Protein Structure

1) Prediction → Secondary structure — 80% — 85%
   → Tertiary Structure
      → MD (Molecular Dynamic Simulation) ≤ 100 length
      → Data-based ≤ 30%

2) Structural Alignment
   → minimize the coordinate RMSD
   → Internal distance based
   → Geometric Hashing

Given : protein A & protein B

$A = \{a_1, a_2, \dots, a_n\}$ $\quad B = \{b_1, b_2, \dots, b_m\}$

$|A| = n$ $\quad |B| = m$

$a_i \rightarrow C_\alpha^i$ (Carbon α)

$\vec{a_i} \in \mathbb{R}^3 = (a_i^x, a_i^y, a_i^z)$ $\quad \vec{b_j} \in \mathbb{R}^3$

Coordinate Root Mean Squared Distance

Given an Equivalence $\mathcal{E}$, $|\mathcal{E}| = k$

$$RMSD_c(\mathcal{E}) = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(a_i - b_i)^2}$$

Given an equivalence, i.e. a 1-1 mapping between positions in A & B

$a_1 \rightarrow b_2$

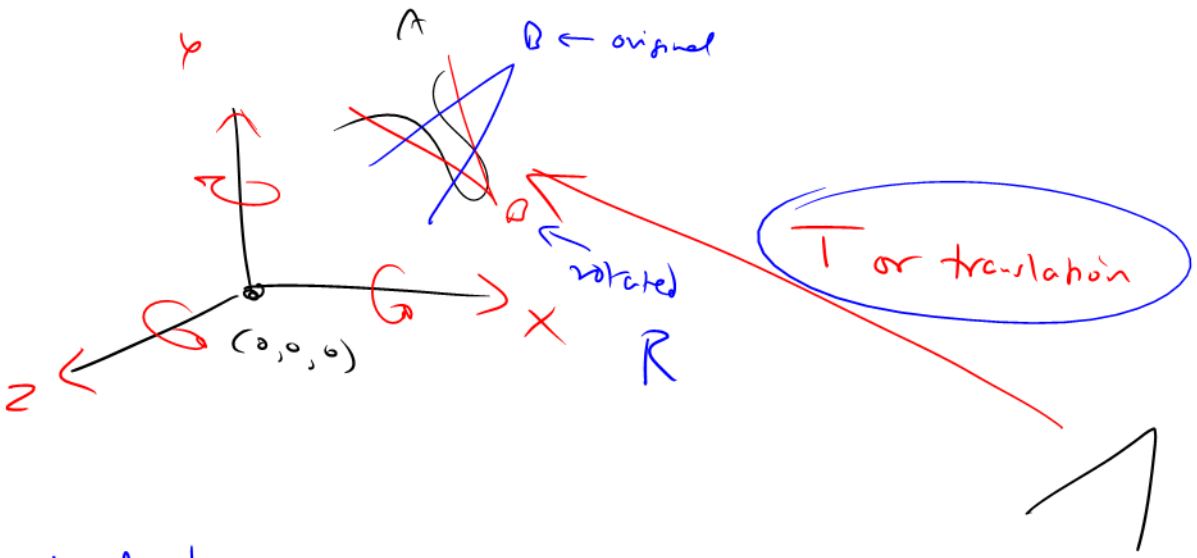
$a_3 \rightarrow b_4$ $\quad k = 4$

$a_4 \rightarrow b_5$

$a_{10} \rightarrow b_6$

$\mathcal{E} = \{ (a_1, b_2)\ (a_3, b_4)\ (a_4, b_5)\ (a_{10}, b_6) \}$

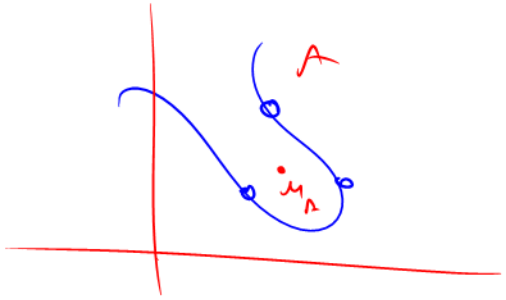
Goal: find a rotation & translation of B that minimizes $RMSD_c$ & maximizes $|\mathcal{E}|$

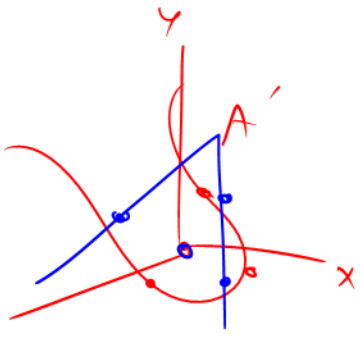


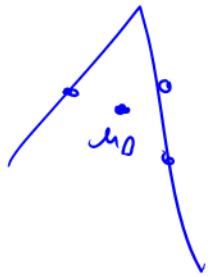
optimal translation

T: Compute $\vec{\mu}_A = \frac{1}{k}\sum_{i=1}^{k}\vec{a}_i = (\mu_A^x, \mu_A^y, \mu_A^z)^T$

$\vec{a}_i' = (\vec{a}_i - \vec{\mu}_A)$







T. ✓

1) find mean μ_A, μ_B from $\in$

2) Subtract μ_A from all $\vec{a}_i \in A$

μ_B " " $\vec{b}_i \in B$

---

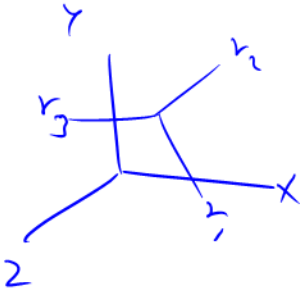
2) find the optimal rotation

Every rotation can be described by a square matrix $R \in \mathbb{R}^{3\times3}$ which satisfy 2 properties

(in 3D)

R: 3x3 matrix

1) R must be orthogonal



$$R = \begin{pmatrix} | & | & | \\ r_1 & r_2 & r_3 \\ | & | & | \end{pmatrix}$$

$r_i^T r_j = 0$ ← new axes are 90° to each other

$r_i^T r_i = 1$ ← axes are normalized to be unit vectors

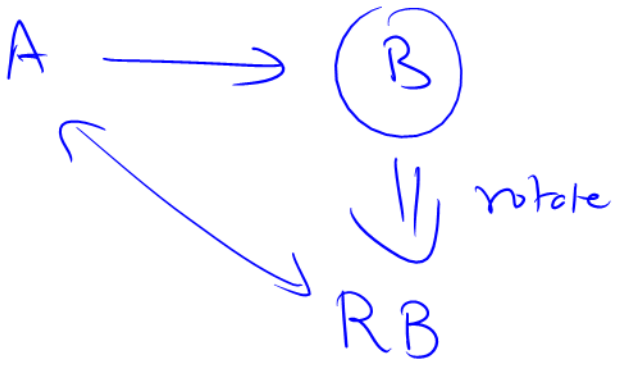
$\Rightarrow \underline{R^{-1} = R^T}$, the inverse is simply the transpose

$$R^T R = I$$

2) $\det(R) = +1$

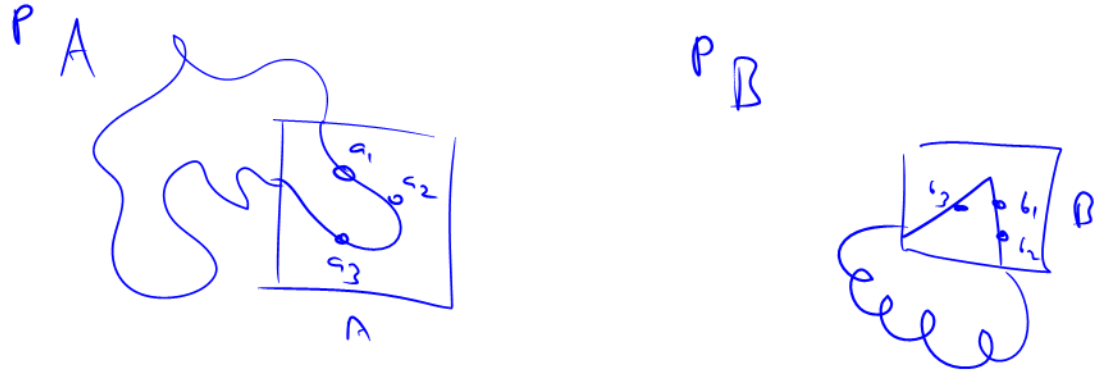
Optimal rotation R after centering

↘ i.e. $\min_R RMSD_C(E)$



$$R^* = \min_R RMSD_C(\Theta, R) = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\vec{a}_i - R\vec{b}_i)^2}$$

The matrix A = 3 x k matrix



$$\mathcal{E} = \{(a_1 b_1)\ (a_2 b_2)\ (a_3 b_3)\}$$

A: $3 \times k$ matrix $\begin{pmatrix} | & | & | & & | \\ a_1 & a_2 & a_3 & \cdots & a_k \\ | & | & | & & | \end{pmatrix}$

B = $3 \times k$ matrix $\begin{pmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_k \\ | & | & & | \end{pmatrix}$

$$C = B A^T$$

$3 \times k$ $\quad k \times 3$

3×3

$$\begin{pmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_k \\ | & | & & | \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$$

B $\qquad$ A^T

singular vectors ↘ ↙ singular values

$$C = U \Delta V^T$$

do SVD: Singular Value Decomp

Δ is a 3×3 diagonal matrix $\begin{pmatrix} \lambda_1 & 0 & \\ & \lambda_2 & \\ 0 & & \lambda_3 \end{pmatrix}$

$$R = V U^T$$

Is $\det R = -1$ possible?

$$R' = VWU^T$$ ← final rotation matrix

$$= V \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \text{sign} \end{pmatrix} U^T$$

sign $\propto$ det (R)

$\therefore$ if sign $= +1$, then $R' = VIU^T = VU^T$

sign $= -1$, then R' will have +ve det.

Iterative Superposition & Dynamic Programming

assume k is given

0) Guess a starting $\mathcal{E}_0$: randomly ✓

Choose k pairs (a_i, b_i) at random

$t = 0$

Given $\mathcal{E}_t$

1) find the optimal R after centering

Compute $RMSD_C(\mathcal{E}_t)$ using R^*

↳ $C = BA^T$

$R^* = VWU^T$

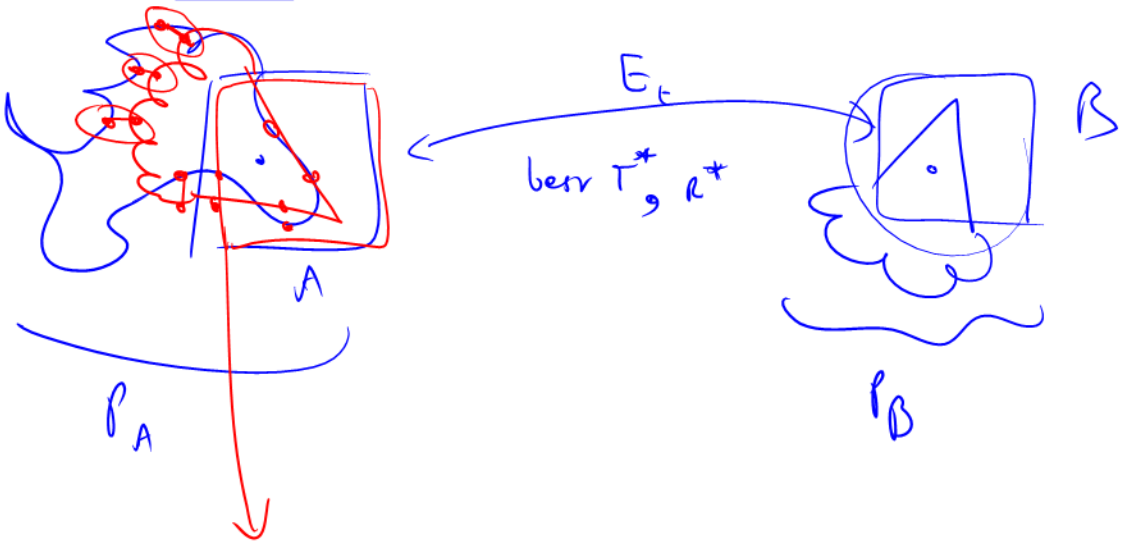
Now rotate the entire protein P_B using R^*

2) Given that P_A & P_B have been superimposed optimally given $\mathcal{E}_t$

find a new set of Equivalences $\mathcal{E}_{t+1}$

done using DP: Dynamic Programming

3) repeat 1) & 2) until convergence or after some # of steps



select $\mathcal{E}_{t+1}$ using DP.