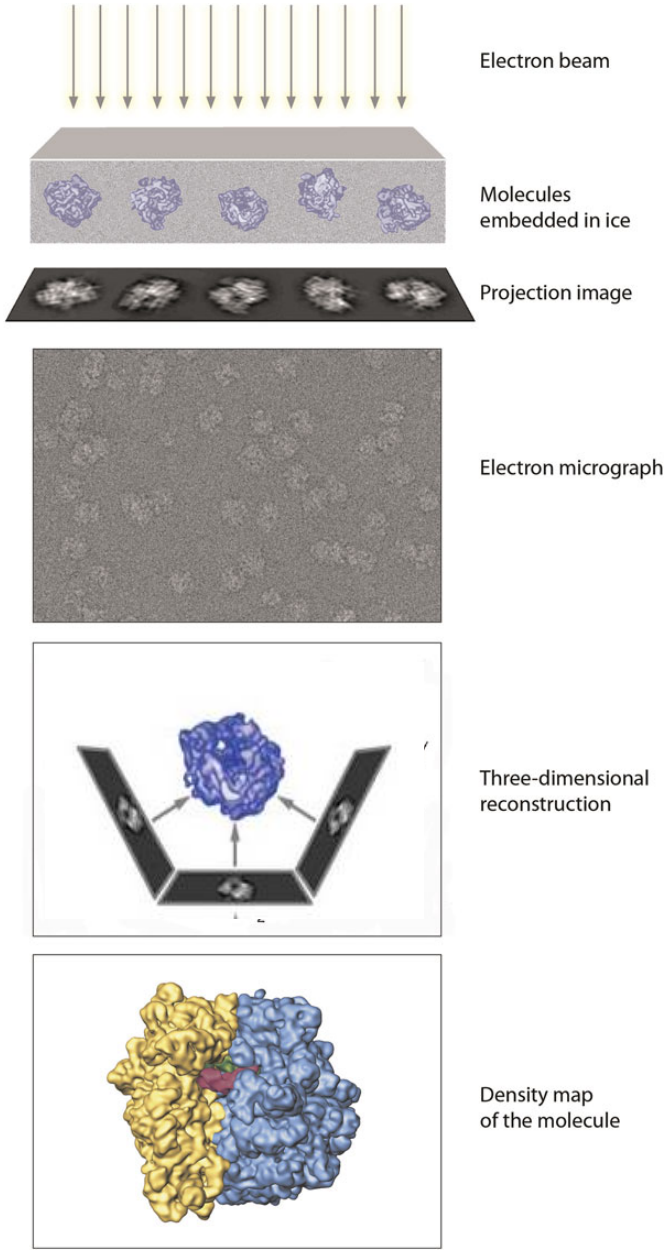
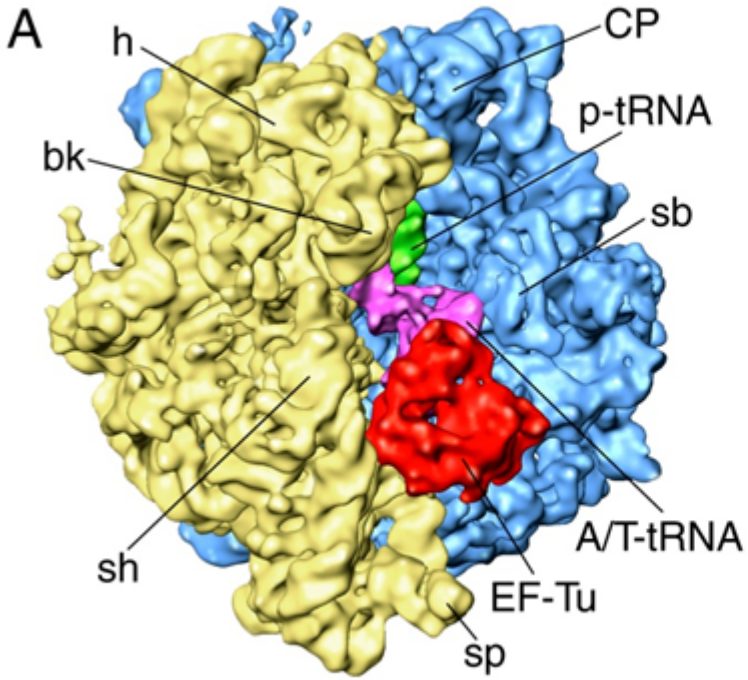
Macromolecular dynamics: Flexible fitting into cryo-EM maps

Jayati Sengupta, CSIR-IICB, Kolkata, India

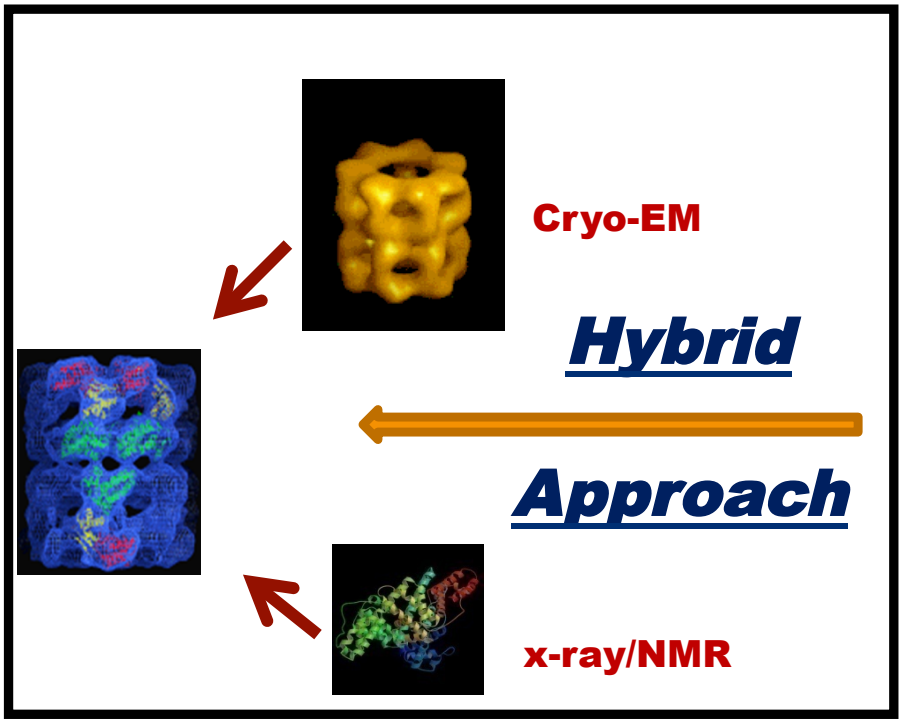


Hum Genomics. 2010

The aim of structural biology is to achieve atomic resolution structure .....



The atomic structures by X-ray crystallography brings out the full potential of cryo-EM, and a hybrid approach where cryo-electron microscopy and X-ray crystallography are combined together has been proved to be useful.

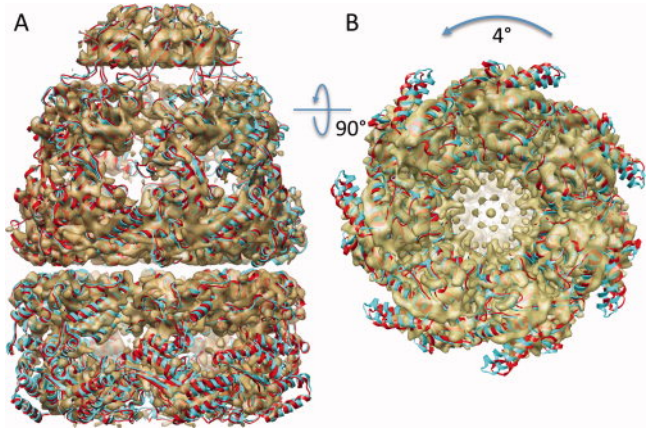


Better Structures Through Synergy

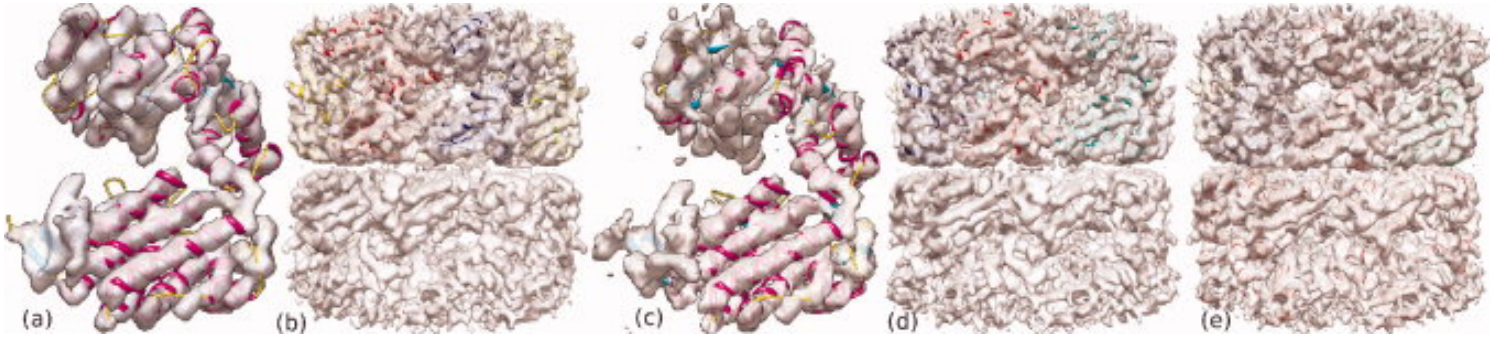
A cryo-EM reconstruction can be interpreted at a level of detail that is greater than the experimental resolution if high-resolution structures of components are available. This is because an atomic model can be placed into a moderate-resolution cryo-EM reconstruction using constrained fitting, with an improved accuracy of placement over the experimental resolution of the cryo-EM map

M.G. Rossmann, **Fitting atomic models into electron-microscopy maps**
Acta Crystallog. (2000)

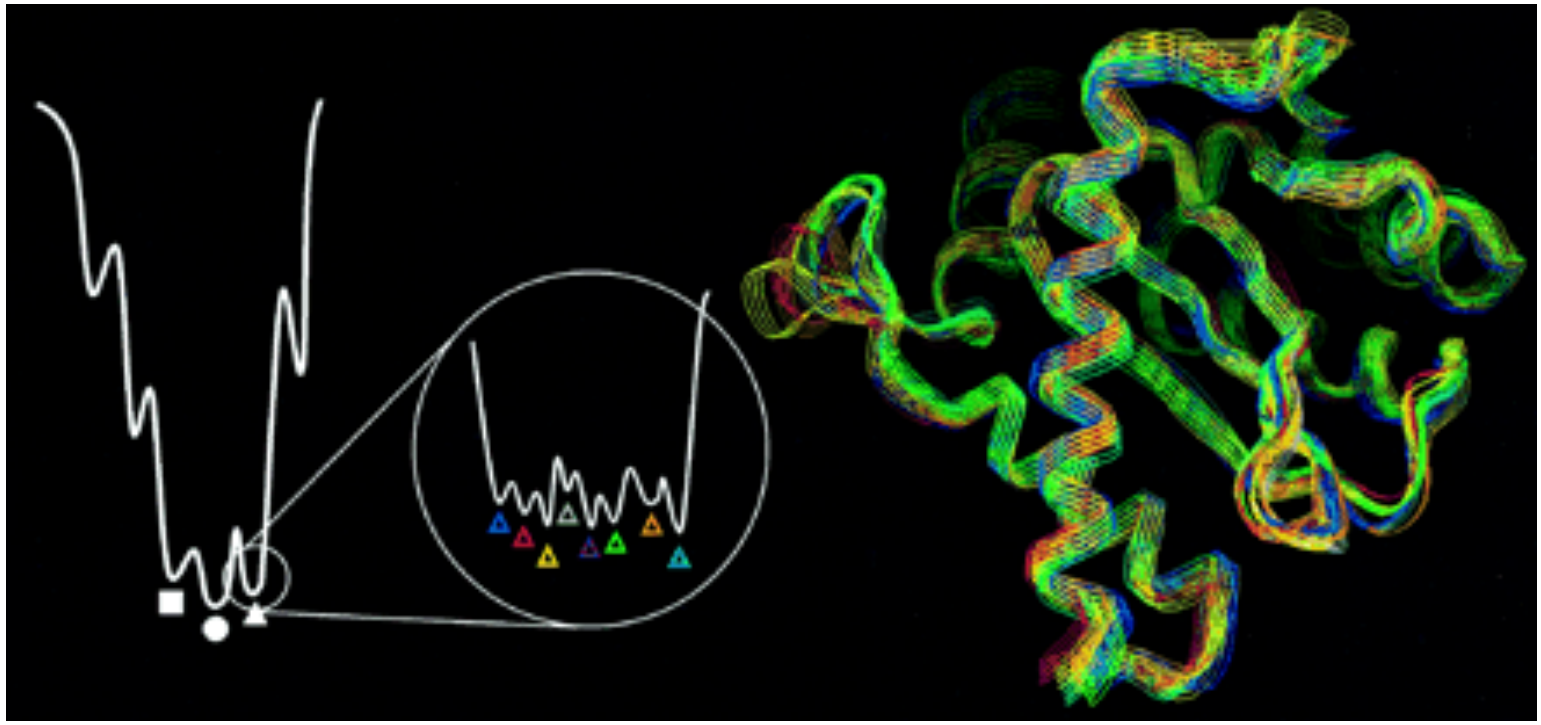
Mark Points of Resolution in Structural Information of Proteins (Ref. Fujiyoshi Y, 1998)



Resolution (Å)	Structural features observable on a good map
20.0	Overall shape of the molecule
7.0	Helices as rods of strong intensity
3.5	The main chain with some ambiguities
3.0	The side chains partly resolved
2.5	Side chain well resolved; Atoms located to about ± 0.04 Å
1.5	Atoms located to about ± 0.01 Å
1.0	Anisotropic structure of atomic potential



The density map generated by the cryo-EM represents the Coulomb potential distribution of the object and can thus be readily and quantitatively compared with the electron density maps obtained by X-ray crystallography, providing the basis for molecular docking.

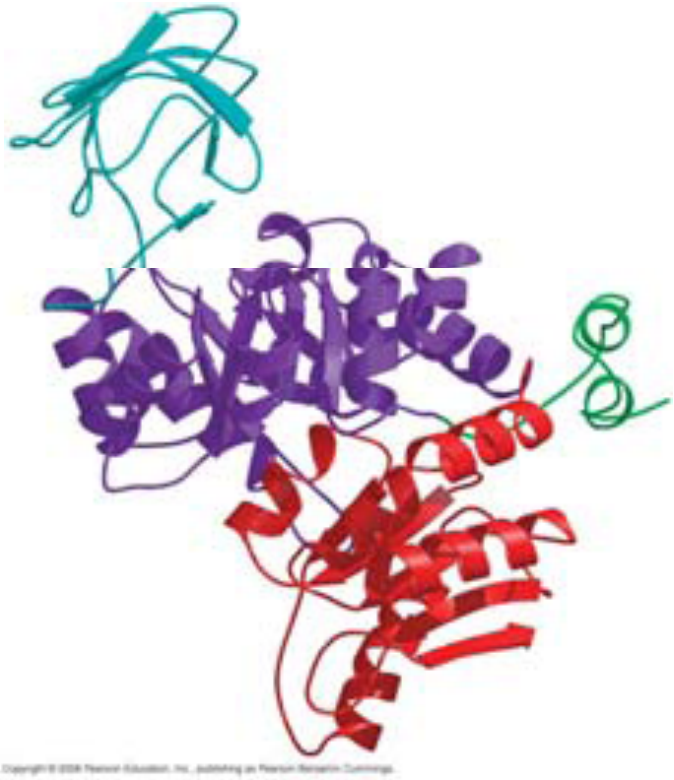


This conformational funnel demonstrates multiple states for a protein (represented by their weighted-average structures: ▪, ●, and ▲). The flexibility inherent in a folded state (▲) is described by a subensemble of conformations (shown here as a collection of colored △).

Mol Pharmacol. 2000

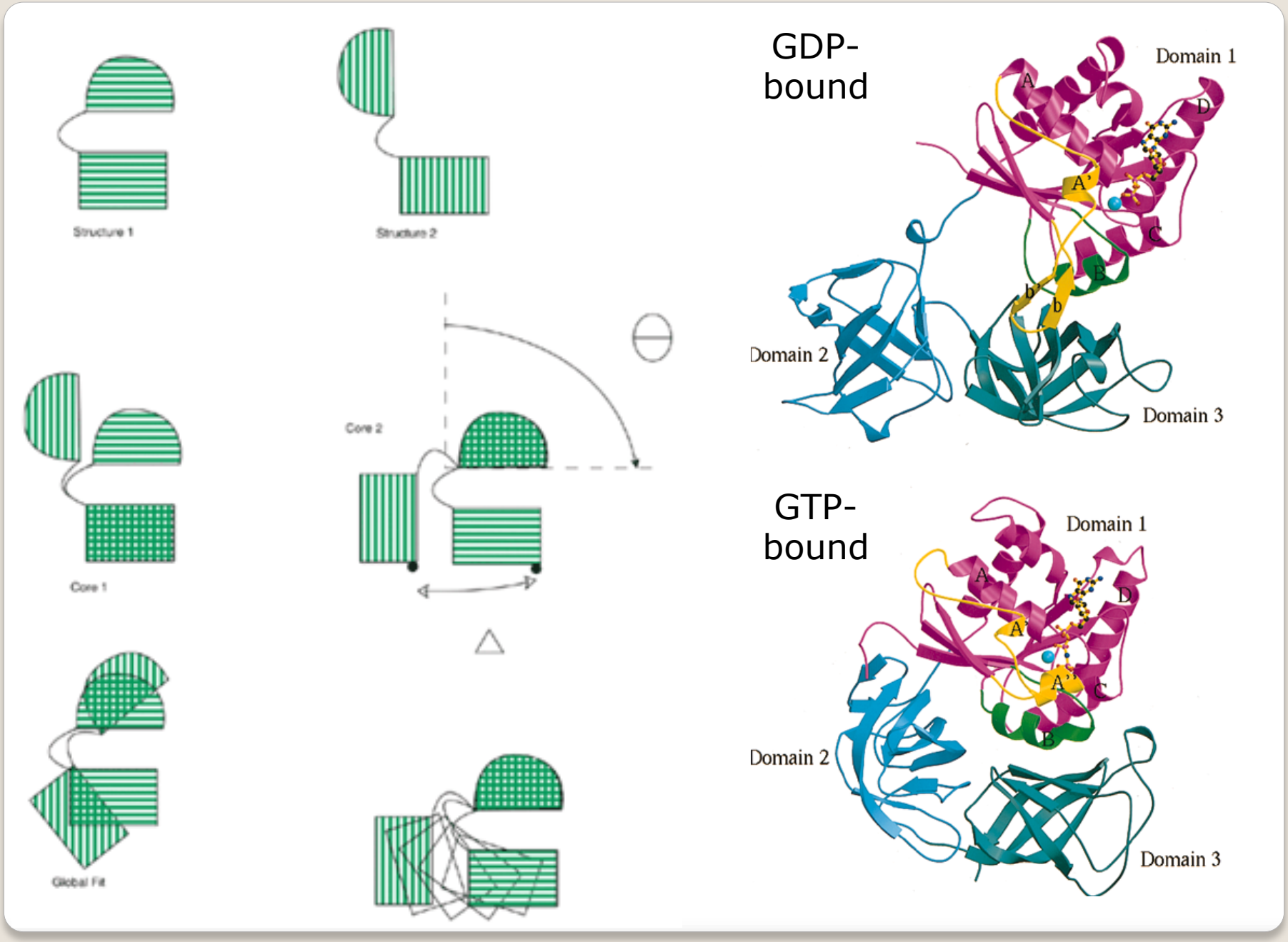
Modular Structure of Proteins

♣ Most proteins consist of multiple independent "domains".
♣ Protein Domain = a part of the protein structure that appears independent of the rest of the structure and that often maintains this structure entirely on its own (e.g. when only the specific domain sequences are expressed in cells, via recombinant DNA)



Different domains in Pyruvate kinase

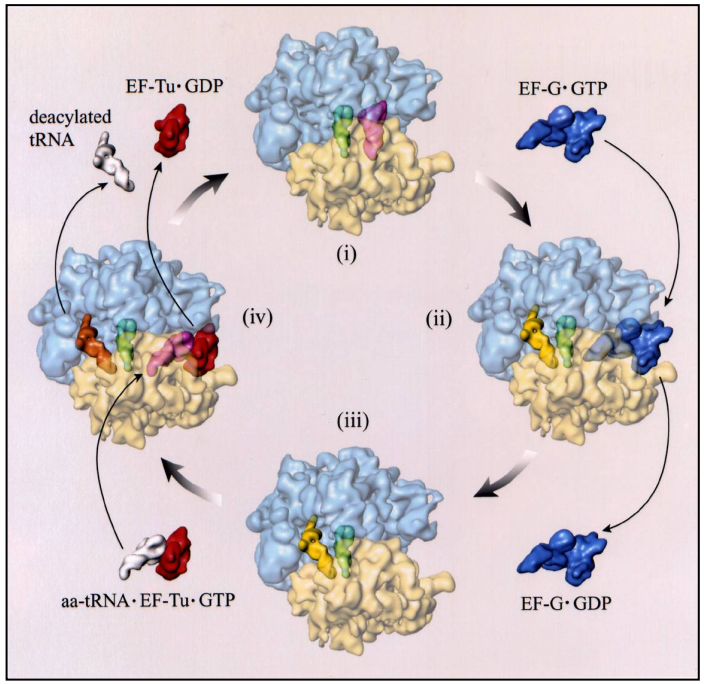
♣ Protein domains often encode specific functions ("functional domain")
♣ Examples:
– Leucine zipper (DNA binding/transcription factor)
– RING-HC (protein-binding/E3 ubiquitin ligase)
– Kinase domain (phosphorylates proteins)
– cAMP binding domain in CAP protein
– DNA binding domain in CAP protein


Structure 1
Structure 2
Core 1
Core 2
Global Fit
GDP-bound
Domain 1
Domain 2
Domain 3
A
D
A'
C
B
b'
b
GTP-bound
Domain 1
Domain 2
Domain 3
A
D
A'
A''
C
B

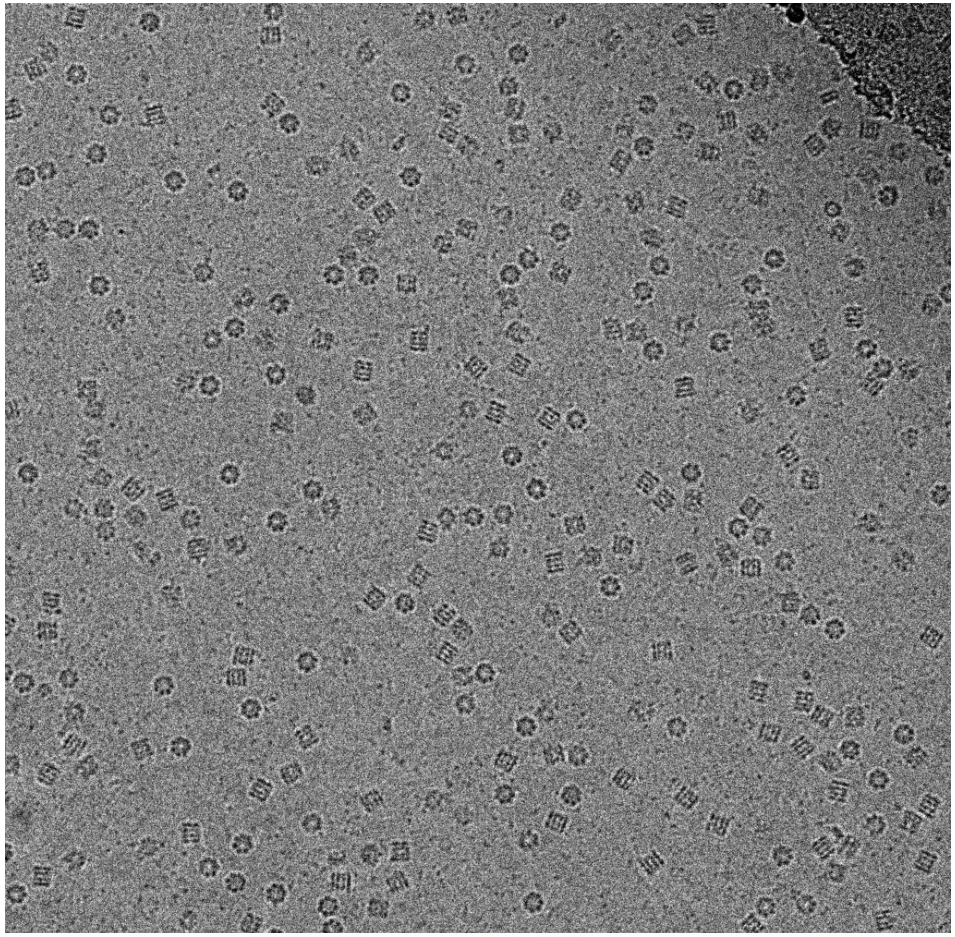
Studying the Dynamics of a Molecular Machine

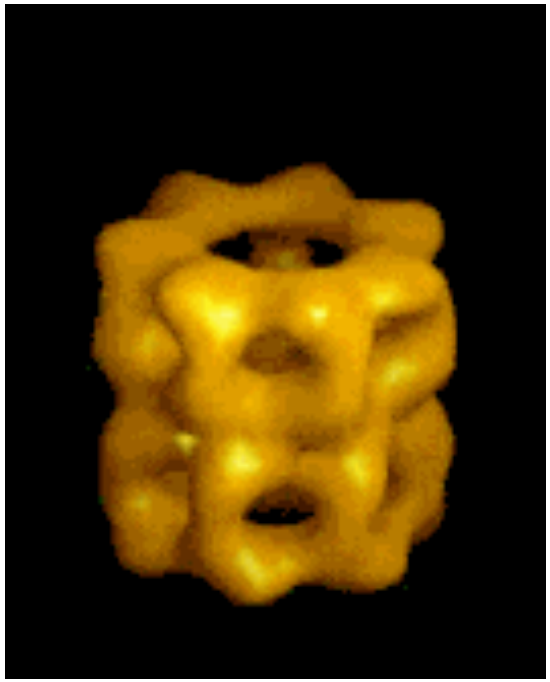
How do we visualize a machine in motion?

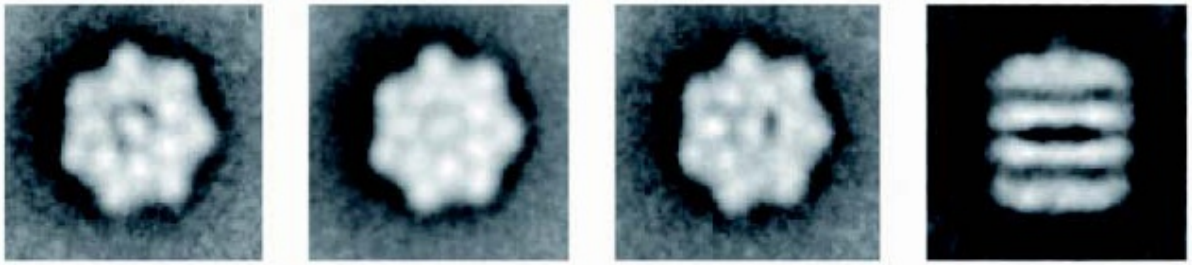
--by stopping it at precisely defined times to take "snapshots"



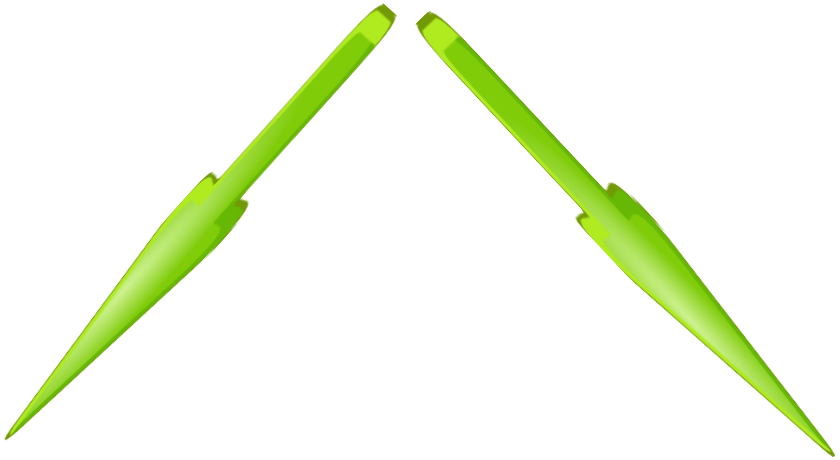
Fitting/docking of X-ray structures into the cryo-EM map provides quasi-atomic models ("hybrid methods of structure research")







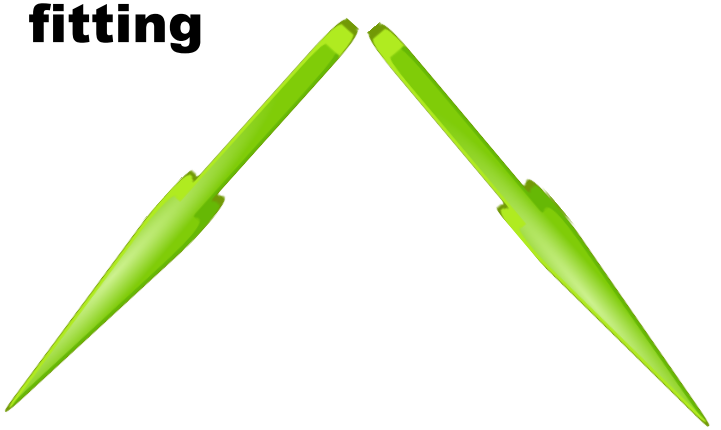
Fitting atomic models into cryo-EM maps



Interactive manual fitting

- 'O' : (http://xray.bmc.uu.se/~alwyn/)
- Quanta: Accelrys, Inc., San Diego, USA
- PyMol (http://pymol.sourceforge.net/)
- Coot (http://www.biop.ox.ac.uk/coot/)

Computerized, automated fitting



Rigid body

Flexible fitting

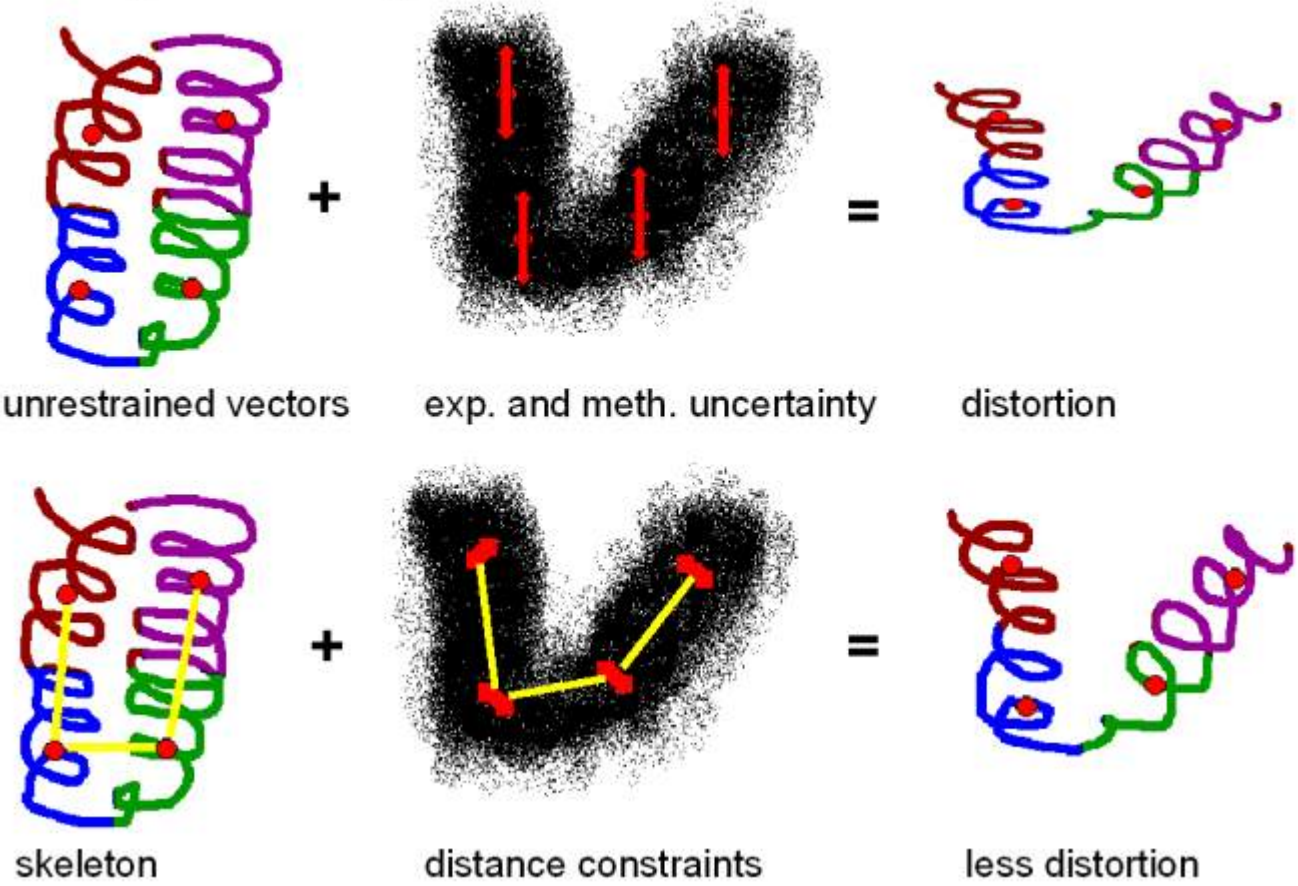
Flexible Docking

- Large macromolecular assemblies often undergo large functional rearrangements
- In some cases, the X-ray structure does not correspond to the EM structure
- This complicates the fitting, as not only orientation, but conformational rearrangements must be considered

Flexible fitting approach

* SITUS (Flexible Fitting with Molecular Dynamics)
* RsRef (real-space refinement) technique
* Normal mode-based flexible fitting (NMFF)
* The molecular dynamics flexible fitting (MDFF) method

freezing inessential degrees of freedom:



The "skeleton" based approach is related to 3D motion capture technology used in the entertainment industry and in biomechanics.

RsRef (real-space refinement technique)

-- to achieve multiple-domain fitting in a quantitative way where following a least-squares optimization, the positions of the structural elements are refined, and, at the same time, stereo-chemical conflicts are minimized.

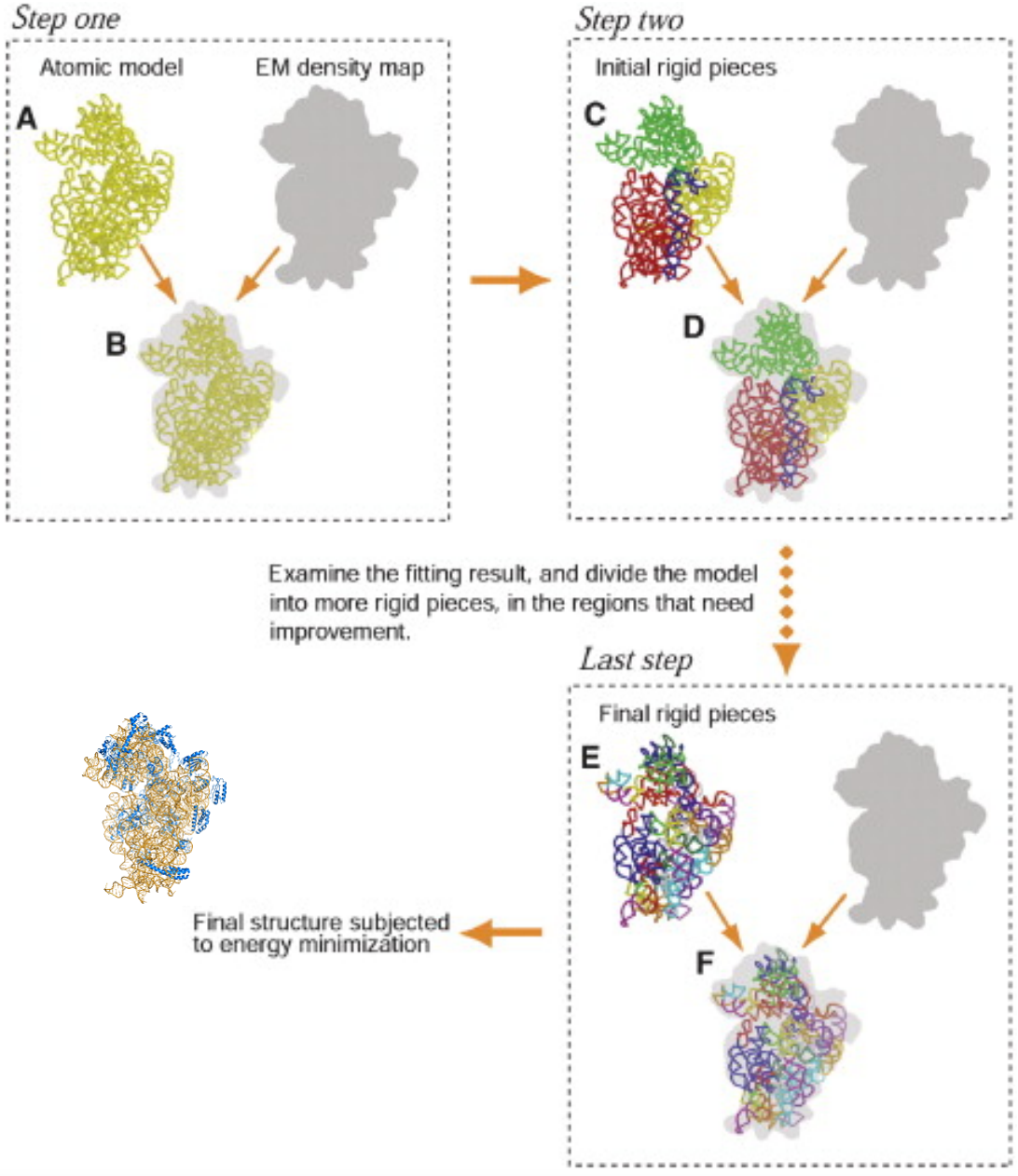
http://www.sb.fsu.edu/~rsref/Distribution/software_distribution.htm

RSRef2000

Runs as a module for CNS, supporting simulated annealing or least-squares optimization with Cartesian or torsion angle parameterization of the model. Available for UNIX and Linux.

RSRef2.0

Runs as a module for TNT, supporting least-squares refinement with Cartesian model parameterization. Available for UNIX.

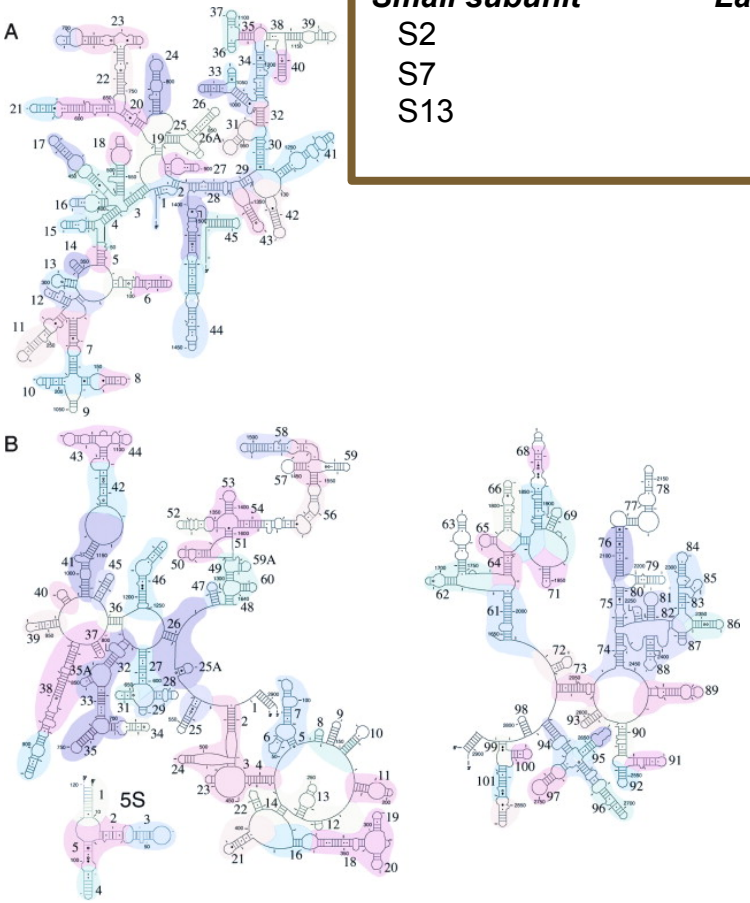


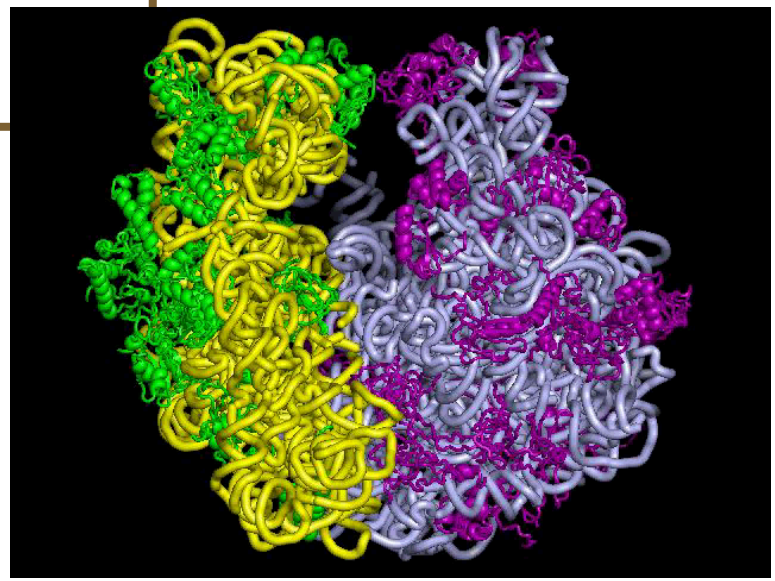
The RSRef package compares/refines the agreement between an atomic model and an electron density map.

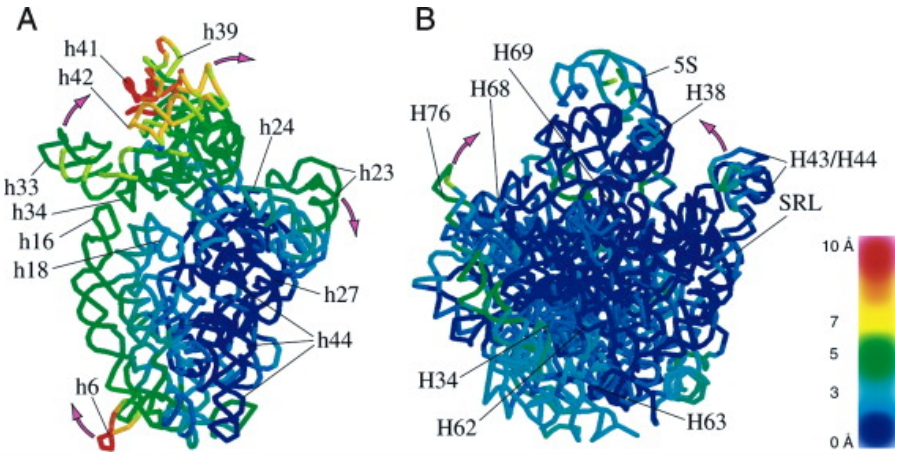
Embedded within CNS, stereochemical restraints are available, and a number of optimizers are supported, including Cartesian or torsion angle dynamics.

The following proteins are cut into more than one rigid body during the fitting:

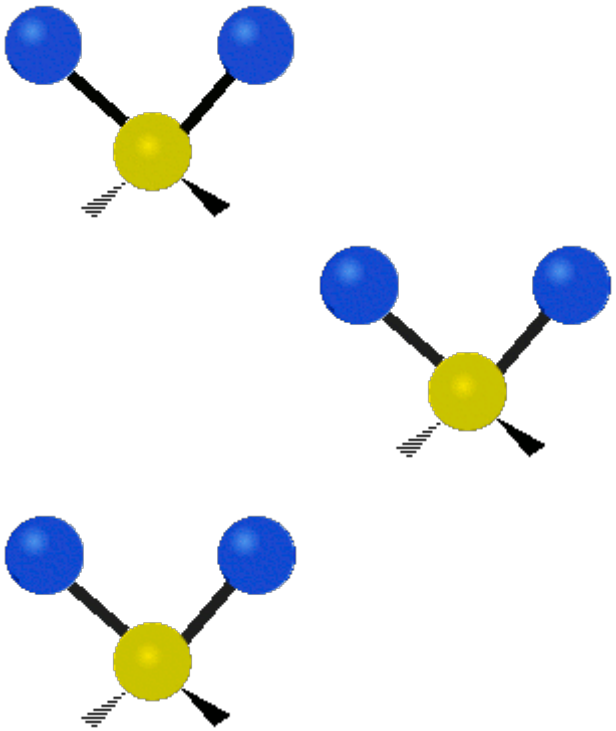
Small subunit	*Large subunit*
S2	L2; L3
S7	L5: L6
S13	L9: L11
	L18; L24







Gao H. *et al* (2003) *Cell*



Normal Mode Analysis

MM Approximate the complex energy landscape by harmonic potentials

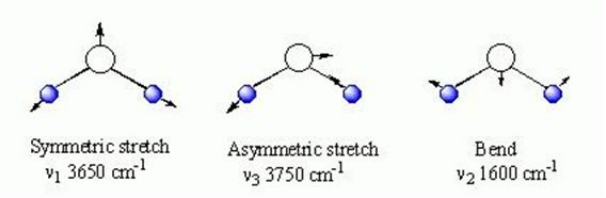
QM Force Constant Matrix: Hessian $H_{ij} = \frac{\partial^2 V(\mathbf{r})}{\partial r_i \partial r_j}$ at the energy minimum

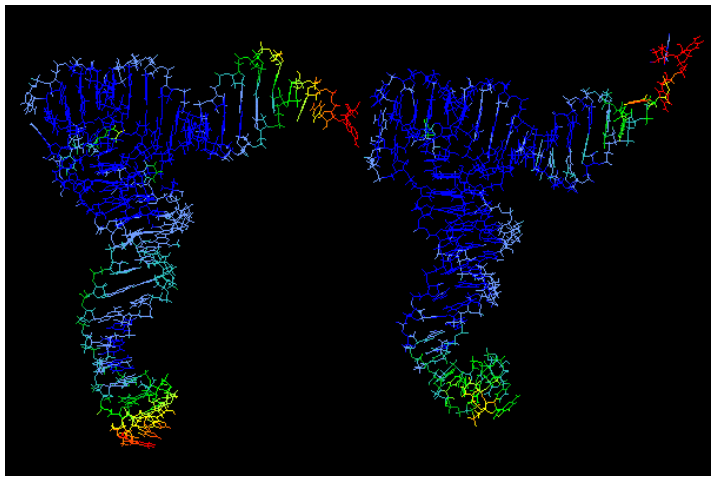
$\Rightarrow$ Normal Modes → vibrational frequencies → energy

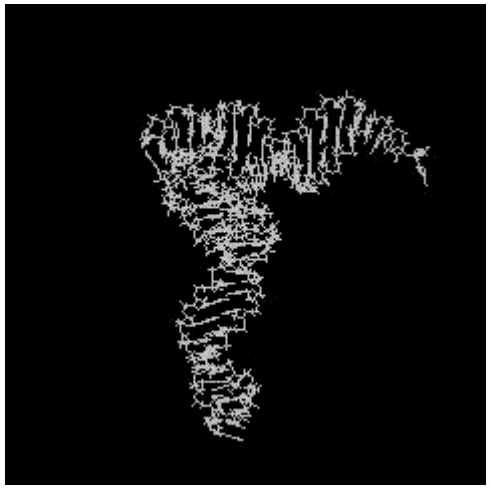
→ eigenvectors → internal motions

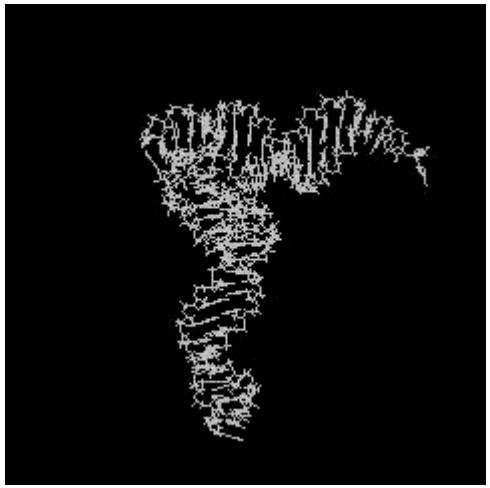
Water

Normal Modes







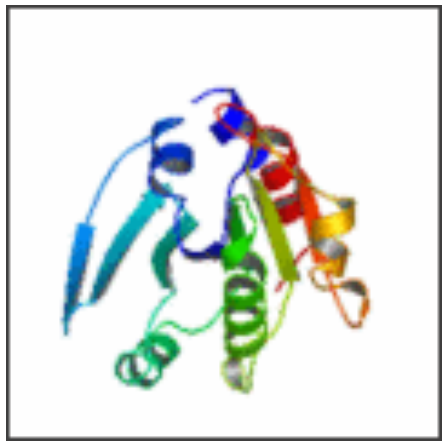


Flexible Fitting: Normal Mode Analysis

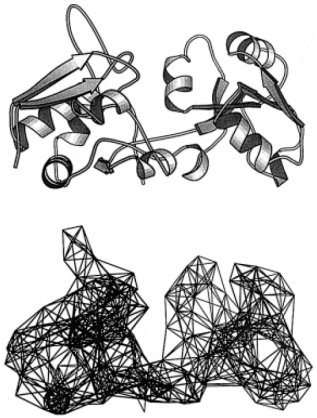
The theory and methods are based on searching along a few lowest frequency normal mode vectors, constructed from a multi-resolution elastic network representation of the atomic structure of interest, to maximize the correlation between the computed electron density for the flexible model and the experimental density.

To increase the correlation coefficient, the fitting is performed by deforming the structure along a set of low-frequency normal modes.

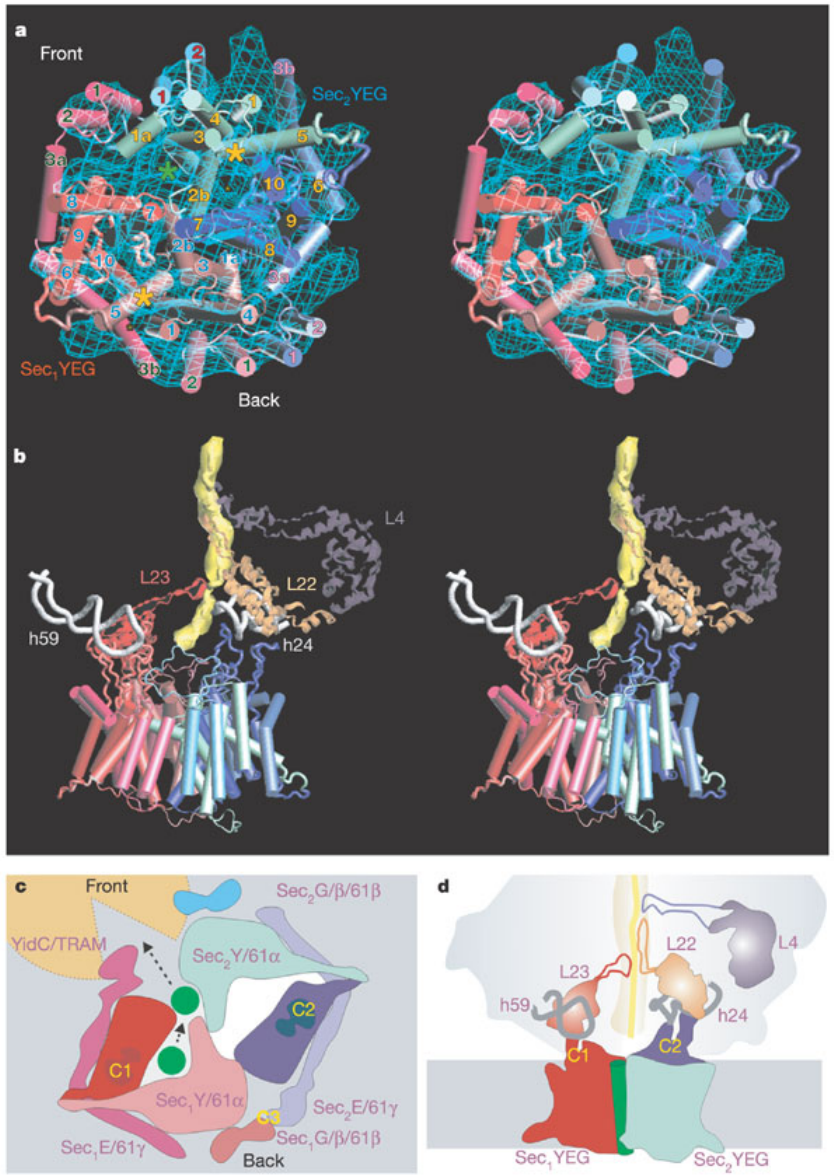
NMA uses a simplified elastic network representation where the mechanical system is modelled as a network of mass points connected with springs that represent inter-residue interactions.

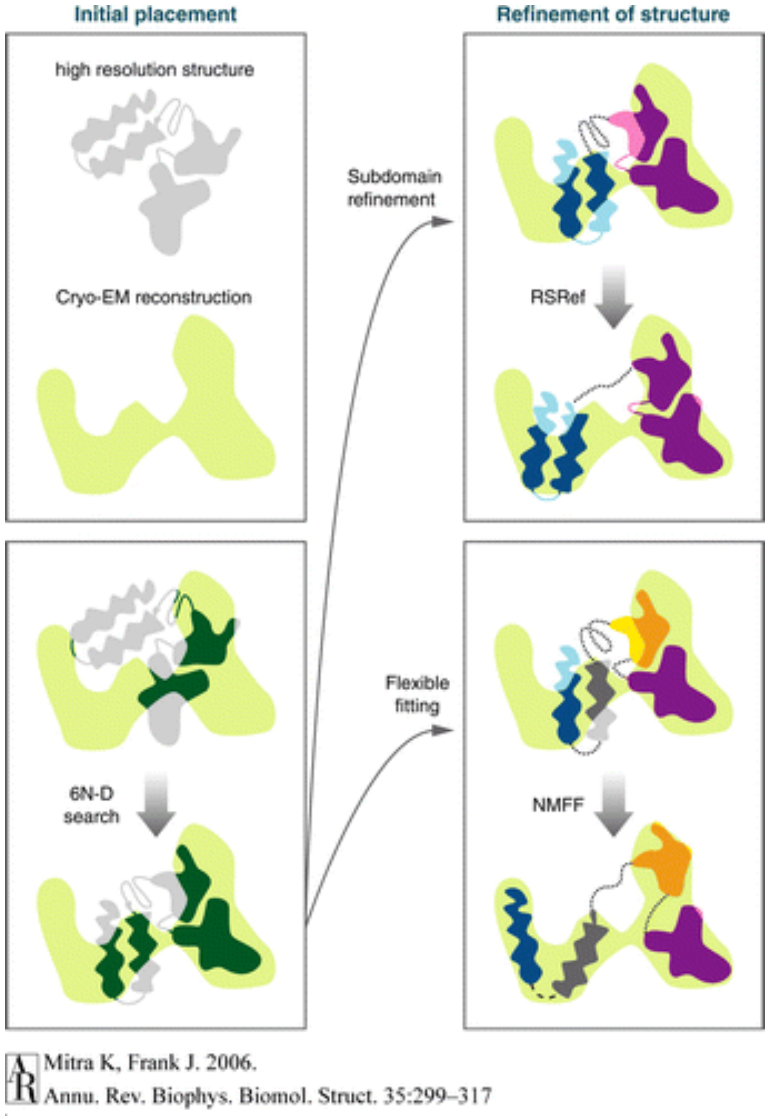


The Elastic Network Model



(figure from Tama and Sanejouand, 2001).





Mitra K, Frank J. 2006.
Annu. Rev. Biophys. Biomol. Struct. 35:299–317

NORMA:

http://www.sciences.univ-nantes.fr/elnemo/NORMA/

NORMA is a freely available software suite that allows to model large conformational changes of 3-D protein structures under the constraint of a low resolution electron density map. Typical applications are the interpretation of electron microscopy data using atomic scale resolution structural models.

NMFF:

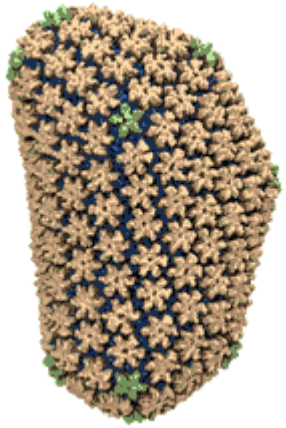
https://mmtsb.org/software/nmff.html

Normal Mode Flexible Fitting **(**NMFF) is an evolving package of programs and methods that enable the flexible multi-resolution fitting of large atomically detailed structures into electron density maps from cryoEM, tomography and related lower resolution methods.

The molecular dynamics flexible fitting (MDFF) method

--incorporate the EM data as an external potential added to the molecular dynamics force field, allowing all internal features present in the EM map to be used in the fitting process, while the model remains fully flexible and stereochemically correct.

http://www.ks.uiuc.edu/Research/mdff/software.html



In the MDFF method, an external potential derived from the EM map is introduced into an MD simulation to steer the atoms into high-density regions. The stereochemical quality of the structure is preserved by the MD force field, and also through harmonic restraints applied to enforce the integrity of secondary structure. The method, therefore, adds two extra terms to the potential energy function of an MD simulation

Utotal = UMD+UEM+USS

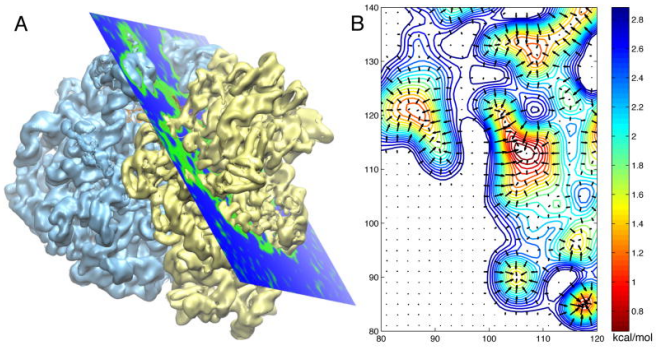
where U_{MD} is the conventional MD potential energy function, U_{EM} corresponds to a potential derived from the EM data, and U_{SS} is a potential that aims to preserve the secondary structure of protein and nucleic acids.

The data provided by cryo-EM reconstructions represent the Coulomb potential of the macromolecule; the dependence of this potential on the atomic number of the composing atoms makes it roughly proportional to the mass density of the macromolecule. It is then sensible to define a potential so that when the atomic structure is placed in it, the atoms are driven through application of forces into high-density areas and away from low-density areas. This potential can be defined on a grid, thus preserving all the information contained in the EM density map.

MDFF method

❑ Incorporating EM data into simulation

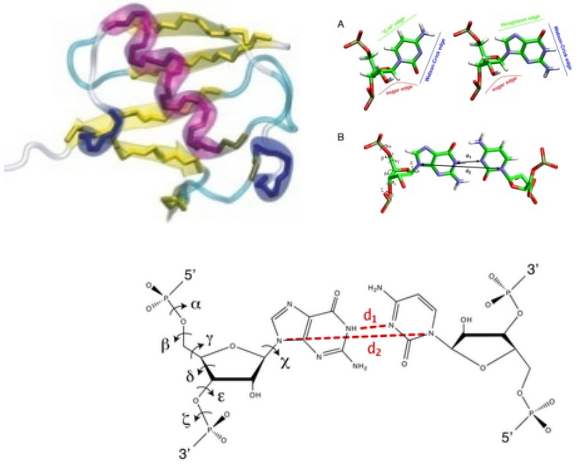
The method incorporates the EM density map as a potential so that high density areas in the map correspond to energy minima, so that the atoms in the structure are subject to forces proportional to the gradient of the EM map.



Volume slice of a 6.7-Angstrom EM map of the ribosome. The arrows show the gradient of the external field that directs the structure into the density.

❑ Preventing overfitting

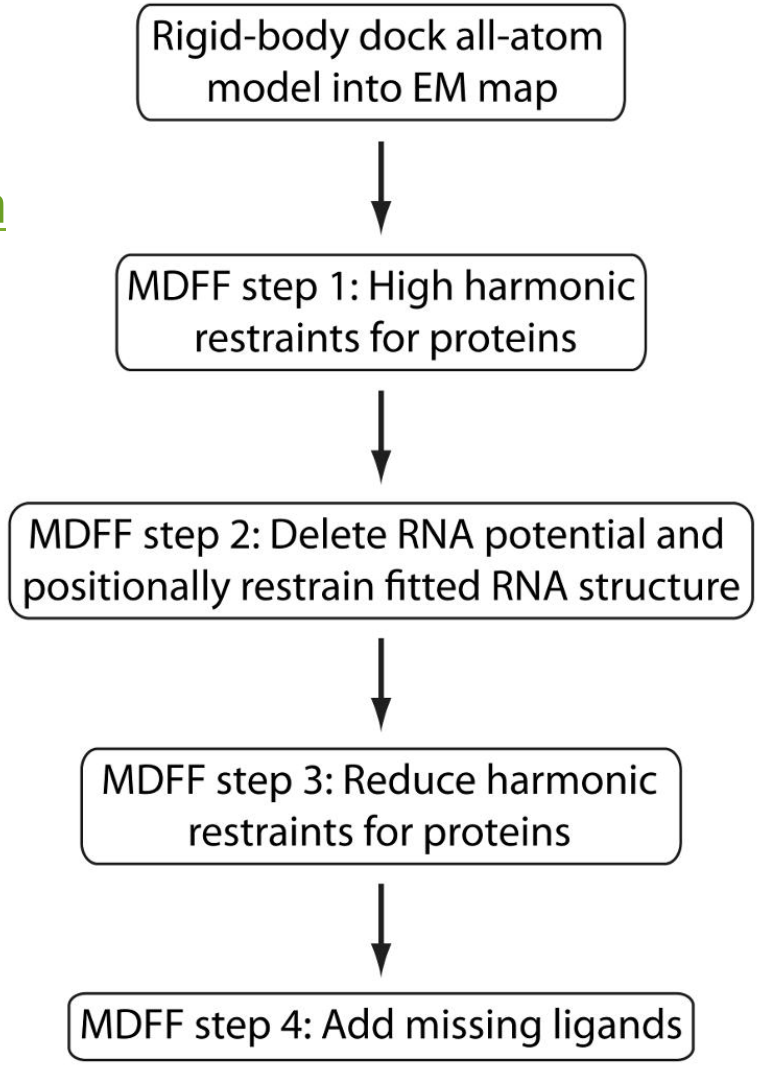
For proteins, the dihedral angles of residues are restrained to their initial positions. For nucleic acids, restraints are imposed to seven dihedral angles, as well as two interatomic distances between base pairs.



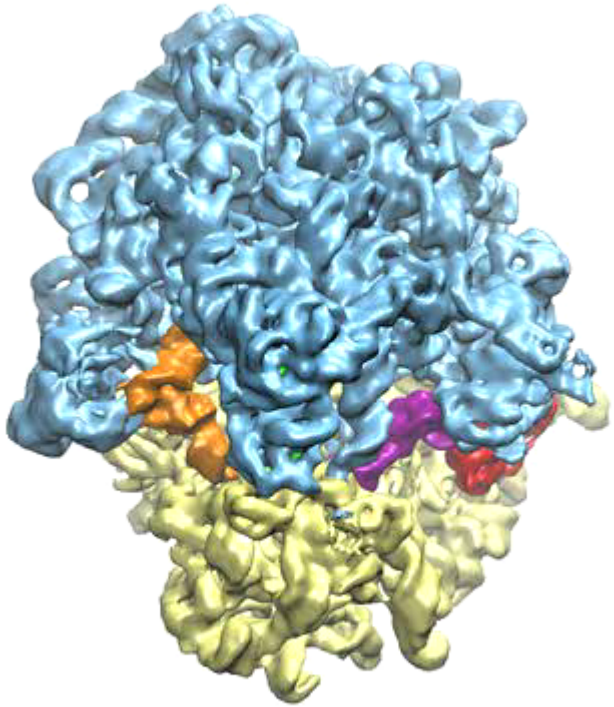
Harmonic restraints are imposed to preserve secondary structure of proteins and nucleic acids.

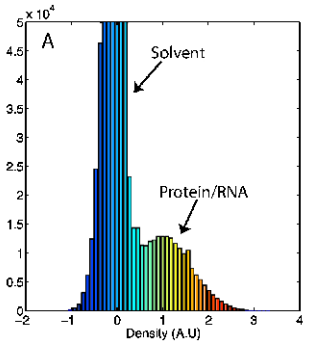
MDFF Method

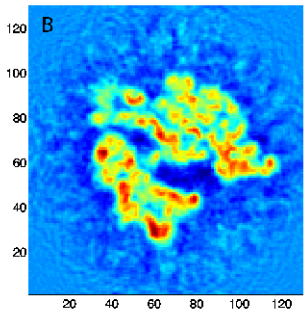
- Incorporating EM data into simulation
- Preventing overfitting
- Validation: Crystal structures of the same
- Biomolecule captured in two conformational states
- Application: The bacterial ribosome
- Symmetry-restrained MDFF for symmetric EM data

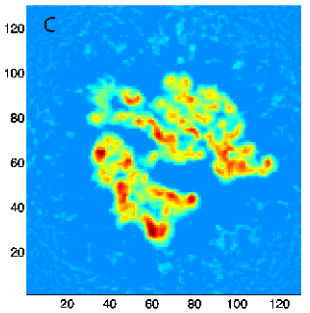


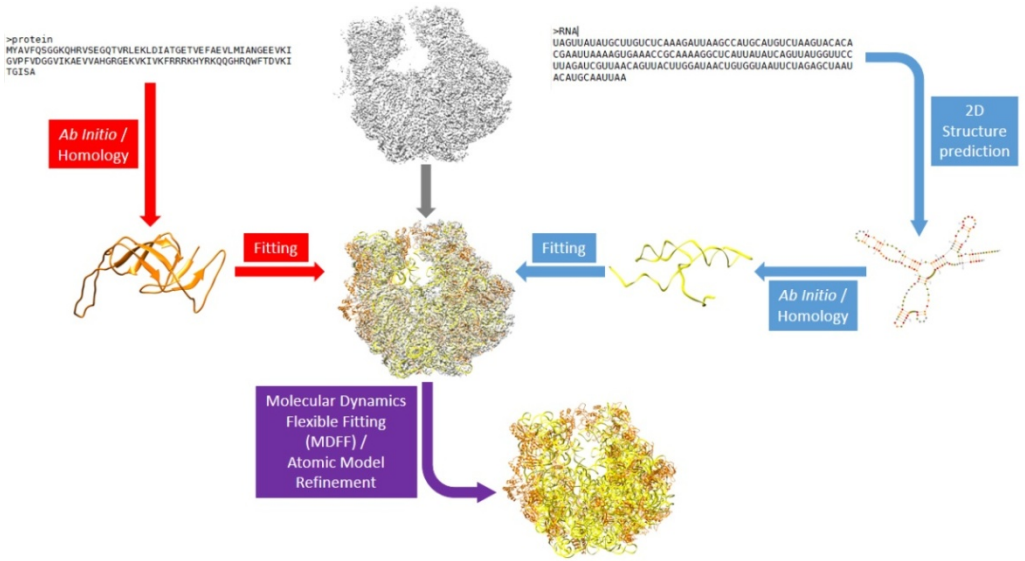
Cryo-EM map of the *E. coli* ribosome at 6.7-Å resolution











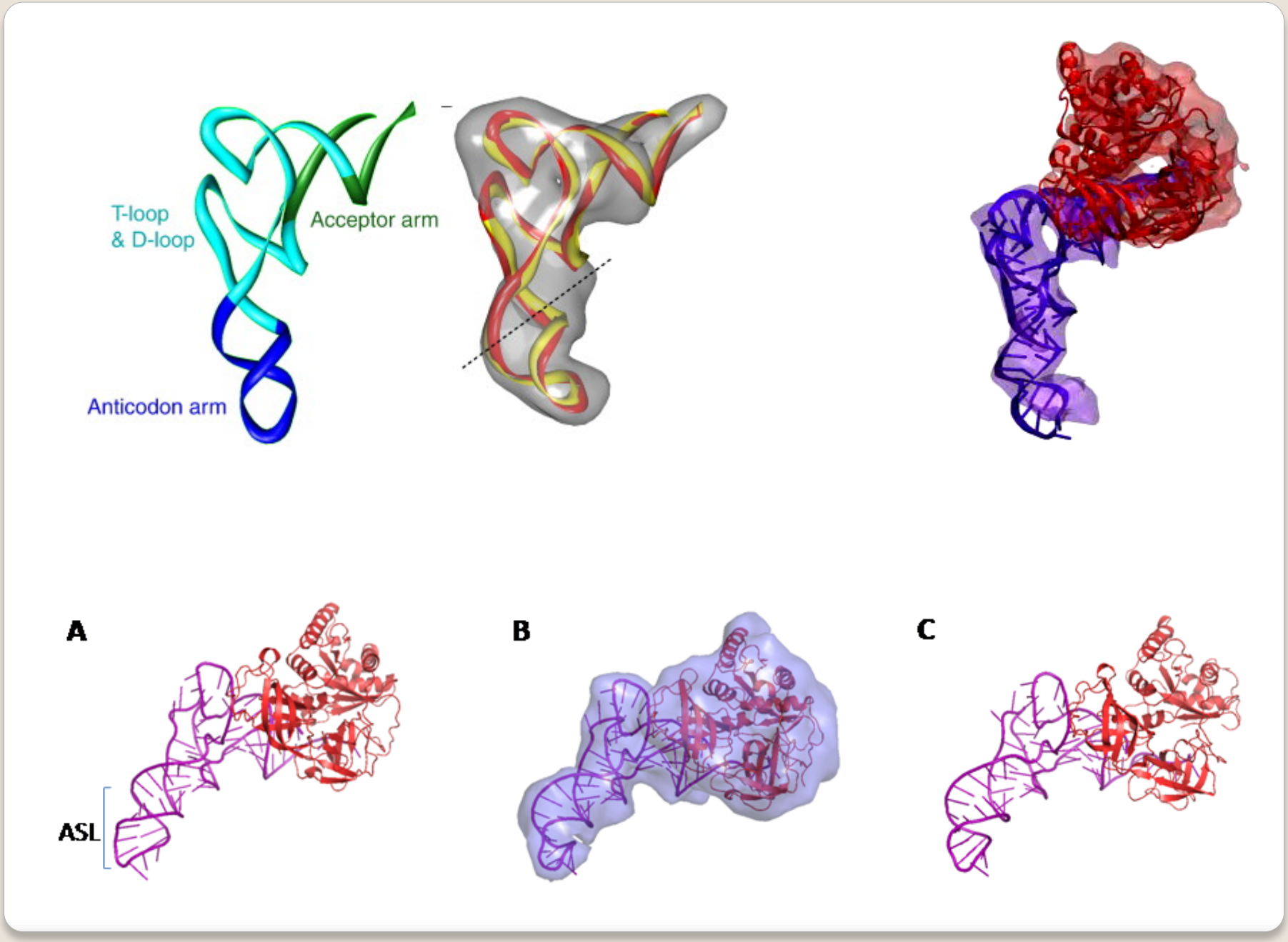
Multistep flexible fitting protocol applied to ribosome cryo-EM maps

To quantify the goodness of the fit, a simulated map can be generated from the fitted atomic structure with the same target resolution as the EM map.

The Pearson's correlation coefficient (usually referred to in the EM literature as the cross-correlation coefficient) between these two data sets, *i.e.*, the simulated (*S*) and the experimental (*E*) 3-D maps, can be used as a measure of similarity between them, and is given by

$$\rho_{SE} = \frac{\langle (S - \langle S \rangle)(E - \langle E \rangle) \rangle}{\sigma_S \sigma_E},$$

where $\langle S \rangle$ and $\langle E \rangle$ correspond to the average voxel values of the simulated and experimental maps, respectively,


T-loop
& D-loop
Acceptor arm
Anticodon arm
A
ASL
B
C


Thank you!