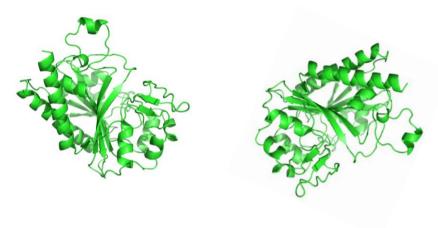
Lecture 14-15

Measuring the goodness of the modeled Protein structures

Structure Alignment

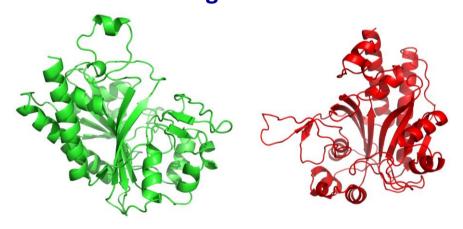


Why do we need structure alignment?



Frame of reference is different.

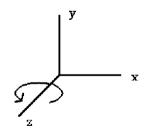
Why do we need structure alignment?



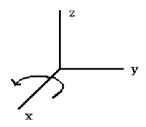
Frame of reference is different.

How?

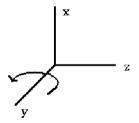
Rotation About Z-Axis in 3D



Rotation About X-Axis in 3D



Rotation About Y-Axis in 3D



$$Ry(q) = (0 & 0 & \sin q & 0) \\ (-\sin q & 0 & \cos q & 0) \\ (0 & 0 & 0 & 1)$$

Rotation About an Arbitrary Axis in 3D

- (1) Translate space so that the rotation axis passes through the origin.
- (2) Rotate space about the z axis so that the rotation axis lies in the xz plane.
- (3) Rotate space about the y axis so that the rotation axis lies along the z axis.
- (4) Perform the desired rotation by θ about the z axis.
- (5) Apply the inverse of step (3).
- (6) Apply the inverse of step (2).
- (7) Apply the inverse of step (1).

Rotation About an Arbitrary Axis in 3D

The matrices for rotation by α around the x-axis, β around the y-axis, and γ around the z-axis

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha & 0 \\ 0 & \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_y(\beta) = \begin{bmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_{z}(\gamma) = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 & 0\\ \sin \gamma & \cos \gamma & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Rotation About an Arbitrary Axis in 3D

The general rotation matrix depends on the order of rotations. The first matrix rotates about x, then y, then z; the second rotates about z, then y, then x.

$$R_{z}R_{y}R_{x} = \begin{bmatrix} \cos\beta\cos\gamma & \cos\gamma\sin\alpha\sin\beta - \cos\alpha\sin\gamma & \cos\alpha\cos\gamma\sin\beta + \sin\alpha\sin\gamma & 0\\ \cos\beta\sin\gamma & \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & -\cos\gamma\sin\alpha + \cos\alpha\sin\beta\sin\gamma & 0\\ -\sin\beta & \cos\beta\sin\alpha & \cos\alpha\cos\beta & 0\\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_{x}R_{y}R_{z} = \begin{bmatrix} \cos\beta\cos\gamma & -\cos\beta\sin\gamma & \sin\beta & 0\\ \cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma & \cos\alpha\cos\gamma - \sin\alpha\sin\beta\sin\gamma & -\sin\alpha\cos\beta & 0\\ \sin\alpha\sin\gamma - \cos\alpha\sin\beta\cos\gamma & \sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma & \cos\alpha\cos\beta & 0\\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_x R_y R_z = \begin{bmatrix} \cos \beta \cos \gamma & -\cos \beta \sin \gamma & \sin \beta & 0\\ \cos \alpha \sin \gamma + \sin \alpha \sin \beta \cos \gamma & \cos \alpha \cos \gamma - \sin \alpha \sin \beta \sin \gamma & -\sin \alpha \cos \beta & 0\\ \sin \alpha \sin \gamma - \cos \alpha \sin \beta \cos \gamma & \sin \alpha \cos \gamma + \cos \alpha \sin \beta \sin \gamma & \cos \alpha \cos \beta & 0\\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Limitations

- Computationally slow
- Not recommended for large scale application
- Alternative is Quaternion based method.

$$\begin{bmatrix} c + a_x^2(1-c) & a_x a_y(1-c) - a_z s & a_x a_z(1-c) + a_y s \\ a_y a_x(1-c) + a_z s & c + a_y^2(1-c) & a_y a_z(1-c) - a_x s \\ a_z a_x(1-c) - a_y s & a_z a_y(1-c) + a_x s & c + a_z^2(1-c) \end{bmatrix}$$

Homework

- Develop a faster method compared to this.
 - Hint: Can use Quaternion.

Measure of Structure Alignment



RMSD – Root Mean Square Deviation

> All atom RMSD Backbone RMSD Backbone-trace RMSD

Measure of Structure Alignment

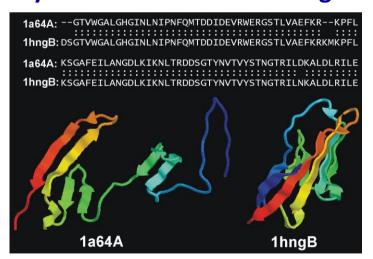


RMSD – Root Mean Square Deviation

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2}$$

All atom RMSD Backbone RMSD Backbone-trace RMSD

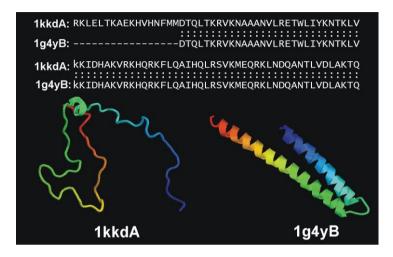
Why do we need structure alignment?



High sequence identity, but different fold.

Zhang & Skolnick (2005). NAR 33:2302-2309

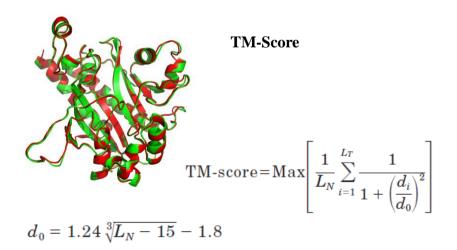
Why do we need structure alignment?



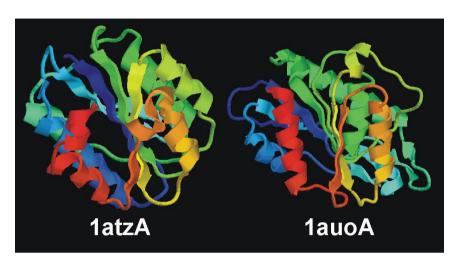
High sequence identity, but different fold.

Zhang & Skolnick (2005). NAR 33:2302-2309

Measure of Structure Alignment

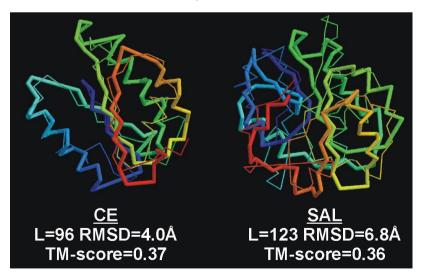


Comparison



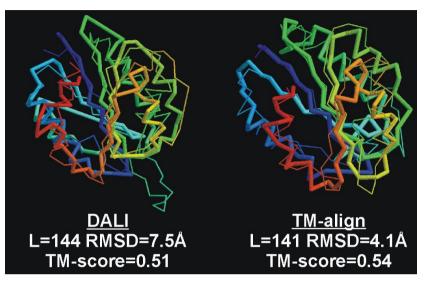
Zhang & Skolnick (2005). NAR 33:2302-2309

Comparison



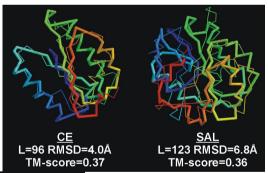
Zhang & Skolnick (2005). NAR 33:2302-2309

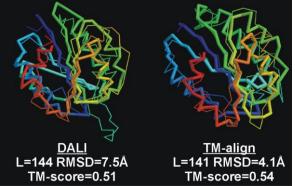
Comparison



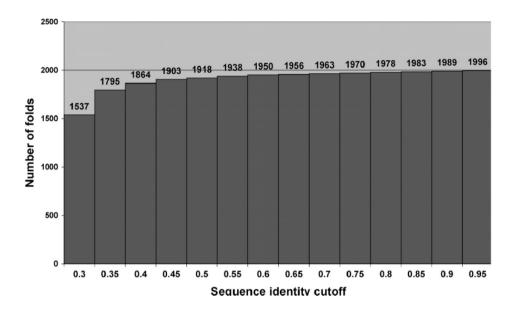
Zhang & Skolnick (2005). NAR 33:2302-2309

Comparison



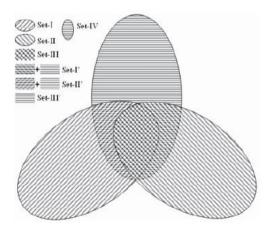


Zhang & Skolnick (2005). NAR 33:2302-2309



Zhang & Skolnick (2005). NAR 33:2302-2309

Significance of TM-Score



Set-I: 746 420 same Fold domain pairs generated from 11 239 protein domains in SCOP.

Set-II: 2 769 868 same Topology domain pairs generated from 14 830 protein domains in CATH.

Set-III: overlap part of Set-I and Set-II, which includes 186 359 pairs from 5105 consensus domains.

Set-IV: 13 027 960 all-to-all pairs from the 5105 consensus domains.

Set-I': different fold set for SCOP, generated by subtracting a subset of Set-I from Set-IV.

Set-II': different fold set for CATH, generated by subtracting a subset of Set-II from Set-IV.

Set-III': different fold set for Set-III and obtained by subtracting subsets of Set-I and Set-II from Set-IV.

Measuring the goodness of the decoy structures

- LRMSD
- iRMSD
- Fnat score

