Methods of Protein Structure Elucidation

Yaroslav Ryabov

Main goals of Computational Biology

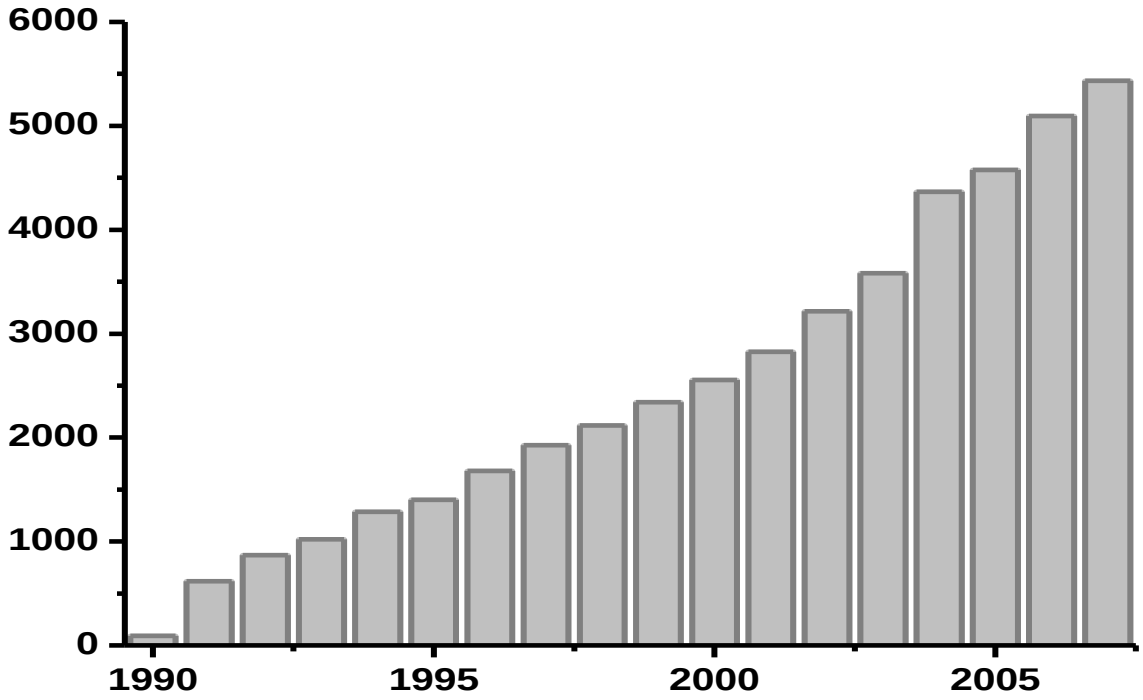
- Advance understanding of biological processes
- And eventually provide new drugs

By using modern computational technologies

There is still no a single drug on the market that was designed exclusively by a computer

But …

Publications on Computer Drug Discovery subject



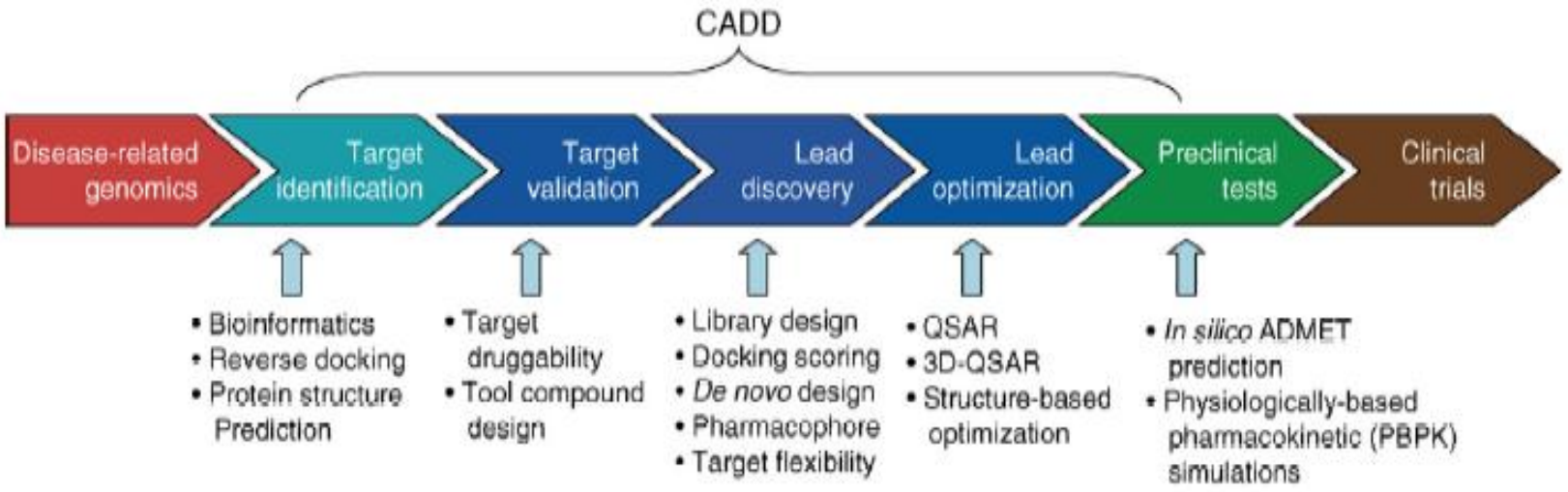
Utility of computational methods in drug discovery

The cost of drug discovery and development

$800 million dollars

Time to market approximately **12** years

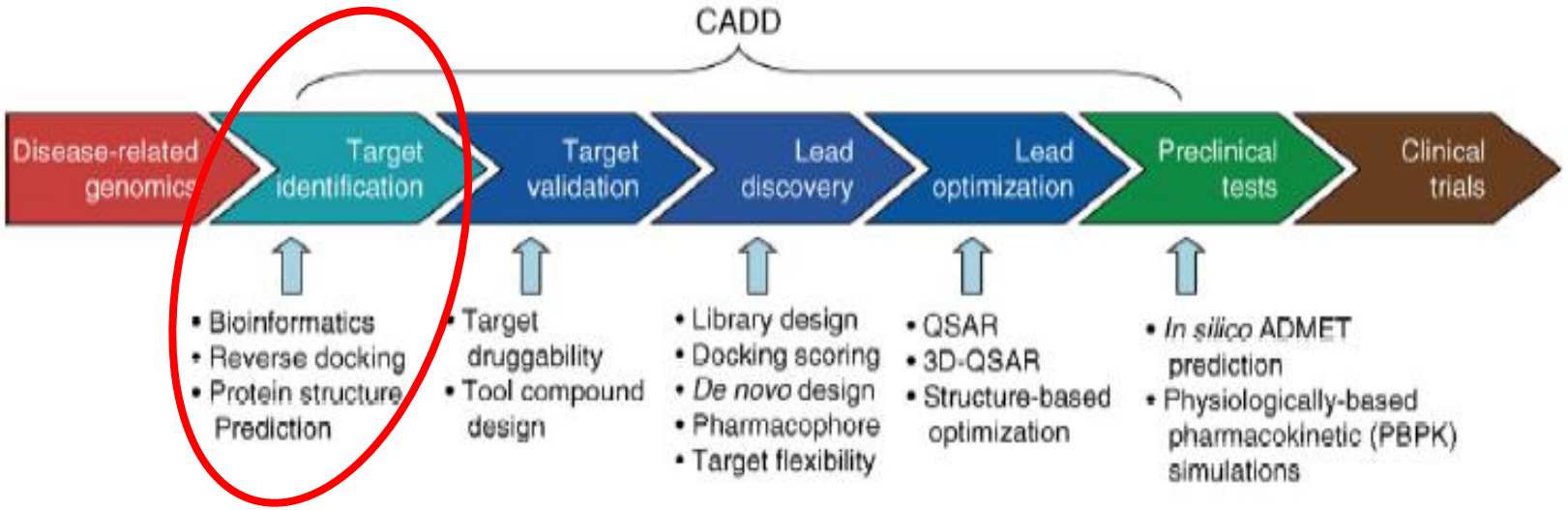
Drug discovery pipeline
And Computer Assisted Drug Design tools.



QSAR Quantitative structure-activity relationship

ADME Absorption, Distribution, Metabolism, Excretion and Toxicology

Drug discovery pipeline
And Computer Assisted Drug Design tools.



QSAR Quantitative structure-activity relationship

ADME Absorption, Distribution, Metabolism, Excretion and Toxicology

Getting protein structures

Imaging methods

X-ray crystallography, Cryo-Electron microscopy

Computational methods

Structure Predictions and Structure Elucidation

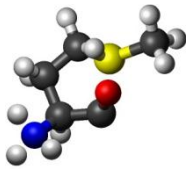
Protein structure prediction

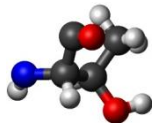
Sequence of amino-acid residues for streptococcal Protein GB1

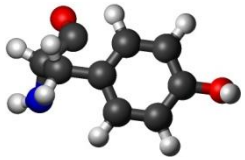
MET THR TYR LYS LEU ILE LEU ASN GLY
LYS THR LEU LYS GLY GLU THR THR THR
GLU ALA VAL ASP ALA ALA THR ALA GLU
LYS VAL PHE LYS GLN TYR ALA ASN ASP
ASN GLY VAL ASP GLY GLU TRP THR TYR
ASP ASP ALA THR LYS THR PHE THR VAL
THR GLU

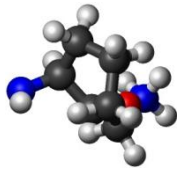
Protein structure prediction

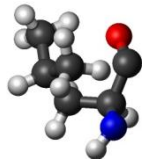
Structures of amino-acid residues are known

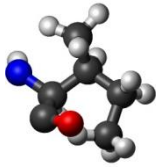
MET 

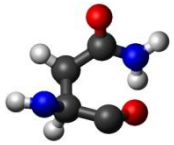
THR 

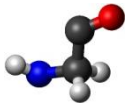
TYR 

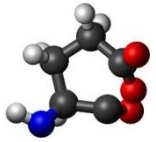
LYS 

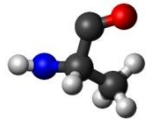
LEU 

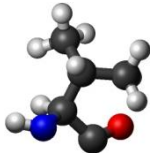
ILE 

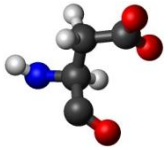
ASN 

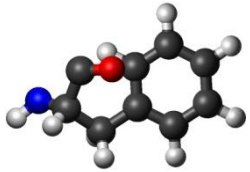
GLY 

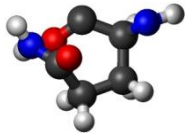
GLU 

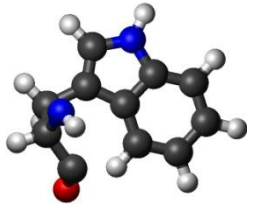
ALA 

VAL 

ASP 

PHE 

GLN 

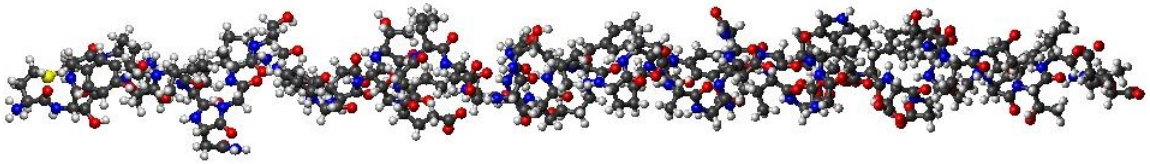
TRP 

Protein structure prediction

The Goal

Use Extended strand

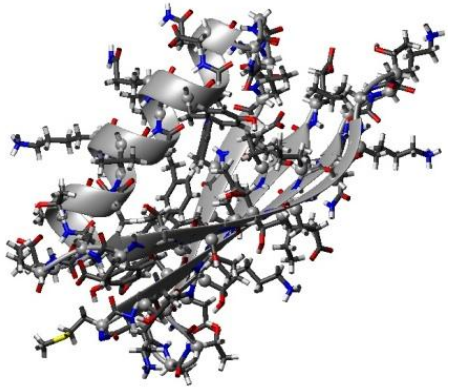
Which is build using structures of amino-acid residues



+

Model inter molecular Force fields (CHARMM)

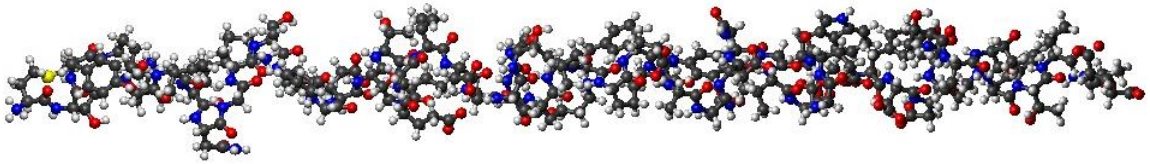
=



Protein structure prediction

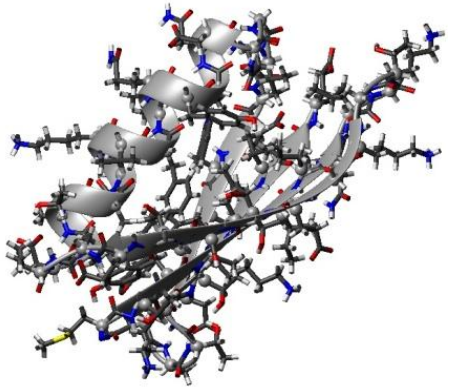
The Goal

Use Extended strand
Which is build using structures of amino-acid residues



+

Model inter molecular Force fields (CHARMM)

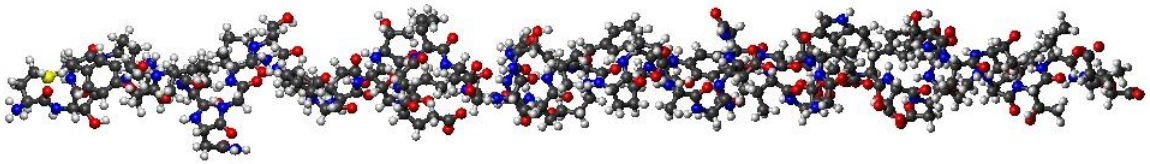


Is yet to accomplish

Structure elucidation

Use Extended strand

Which is build using structures of amino-acid residues



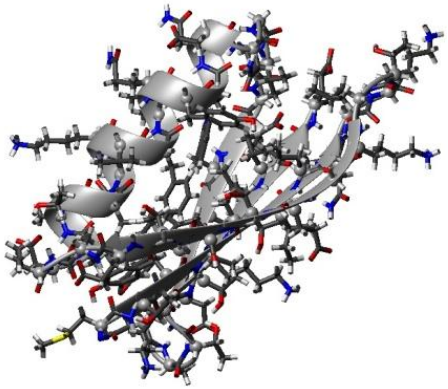
+

Model inter molecular Force fields (CHARMM)

+

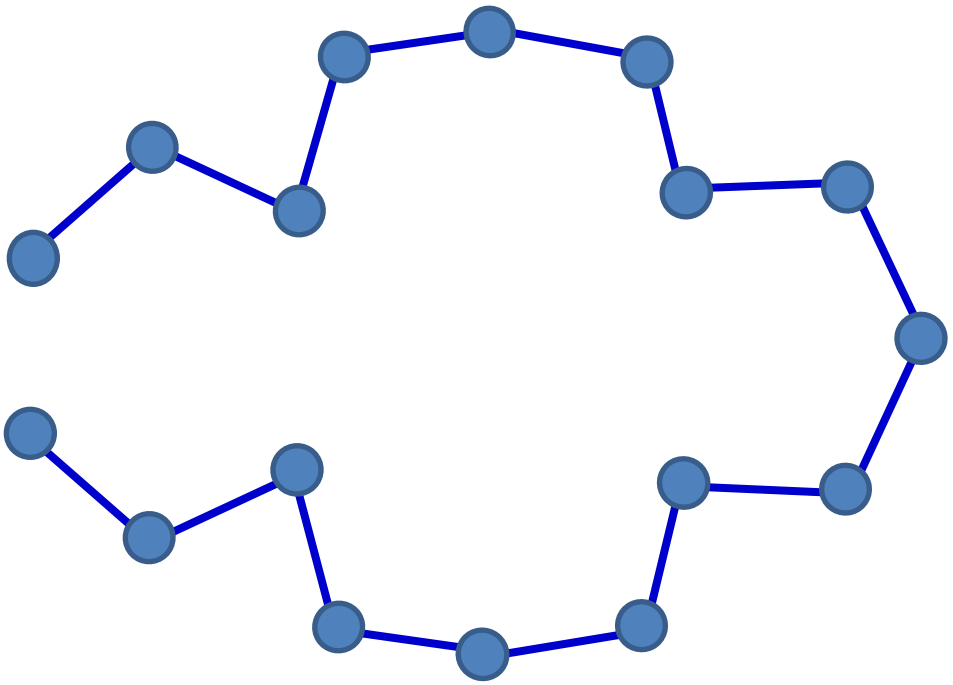
Experimental restraints: NOE, RDC, T1 and T2, etc.

=



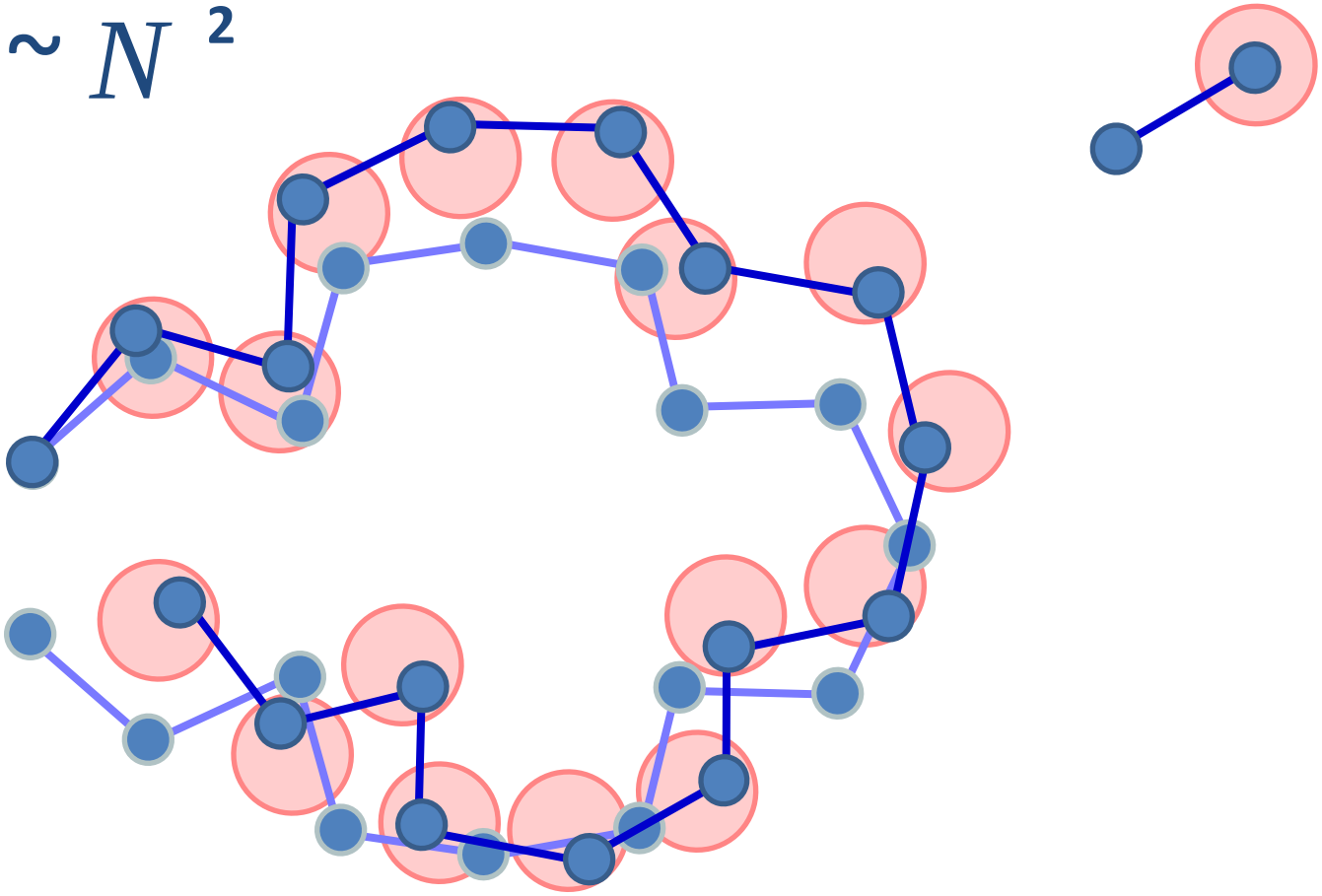
Local and Global restraints

Ideal structure



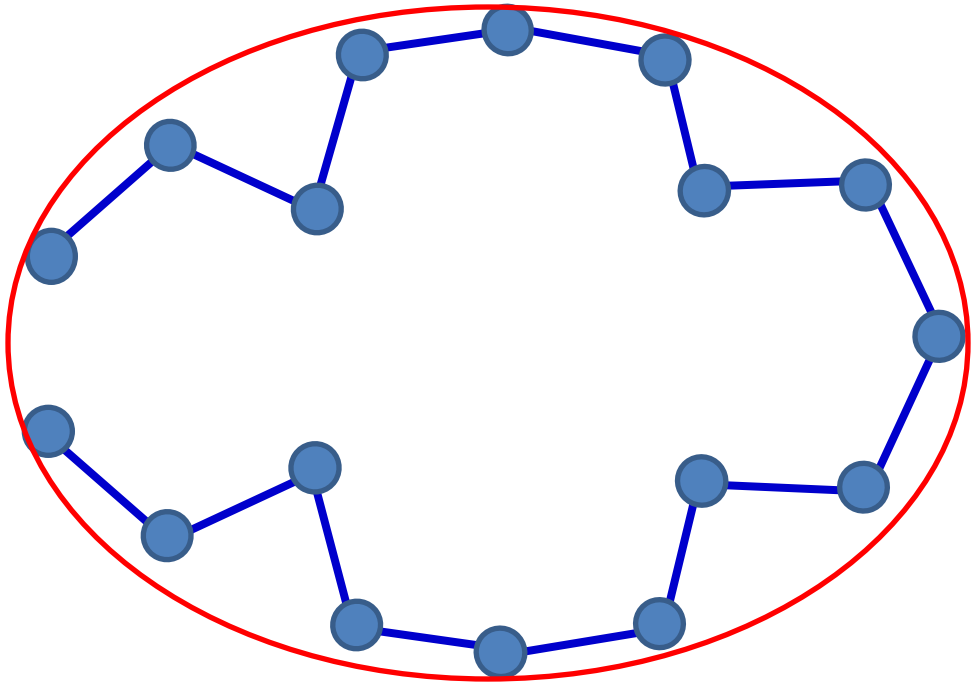
Local restraints

$$\Delta \sim N^{2}$$

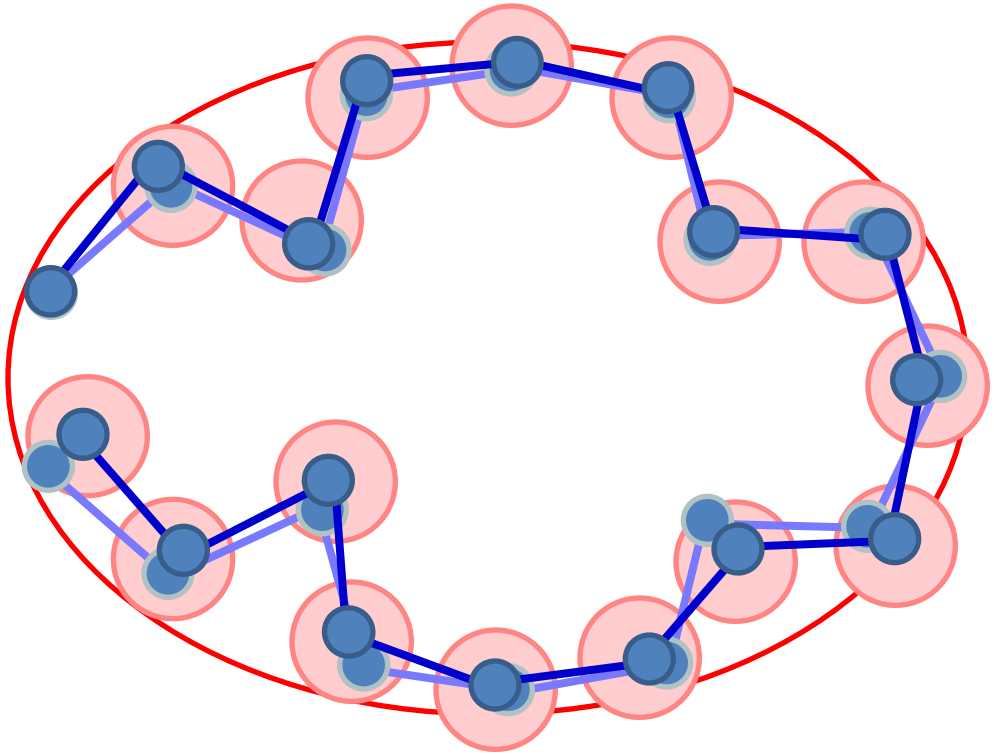


Global restraints

Overall shape

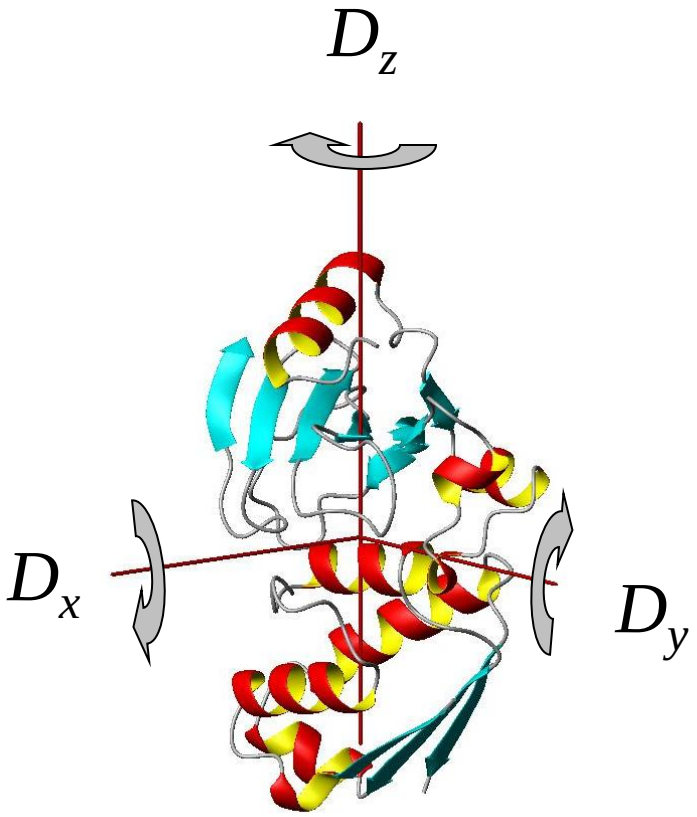


Local and Global restraints



Overall shape restraints from protein dynamics

$$\begin{bmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_z \end{bmatrix}$$



3 Euler angles for
Diffusion Tensor PAF

Diffusion Properties of Proteins

from ellipsoid model

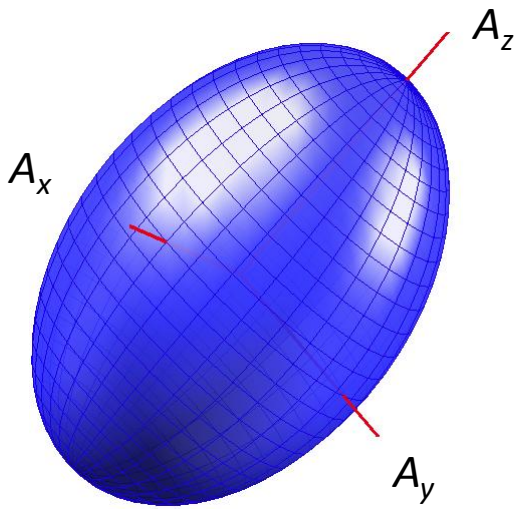
Why an ellipsoid model ?

Diffusion Tensor

$$\begin{bmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_z \end{bmatrix}$$

One-to-**One**
mapping

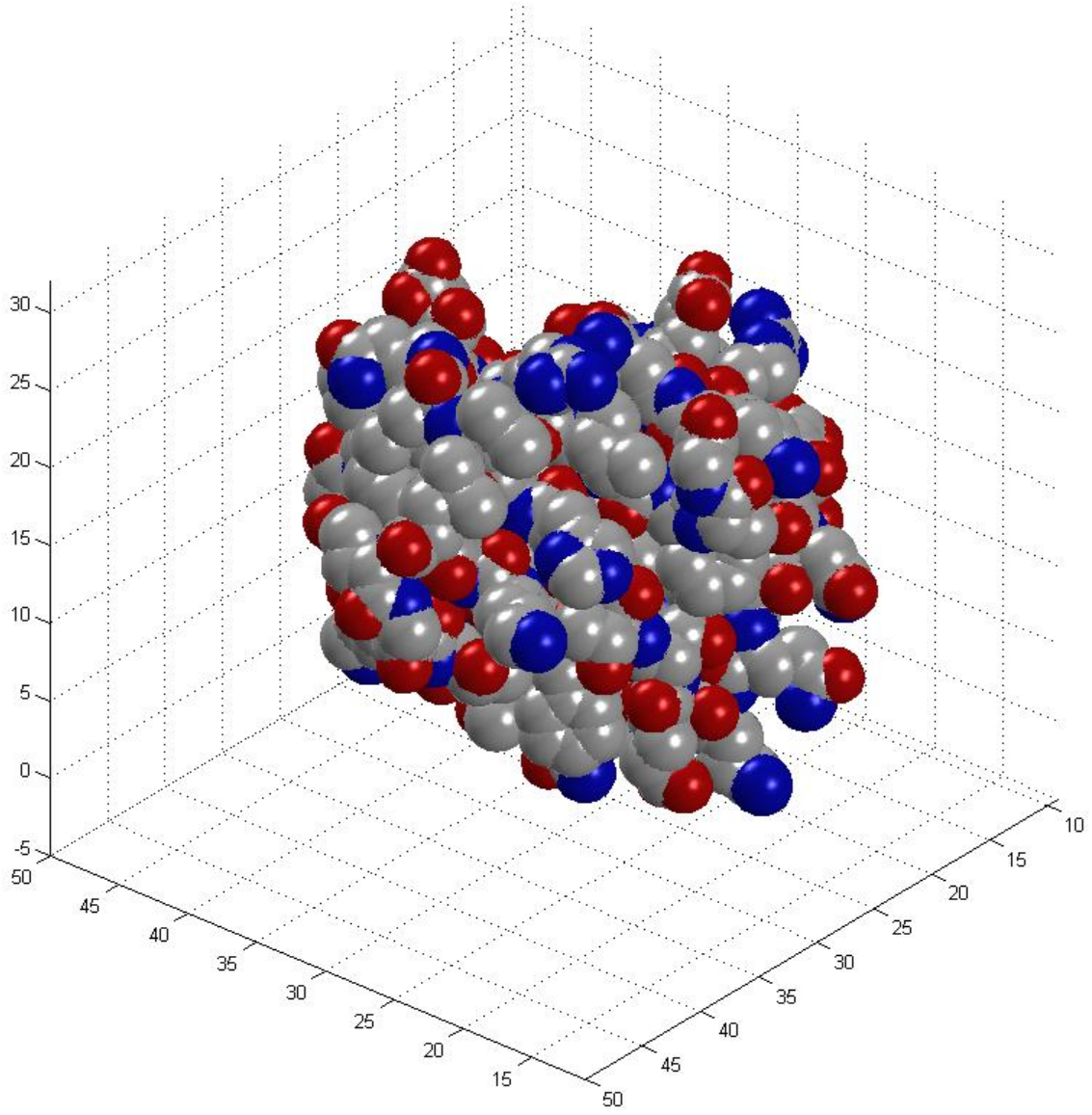
Ellipsoid Shell



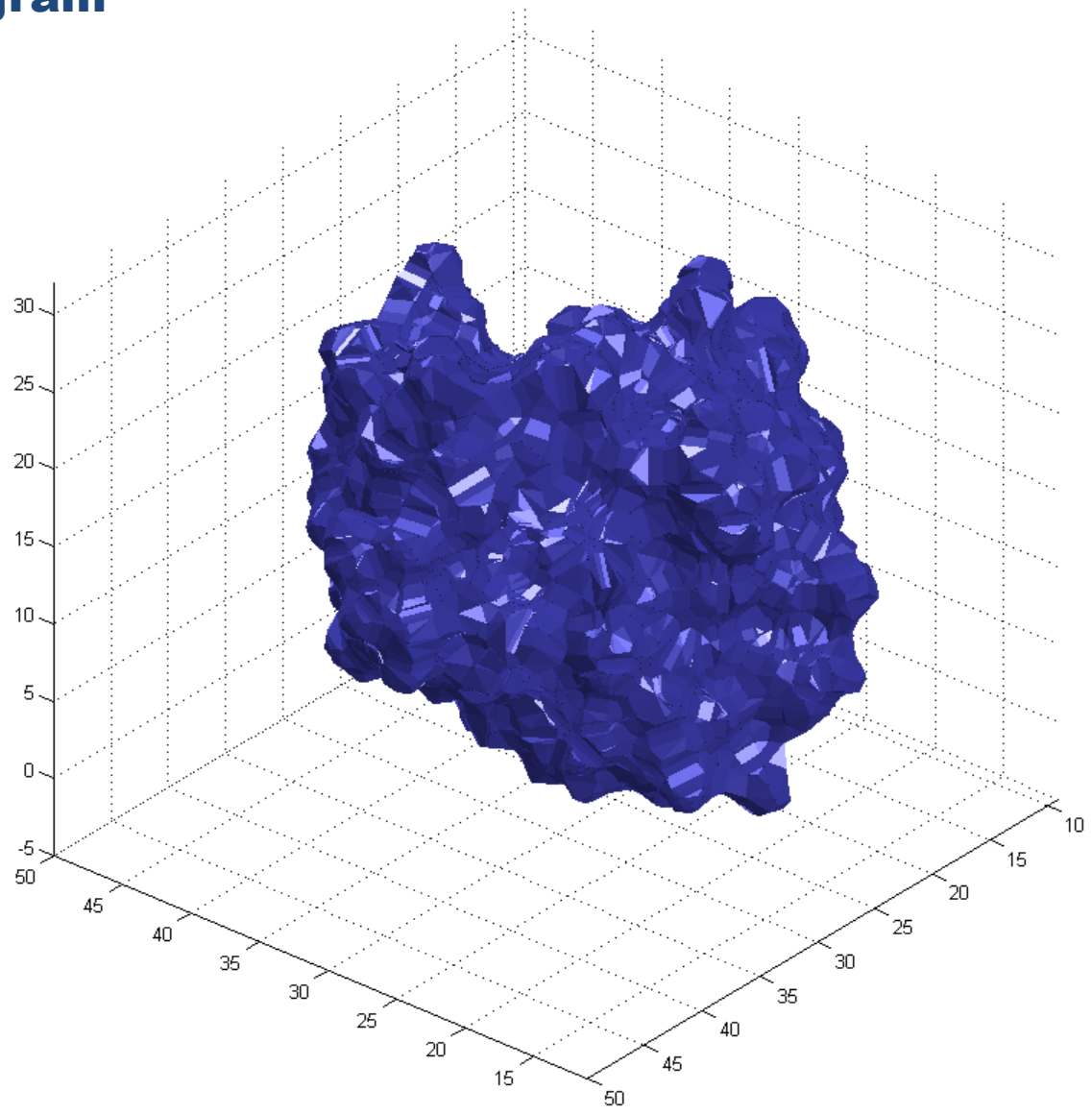
3 Euler angles for
Diffusion Tensor PAF

3 Euler angles for
Ellipsoid orientation

Mapping protein surfaces



Mapping protein surfaces
SURF program



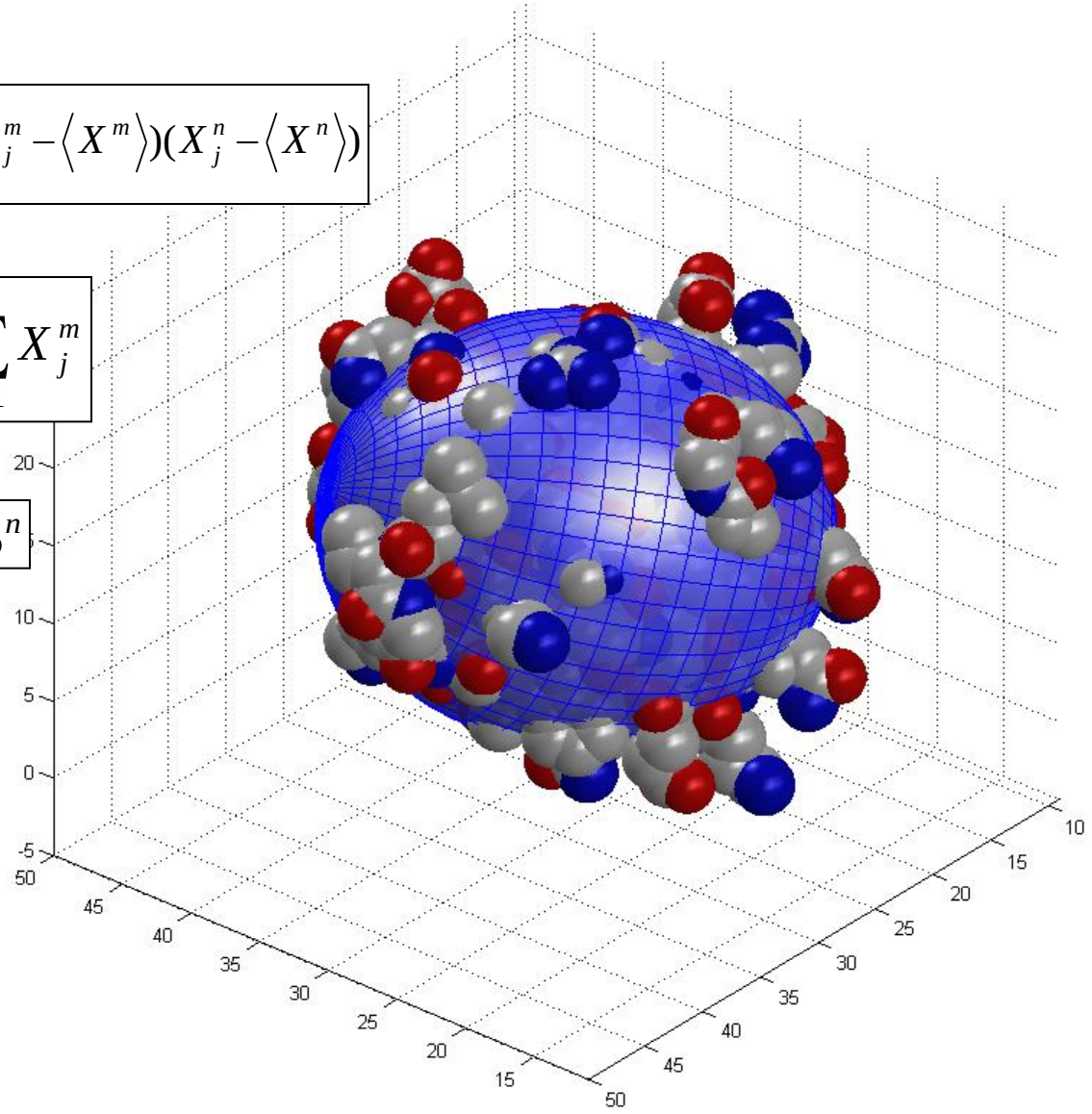
Build equivalent ellipsoid
Principal Component Analysis (PCA)

$$Cov_{m,n} = \frac{1}{N}\sum_{j=1}^{N}(X_j^m - \langle X^m \rangle)(X_j^n - \langle X^n \rangle)$$

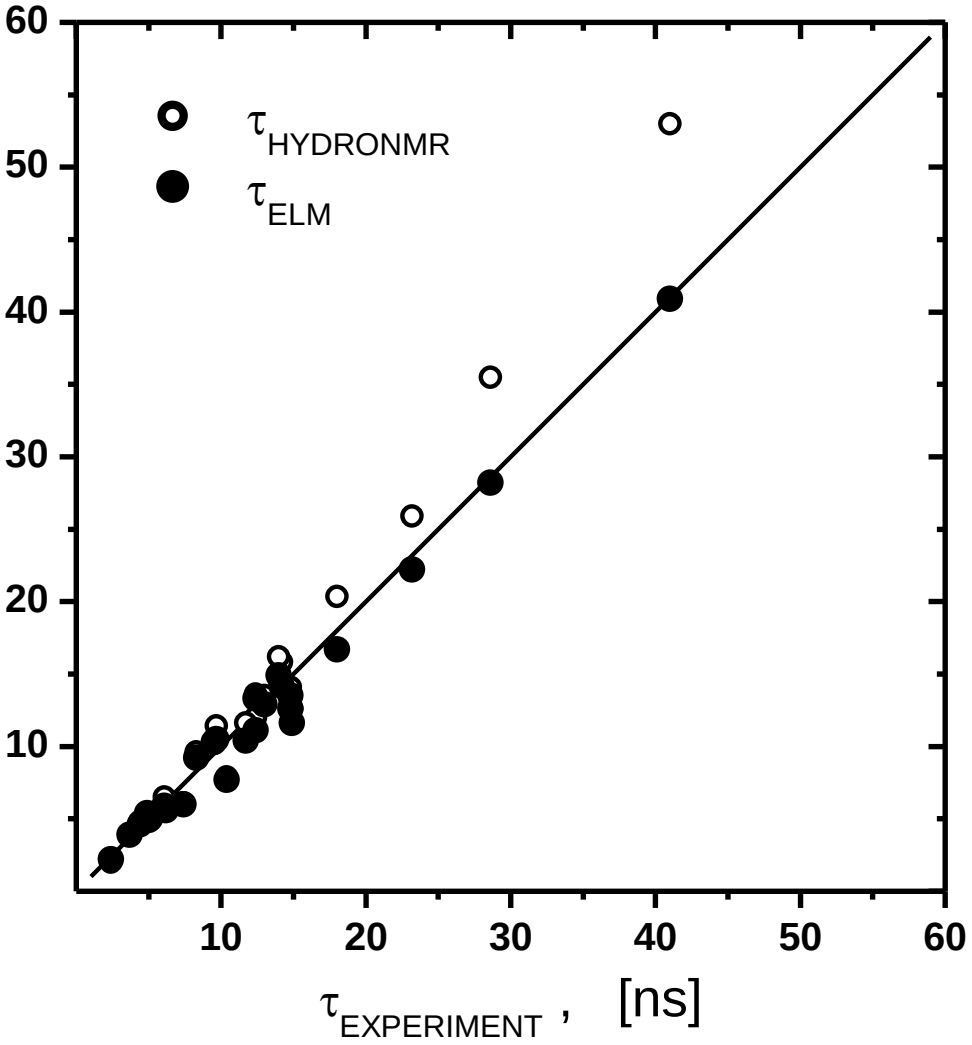
$$< X^m >= \frac{1}{N}\sum_{j=1}^{N} X_j^m$$

$$\mathbf{Cov}\,\mathbf{S}^n = E^n \mathbf{S}^n$$

$$a_n = \sqrt{3E^n}$$

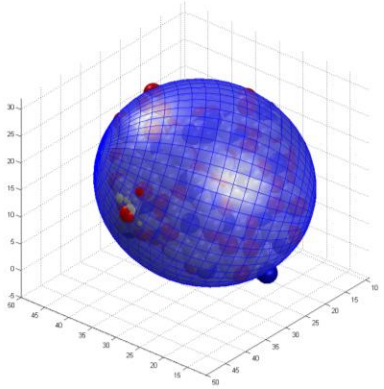


Comparison with the experimental data



A Very General Concept

- During the course of structure elucidation build an equivalent ellipsoid for every snapshot of protein structure conformation



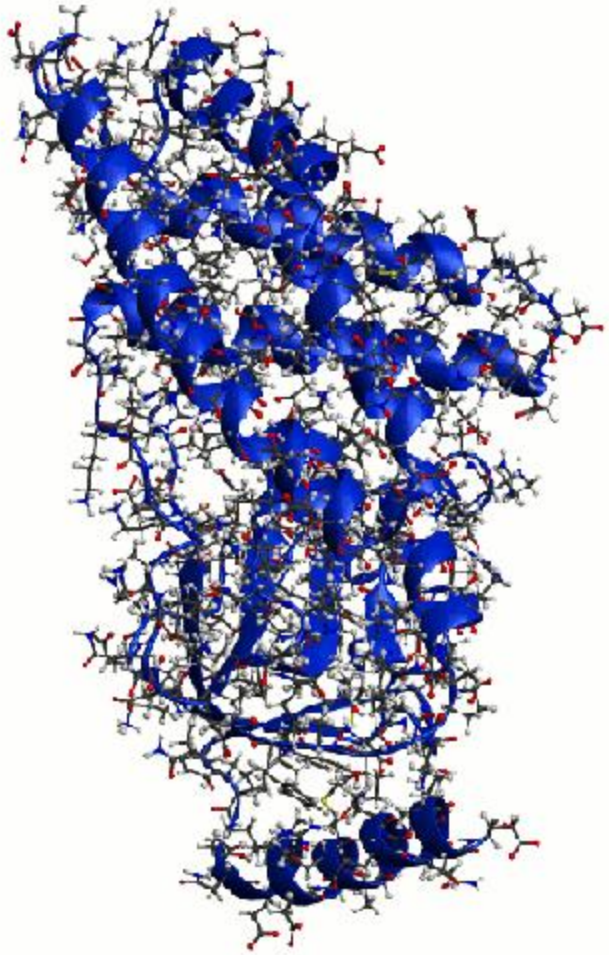
- Then calculate parameters of protein diffusion tensor using the equivalent ellipsoid shape

$$\begin{bmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_z \end{bmatrix}$$

- Compare calculated diffusion tensor parameters with those which were derived from the experimental data and establish a pseudo energy term proportional to the sum of square differences between components of calculated and experimental diffusion tensors

$$\chi^2 = \sum_{\substack{i=1,3 \\ j=i,3}} \left(D_{i,j}^{calc} - D_{i,j}^{exp} \right)^2$$

Refinement of a protein structure with Xplor-NIH using overall shape restraints from diffusion tensor



N terminal domain from Enzyme I (EIN)

Standard Xplor-NIH simulated annealing protocol
started from 3000K down to 25 K
with 12.5 K steps

Experimental restraints:

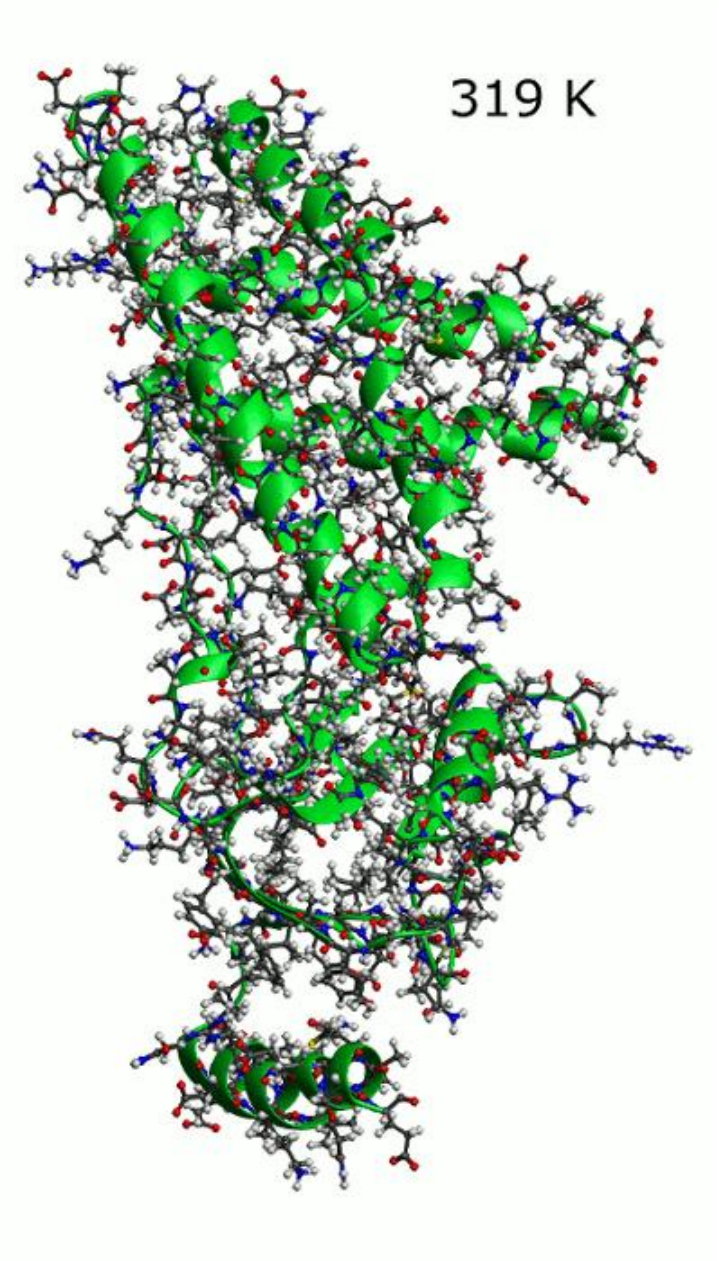
distance restraints derived from NOE

and

Components of Rotation Diffusion Tensor

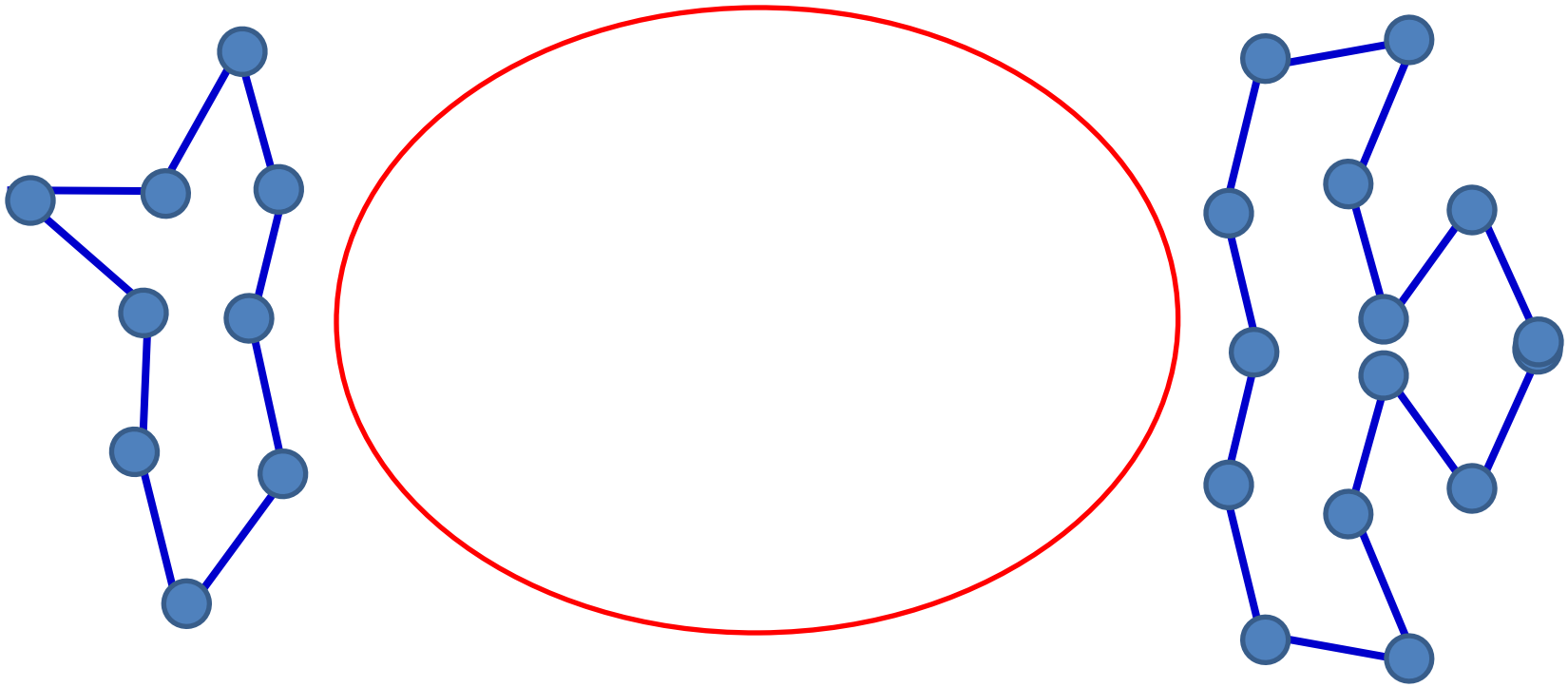
10 lowest energy structures: Blue with diffusion tensor restraints
Green without diffusion tensor restraints

Effect of temperature variations on protein structure

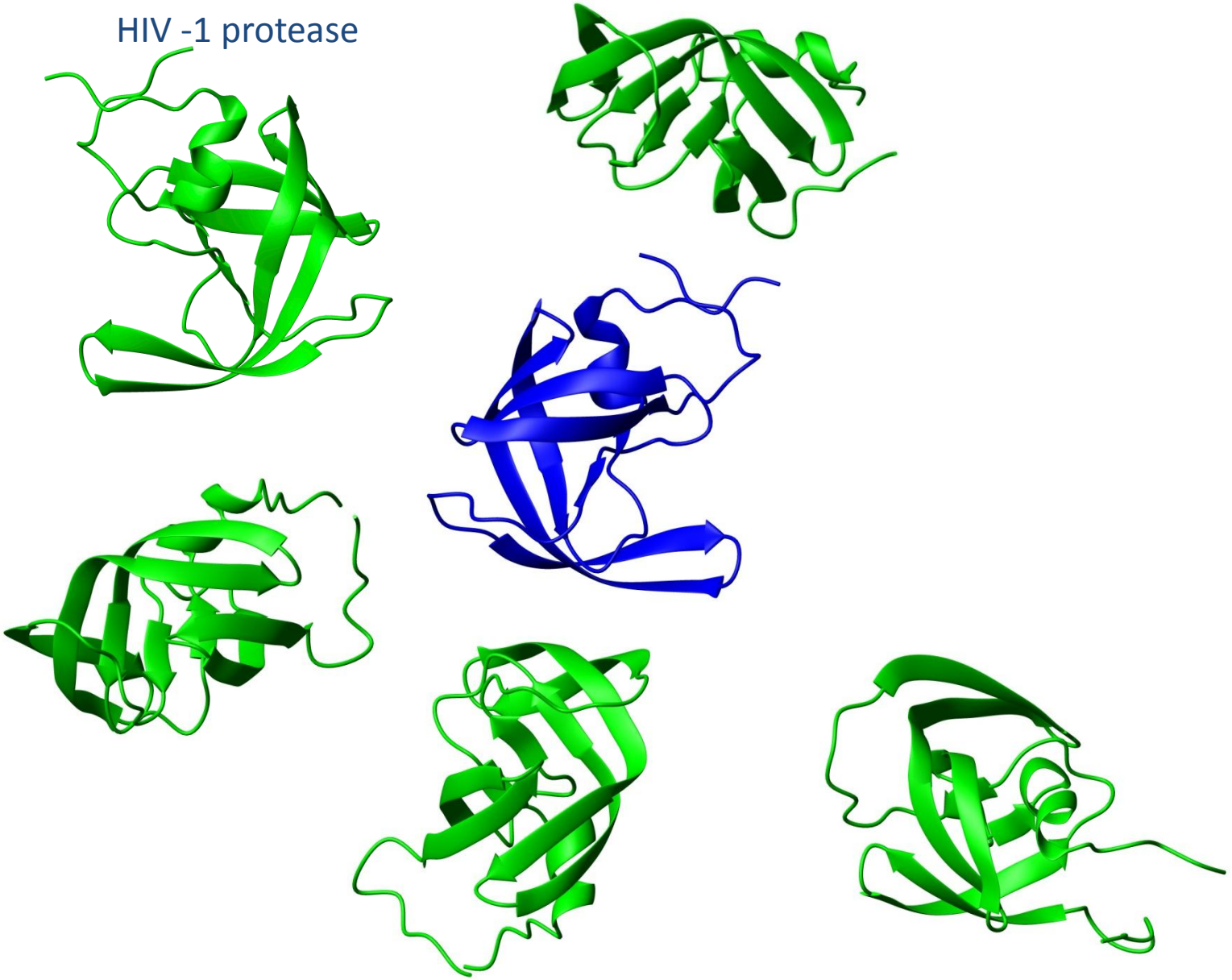


Assembling structures of multi domain proteins using the overall shape restraints provided by the diffusion tensor

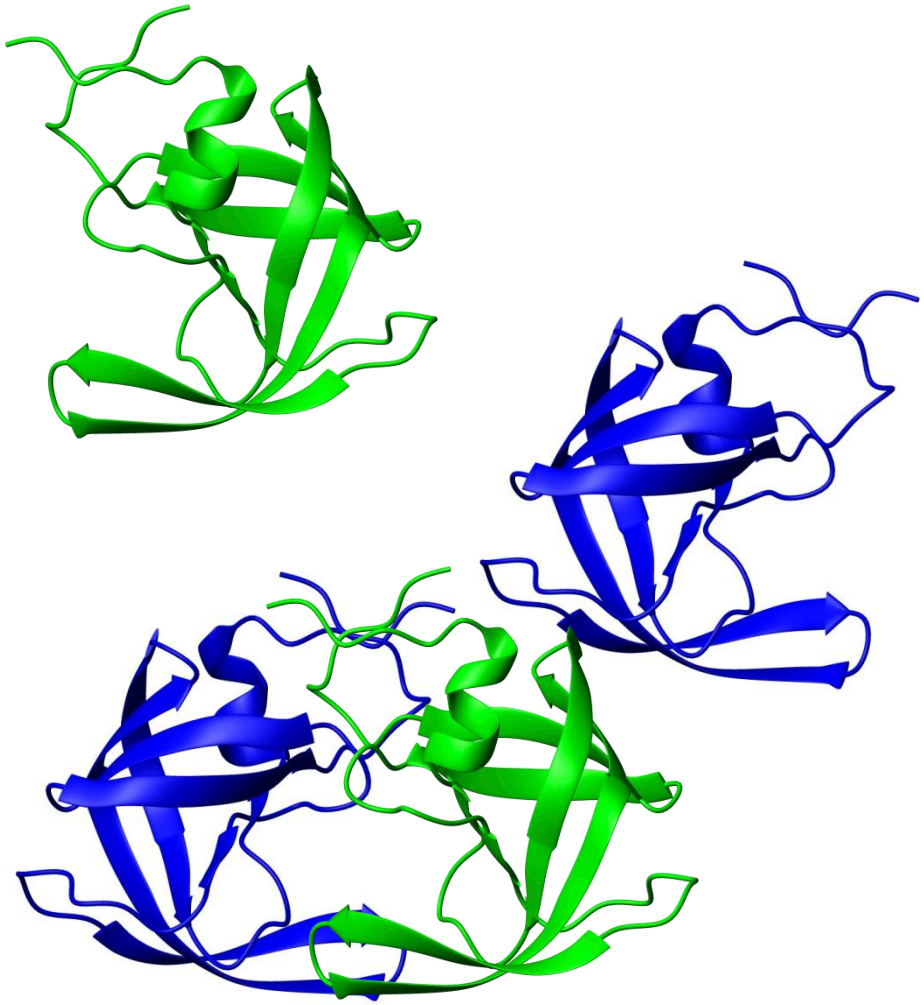
Global restraints on Overall shape



Assembling structure of a symmetric protein homo dimer



Assembling structure of a symmetric protein homo dimer



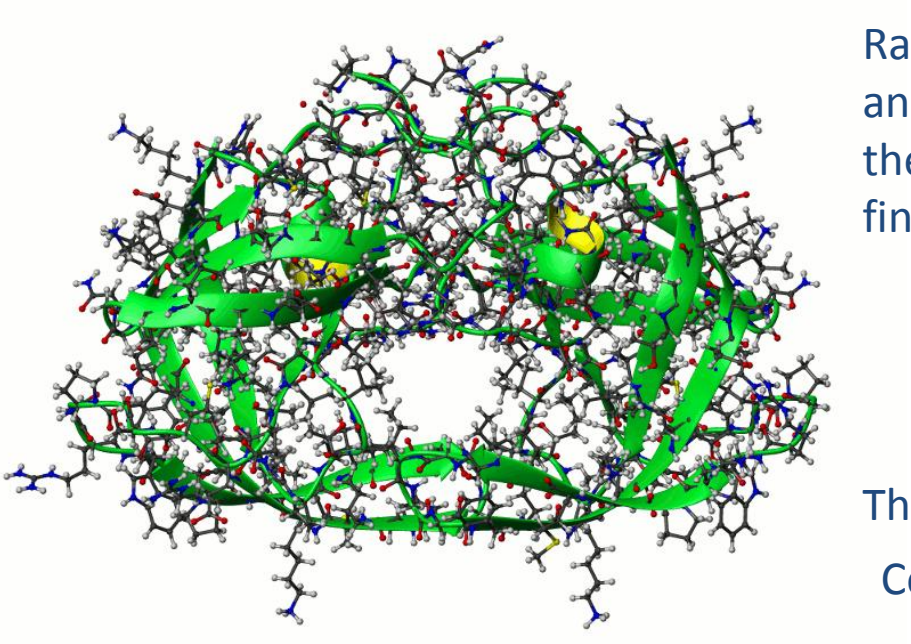
Generic docking protocol

Part I:
Rigid body dynamics
for raw domain positioning.

Part II:
Simulated annealing
with flexible side chains
for final adjustment.

Assembling structure of a symmetric protein homo dimer

HIV -1 protease



10 lowest energy structures

Randomization of domain positions
and Rigid body dynamics repeated 10 times;
then the lowest energy structure submitted to
final simulated annealing part of the protocol

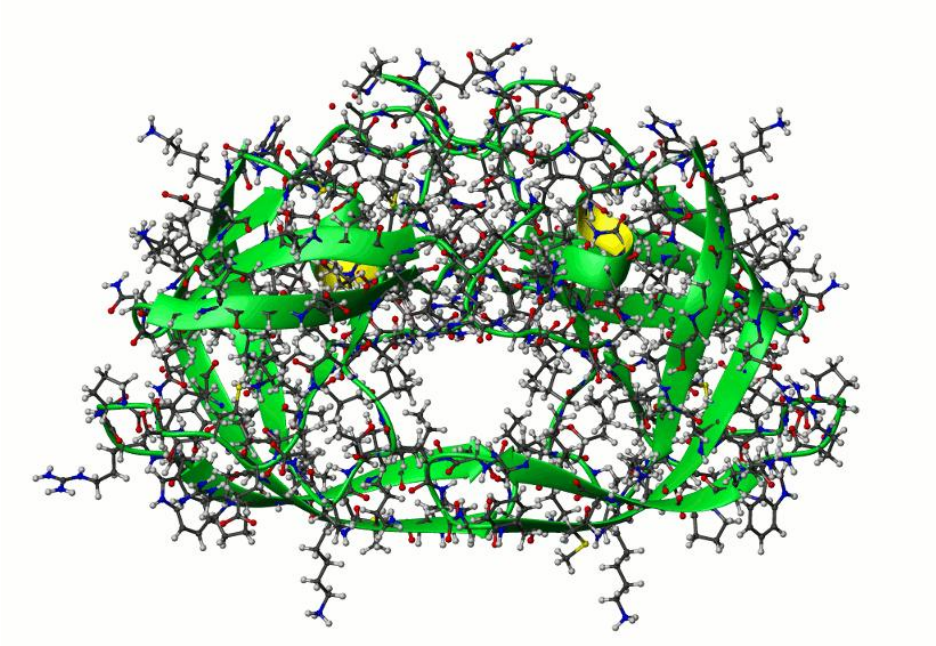
512 structures calculated.

The only experimental restrains are

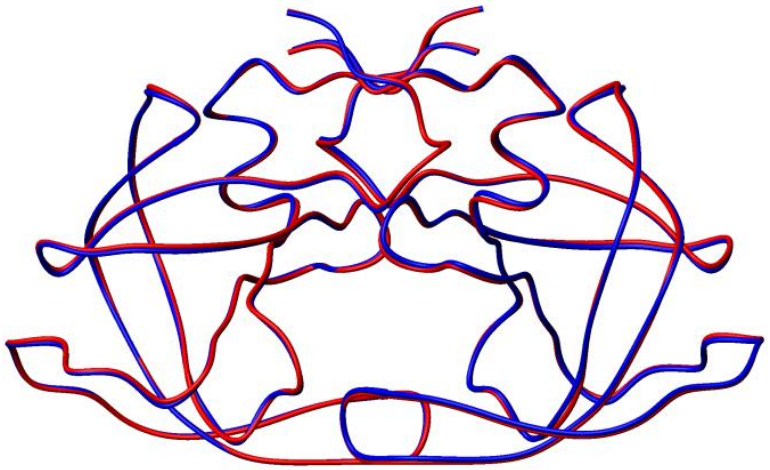
Components of Rotation Diffusion Tensor

Assembling structure of a symmetric protein homo dimer

HIV -1 protease



10 lowest energy structures

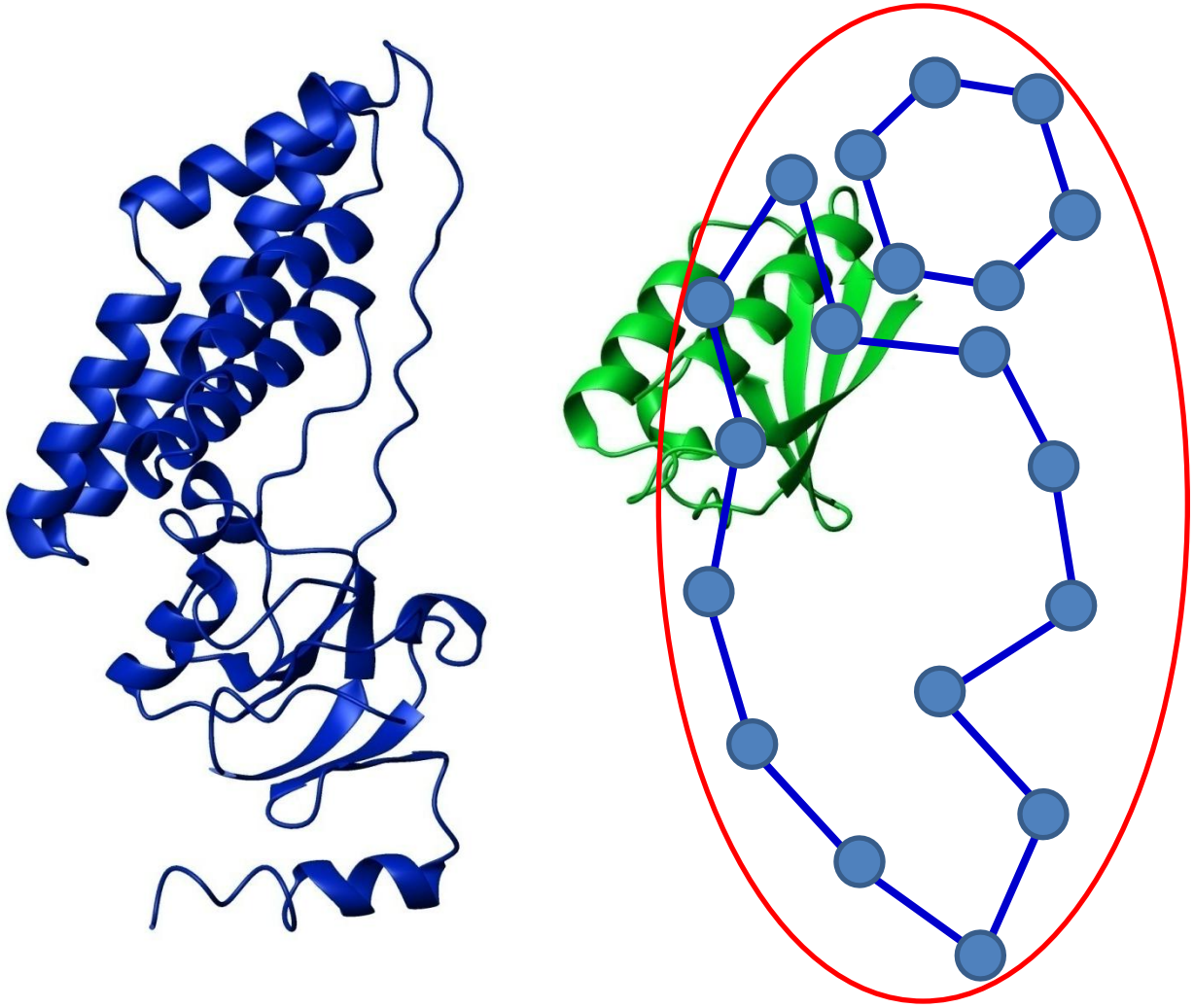


Averaged over 10 lowest energy structures (blue) versus reference (red)

Cα RMSD 0.35 ±0.09 [Å]

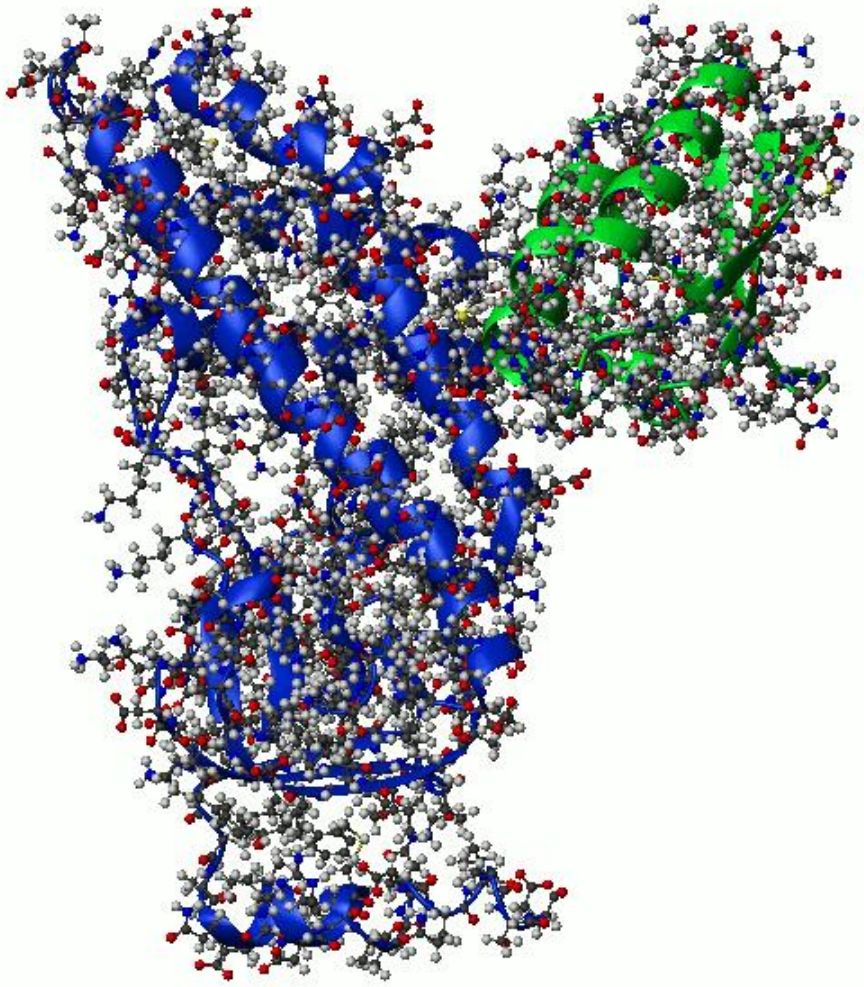
Application to an asymmetric complex

EIN – HPr complex

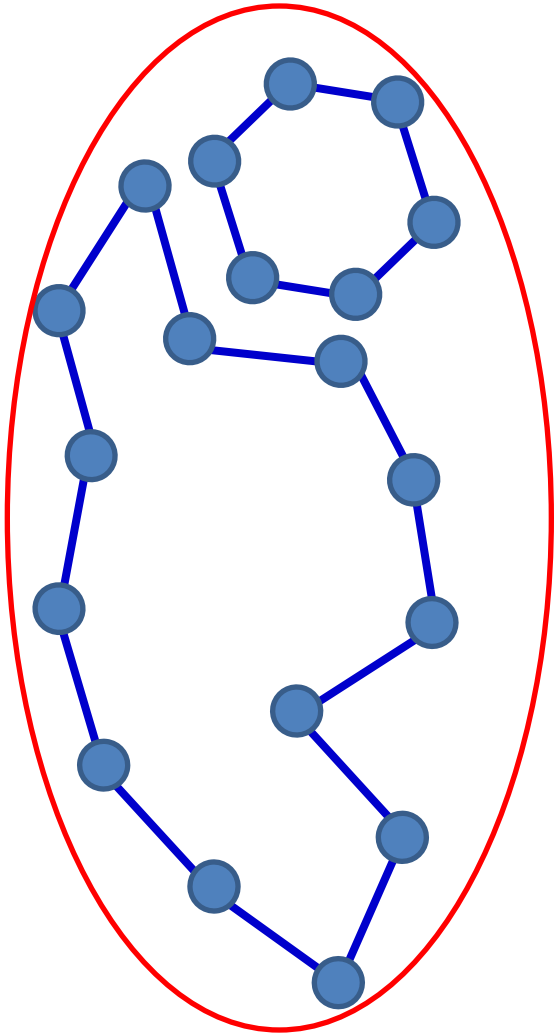


Application to an asymmetric complex

EIN – HPr complex

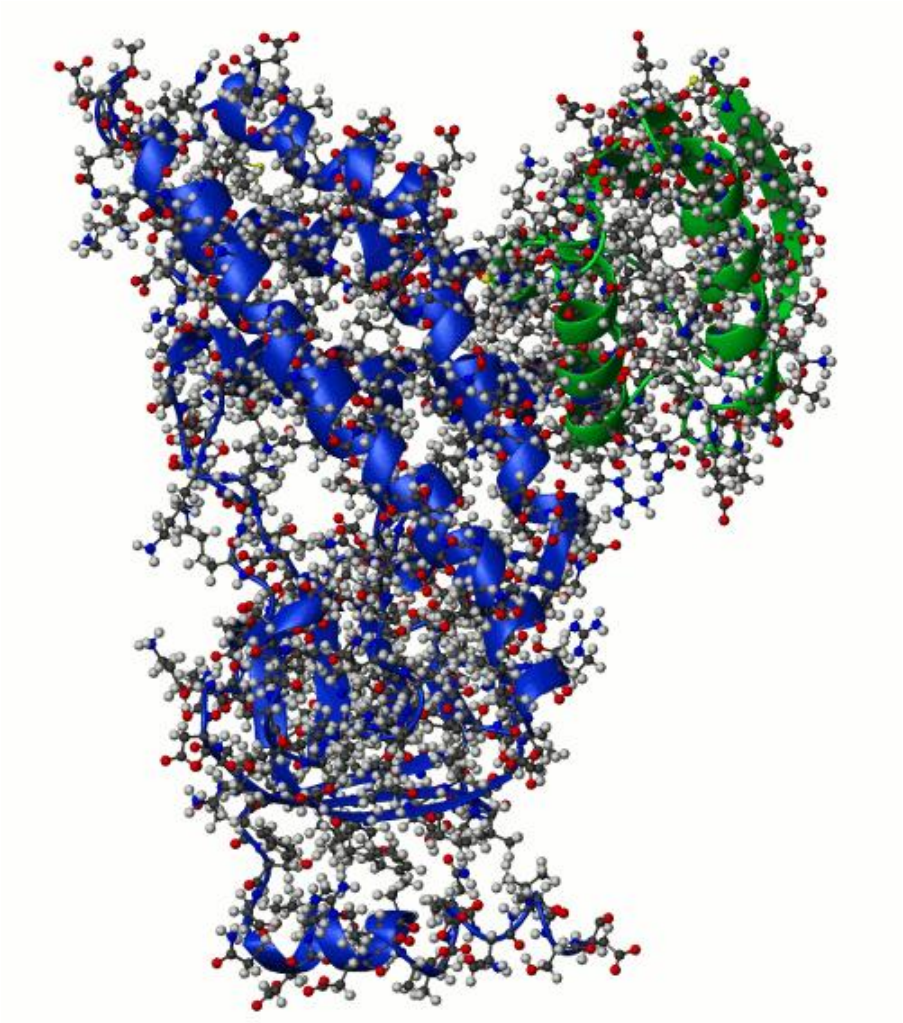


10 lowest energy structures

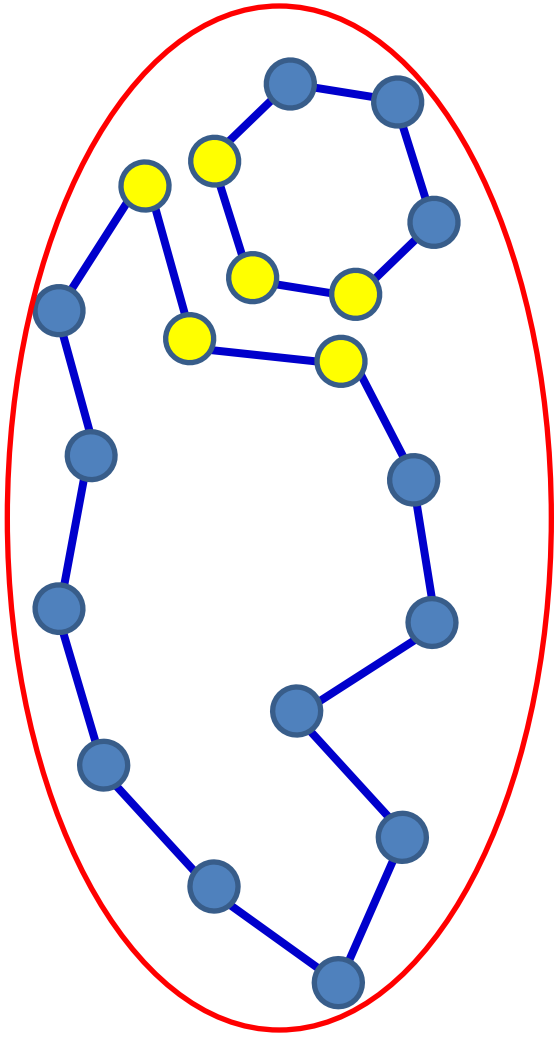


Application to an asymmetric complex

EIN – HPr complex

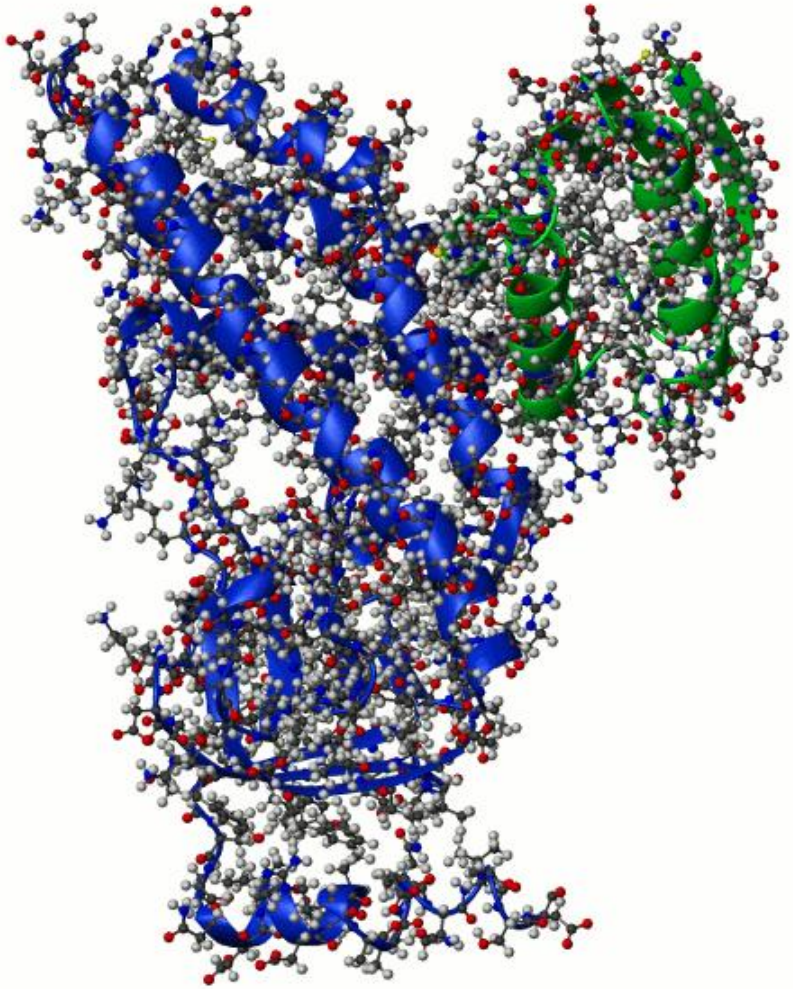


10 lowest energy structures



Application to an asymmetric complex

EIN – HPr complex



10 lowest energy structures

Randomization of domain positions
and Rigid body dynamics repeated 10 times;
then the lowest energy structure submitted to
final simulated annealing part of the protocol

512 structures calculated.

Experimental restrains were

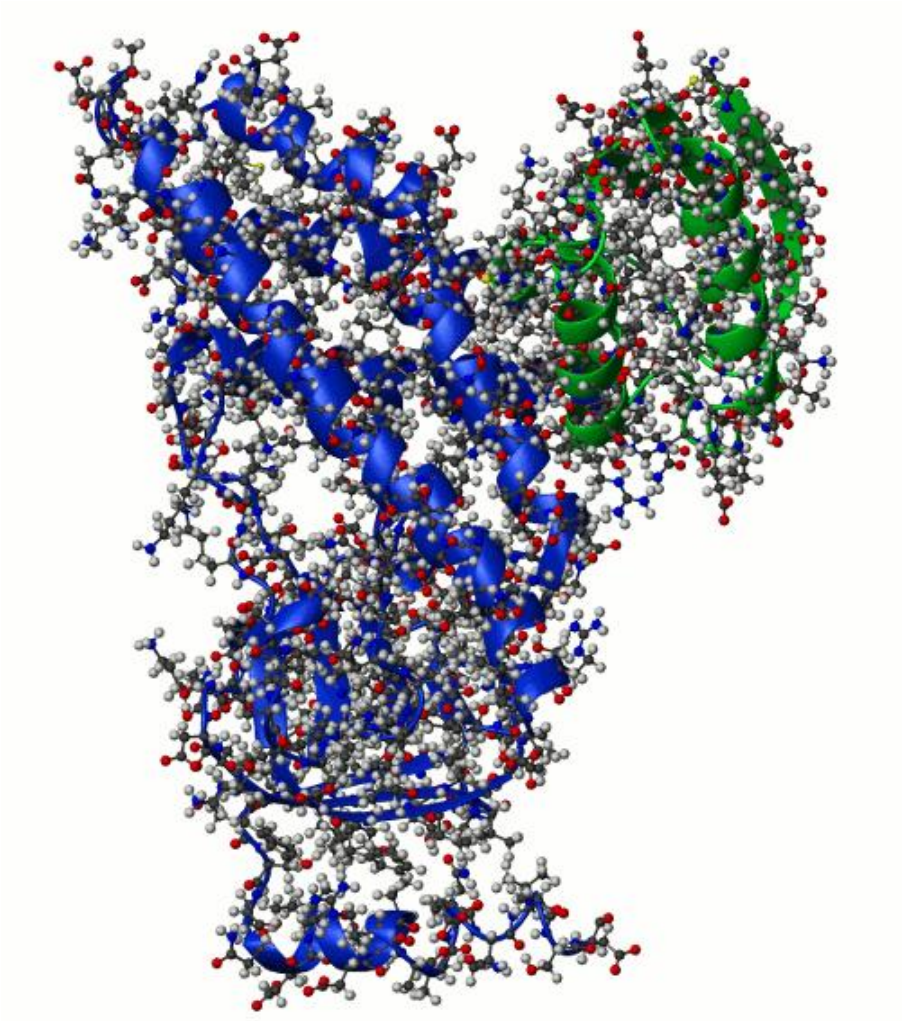
Components of Rotation Diffusion Tensor

and

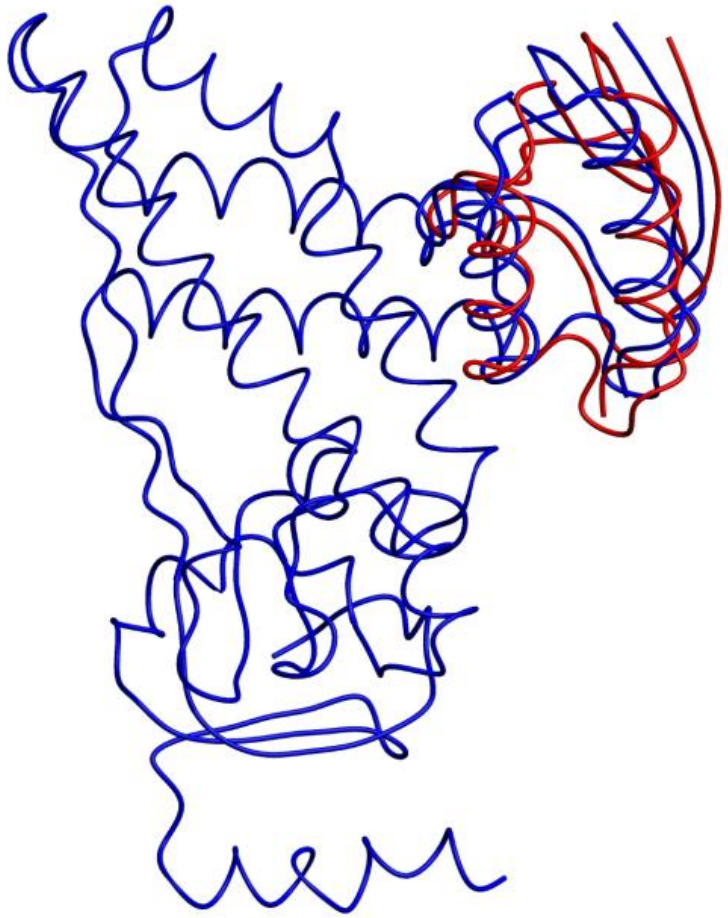
Highly ambiguous distance restraints
from chemical shift perturbation mapping

Application to an asymmetric complex

EIN – HPr complex



10 lowest energy structures



Cα RMSD 1.20 ±0.03 [Å]

XPLOR-NIH structure elucidation platform

Originated from **Axel T Brunger**'s X-PLOR package
to CARMM molecular dynamics engine (1987)

Currently developed by
Charles D. Scweiters , **G. Marius Clore** , **Nico Tjandra**, and **John Kuszewski** @ NIH

Experimental data

Inter molecular force Field

Data base potentials

XPLOR-NIH