NRAMM cryo-EM
Workshop 11/05

NMFF - Flexible fitting of atomic structures into EM maps

Charles L. Brooks III
(Florence Tama)

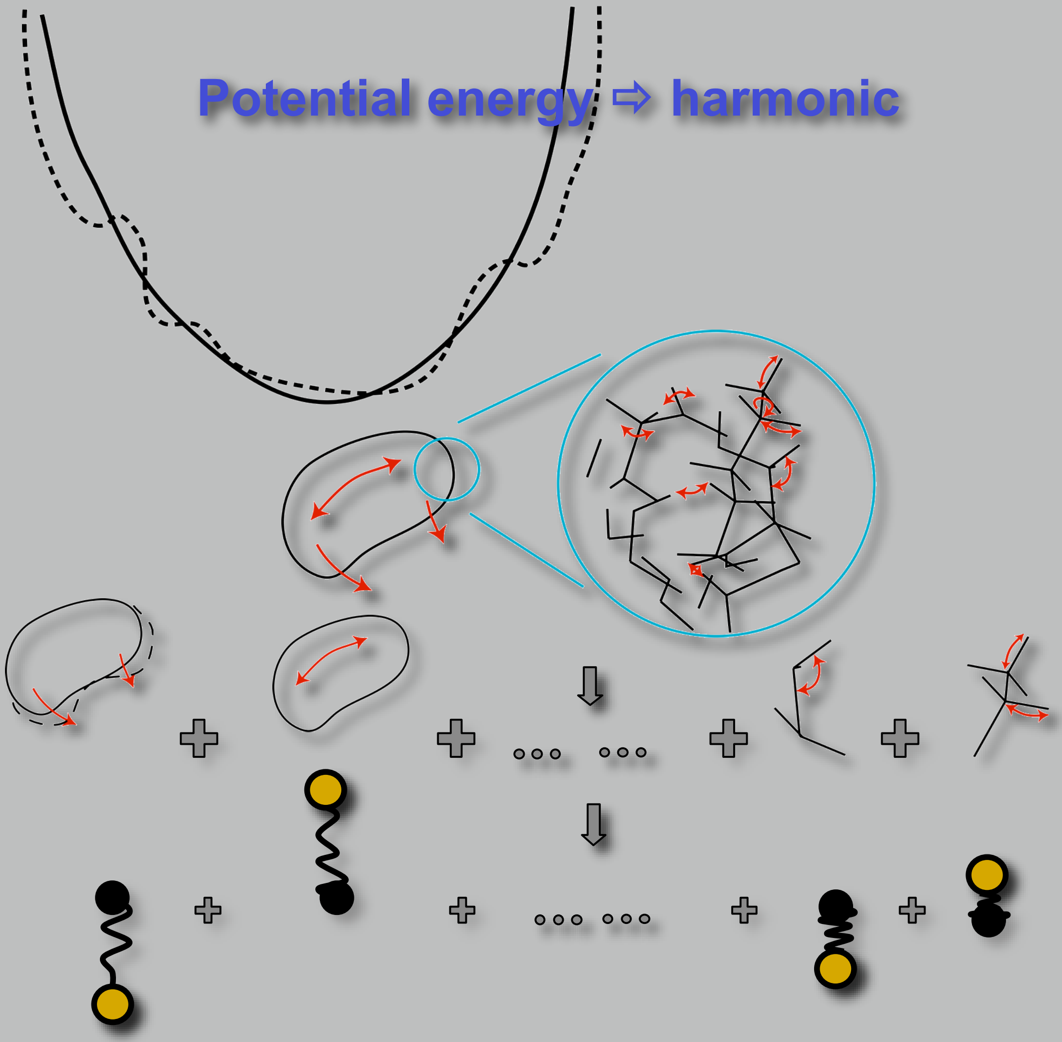
MMTSB

NCRR

THE
SCRIPPS
RESEARCH
INSTITUTE

Elastic network normal mode analysis - a multi-resolution framework for exploration of large-scale conformational changes

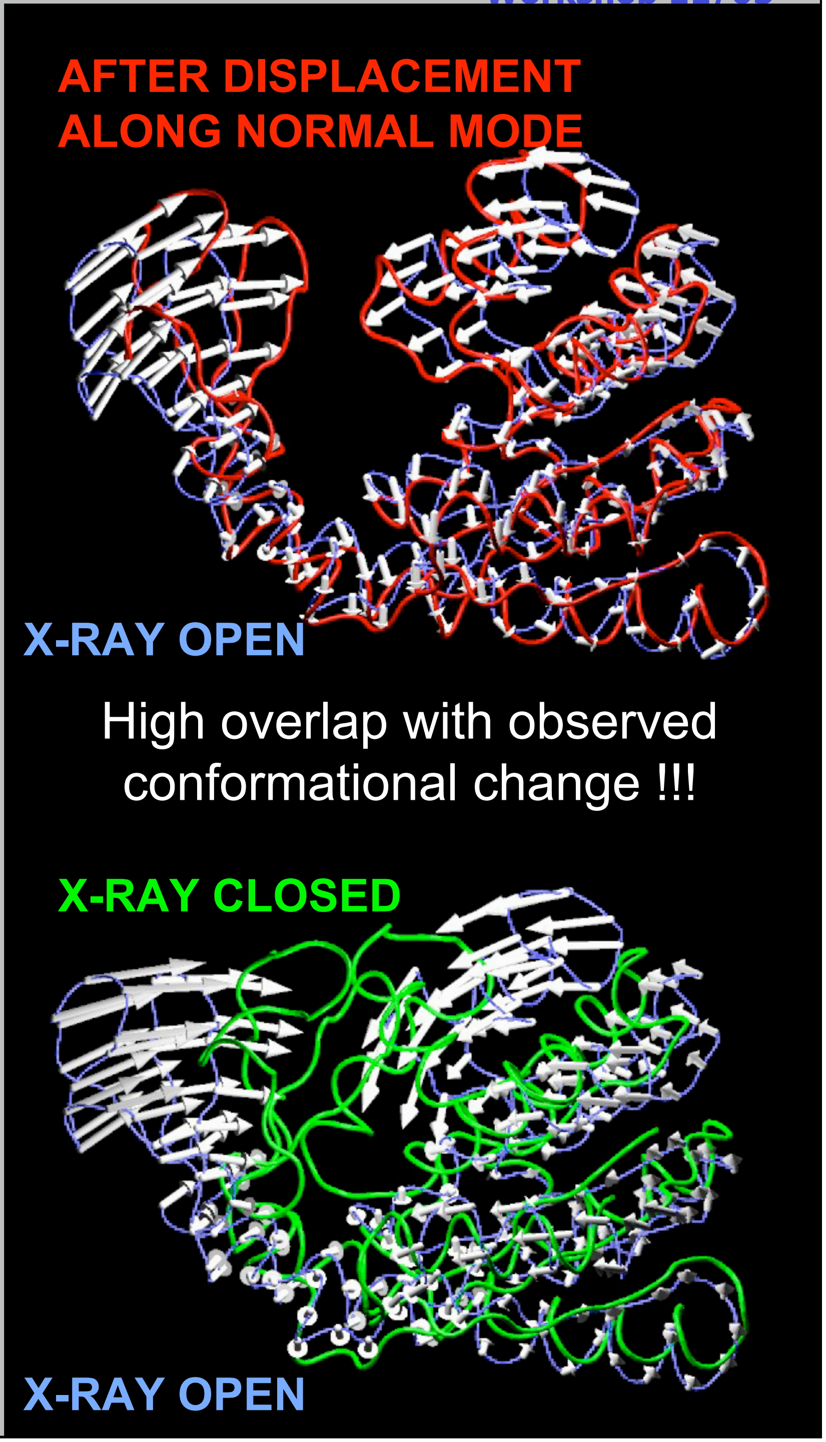
Normal mode analysis



Potential energy ⇨ harmonic

Low frequencies ⟷ High frequencies

Related to functional changes in biological systems



High overlap with observed conformational change !!!

Technical Issues in Normal Mode Computations

$$U(r) \cong \frac{1}{2}\sum_{ij} \left.\frac{\partial^2 U}{\partial r_i \partial r_j}\right|_{r=r_o} (r_i - r_i^0)(r_j - r_j^0) \longrightarrow \frac{1}{2}\sum_{n} \omega_n^2 q_n^2$$

$$\mathbf{H} = \left(\frac{\partial^2 U}{\partial r_i \partial r_j}\right)$$

Hessian: 2nd derivative
of the potential

↓

Eigenvalue problem

$$\mathbf{A}^T\mathbf{H}\mathbf{A} = \mathbf{L}$$

↓

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \omega_1^2 & & 0 \\ & \omega_2^2 & \\ 0 & & \ddots \end{pmatrix}$$

Eigenvector = normal mode

Eigenvalue = frequency

Technical Issues in Normal Mode Computations

$$U(r) \cong \frac{1}{2}\sum_{ij} \left.\frac{\partial^2 U}{\partial r_i \partial r_j}\right|_{r=r_o} (r_i - r_i^0)(r_j - r_j^0) \longrightarrow \frac{1}{2}\sum_{n} \omega_n^2 q_n^2$$

Problems with large biological systems

$$\mathbf{H} = \left(\frac{\partial^2 U}{\partial r_i \partial r_j}\right)$$

Hessian: 2nd derivative of the potential → Minimization

Eigenvalue problem

$$\mathbf{A}^T\mathbf{H}\mathbf{A} = \mathbf{L}$$

→ Size of the system (3Nx3N)

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots \end{pmatrix}$$

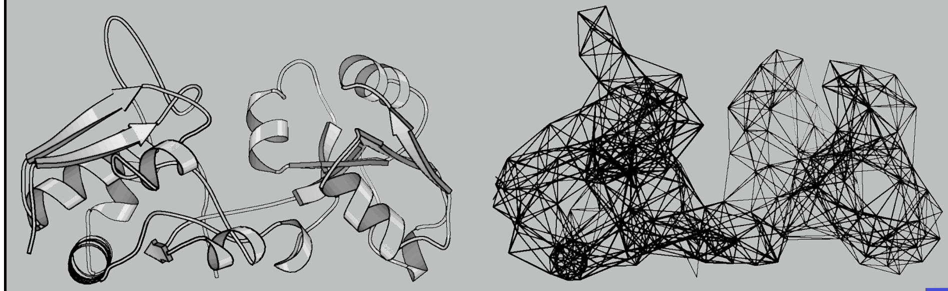
$$\mathbf{L} = \begin{pmatrix} \omega_1^2 & & 0 \\ & \omega_2^2 & \\ 0 & & \ddots \end{pmatrix}$$

Eigenvector = normal mode

Eigenvalue = frequency

Elastic network normal mode analysis

Minimization => Tirion Potential (*)



$$E(r_a, r_b) = \frac{C}{2}\left(\left|r_{a,b}\right| - \left|r^0_{a,b}\right| \right)^2$$

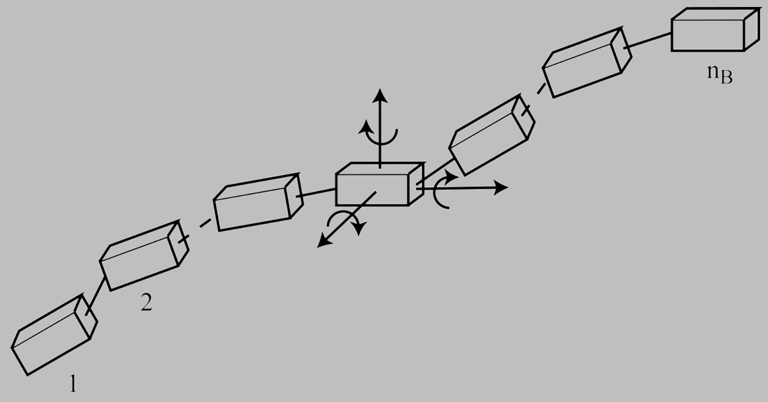
Hookean potential

- ➢ No minimization
- ➢ Coarse grained model ⇨ Cα atoms

$$E_p = \sum_{a,b} E(r_a, r_b) \begin{cases} 1 \rightarrow r^0_{a,b} \le R_{Cut} \\ 0 \rightarrow r^0_{a,b} \ge R_{Cut} \end{cases}$$

Cutoff for network elastic bonds

Diagonalization of Hessian => RTB (Rotation Translation Blocks) method (**)



- block = 1 or several residues treated as rigid body
- rotation + translation of block ⇨ new basis
- expression of Hessian in this new basis
- Diagonalization of a matrix $6n_B*6n_B$

* Tirion MM (1996) *Phys Rev Lett*. **77**, 1905-1908

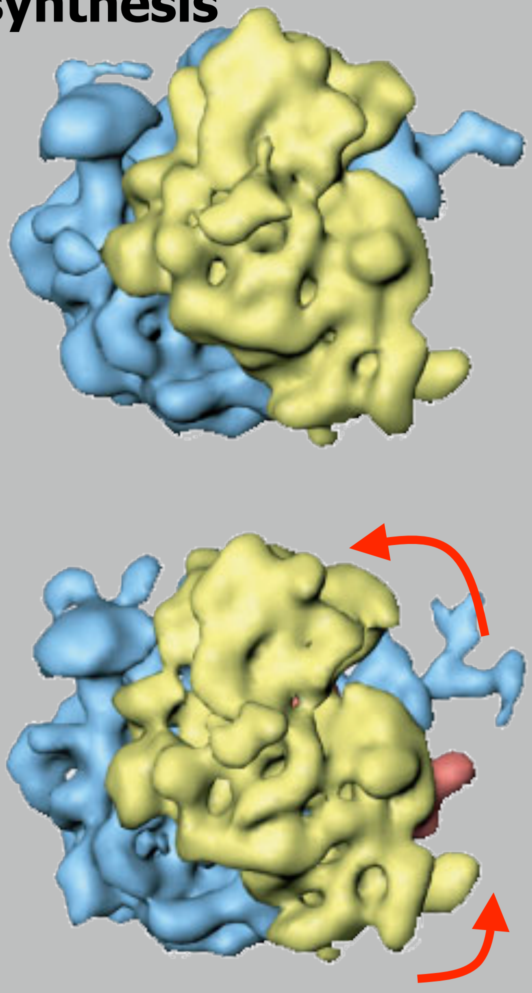
** Tama et al. (2000) Proteins

Elastic network normal mode analysis

- Determining the constants for ENNMA
  - R_{cut} → 2^{nd} minimum in (pseudo)atom - atom(pseudo) distribution function
    - 4.5 Å heavy atoms, 7-8 Å $C\alpha$-$C\alpha$, 10-12 Å P-P (DNA/RNA), 10-12 Å $C\alpha$-P (protein-na)
    - Can be as large as 15-25 Å for really coarse-grained models
  - Level of RT-block coarsening
    - Varies depending on system
      - Residue for proteins (aa), 1-5 (or more) for $C\alpha$/P in small complexes, larger, e.g., 1 per protein, in large structures like viruses

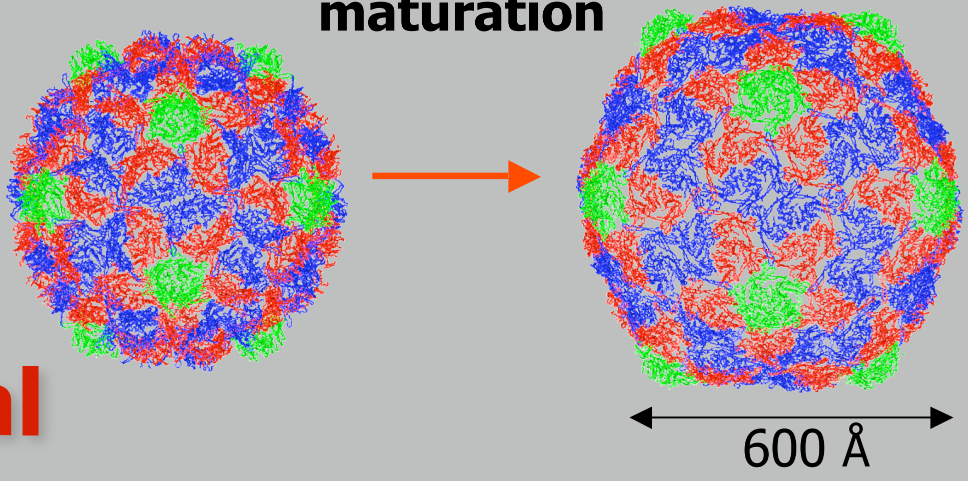
Large-scale conformational changes

Ribosome: protein synthesis

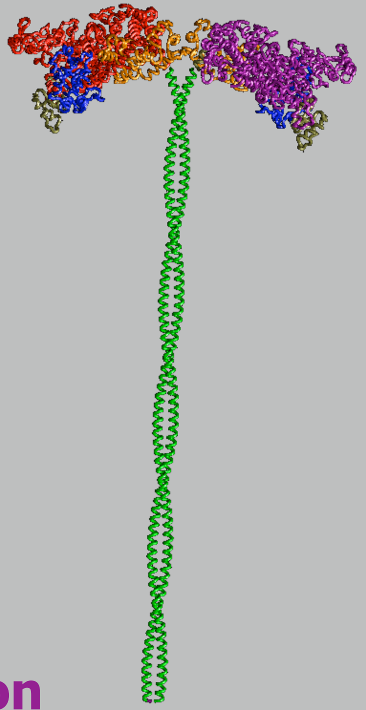


Functional motions

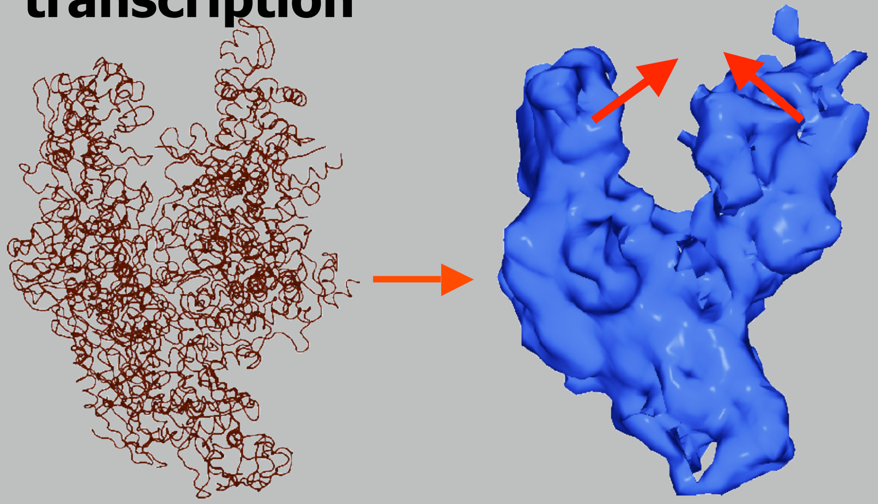
Virus maturation



Myosin II inhibition



RNA polymerase: transcription

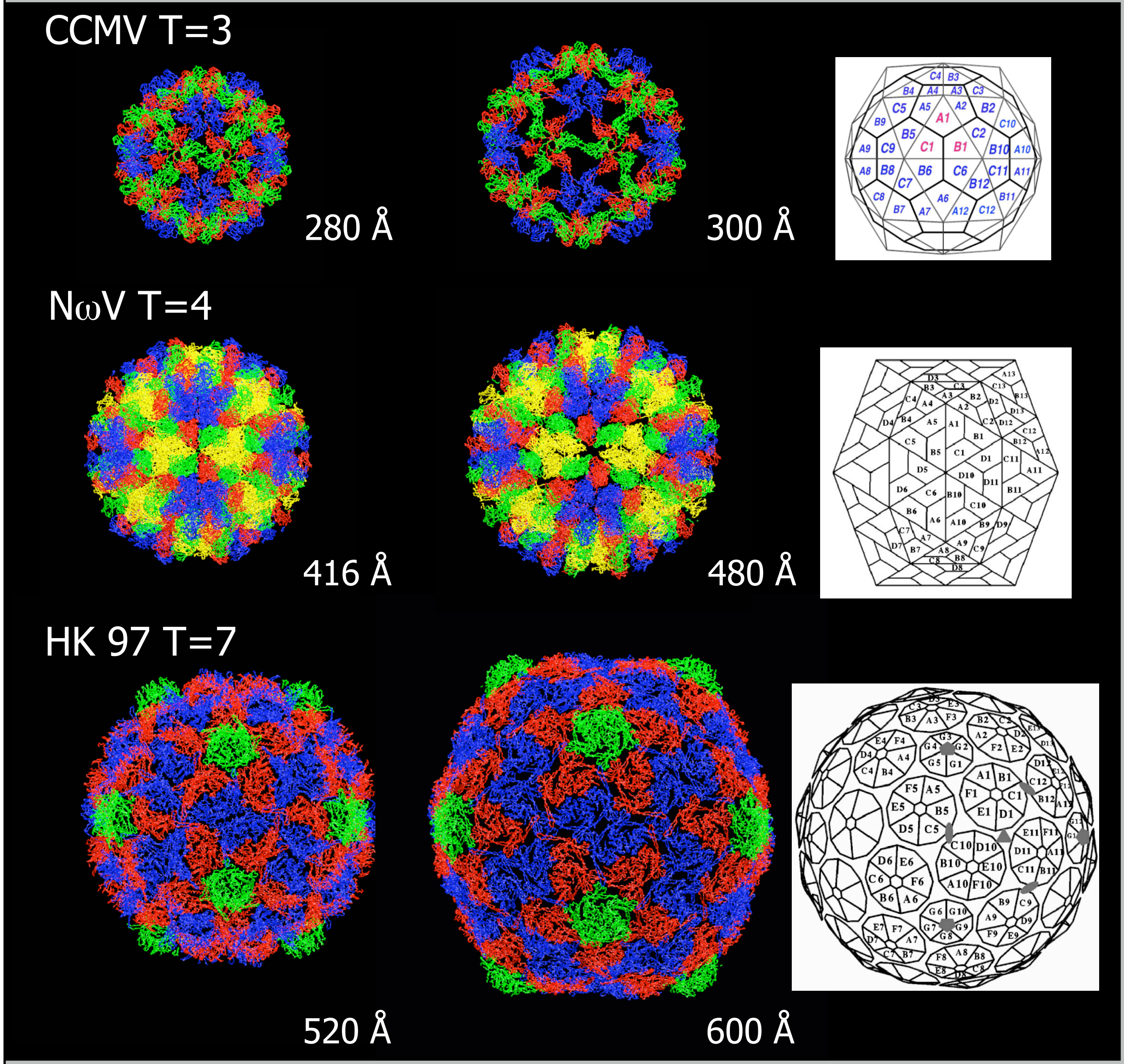


- Time scale (> ms)
- Structural information: from low (cryo-EM) to high resolution

Exploring macromolecular machines with ENNMA - virus capsids

Tama & Brooks, *JMB*, (2002); *ibid* (2005).

Exploring large-scale conformational changes in virus maturation



- Large conformational changes observed for several viruses
- Icosahedral symmetry
- Normal mode analysis

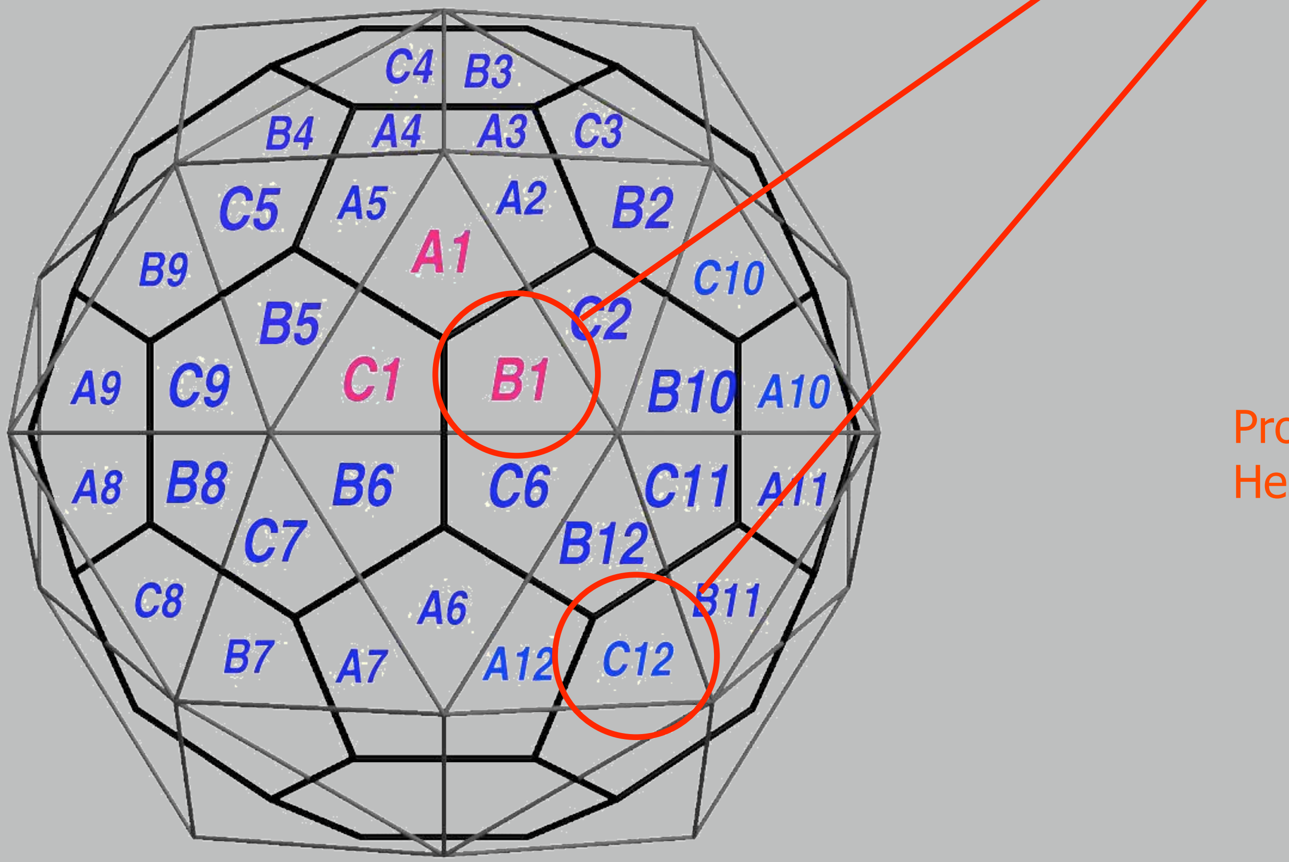
- well reproduces conformational change of a small virus CCMV

- do different motions characterize dynamics of viruses with different quasi-equivalent symmetries?

http://viperdb.scripps.edu

Normal mode analysis applied to viruses

Coarse grained model: Cα atoms + RTB method → one protein = one block



Rotation + Translation of blocks ⇨ new basis

Projection of the Hessian

Diagonalization of matrix

T=3 ⇨ 1080 x 1080

T=4 ⇨ 1440 x 1440

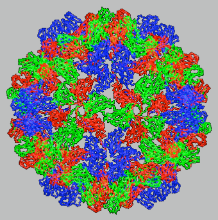
T=7 ⇨ 2520 x 2520

T=13 ⇨ 5040 x 5040

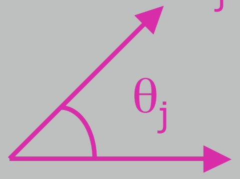
Tama & Brooks (2002) J. Mol. Biol.

How well is the conformational change described?

Native

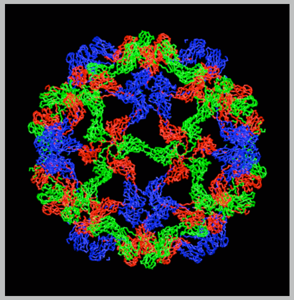


Normal Mode $\mathbf{a}_j$



Displacement vector $\Delta\vec{r}$

Procapsid

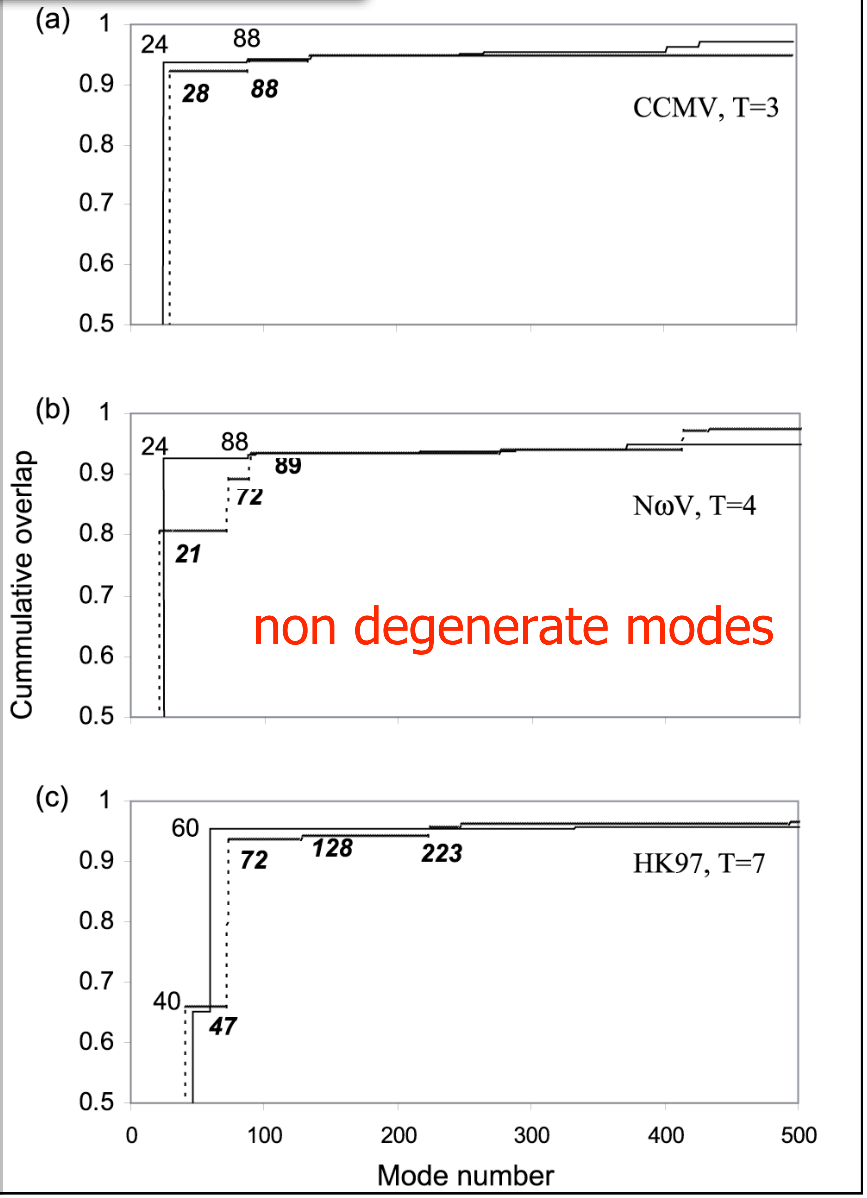


$$overlap_j = \frac{\mathbf{a}_j \cdot \Delta\mathbf{r}}{|\Delta\mathbf{r}||\mathbf{a}_j|}$$

Σ overlap^2

- ➢ Represents 95% of the conformational change
- ➢ CCMV and native NωV ⇨ one predominant mode accounts for more than 90 % of the conformational change.
- ➢ HK97 ⇨ first mode only 65 %.
- ➢ First mode is well conserved between the two states

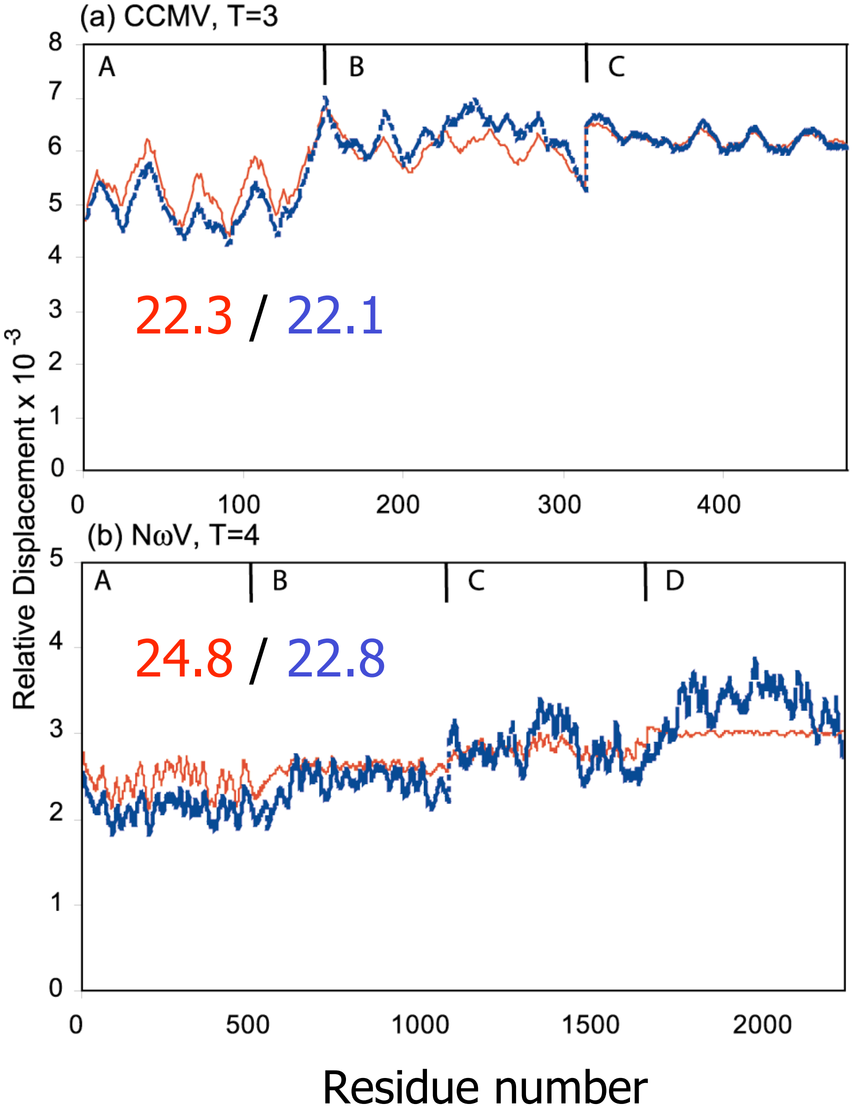
Overlap	CCMV ⇨	0.99
	NωV ⇨	0.91
	HK97 ⇨	0.97

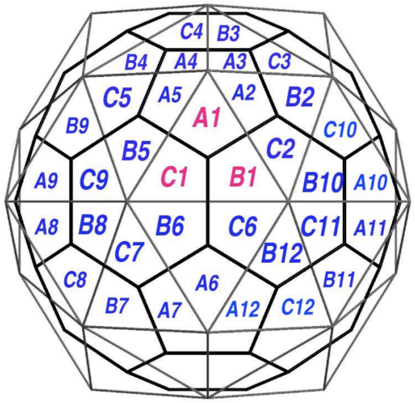


non degenerate modes

NRAMM cryo-EM Workshop 11/05

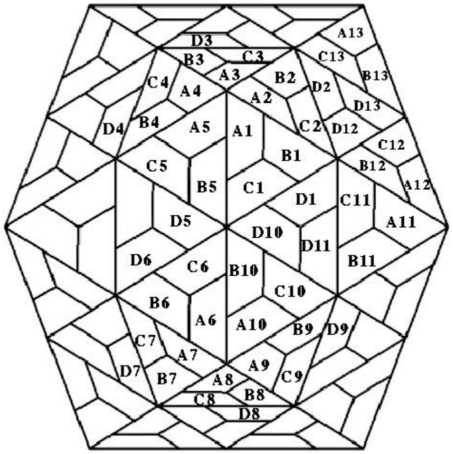
Nature of the low-frequency normal modes T=3/T=4





Displacement along mode

$$\mathbf{r} = (\mathbf{a}_j \cdot \Delta\mathbf{r})\,\mathbf{a}_j + \mathbf{r}_0$$



Translation (Å) experimentally observed

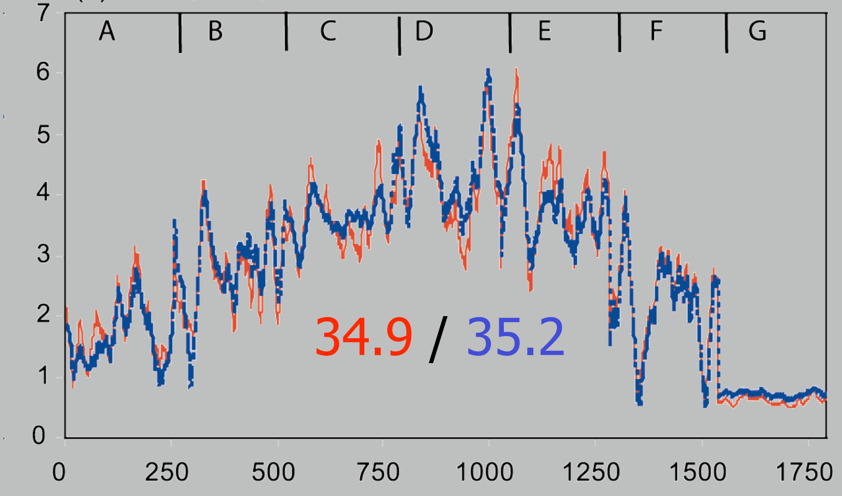
CCMV:	21.7
NωV:	24.7

Lowest frequency mode captures the overall translation of the asymmetric unit

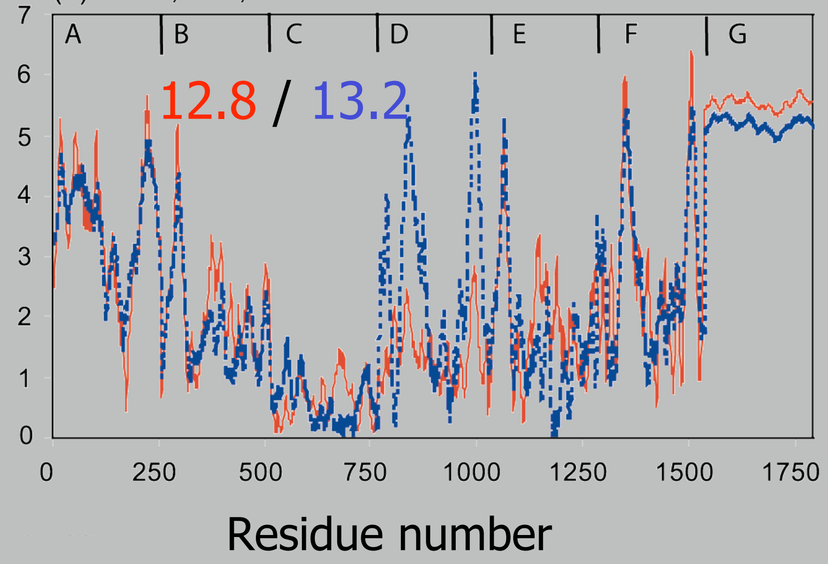
Nature of the low-frequency normal modes: HK97

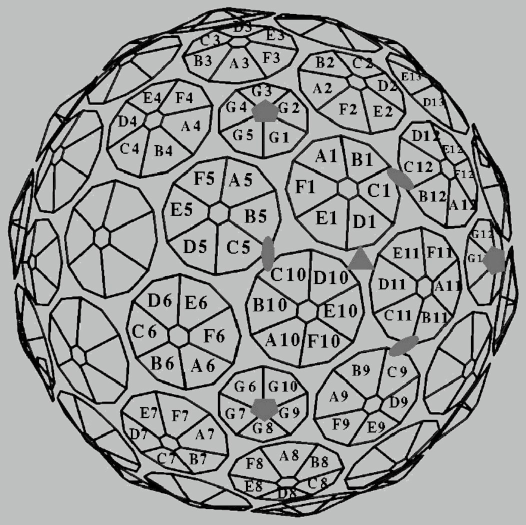
Experimentally observed translation - 47 Å

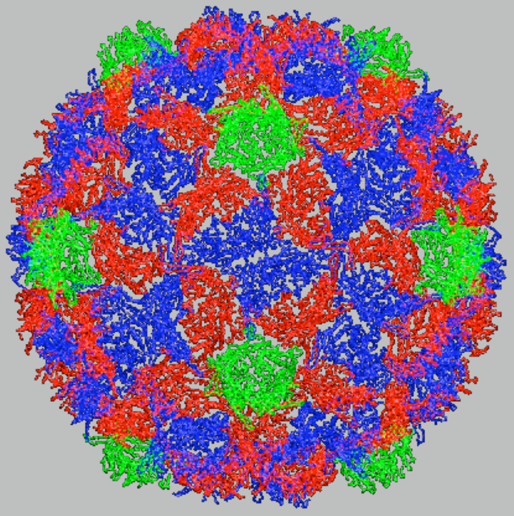
(c) HK97, T=7, mode 1



(d) HK97, T=7, mode 2





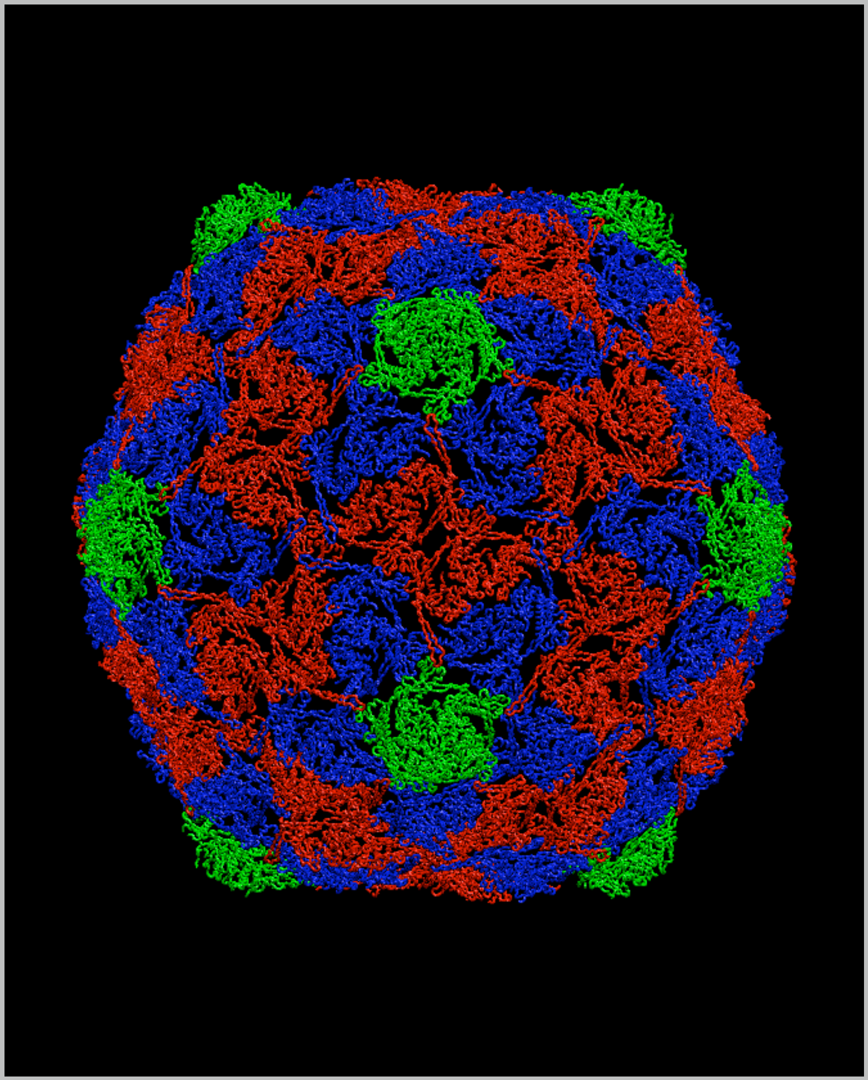


HK97 ⇨ 2 modes to describe the whole translation

HK97

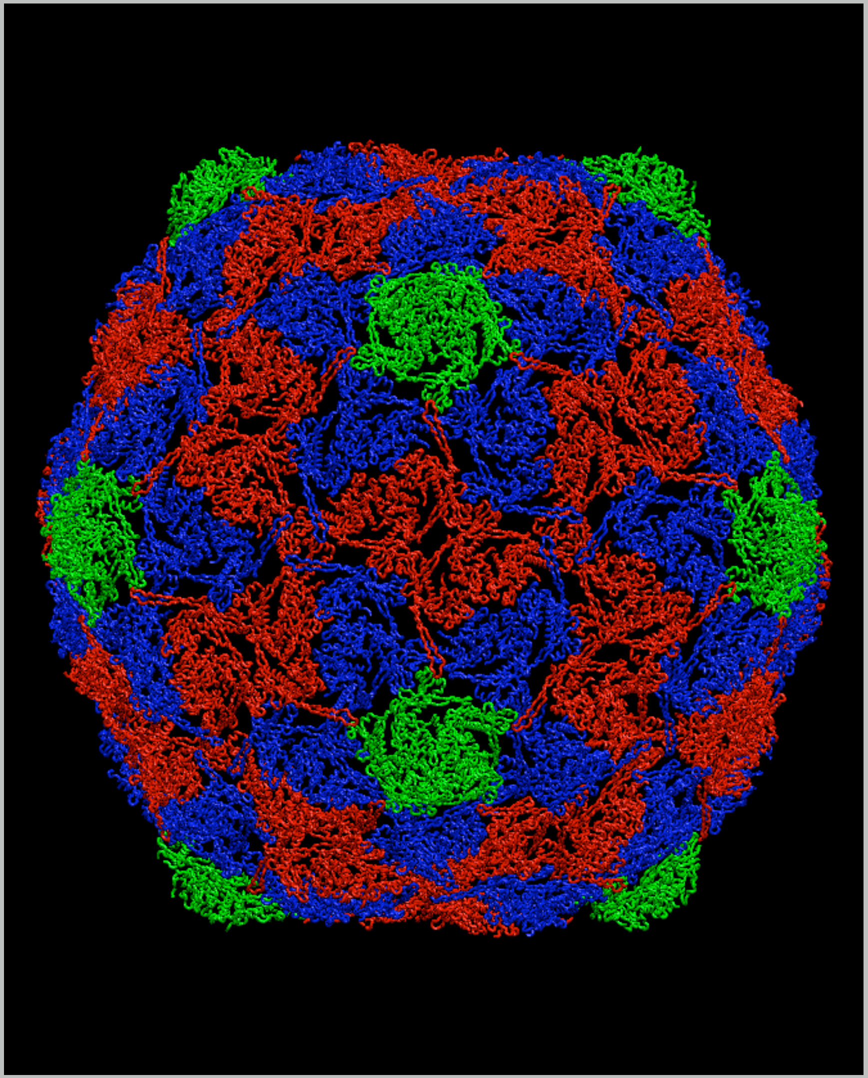
NRAMM cryo-EM
Workshop 11/05

Prohead II: Mode 1



Hexamers

Head II: Mode 2



Pentamers

Necessary to achieve the shape transition

Summary - virus dynamics

- ➢ HK97: a pair of low frequency normal modes is necessary to produce the non-uniform conformational transition
- ➢ NωV and CCMV: one normal mode provides the nearly uniform overall translation associated with the conformational transition
- ➢capsid shell is not mechanically uniform, especially for viruses of higher complexity such as T=7 and T=13 viruses
- ➢ pentameric units display higher flexibility

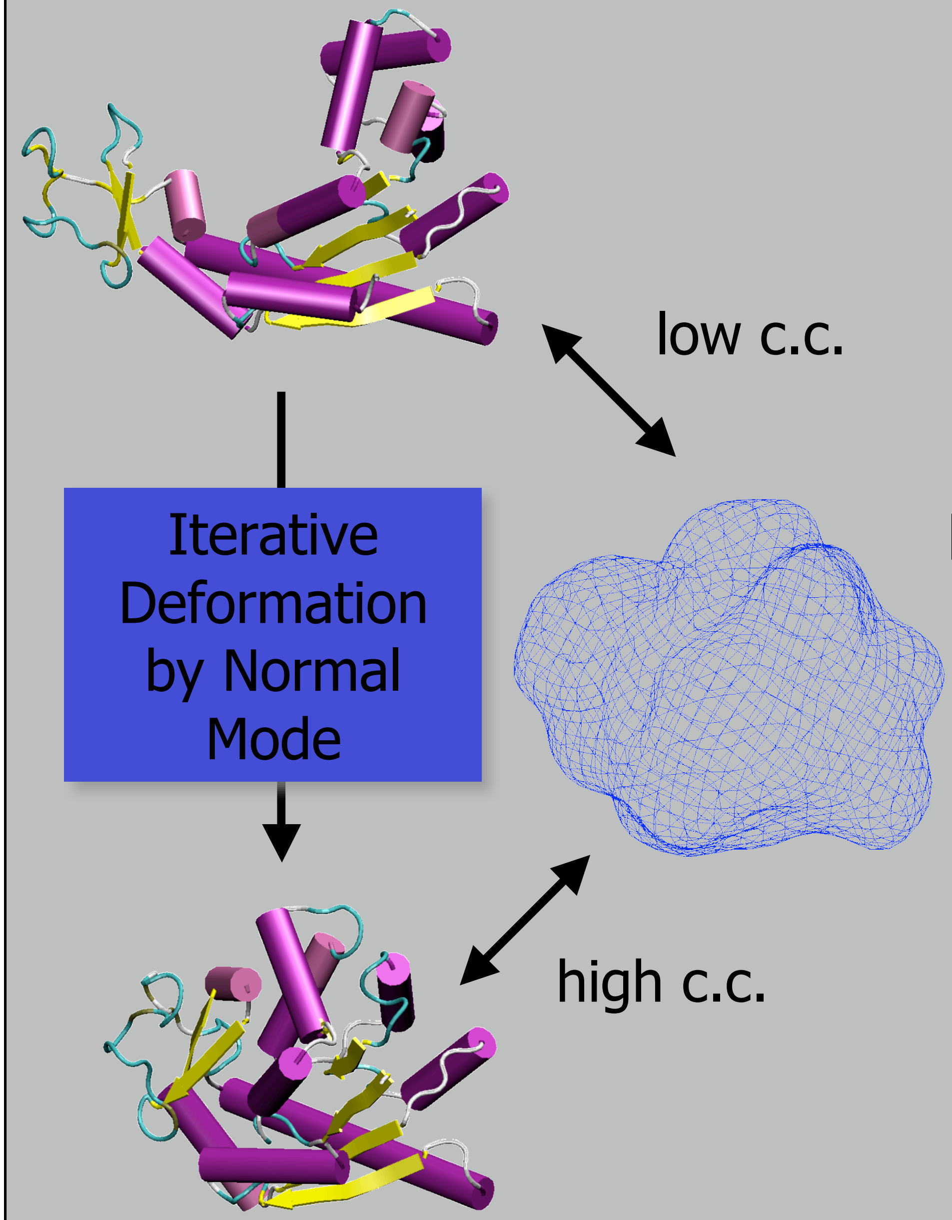
“Practical” tools for structural biology based on ENNMA

- By exploiting the low-dimensionality of the space required to achieve functionally relevant spatial reorganization we can develop lower-resolution structure refinement/fitting methodologies
- Normal Mode Flexible Fitting for flexibly fitting atomic models into low resolution structural data from cryo-EM

Normal Mode Flexible Fitting

NRAMM cryo-EM Workshop 11/05

$$\mathrm{c.c.} = \frac{\sum_{ijk} \rho^{\exp}(i,j,k)\rho^{\mathrm{sim}}(i,j,k)}{\sqrt{\sum_{ijk} \rho^{\exp}(i,j,k)^2 \sum_{ijk} \rho^{\mathrm{sim}}(i,j,k)^2}}$$

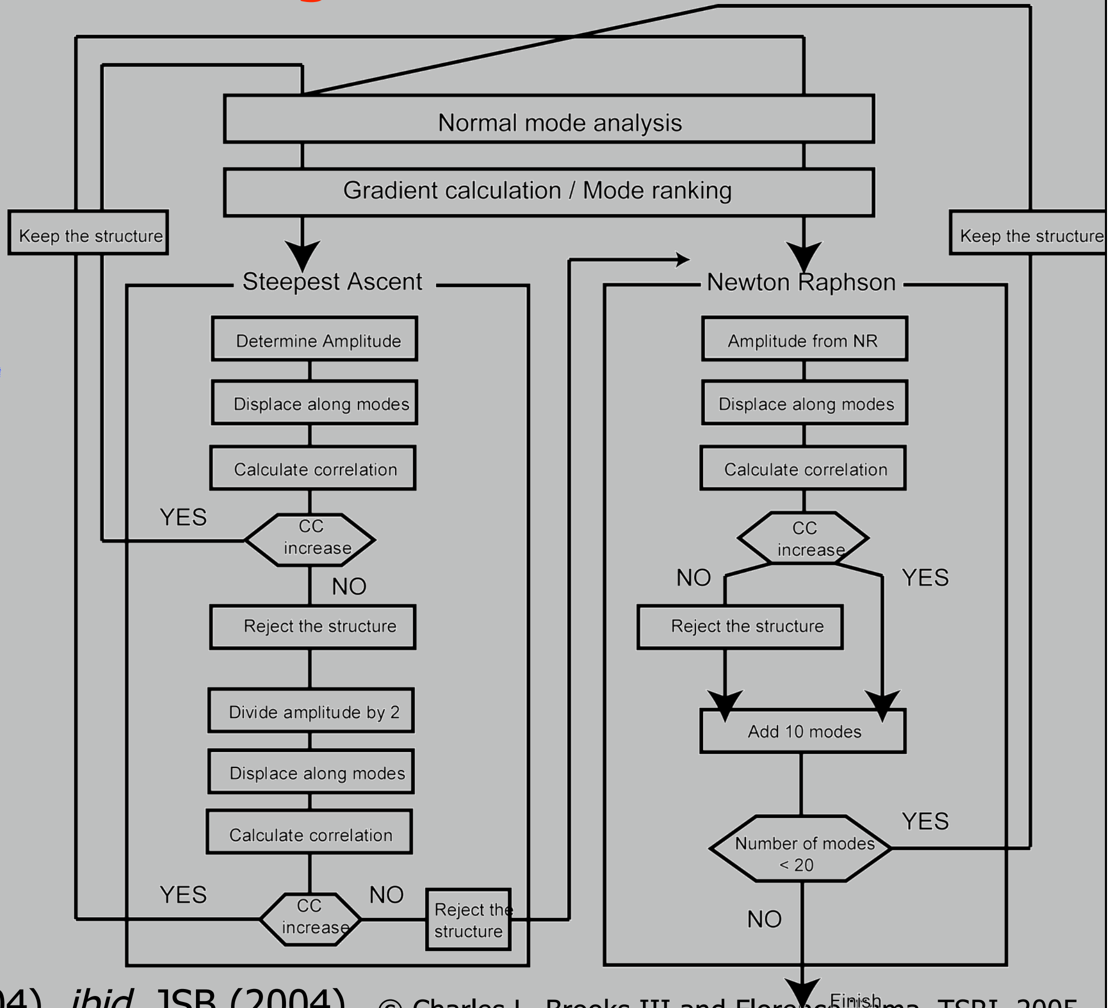


low c.c.

Iterative Deformation by Normal Mode

high c.c.

NMFF Algorithm



Normal mode analysis

Gradient calculation / Mode ranking

Keep the structure

Steepest Ascent

Determine Amplitude

Displace along modes

Calculate correlation

YES

CC increase

NO

Reject the structure

Divide amplitude by 2

Displace along modes

Calculate correlation

YES

CC increase

NO

Reject the structure

Newton Raphson

Amplitude from NR

Displace along modes

Calculate correlation

CC increase

NO

YES

Reject the structure

Add 10 modes

Number of modes < 20

YES

NO

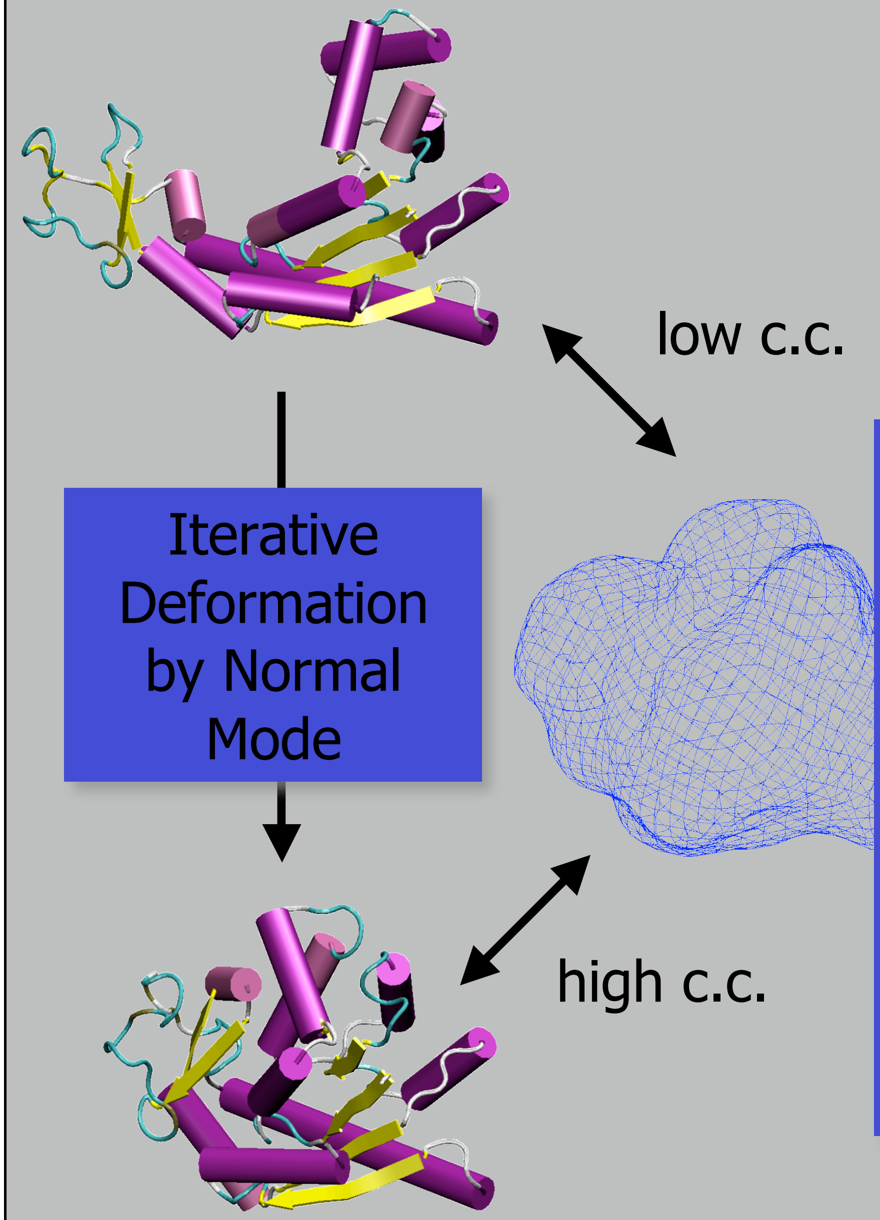
Keep the structure

Finish

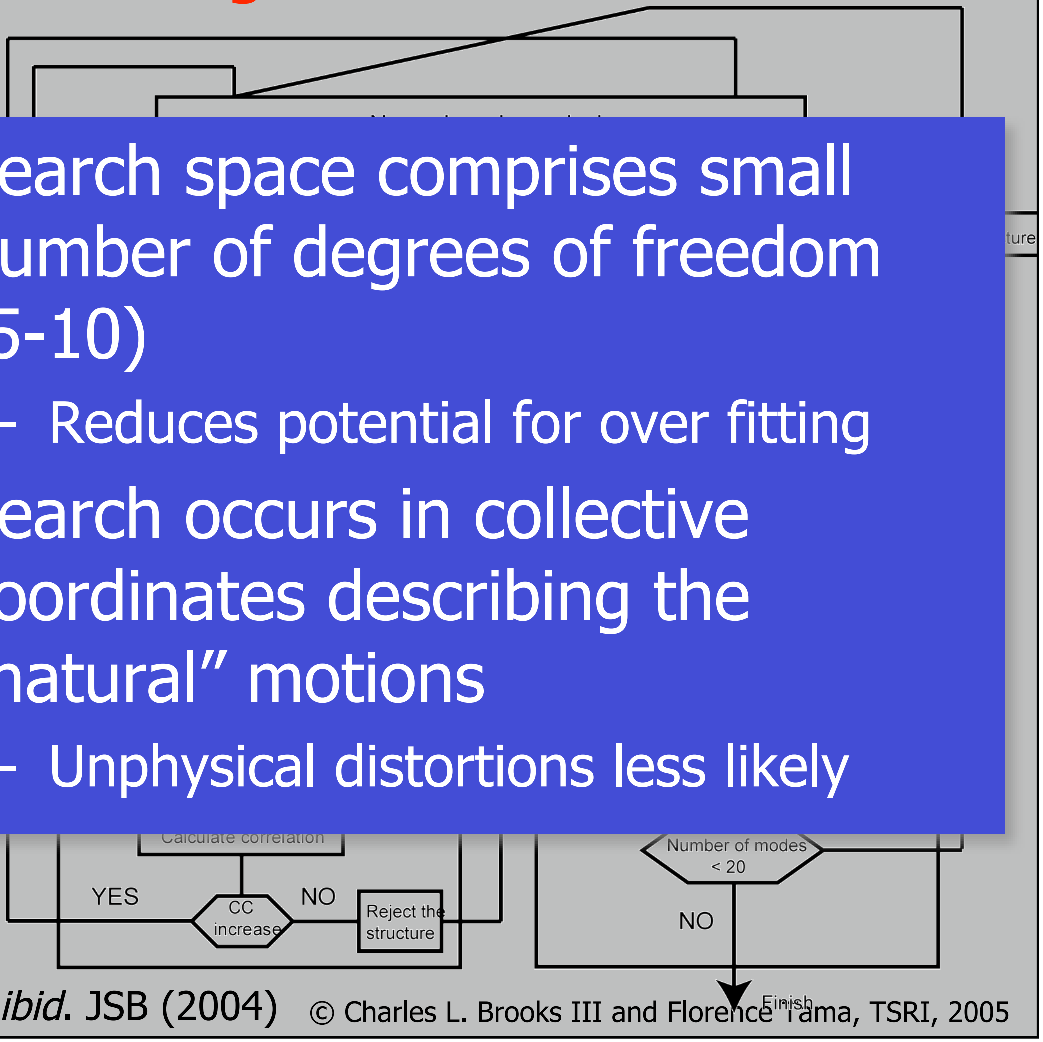
F. Tama *et al.* J. Mol. Biol. (2004), *ibid.* JSB (2004)

Normal Mode Flexible Fitting

$$\text{c.c.} = \frac{\sum_{ijk} \rho^{\exp}(i,j,k)\,\rho^{\text{sim}}(i,j,k)}{\sqrt{\sum_{ijk} \rho^{\exp}(i,j,k)^2 \sum_{ijk} \rho^{\text{sim}}(i,j,k)^2}}$$



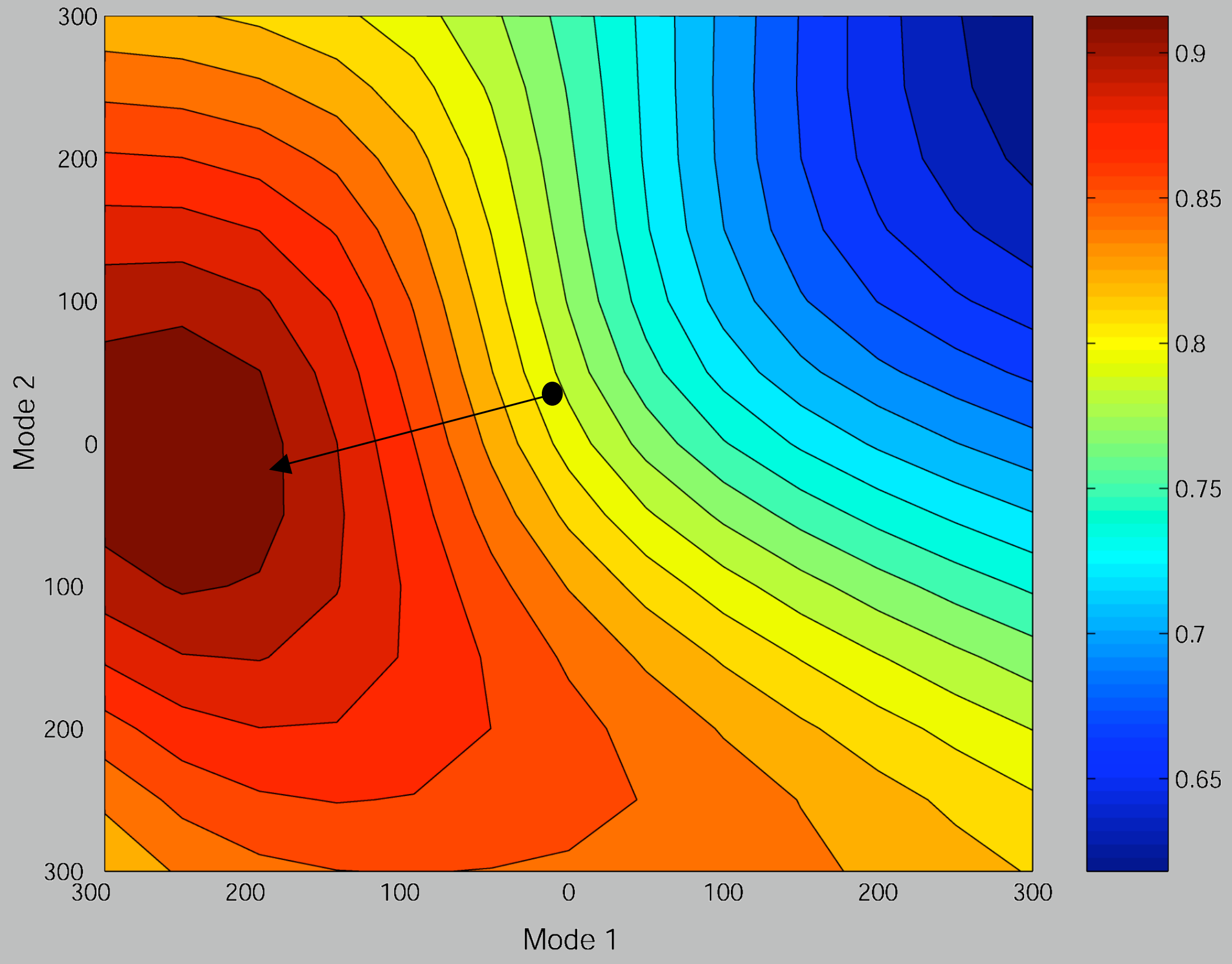
NMFF Algorithm



- Search space comprises small number of degrees of freedom (5-10)
  - Reduces potential for over fitting
- Search occurs in collective coordinates describing the “natural” motions
  - Unphysical distortions less likely

F. Tama *et al.* J. Mol. Biol. (2004), *ibid.* JSB (2004)

Correlation coefficient: Maximization problem



$$F_k = \frac{\partial \, \text{c.c.}}{\partial q_k}$$

- ☐ Maximization of the correlation coefficient as a function of normal mode coordinate

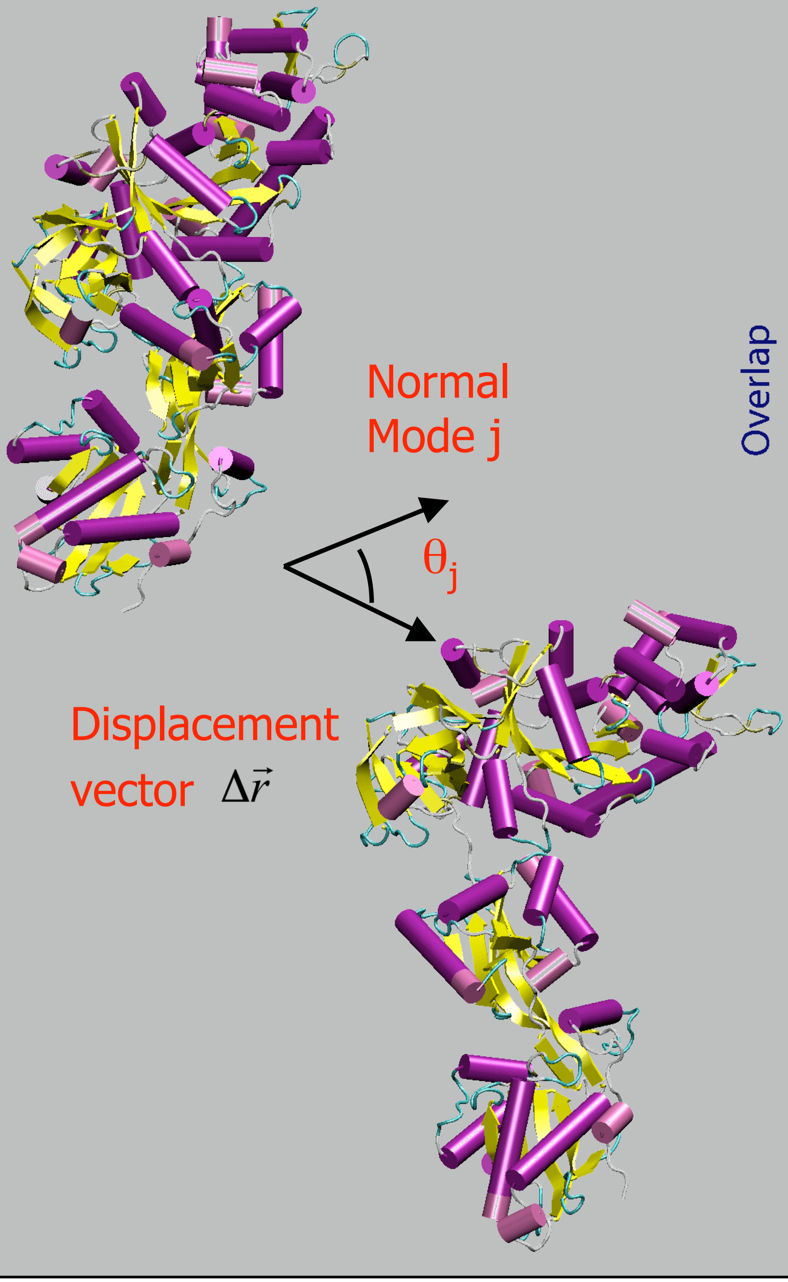
- ☐ Identification of the relevant normal modes

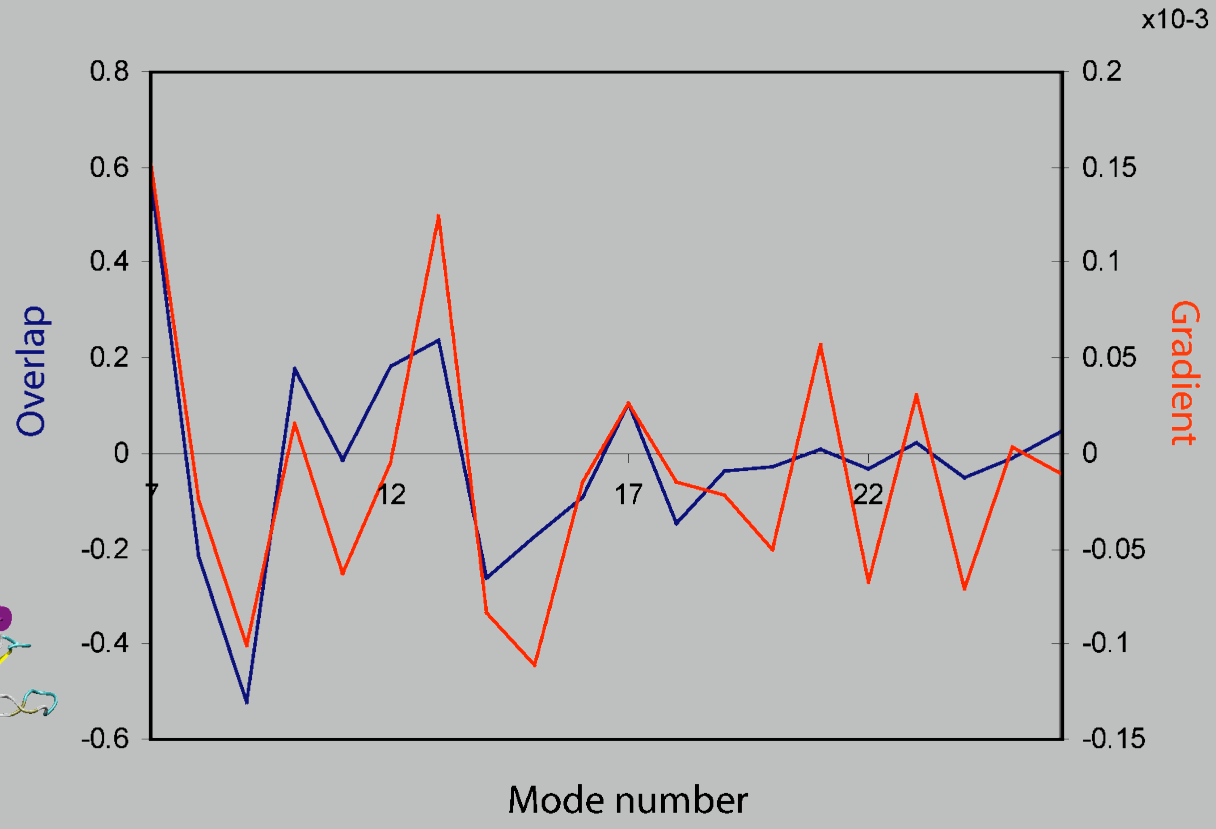
Steepest Descent /Newton Raphson

↓

Gradient/Hessian

Derivative of correlation coefficient by normal mode coordinates





$$F_k = \frac{\partial\, \mathrm{c.c.}}{\partial q_k}$$

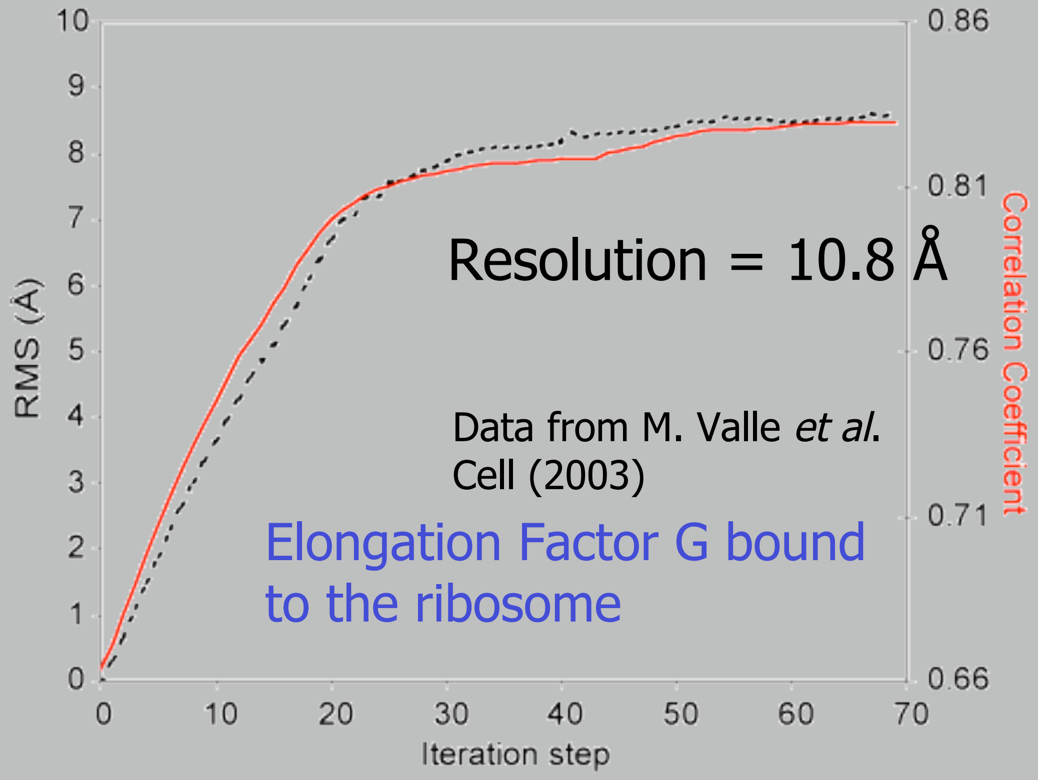
Flexible fitting results

			Final RMSD (Å)		
Initial structure	Atoms included	Resolution (Å)	Lactoferrin	EF2	Ca^{2+}-ATPase
RMSD fitted structure	All atoms	10	0.8 (1.4)	1.8 (2.1)	4.3 (4.9)
		20	1.1 (1.5)	2.1 (2.3)	5.0 (5.6)
		30	1.4 (1.9)	2.6 (3.1)	5.0 (5.5)
	Cα atoms	10	1.0	1.8	5.1
		20	1.3	2.2	4.7
		30	1.8	2.8	5.4
Situs rigid body*	All atoms	10	0.9 (1.4)	2.1 (2.3)	4.5 (5.0)
		20	1.0 (1.5)	2.2 (2.4)	4.9 (5.5)
		30	1.4 (1.8)	2.9 (3.0)	5.2 (5.7)
	Cα atoms	10	1.2	1.9	4.8
		20	1.4	2.2	4.7
		30	2.0	2.6	5.2
Original RMSD (Å)			6.5	14.6	14.4

Simple example flexible refinement

http://mmtsb/scrpps.edu/software/nmff.html

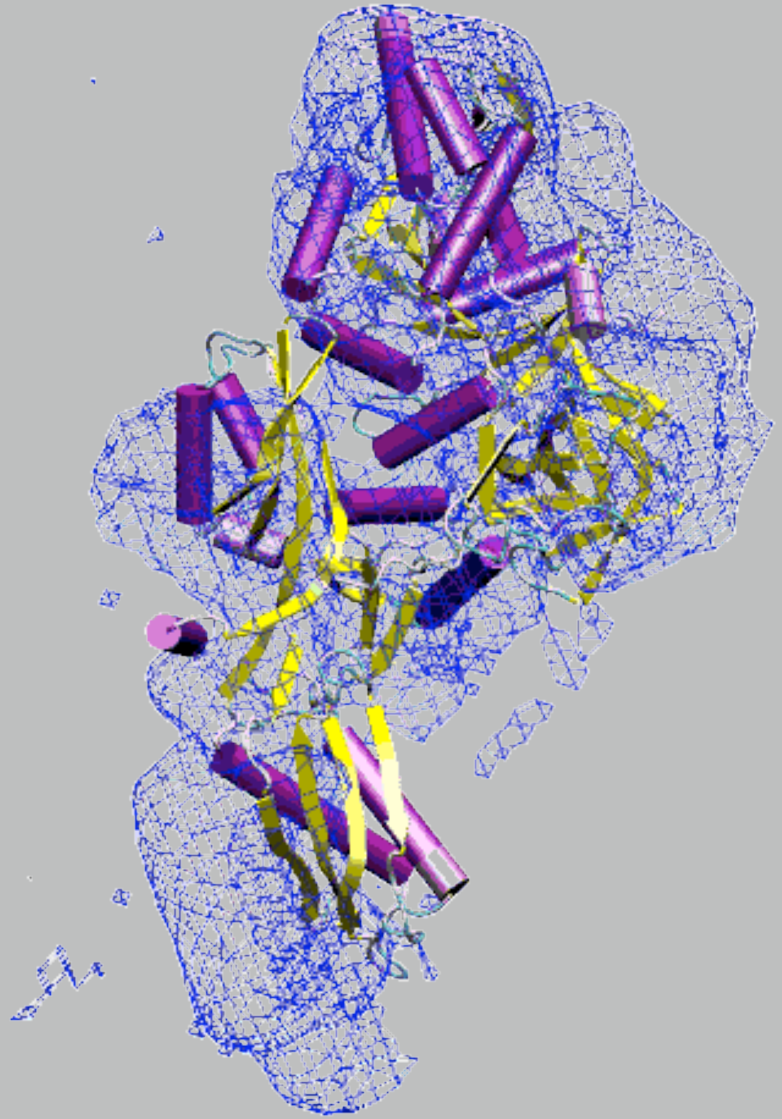
Flexible refinement of atomic structures into low-resolution EM maps using elastic network normal modes



Resolution = 10.8 Å

Data from M. Valle *et al*.
Cell (2003)

Elongation Factor G bound
to the ribosome

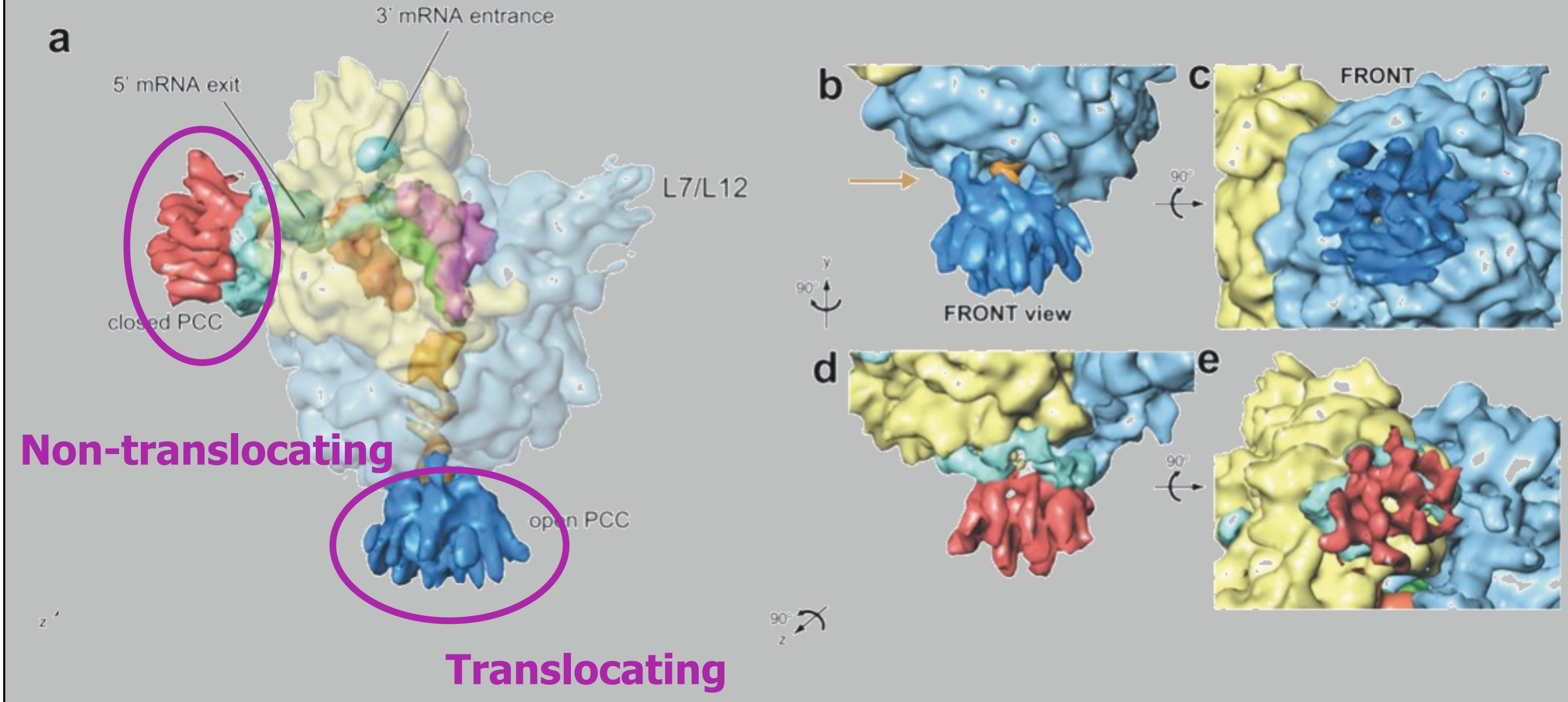


F. Tama *et al.*, J. Struct. Biol. (2004)

Examples of NMFF refinement in model building and interpretation of structural data

E-coli protein conducting channel bound to a translating ribosome

Structure at ~ 12 Å resolution

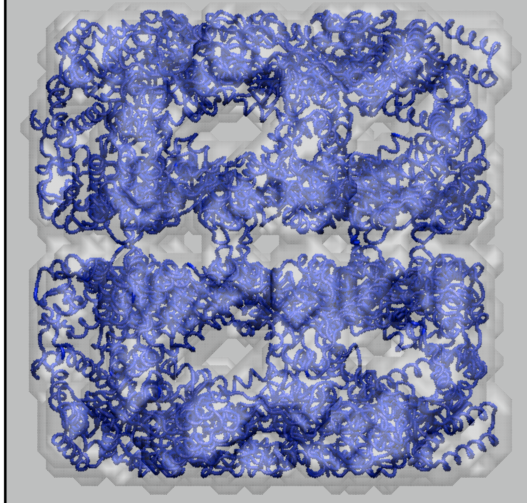


Dimeric structures

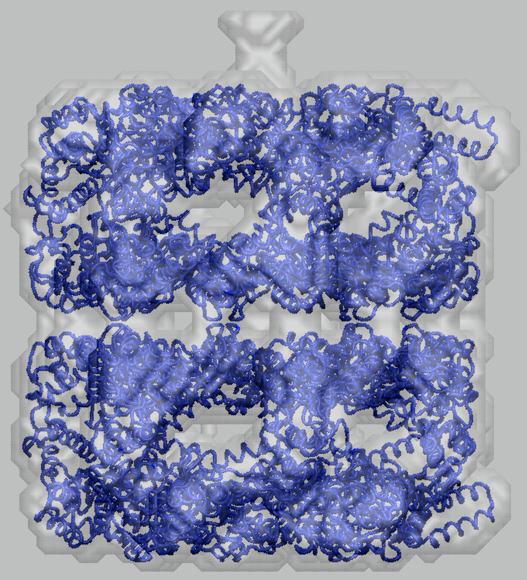
K. Mitra, C. Scaffitzel, T. Shaikh, F. Tama, S. Jenni, CL. Brooks III, N. Ban and J. Frank. **Nature** (2005)

13 Å structure of a chaperonin GroEL-protein substrate complex

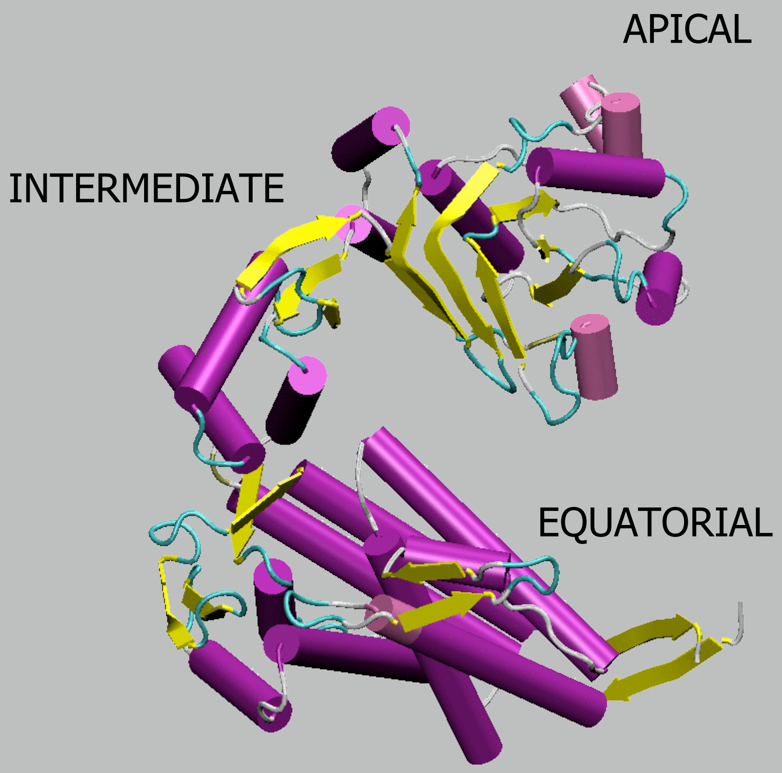
GroEL - substrate
free (map at 12Å)



GroEL with bound E coli
glutamine synthetase (GS)
(map 13Å)



Conformational difference

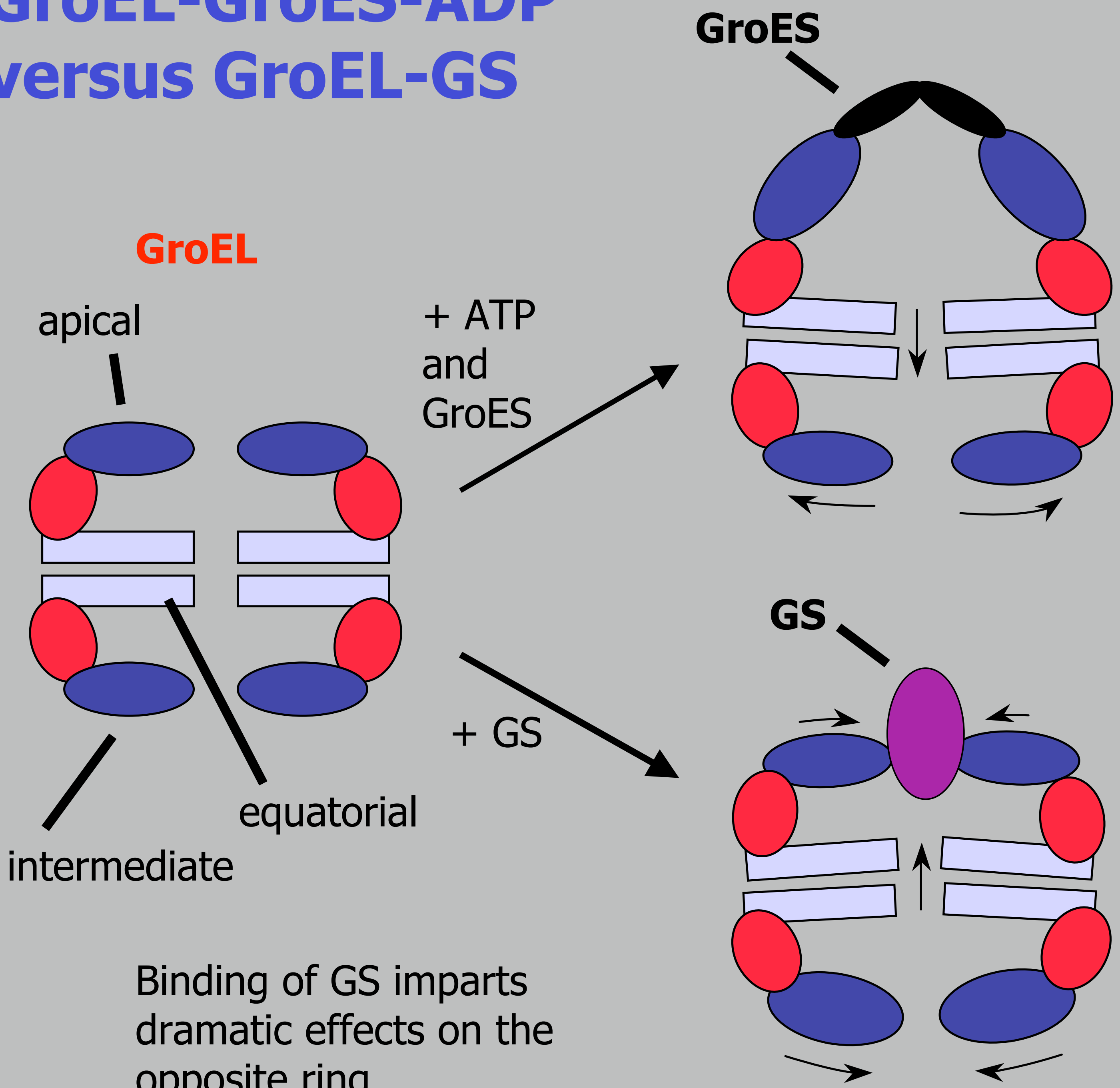


Fitting performed from 1OEL using NMFF
with 7-fold symmetry imposed.

Some residues move by ~ 7Å in the apical domain

S. Falke, F. Tama, CL. Brooks III, EP. Gogol and MT. Fisher. **J. Mol. Biol.** (2005)

GroEL-GroES-ADP versus GroEL-GS



GroEL-GroES-ADP

- equatorial domains maintain contact => the movements in cis lead to movements of the opposite trans apical domain.

GroEL-GS

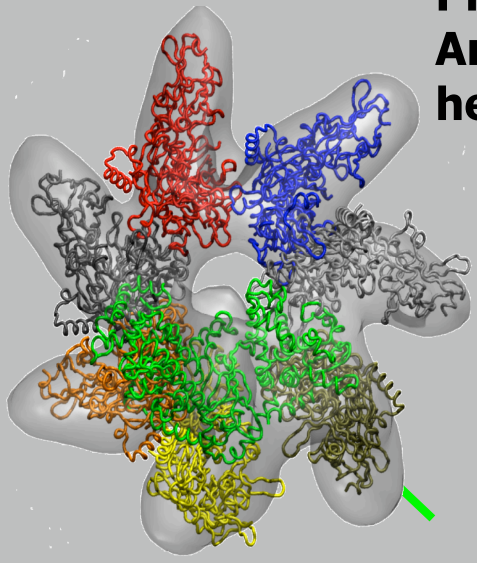
- equatorial domains maintain contact
- movements opposite those observed with GroEL-ES

Binding of GS imparts dramatic effects on the opposite ring

S. Falke, F. Tama, CL. Brooks III, EP. Gogol and MT. Fisher. **J. Mol. Biol.** (2005)

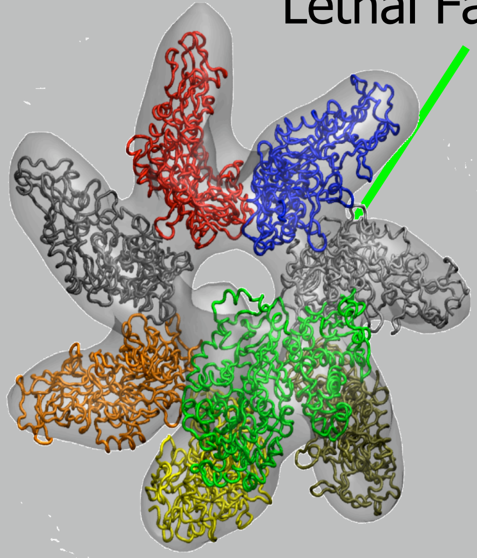
Anthrax toxic complex

Initial fit

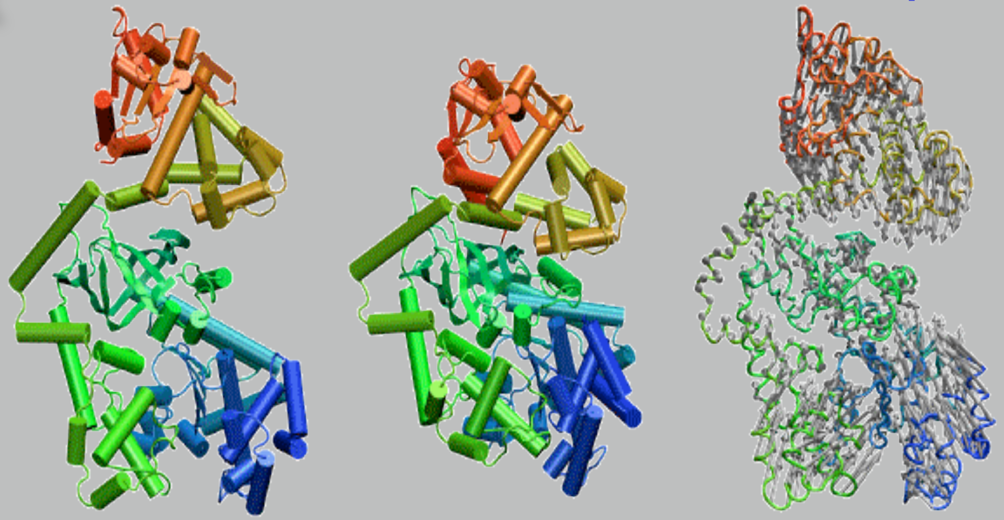


Protective Antigen heptamer

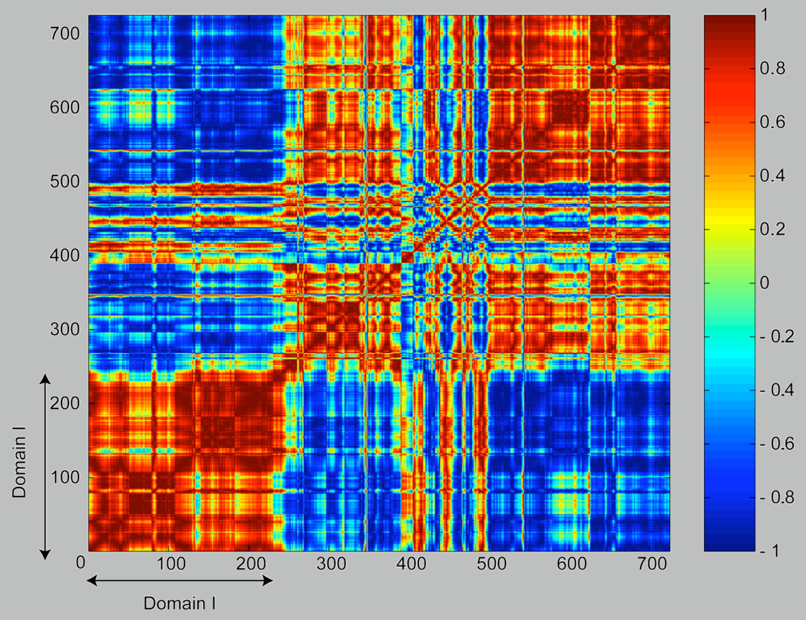
Lethal Factor: LF



Structure obtained with NMFF

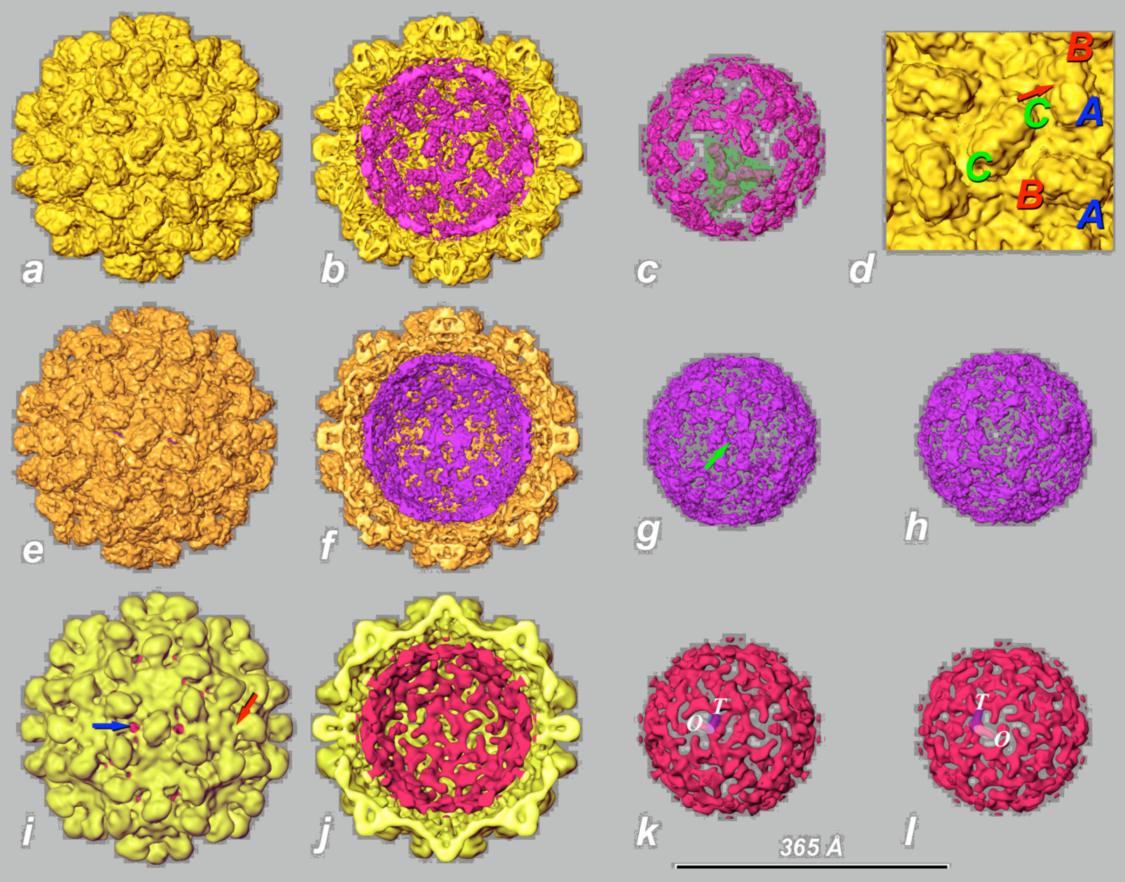


LF becomes smaller which may facilitate its translocation through the lumen



F. Tama, G. Ren, S.H. Leppla, CL. Brooks III and A.K. Mitra
to be submitted

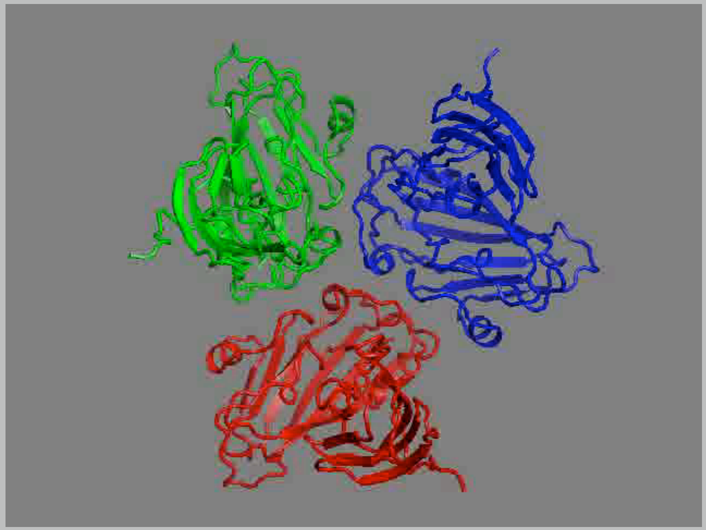
Red Clover Necrotic Mosaic Virus



Homology model

Fitting

Native to EDTA treated virions

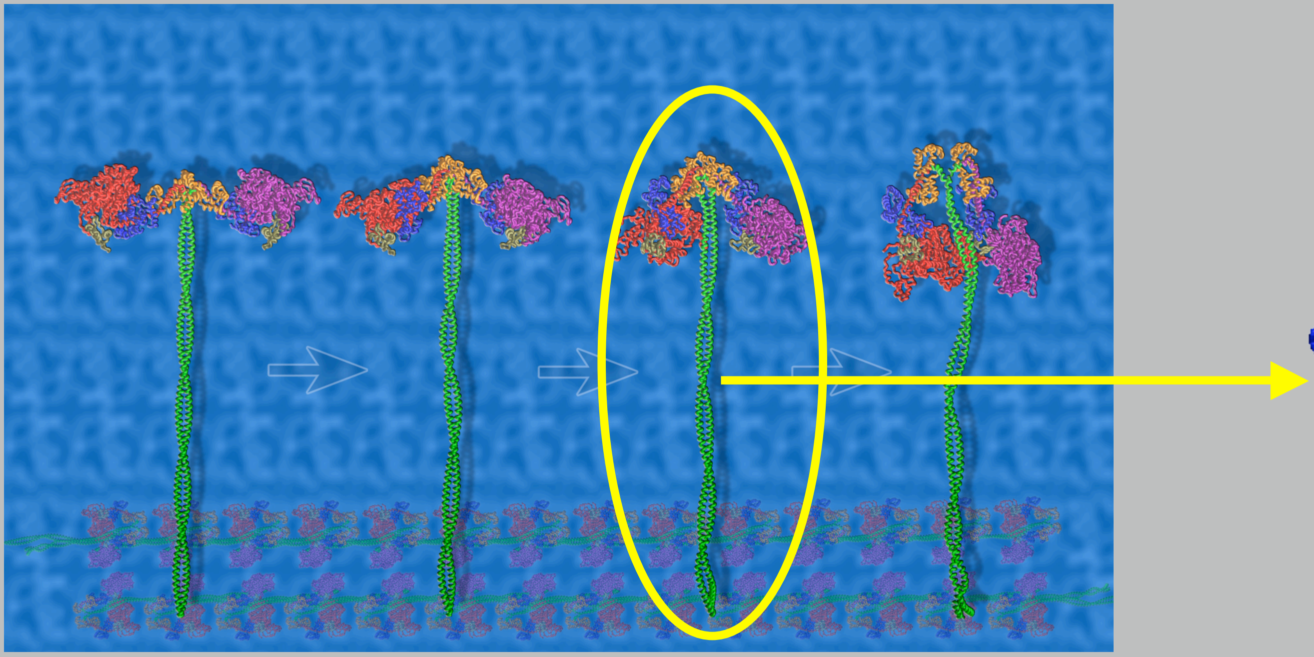


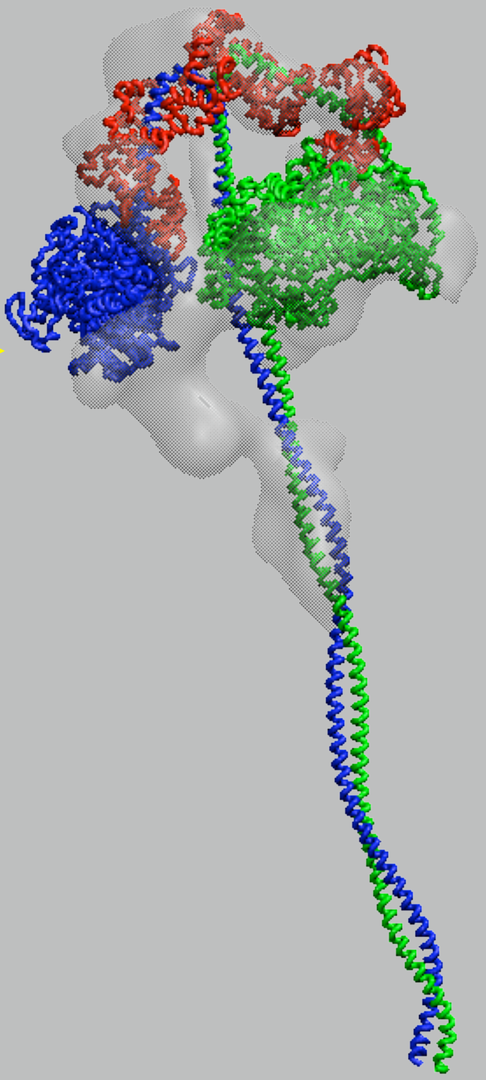
Extraction of ions alters the conformation of the capsid that generates channel through which the genomic RNA is likely to be released

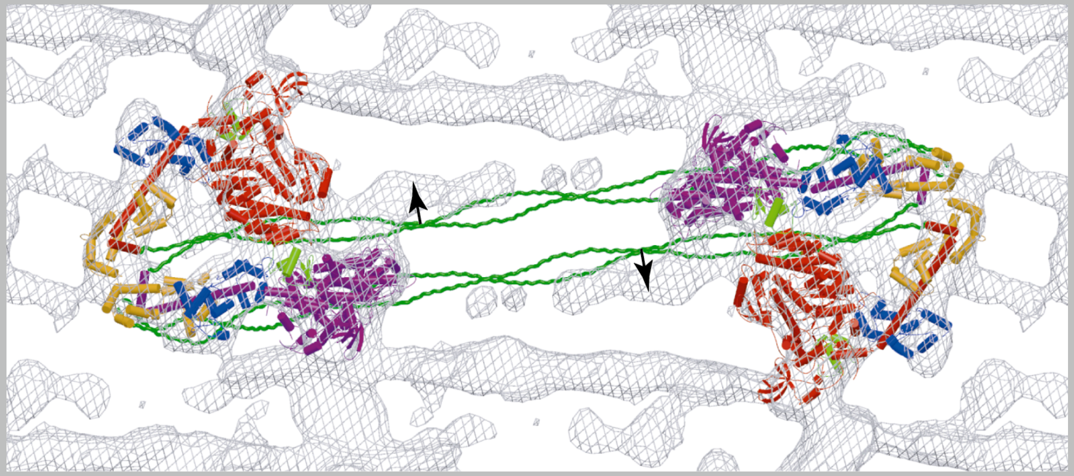
M. B. Sherman, R. H. Guenther, F. Tama, C. L. Brooks, A. M. Mikhailov, E. V. Orlova, T. S. Baker, and S. A. Lommel. *Submitted* to Mol Cell.

Refinement of HMM

Transition between active and inhibited myosin







F. Tama, M. Feig, J. Liu, CL. Brooks III, KA. Taylor J. Mol. Biol. (2005)

Flexible Fitting - Summary

- Uses small number of collective (functionally relevant) independent coordinates to optimize cc
  - Minimizes problems of over-fitting
- Can be used at multiple levels of coarse-graining for optimal model to accommodate date
  - Multi-resolution through RTB as well as pseudo-atomic elastic networks
- Employs symmetry to permit symmetric assemblies to be modeled from asymmetric unit
- Free and available at:

http://mmtsb.scripps.edu/nmff.html

Summary

- Elastic network normal mode analysis provides a multi-resolution approach for exploring functional reorganization of biological assemblies
  - Nature exploits the overall all shape of her biological machines to provide robustness in functional reorganization
- NMFF can be used in conjunction with known atomic level structures and lower resolution data to explore functional rearrangements of biological assemblies as observed by cryo-EM and related low resolution methods

Acknowledgments

Florence Tama (U of AZ), Michael Feig (MSU), Osamu Miyashita (U of AZ), Jack Johnson, Vijay Reddy, Joachim Frank (HHMI), Alok Mitra (Auckland), Mark Fisher (KU), Ed Gogel (UM-KC), Tim Baker (UCSD)