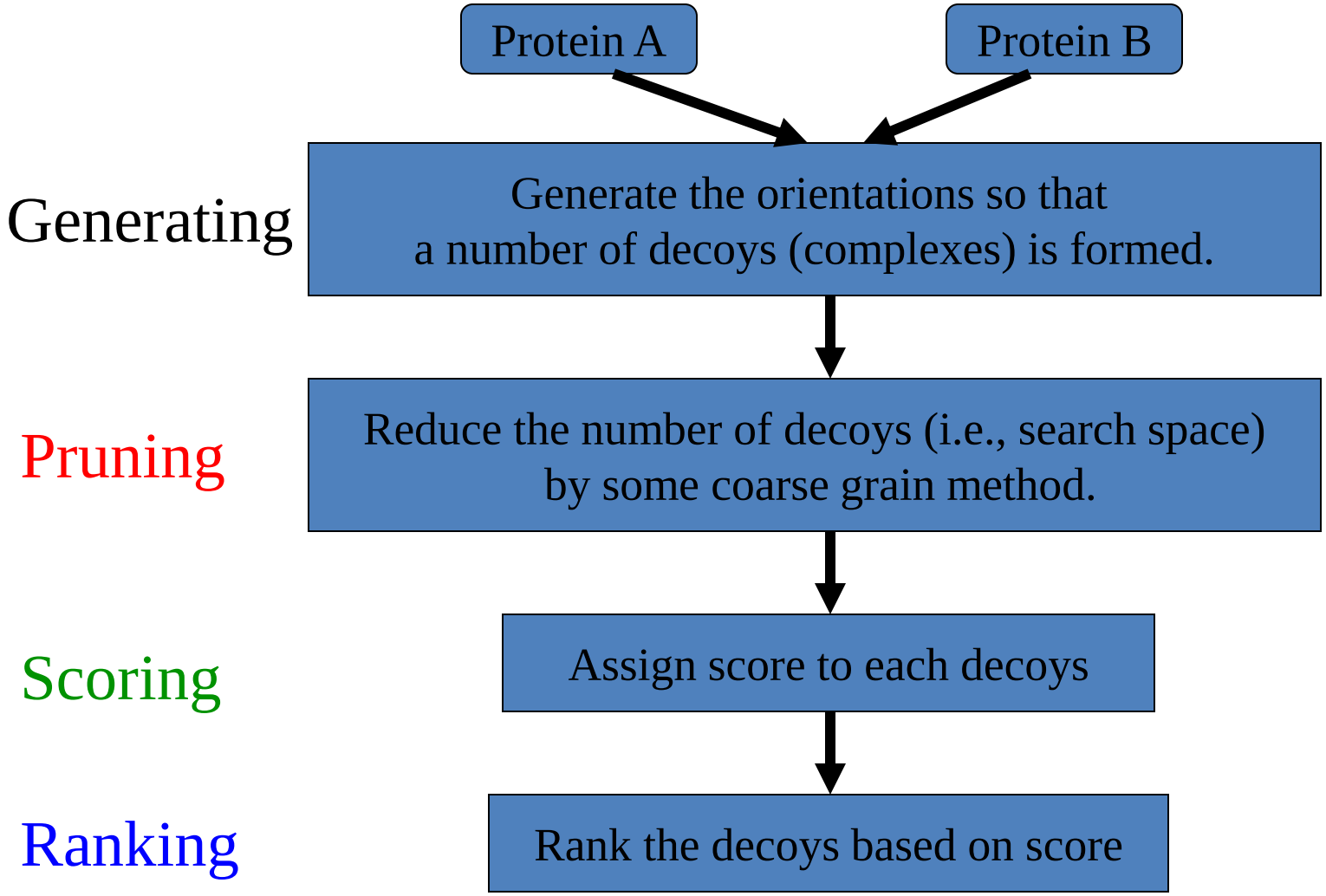
Integrated Scoring

FTDock (Katchalaski-Katzir (1992), PNAS)

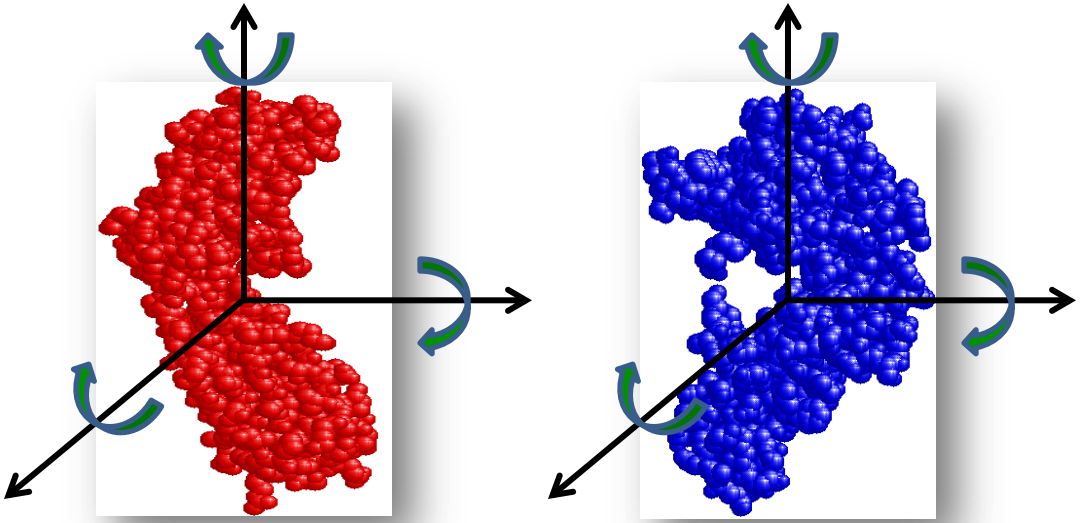
Edge Scoring

PRUNE (Mitra and Pal (2011), Nucleic Acids Res.)

Docking Strategy

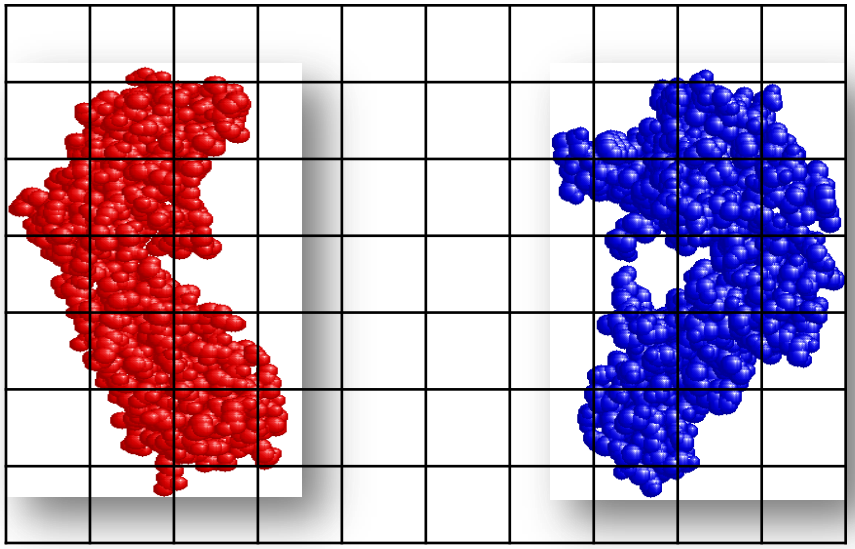


Generation methods



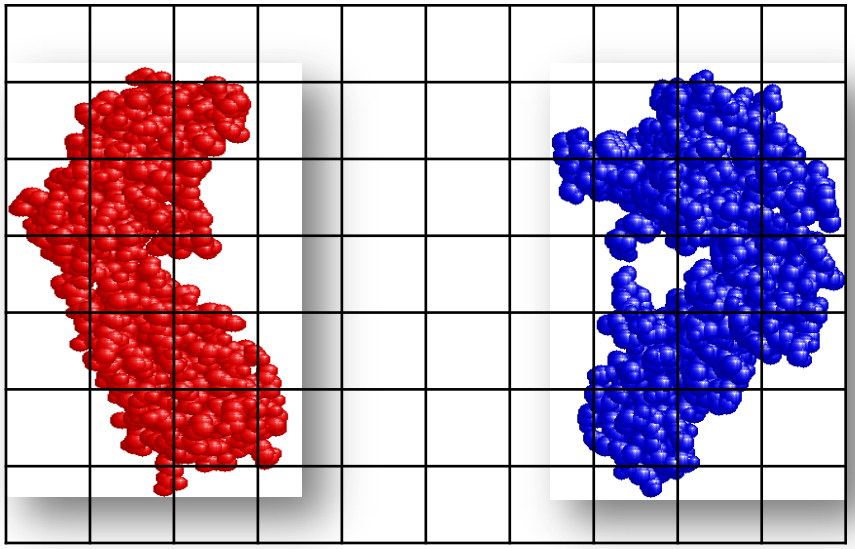
- Tagline – “Higher the decoys; better the possibility of having a hit”
- How many is good?
- Move to discrete space

Generation methods



On an average some brute force method can generate ~10^7 decoys.

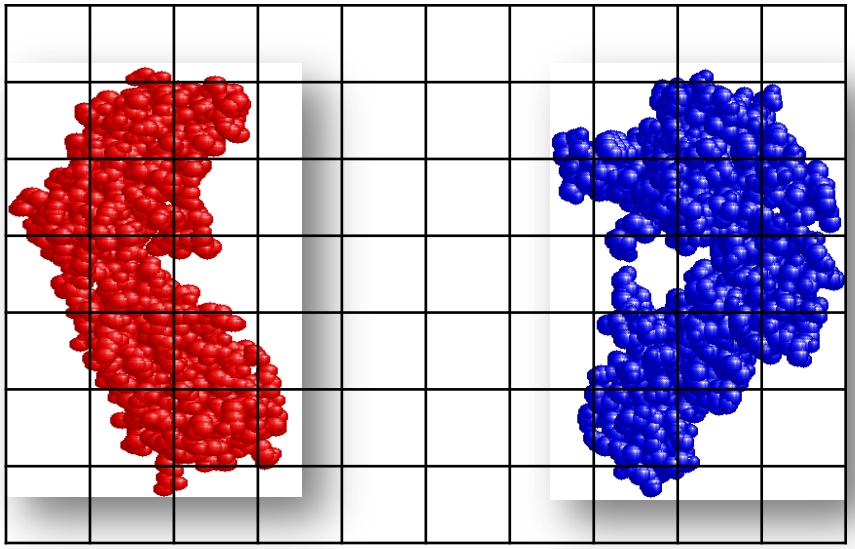
Generation methods



On an average some brute force method can generate ~10^7 decoys.

Assuming processing of each decoy takes 1 sec; total processing time ~115 days.

Fast Fourier Technique

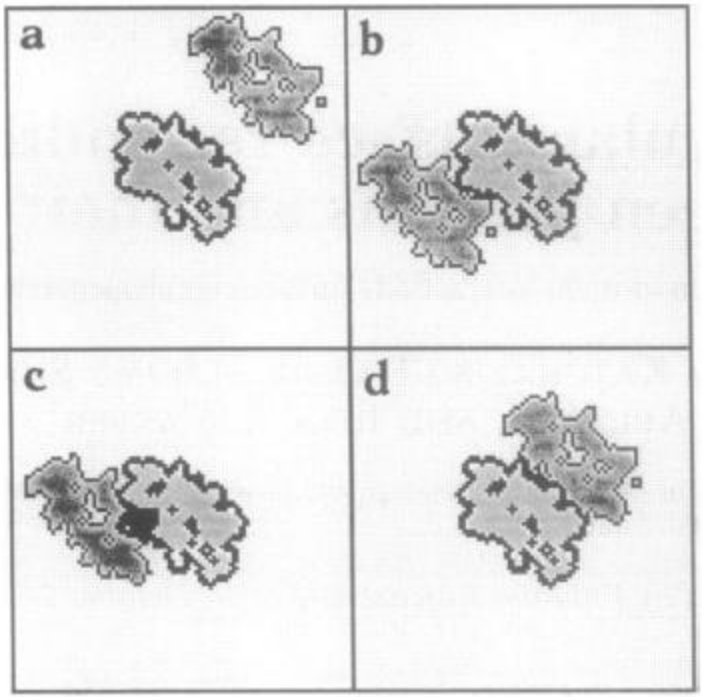


$$\bar{a}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \rho & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases}$$

and

$$\bar{b}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases}$$

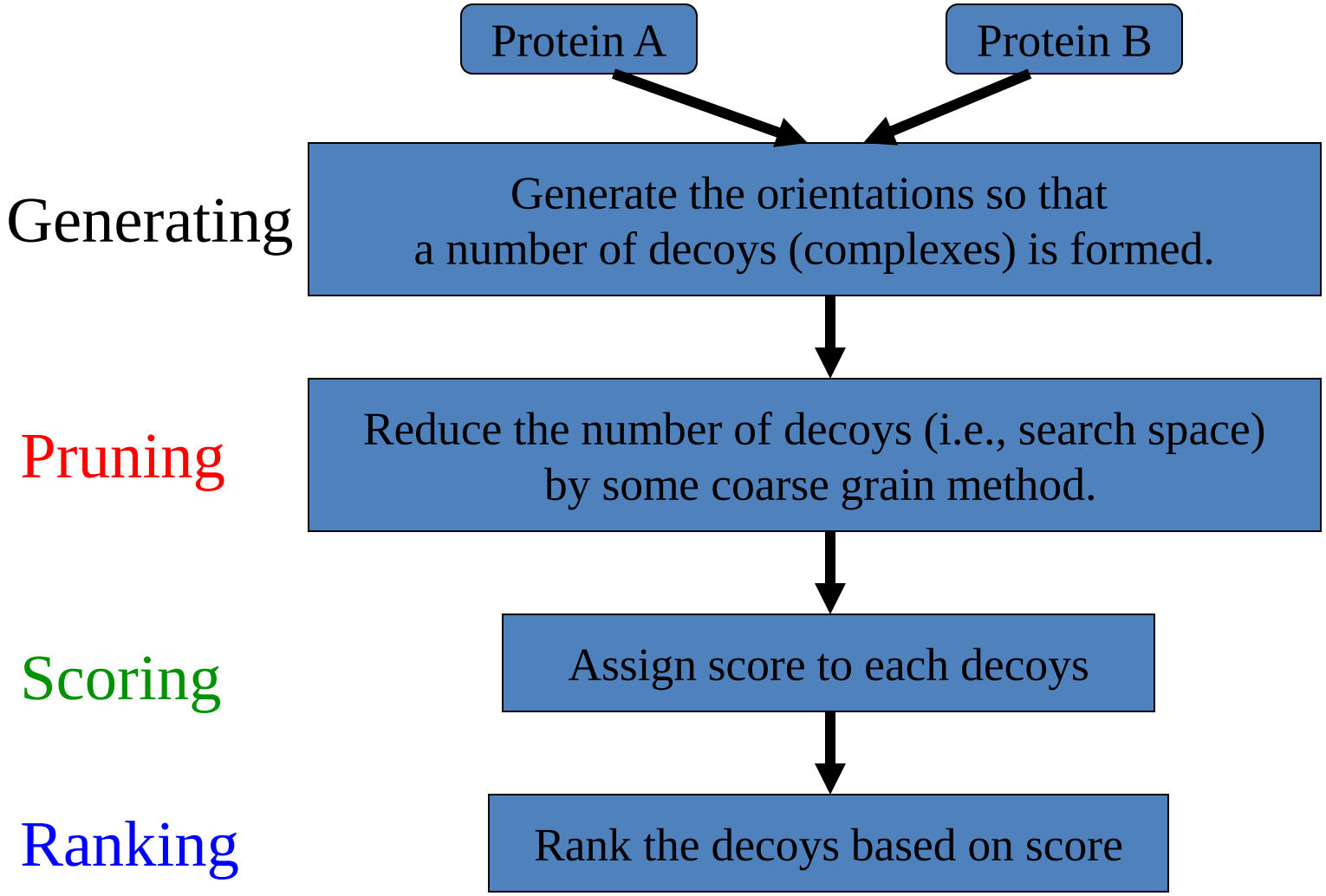
Fast Fourier Technique



$$\bar{C}_{\alpha,\beta,\gamma} = \frac{1}{N^3} \sum_{o=1}^{N} \sum_{p=1}^{N} \sum_{q=1}^{N} \exp[2\pi i(o\alpha + p\beta + q\gamma)/N] \cdot C_{o,p,q}$$

Katchalski-Katzir *et al*, (1992) *PNAS*

Docking Strategy



Scoring methods

Ab initio scoring

- Contact Area
- Contact Packing
- Non-bonded interactions
- Solvation Energy
- Etc.

Template based

Ab initio method

- Interface area (IA)
- Normalized interface packing (NIP)
- Normalized surface complementarity (NSc)
- Non-bonded energy (NE):

$$NE = \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{4\Pi\varepsilon R_{ij}} \right)$$

- Solvation energy (SE):*

$$SE = \sum_{\text{interface atoms}} \Delta\sigma(\text{Atom Type}) \times \Delta\text{ASA}$$

Interface area

- Accessible surface area:
  - The surface area of a molecule which is accessible by the solvent molecules (mostly water, the universal solvent molecule)
- Interface atom:
  - An atom is called as an interface atom if it loses its accessible surface area (ASA) by more than 0.1$Å^2$
- Interface area/Interacting surface:
  - The amount of ASA loses by all the interface atom is the interface area of a complex.