Structure Refinement with Cryo-EM Density Maps

Gunnar F. Schröder
Computational Structural Biology Group

Institute of Complex Systems - Structural Biochemistry (ICS-6)

Forschungszentrum Jülich

and

Heinrich-Heine Universität Düsseldorf

JÜLICH
FORSCHUNGSZENTRUM

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

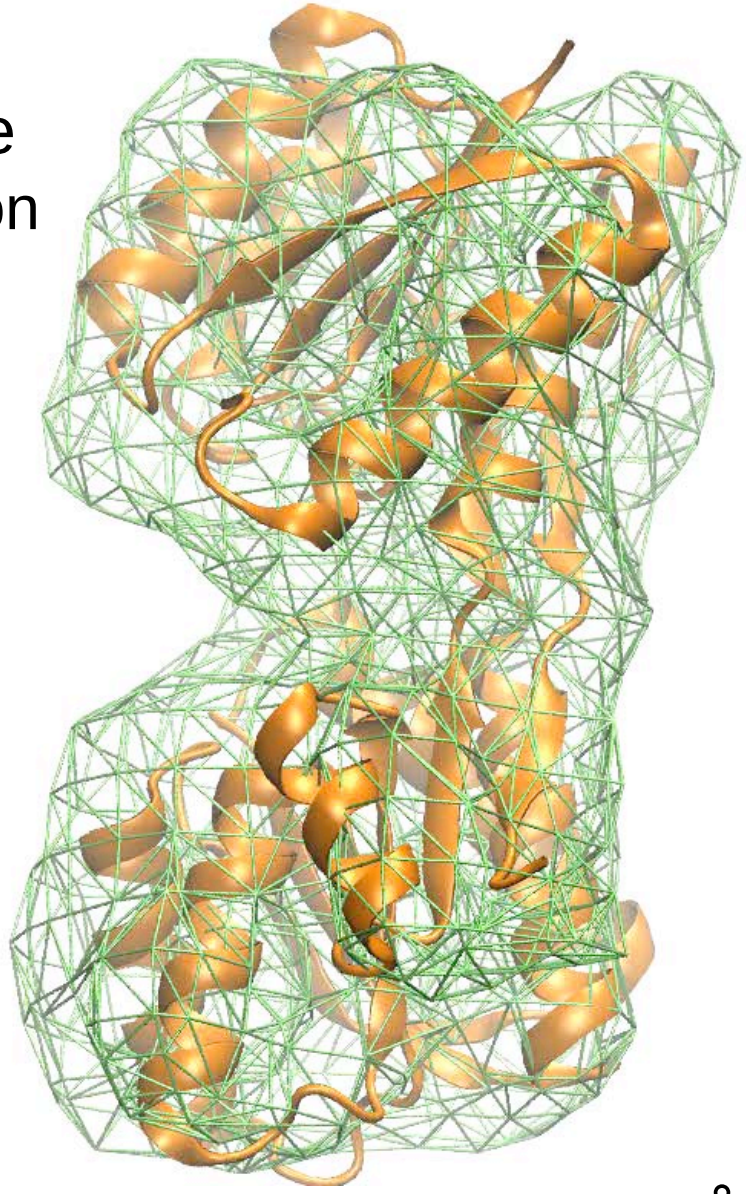
Structure Refinement at Low-resolution

JÜLICH FORSCHUNGSZENTRUM

Assume: We have a starting structure (crystal structure in different conformation or good homology model)

Standard refinement yields a bad structure

How to make use of prior structural information during the refinement?



10 Å

The required X-ray resolution (determinacy point) depends on the number of degrees of freedom and the solvent fraction

Degrees of Freedom & N/N_{res}		***S* (Solvent Volume Fraction)**		
		0.5	**0.6**	**0.7**
All Atoms with H atoms	**48**	2.3 Å	2.5 Å	2.8 Å
All Atoms no H atoms	**24**	2.9 Å	3.2 Å	3.5 Å
All (Φ,Ψ,χ) Torsions	**4**	5.3 Å	5.8 Å	6.3 Å
All (Φ,Ψ) Torsions	**2**	6.7 Å	7.3 Å	8.0 Å
All (α) Torsions	**1**	8.5 Å	9.13 Å	10.1 Å

Deformable Elastic Network (DEN)

JÜLICH FORSCHUNGSZENTRUM

Refine only those degrees of freedom that need to be refined to fit the data, but not more.

Find only the relevant degrees of freedom for which the data actually provide information

Schröder, Brunger & Levitt, *Structure* (2007) **15**:1630

Schröder, Levitt & Brunger, *Nature* (2010) **464**:1218-1222

Deformable Elastic Network

- randomly chosen distance restraints
- deformable distance restraints
- nothing to do with normal modes

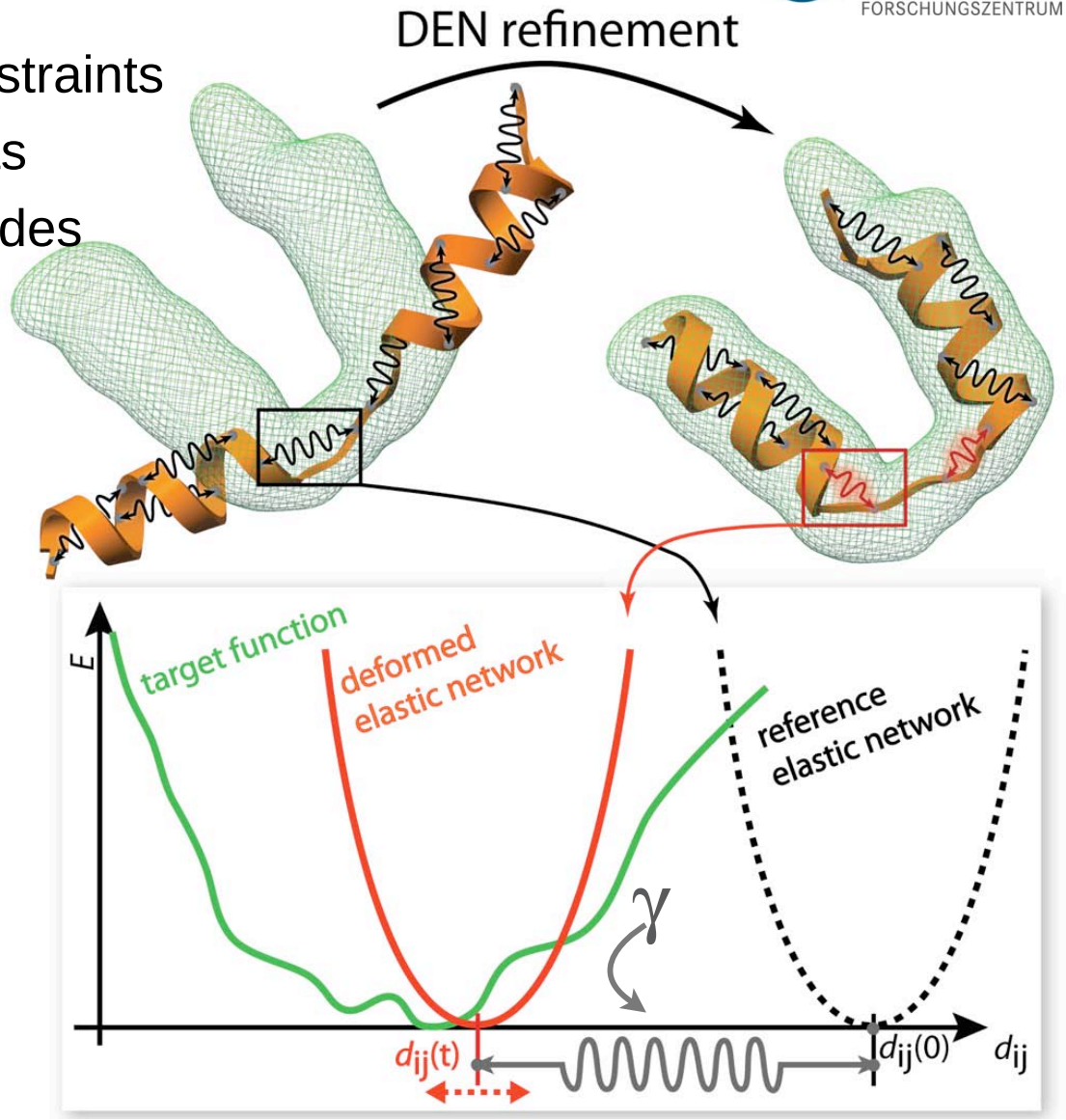
The weight

w_{DEN}

and the

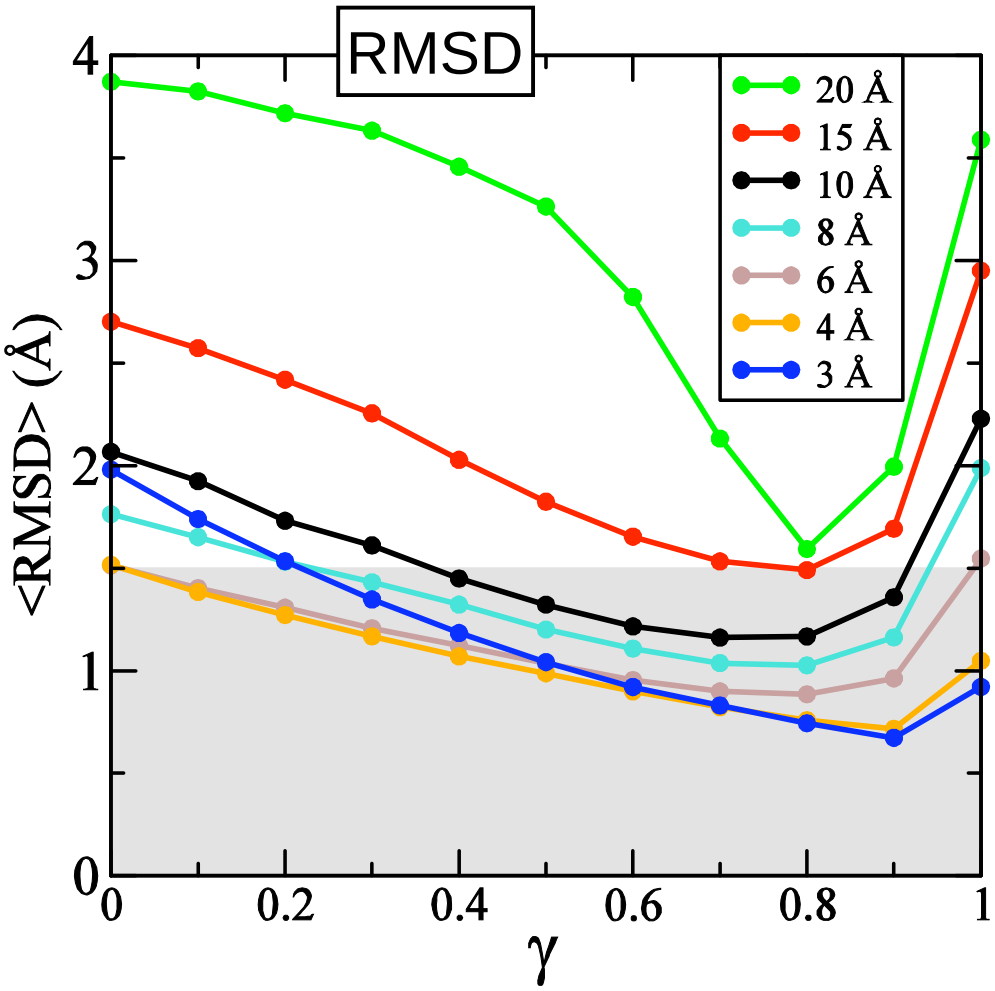
γ-parameter

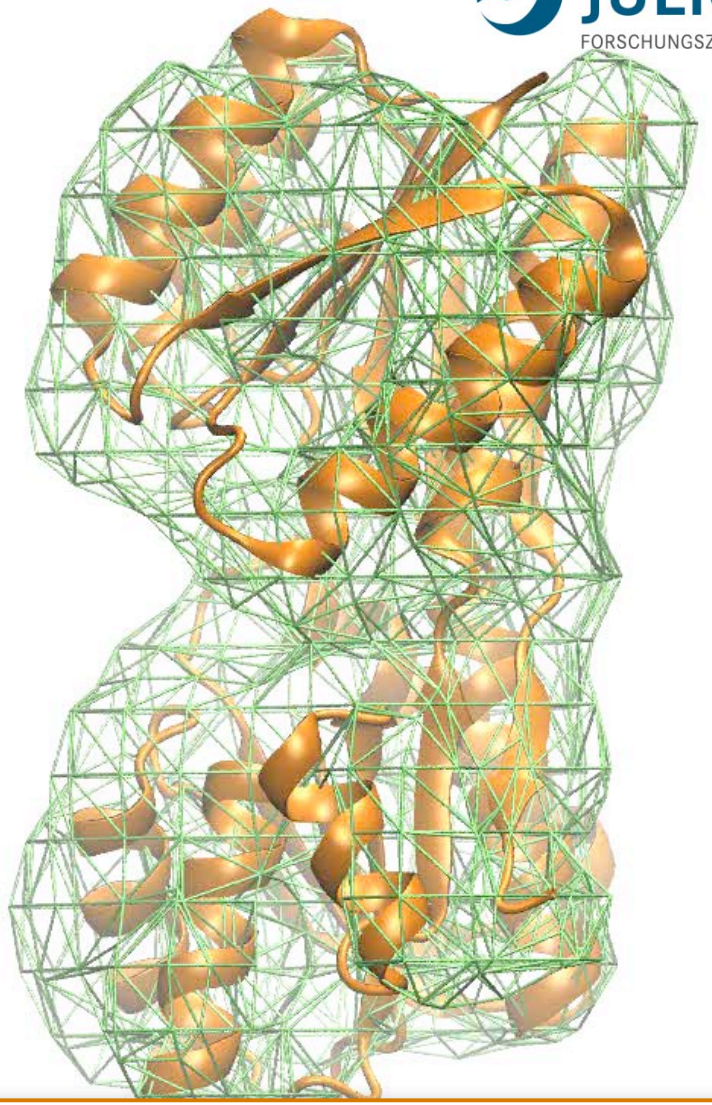
control the flexibility of the restraints



$$E_{Target} = E_{Xray} + E_{Chem} + w_{DEN}E_{DEN}$$

Effect of the γ-parameter





$\gamma = 0$ **increasing deformability** $\gamma = 1$

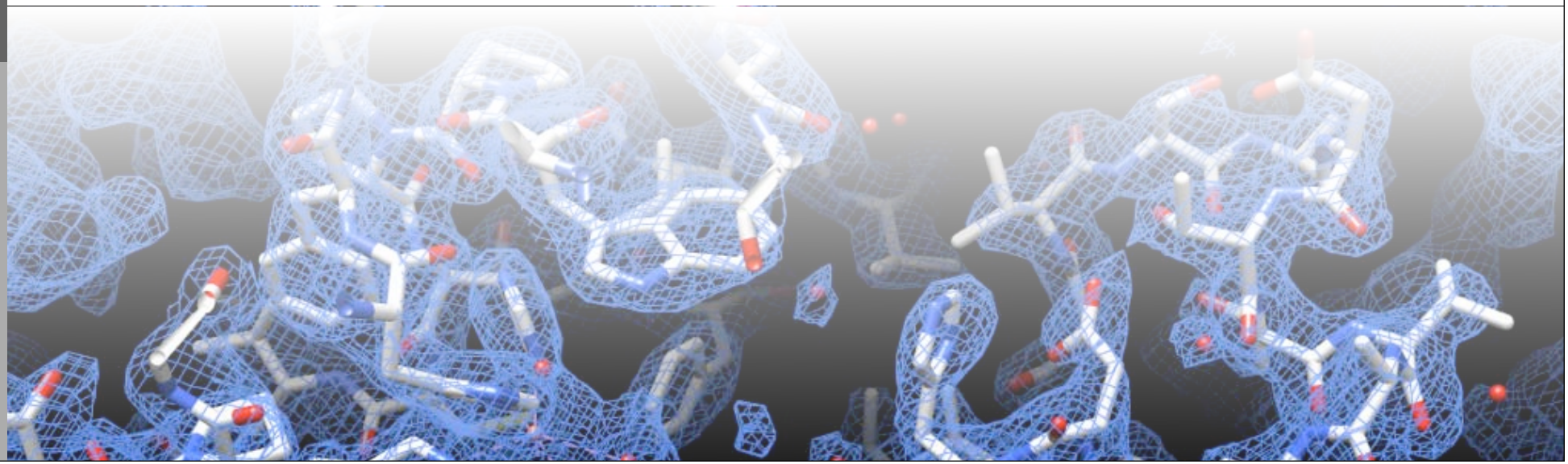
Reference model **Experimental restraints**

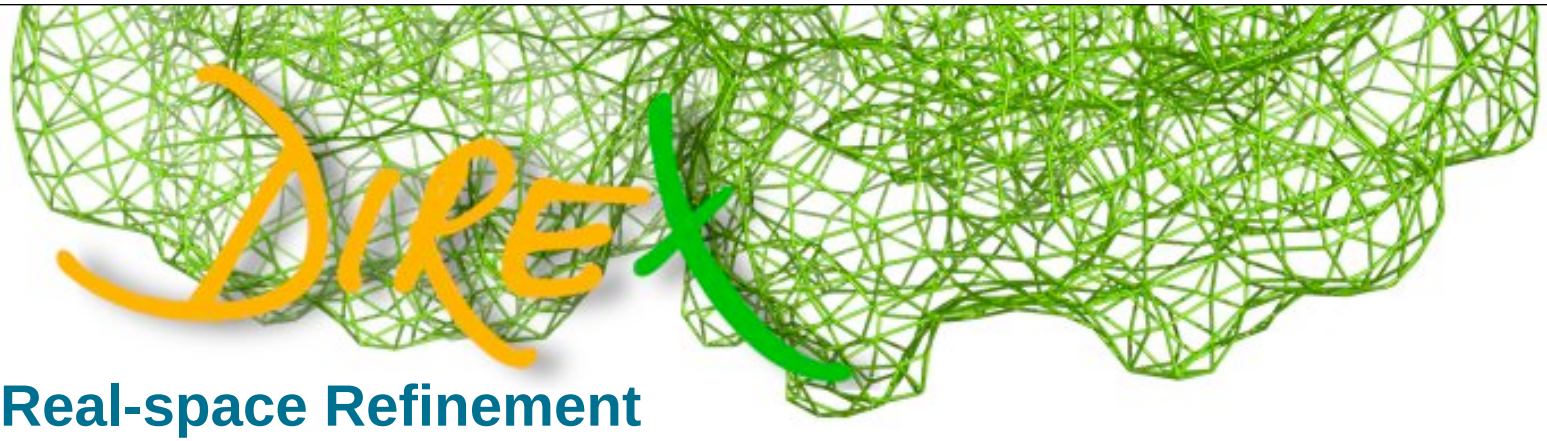
Application of DEN to Reciprocal-space Structure Refinement using X-ray Diffraction Data

JÜLICH FORSCHUNGSZENTRUM

DEN method implemented in

CNS (v1.3)
and
Phenix (v.1.8)





JÜLICH
FORSCHUNGSZENTRUM

Real-space Refinement

- Efficient geometry-based conformational sampling
- Cross-correlation coefficient between model and target density map is optimized
- Cross-validation
- DEN
- Symm
- Distar
- Positi
- Accurate modeling of electron scattering
- Bulk solvent model
- Overall B-factor optimization

- no normal modes !!

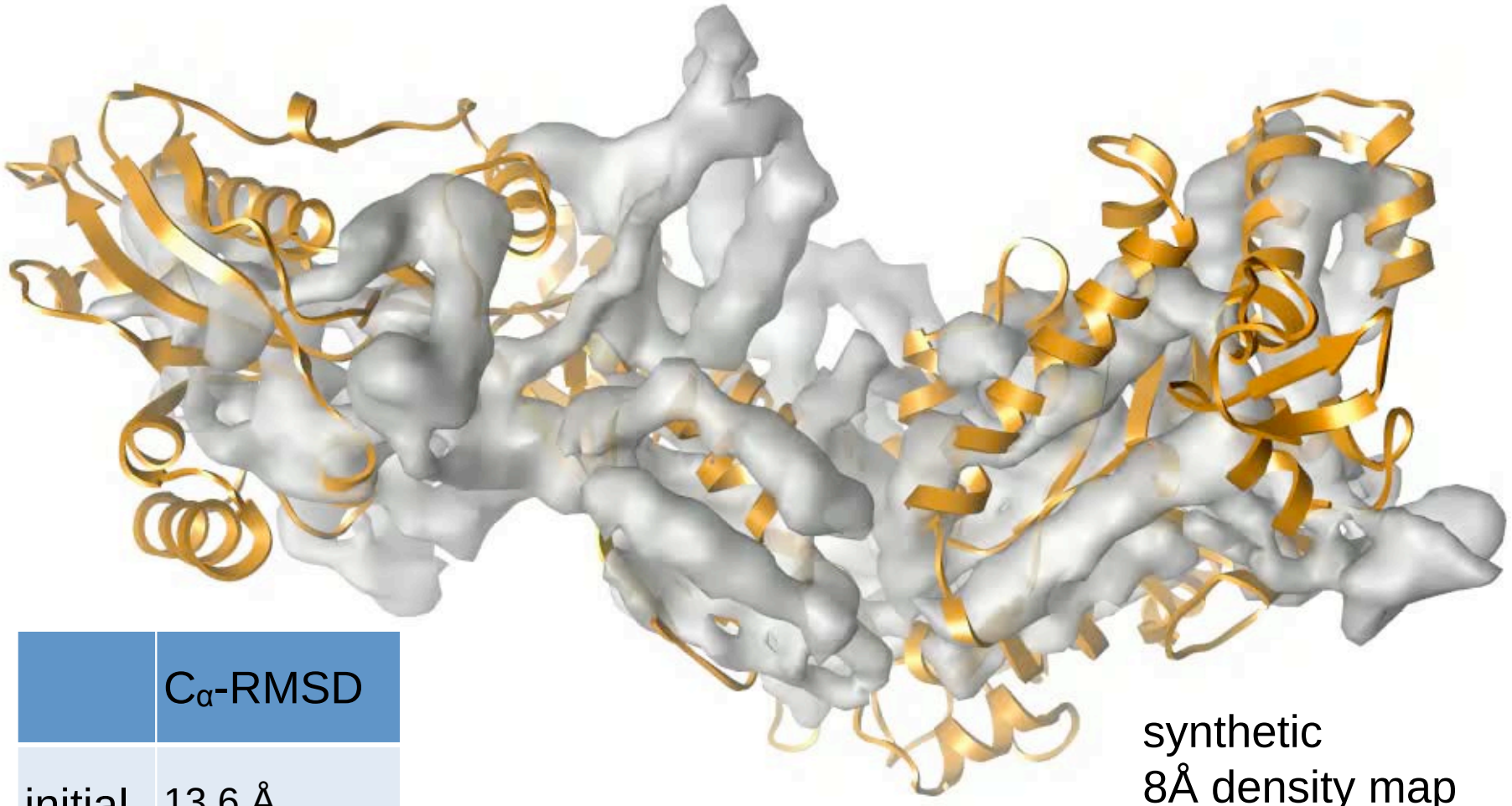
- no coarse-graining (although possible)

http://simtk.org/home/direx/

http://www.schroderlab.org/software/direx/

DireX

Example: Elongation Factor 2 (EF-2)



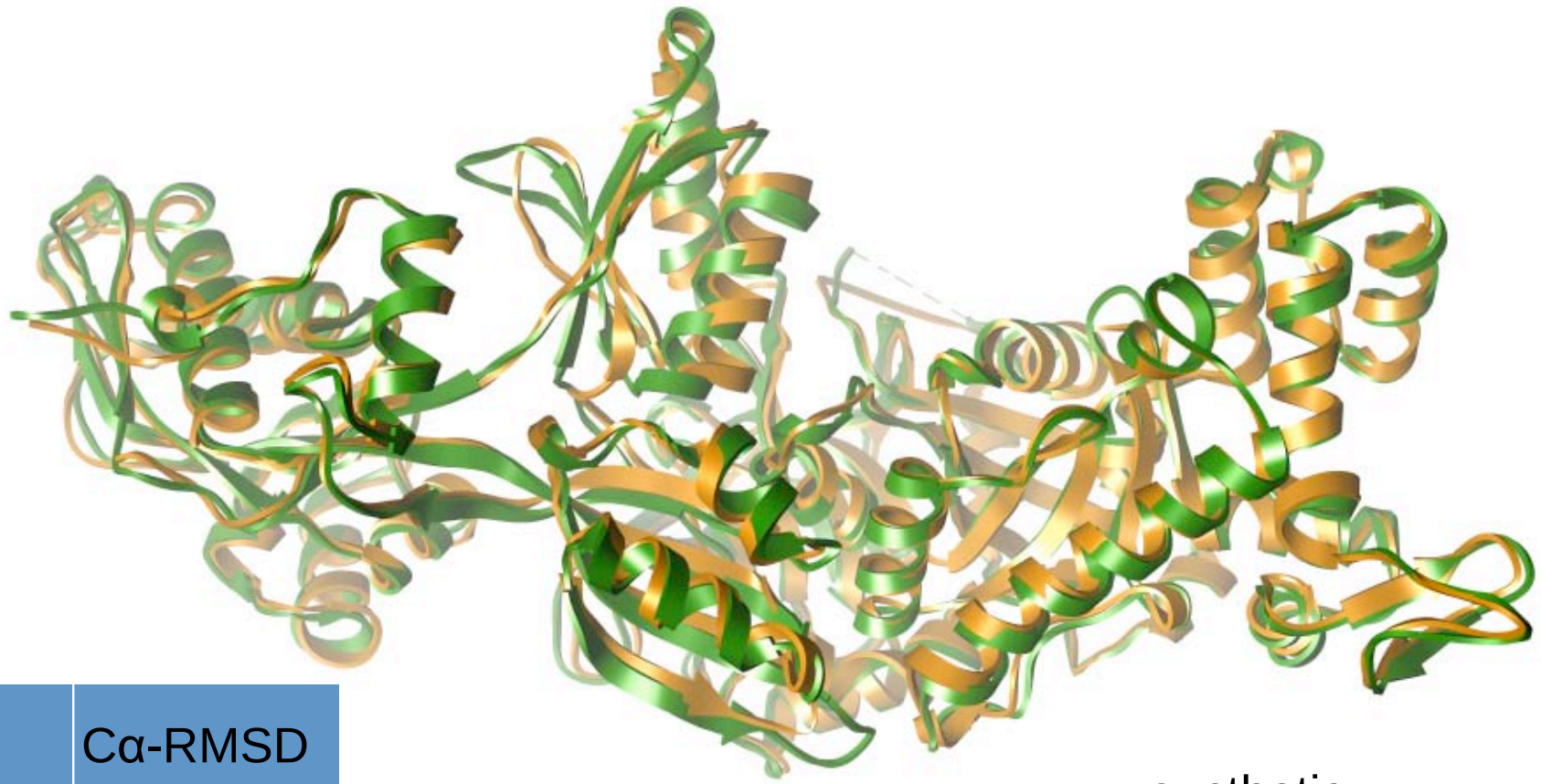
synthetic
8Å density map

	C_α-RMSD
initial	13.6 Å
final	0.8 Å

6300 atoms, 14 steps/min
(3.5 hrs for 3000 steps)

DireX

Example: Elongation Factor 2 (EF-2)



	Cα-RMSD
initial	13.6 Å
final	0.8 Å

synthetic
8Å density map

6300 atoms, 14 steps/min
(3.5 hrs for 3000 steps)

JÜLICH
FORSCHUNGSZENTRUM

There are 3 main forces on atoms in DireX:

1) Concoord restraints: maintain correct stereochemistry (bond lengths, planarity, etc.) and prevent atom overlaps

2) DEN restraints: control the deviation from a reference model

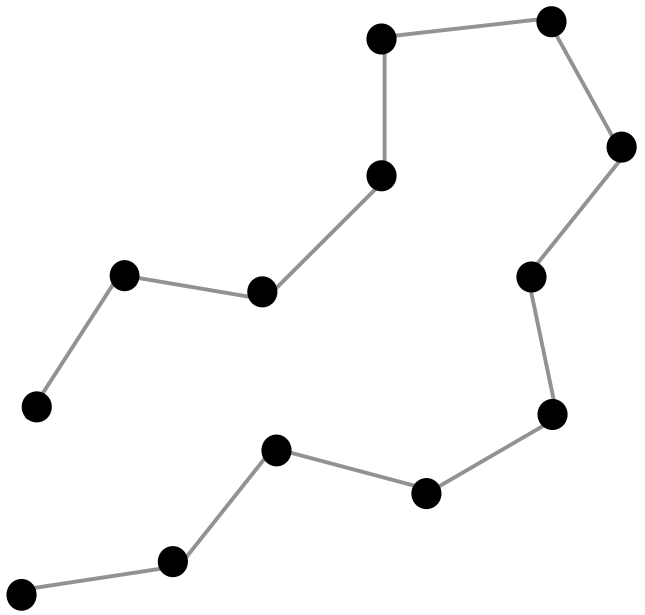
3) Density restraints: fit model into density

DireX:
Geometry-based conformational sampling

based on CONCOORD

B.L. de Groot, et al. Proteins 29: 240-251 (1997)

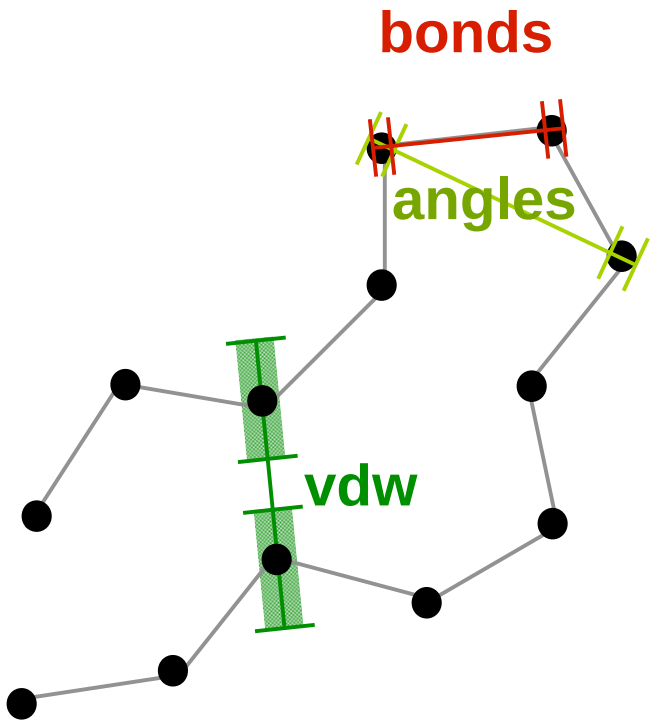
1. Initial model



DireX:
Geometry-based conformational sampling

based on CONCOORD

B.L. de Groot, et al. Proteins 29: 240-251 (1997)



1. Initial model

2. Generate list of
   distance restraints (intervals)

DireX:
Geometry-based conformational sampling

based on CONCOORD

B.L. de Groot, et al. Proteins 29: 240-251 (1997)

1. Initial model

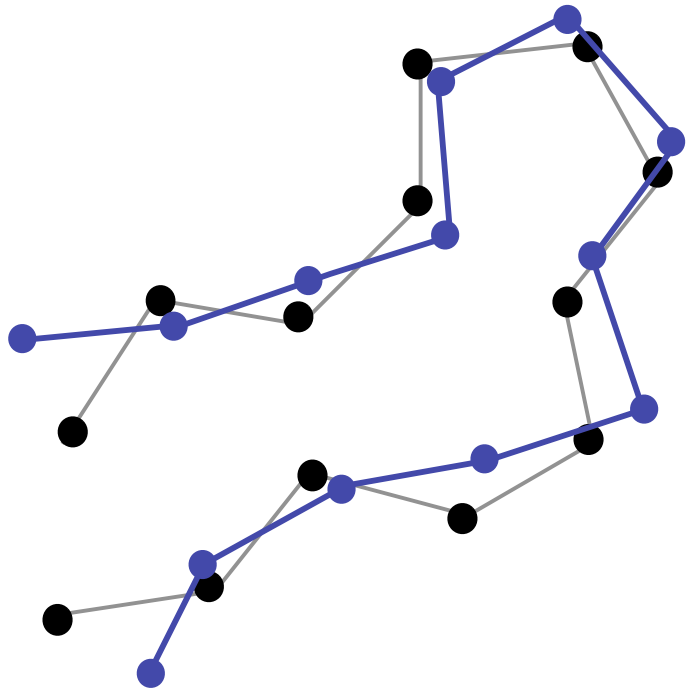
2. Generate list of
   distance restraints (intervals)

3. Perturb coordinates

DireX:
Geometry-based conformational sampling

based on CONCOORD

B.L. de Groot, et al. Proteins 29: 240-251 (1997)



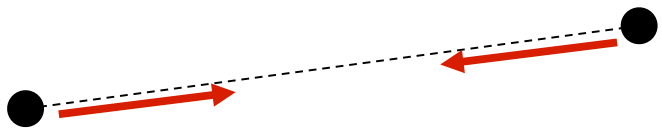
1. Initial model

2. Generate list of
   distance restraints (intervals)

3. Perturb coordinates

4. use CONCOORD algorithm to obtain
   a new structure which also obeys
   all distance restraints

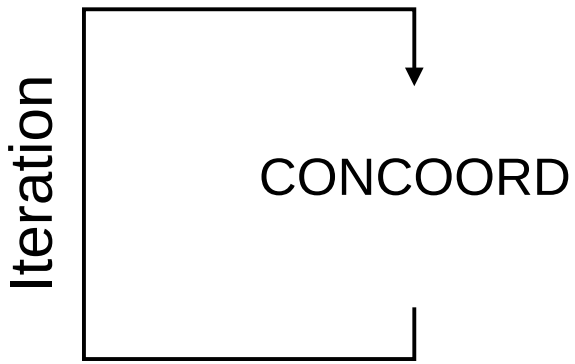
CONCOORD: correct distances
iteratively in a random order

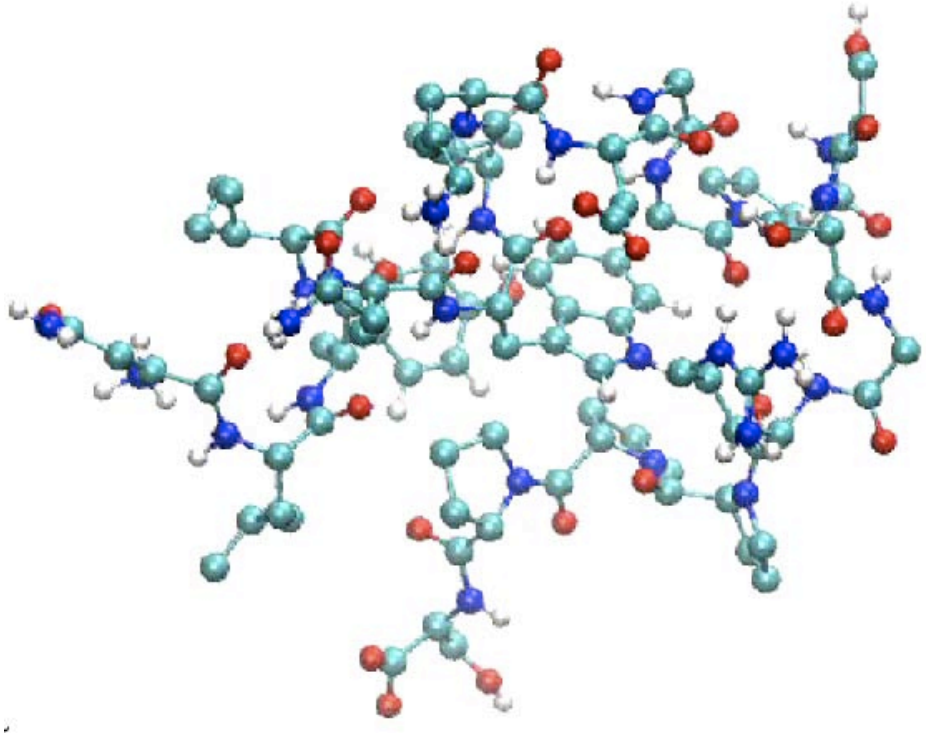


DireX:
Geometry-based conformational sampling

based on CONCOORD

B.L. de Groot, et al. Proteins 29: 240-251 (1997)



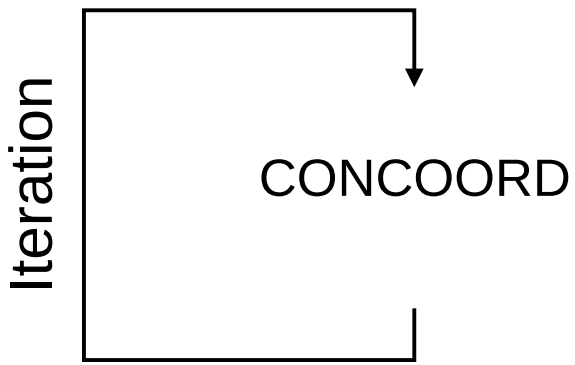


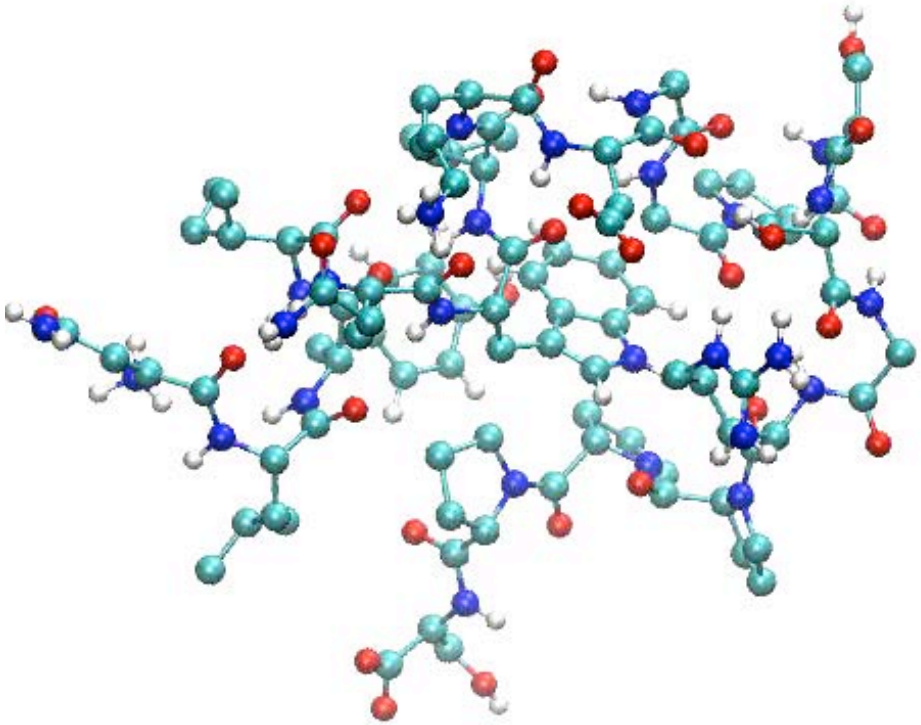
Random walk through conformational space
while maintaining correct stereochemistry
and avoiding atom clashes

DireX:
Geometry-based conformational sampling

based on CONCOORD

B.L. de Groot, et al. Proteins 29: 240-251 (1997)



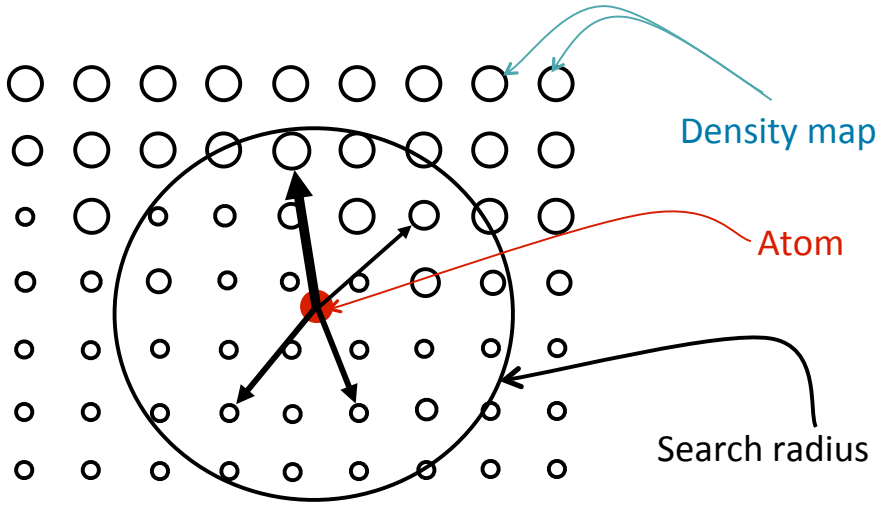


Random walk through conformational space
while maintaining correct stereochemistry
and avoiding atom clashes

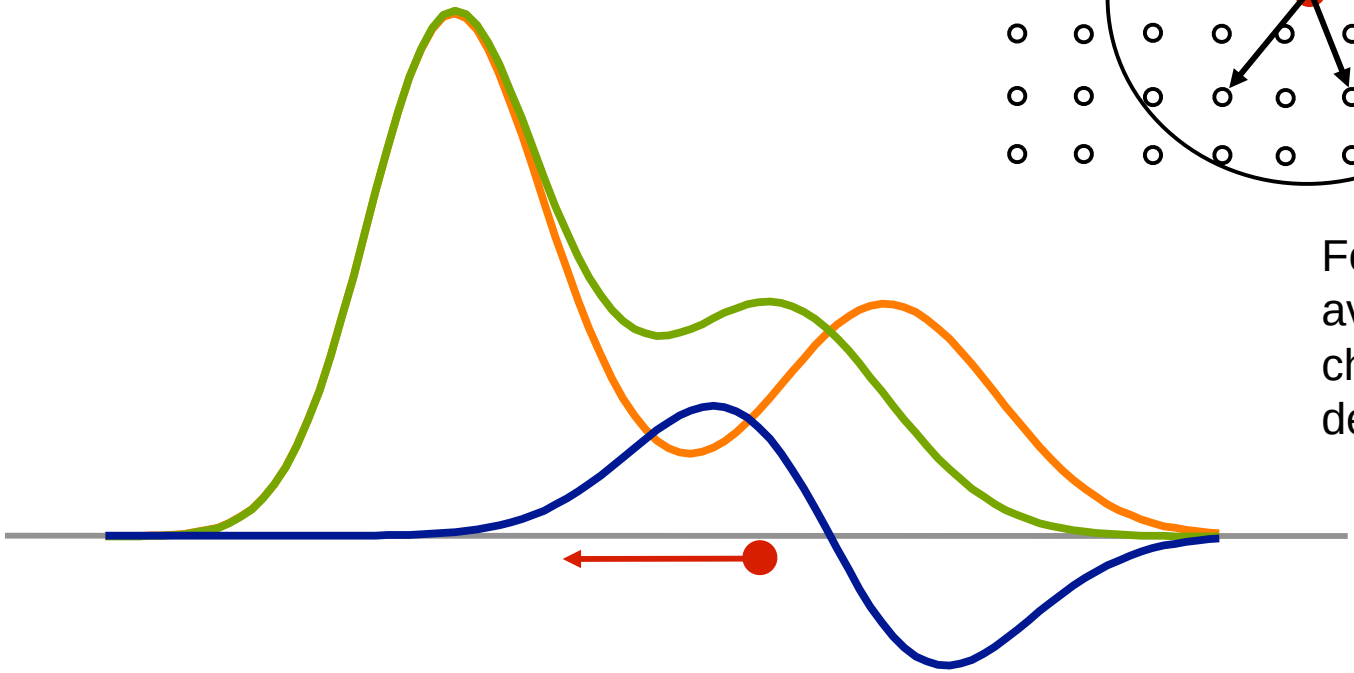
DireX: Forces derived from a density map

$$\rho_{difference}(x) = \rho_{target}(x) - \rho_{model}(x)$$

Stochastic gradient to move atoms into high difference-density regions



Target density
Model density
Difference



For each atom:
average over 10 randomly chosen vectors weighted by density difference

JÜLICH
FORSCHUNGSZENTRUM

Difference between MDFF and DireX

Low-resolution data are (obviously) missing high-resolution information.

The difference is where this missing information comes from:

MDFF uses MD force field, i.e. predicts missing information

DireX takes missing information from crystal structure (or other reference model)

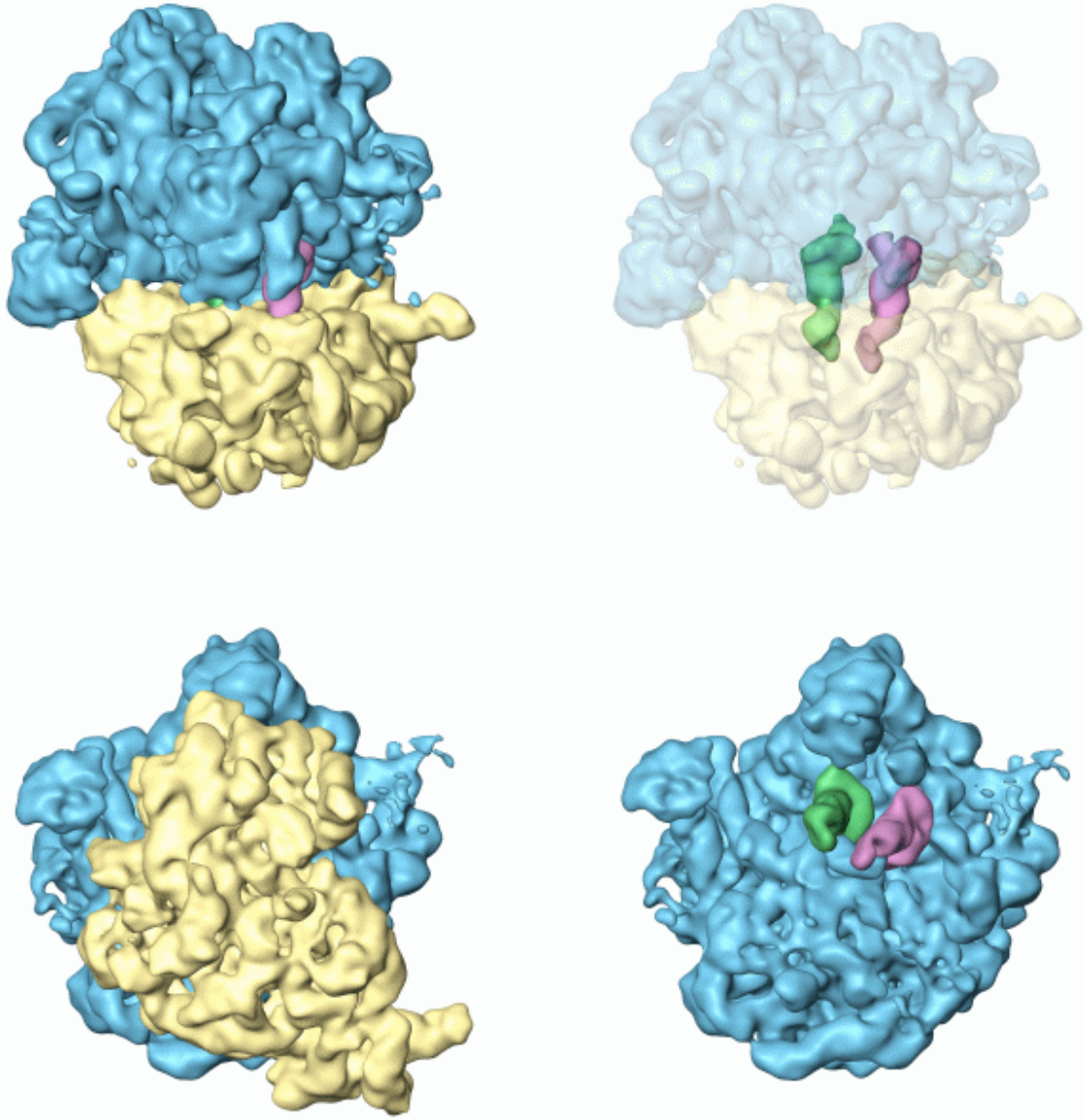
Ribosome

tRNA translocation

2 million single-particle images sorted into 50 conformational substates

Resolution 8 - 15 Å

In collaboration with Holger Stark's lab (MPI Biophysical Chemistry, Göttingen)

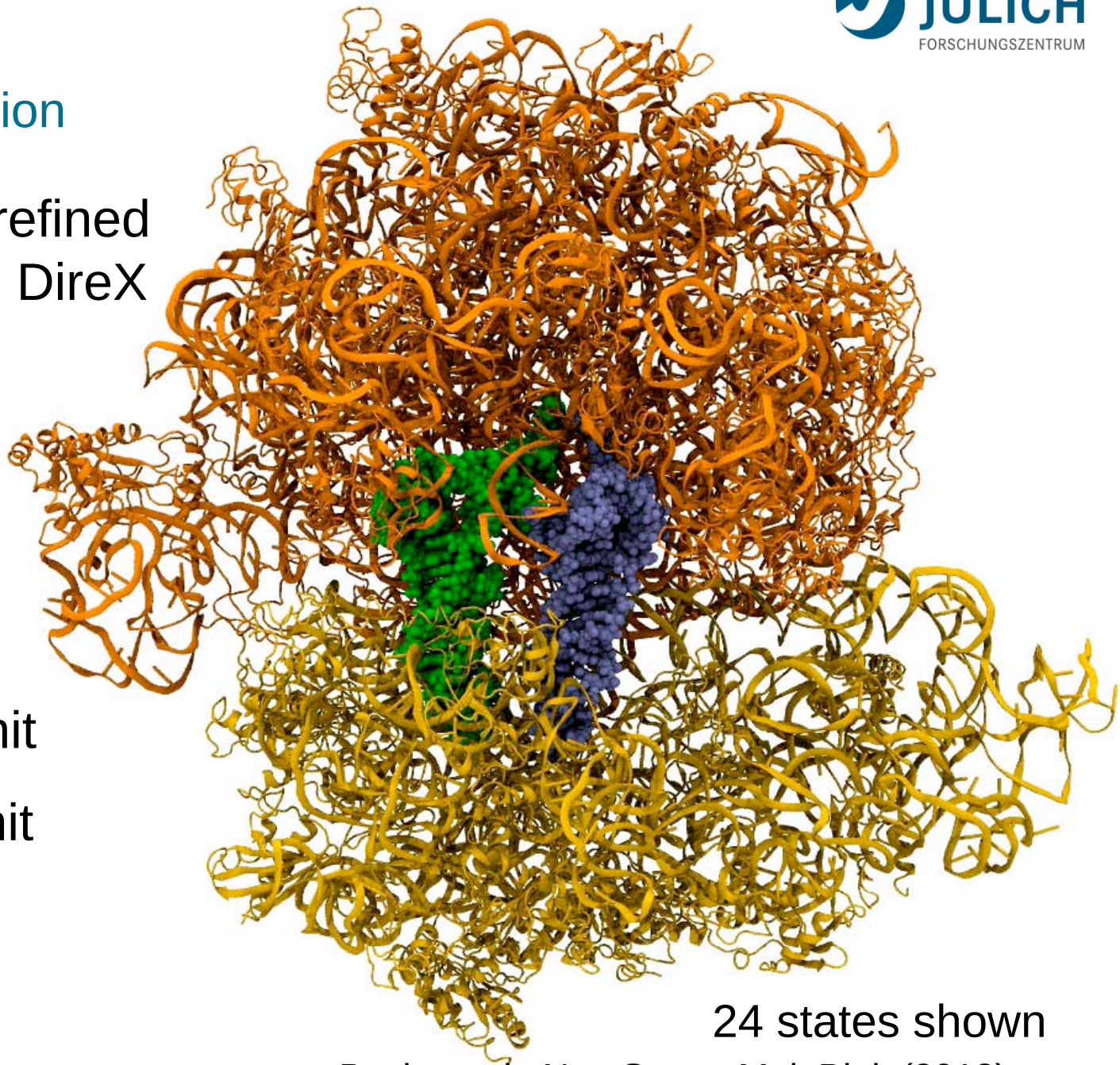


Fischer, Konevega, Wintermeyer, Rodnina & Stark (2010) Nature **466**: 329-333

Ribosome

tRNA translocation

Each state was refined separately using DireX



- Large subunit
- Small subunit
- tRNA
- tRNA

24 states shown

Bock, et al., Nat. Struct. Mol. Biol. (2013)

MD Simulation of Conformational Transitions

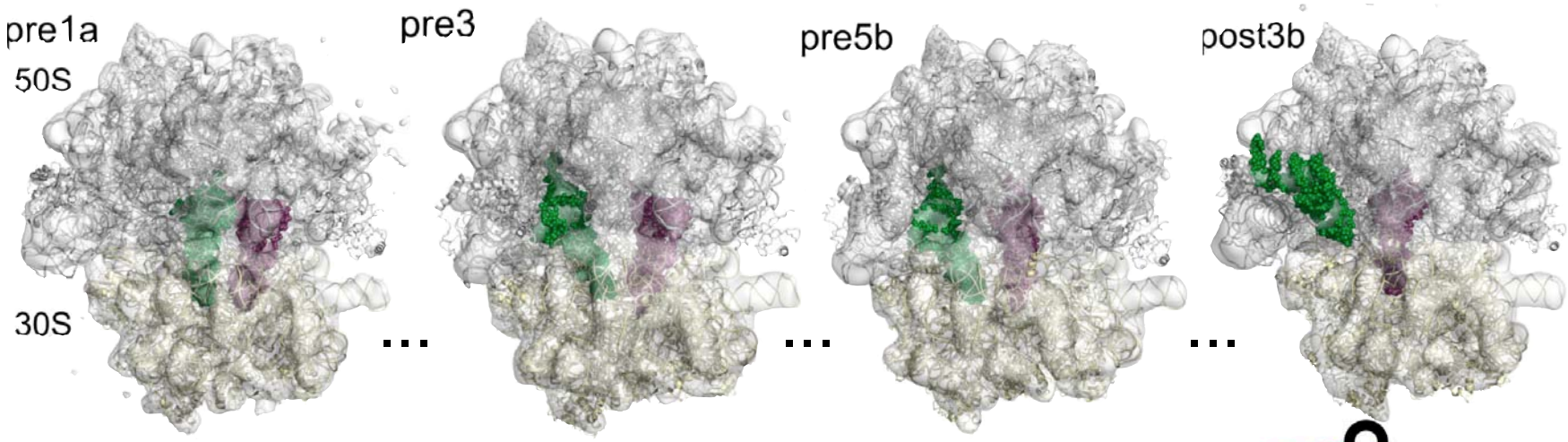
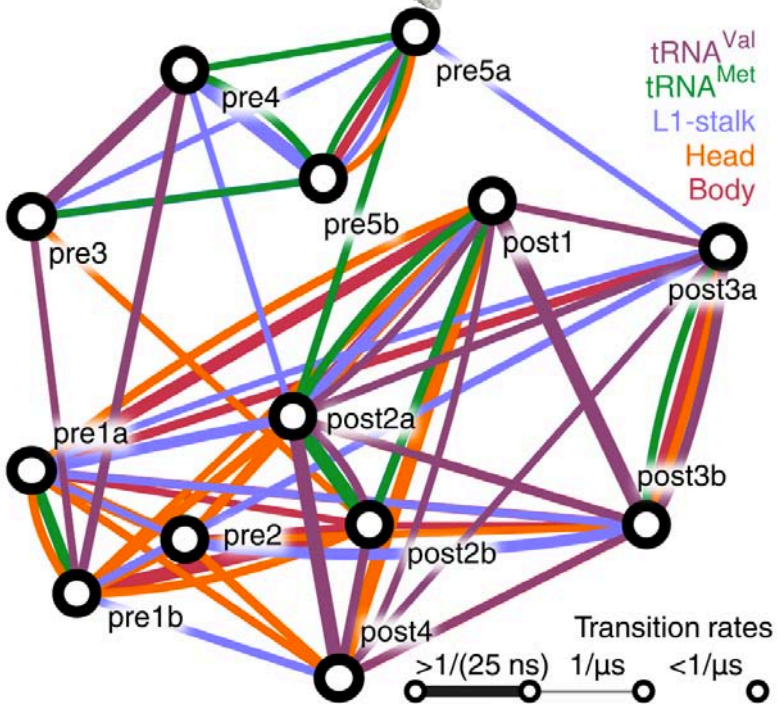


Image classification can yield conformational states that are sufficiently similar to describe the transitions between these states by molecular dynamics simulations



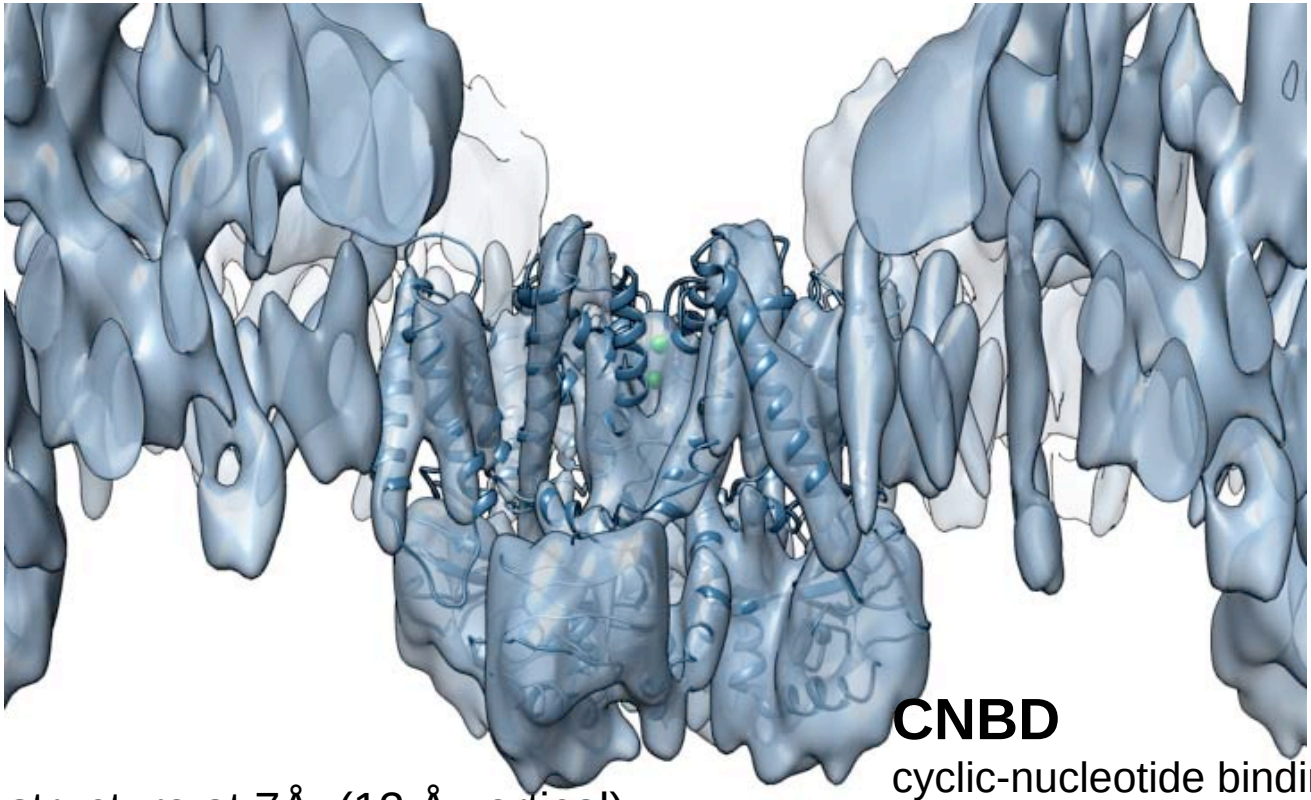
In collaboration with Holger Stark and Helmut Grubmüller (MPI Biophysical Chemistry, Göttingen)

Bock, et al., Nat. Struct. Mol. Biol. (2013)

Ion Channel Gating

The cyclic-nucleotide activated Mlok1 K^+ channel

The channel opens upon binding of cAMP to the CNBD



outside

inside

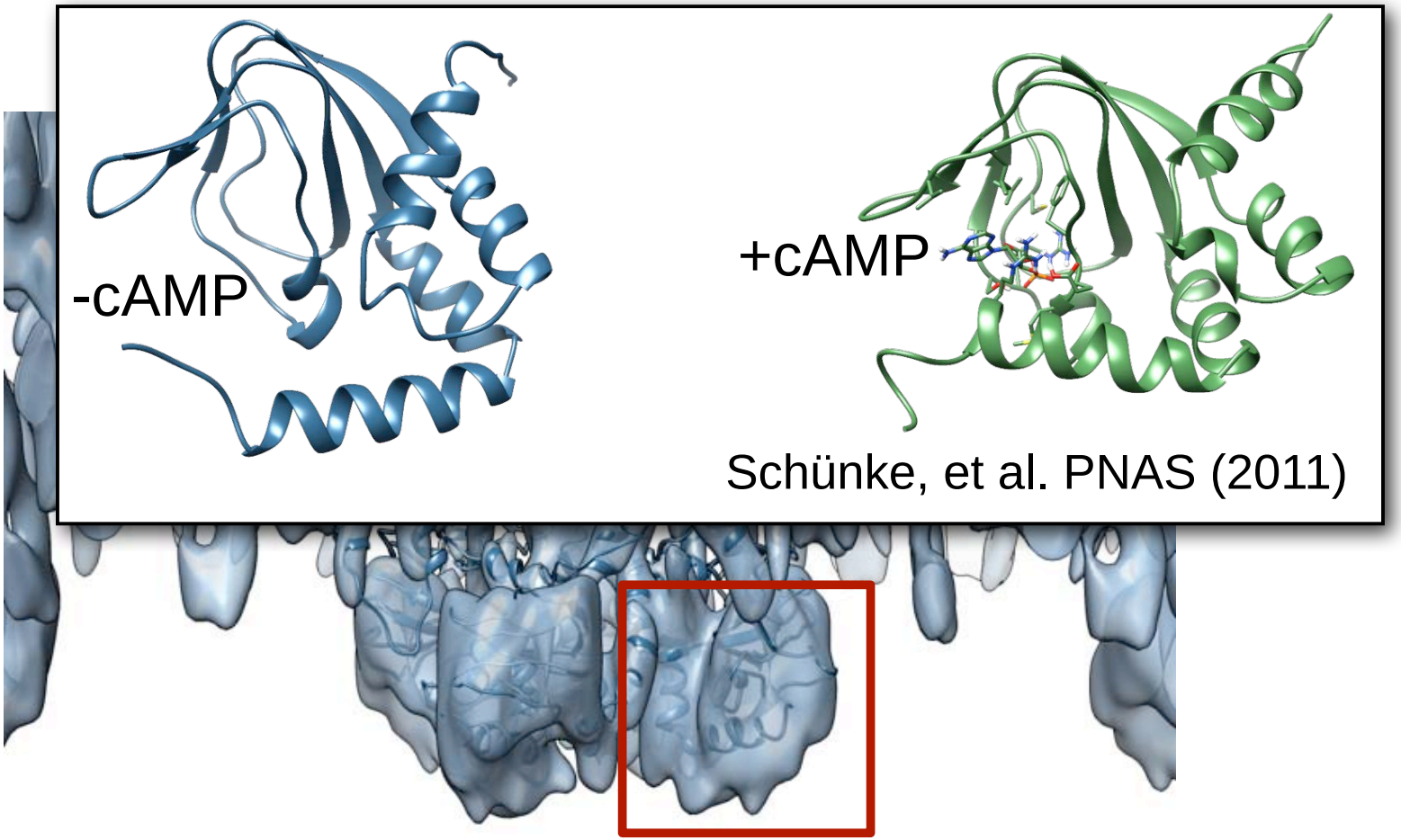
CNBD
cyclic-nucleotide binding domain

Cryo-EM structure at 7Å (12 Å vertical)
in collaboration with H.Stahlberg (Biozentrum Basel) and C.Nimigean (Cornell)

Kowal et al., Nat. Comm. (2013), accepted

Ion Channel Gating

Open and closed conformation of the CNBD determined by NMR in the group of Dieter Willbold



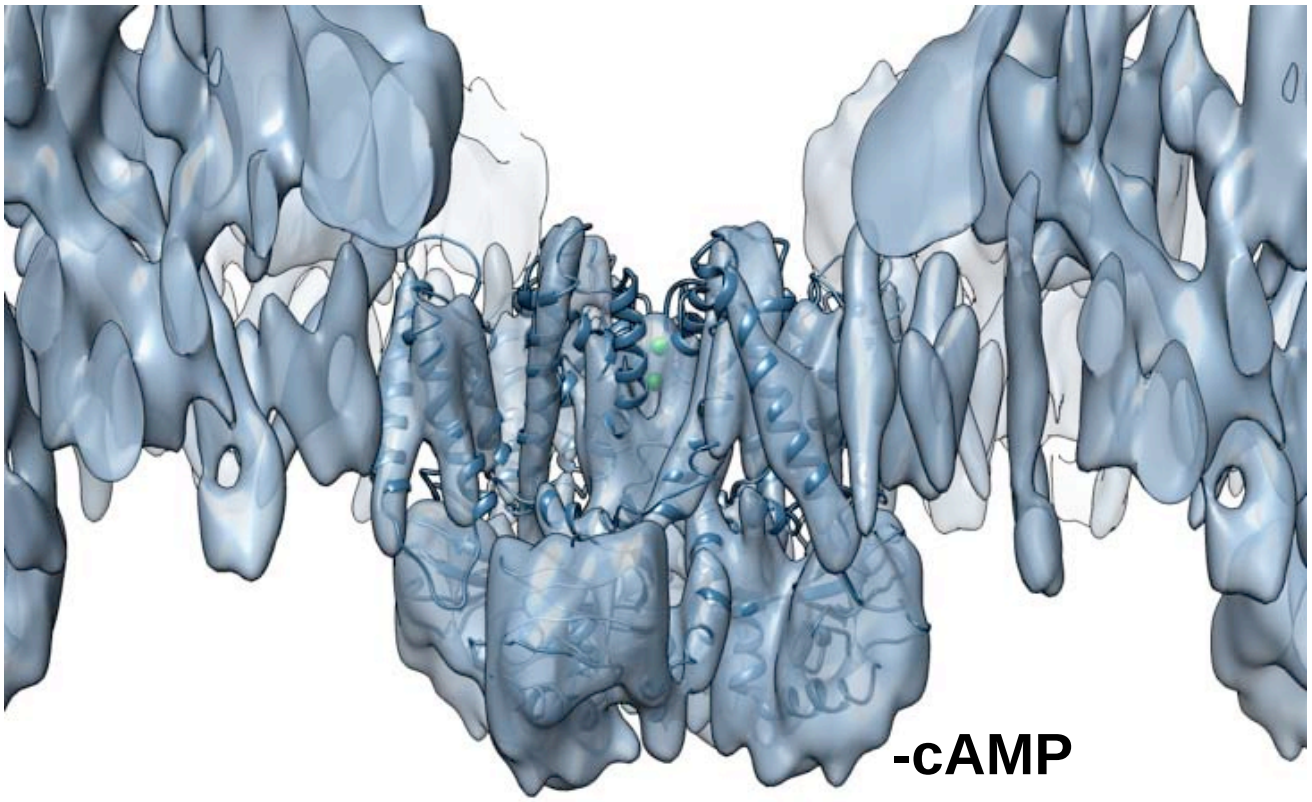
in collaboration with H.Stahlberg (Biozentrum Basel) and C.Nimigean (Cornell)

Kowal et al., Nat. Comm. (2013), accepted

Ion Channel Gating

The cyclic-nucleotide activated Mlok1 K^+ channel

channel closed



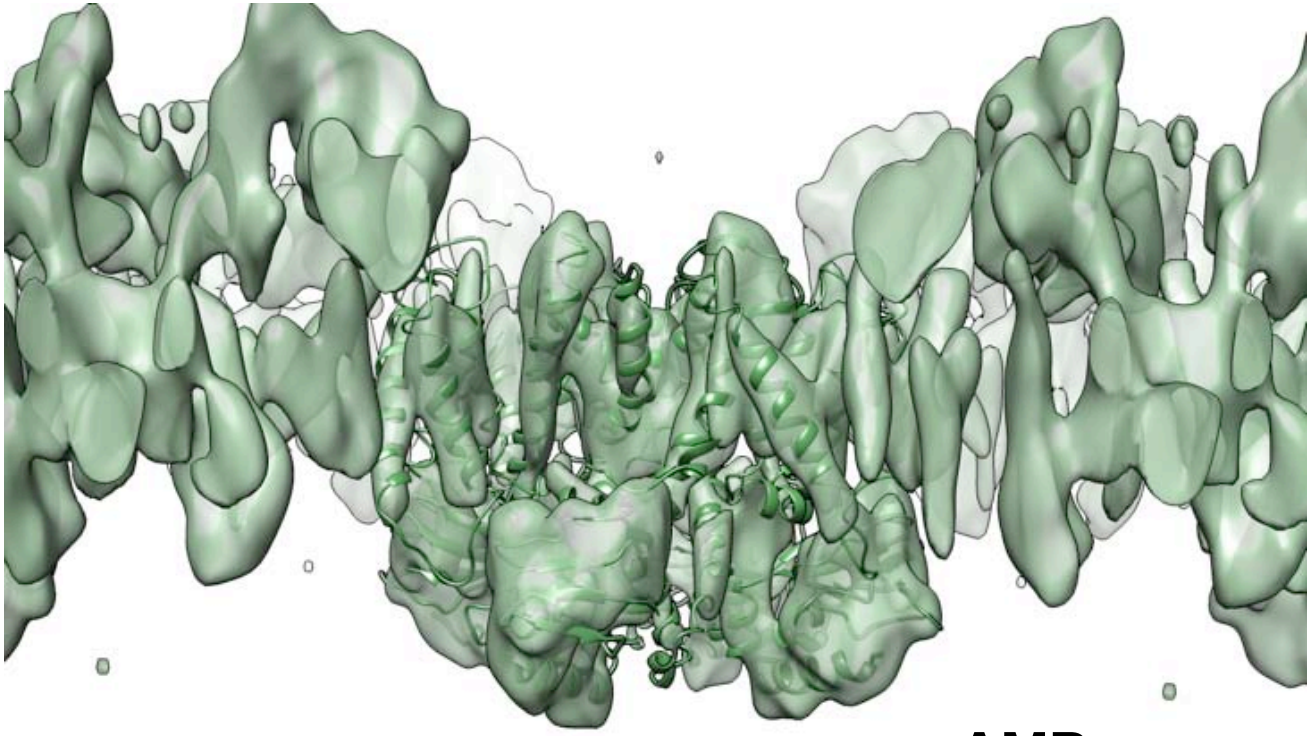
in collaboration with H.Stahlberg (Biozentrum Basel) and C.Nimigean (Cornell)

Kowal et al., Nat. Comm. (2013), accepted

Ion Channel Gating

The cyclic-nucleotide activated Mlok1 K^+ channel

channel open



outside

inside

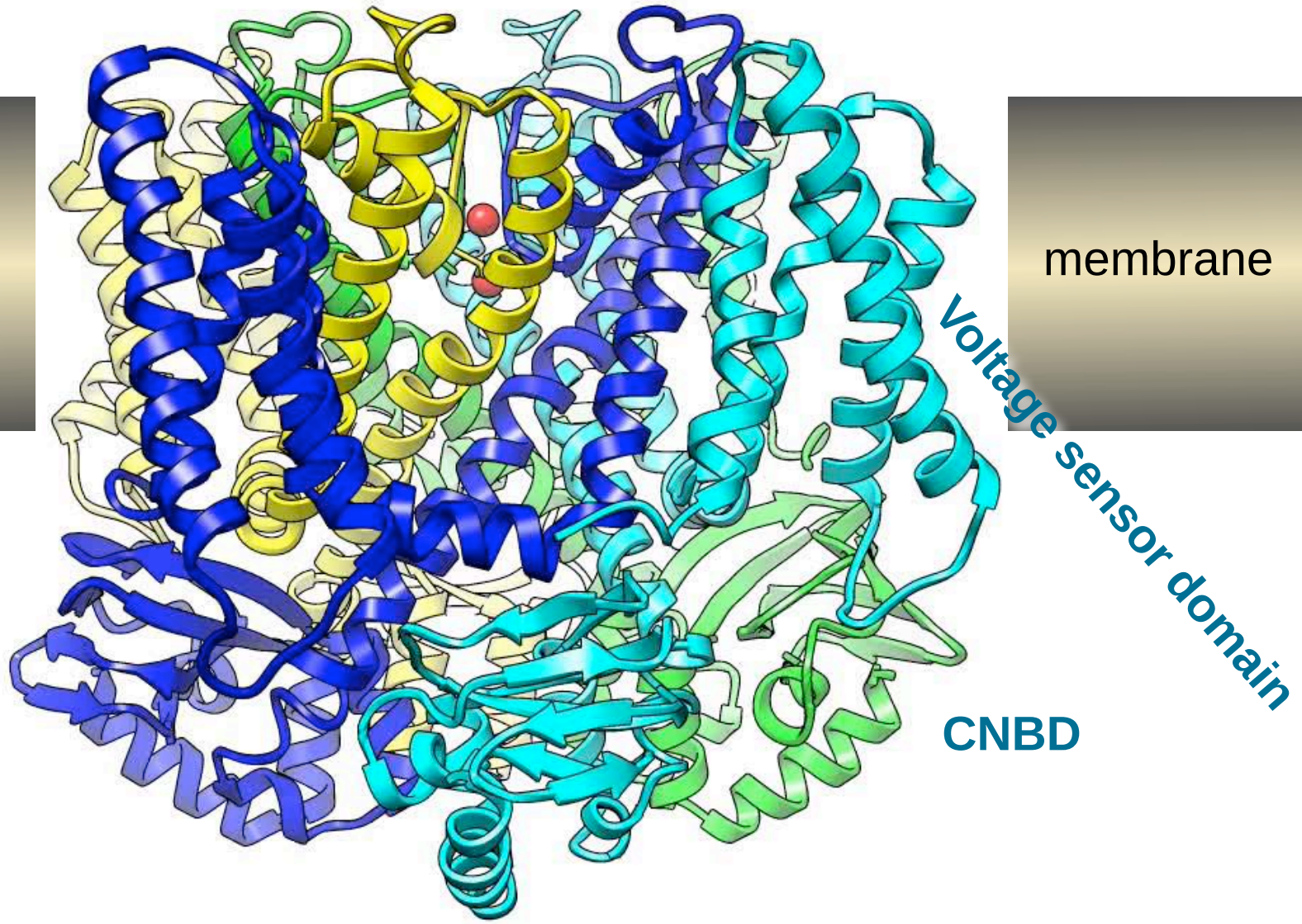
+cAMP

in collaboration with H.Stahlberg (Biozentrum Basel) and C.Nimigean (Cornell)

Kowal et al., Nat. Comm. (2013), accepted

Ion Channel Gating

Transition between open and closed channel conformation



Cross-validation

At low resolution overfitting becomes a serious problem.

Standard procedure in X-ray refinement (Brunger, 1992):

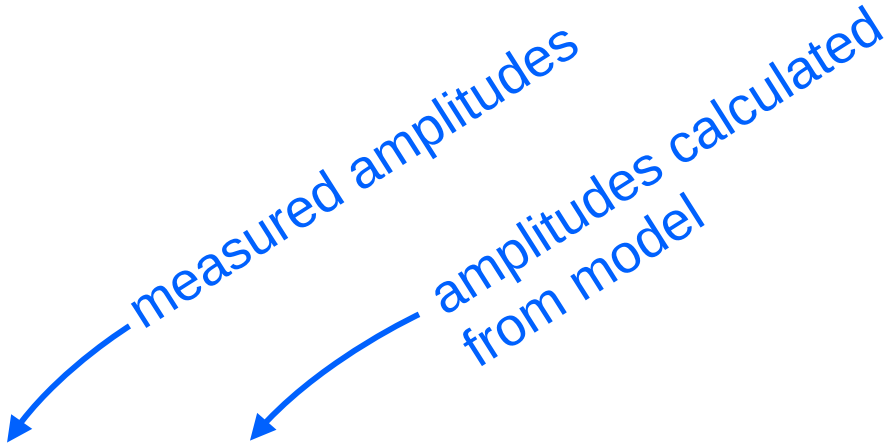
Split data set (randomly) into two sets:

‘**work**’ set and ‘**test**’ set

95% 5%

Refinement is done only with work set.

Test set data are only used for computing R_{free}



$$R = \frac{\sum_{\mathbf{h}} \left| |F_{obs}(\mathbf{h})| - |F_{calc}(\mathbf{h})| \right|}{\sum_{\mathbf{h}} |F_{obs}(\mathbf{h})|}$$

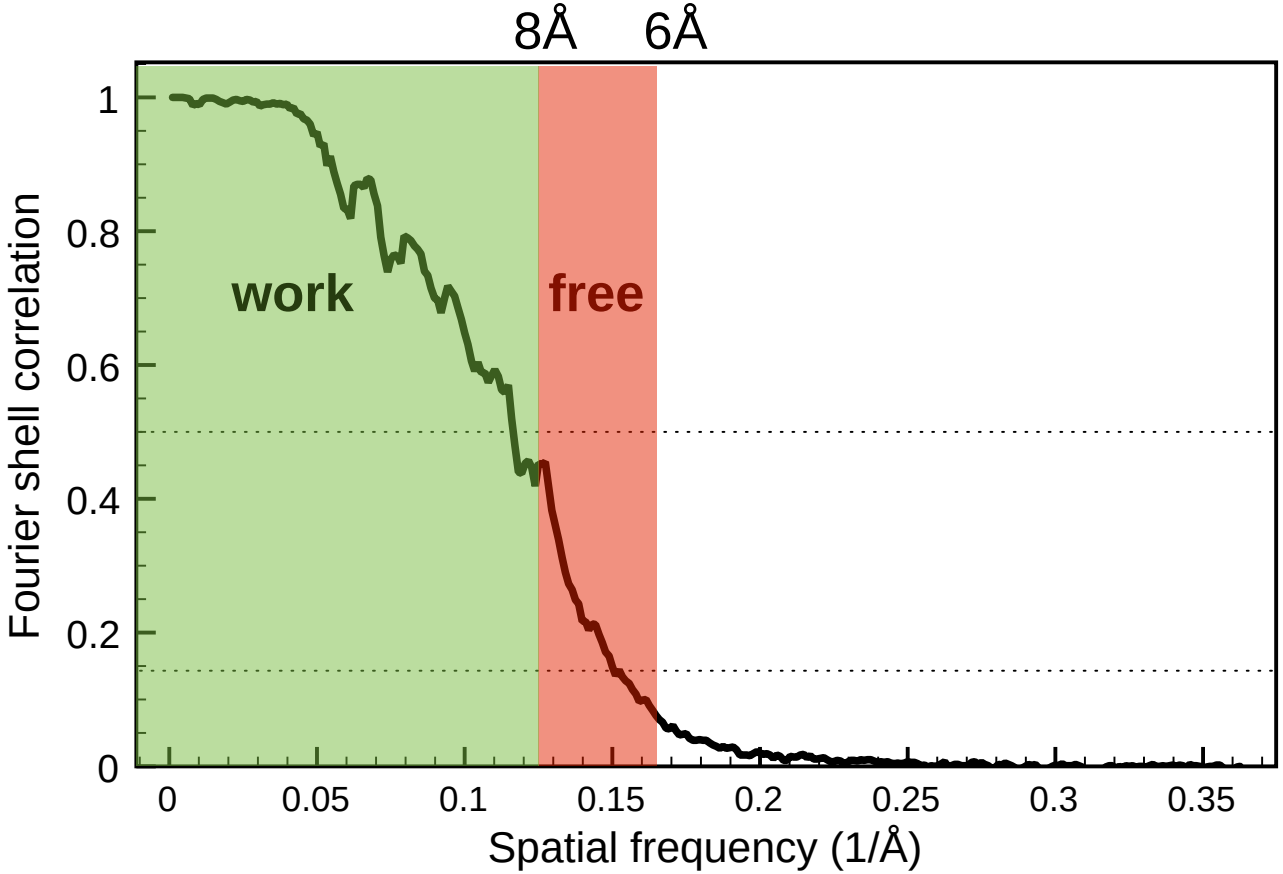
If difference between R_{free} and R_{work} gets too large, model is overfitted.

Cross-validation

Exclude part of the data that is not used for refinement, but only for validation (“test set”).

Neighboring Fourier components are correlated in EM densities, therefore define high-resolution shell as free set

Cfree: density cross-correlation between model and target free maps



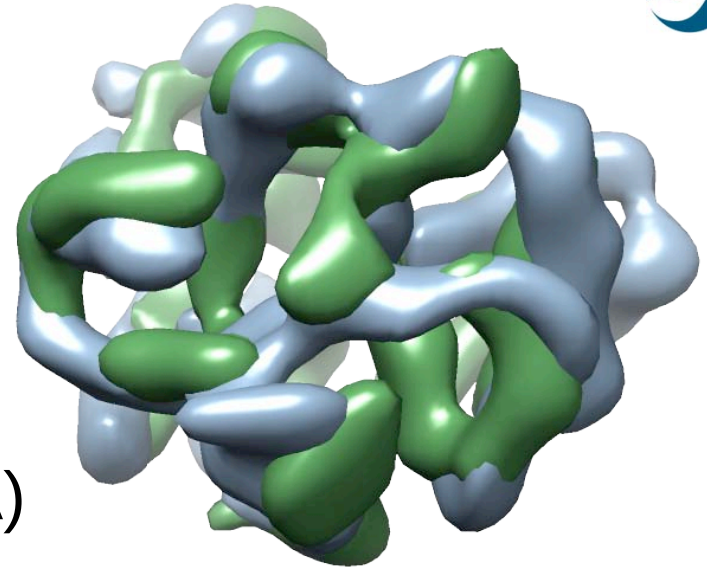
Falkner and Schröder, PNAS (2013)

Cross-validation

EM density

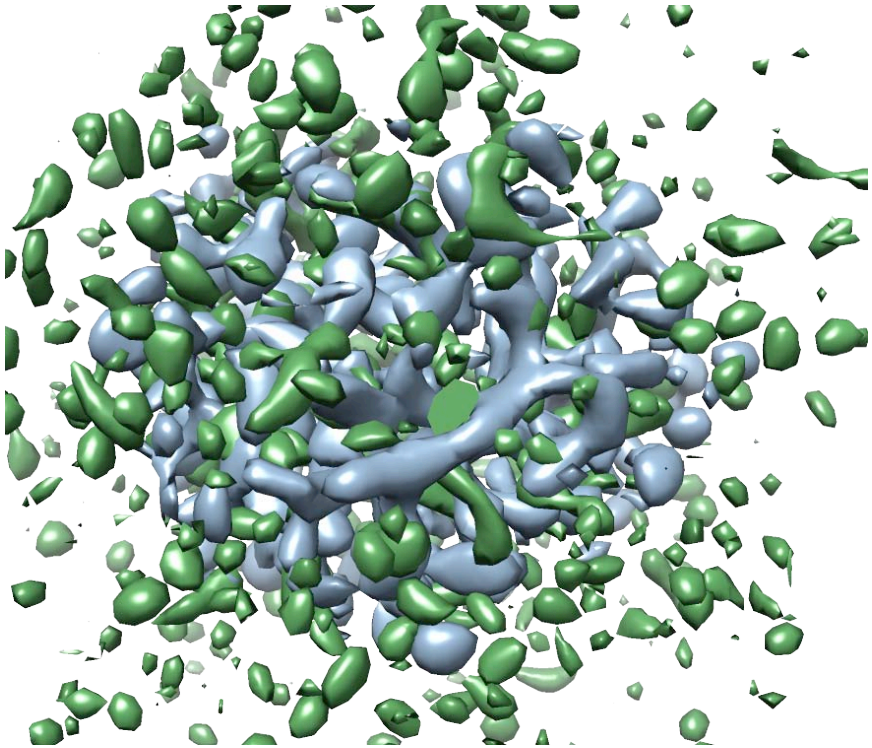
Model density

Density cross-correlation:



Work interval (here 7 - 200 Å)

-> C_{work}



Free interval (here 5 - 7 Å)

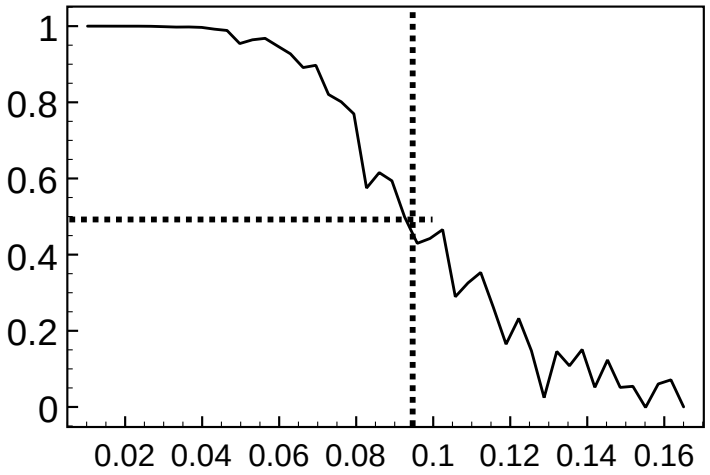
-> C_{free}

Cross-validation

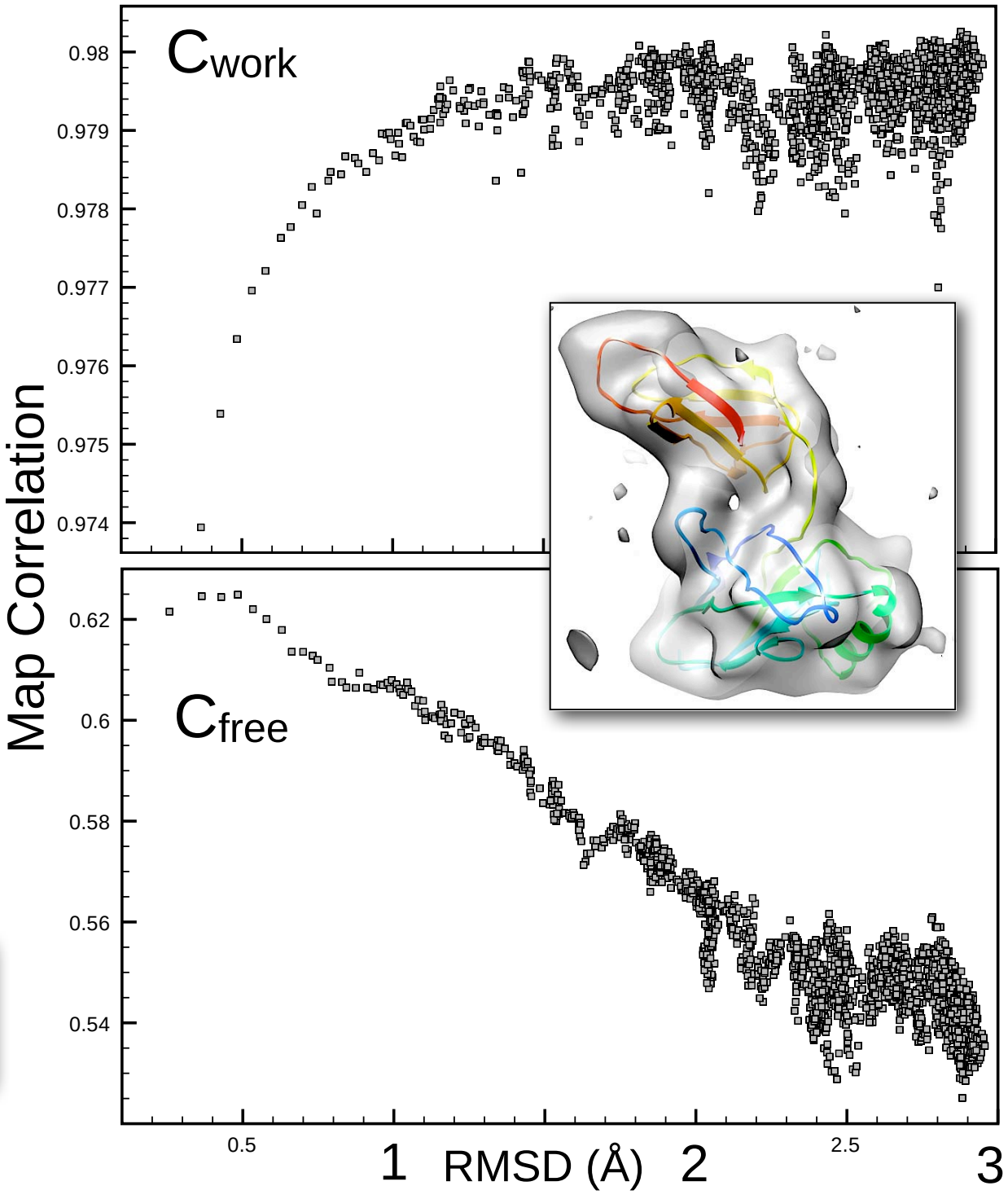
Starting from correct structure (1ikn)

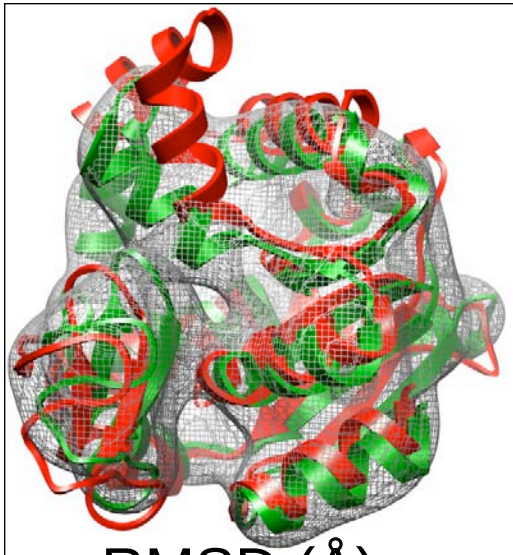
Synthetic density map with added noise

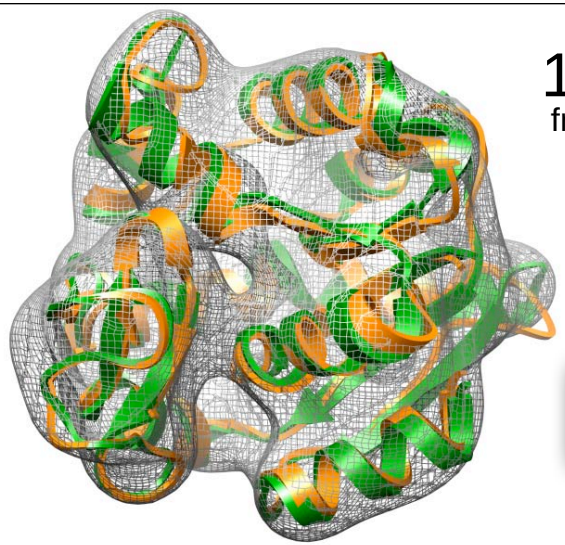
FSC (0.5) ~11 Å



Free resolution range:
10 - 13 Å







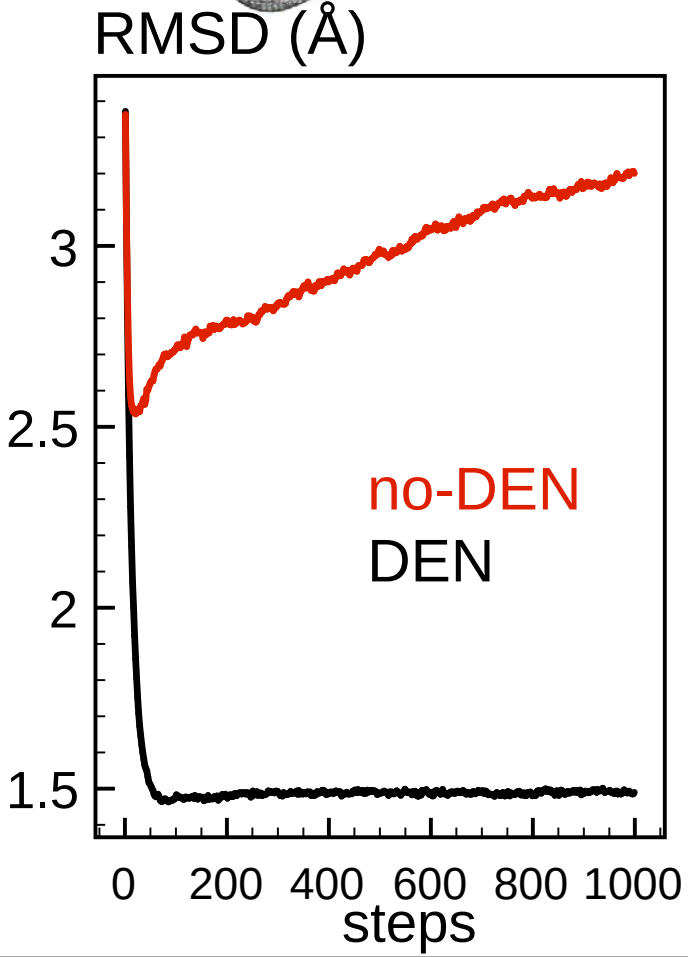
1ake at ~11 Å

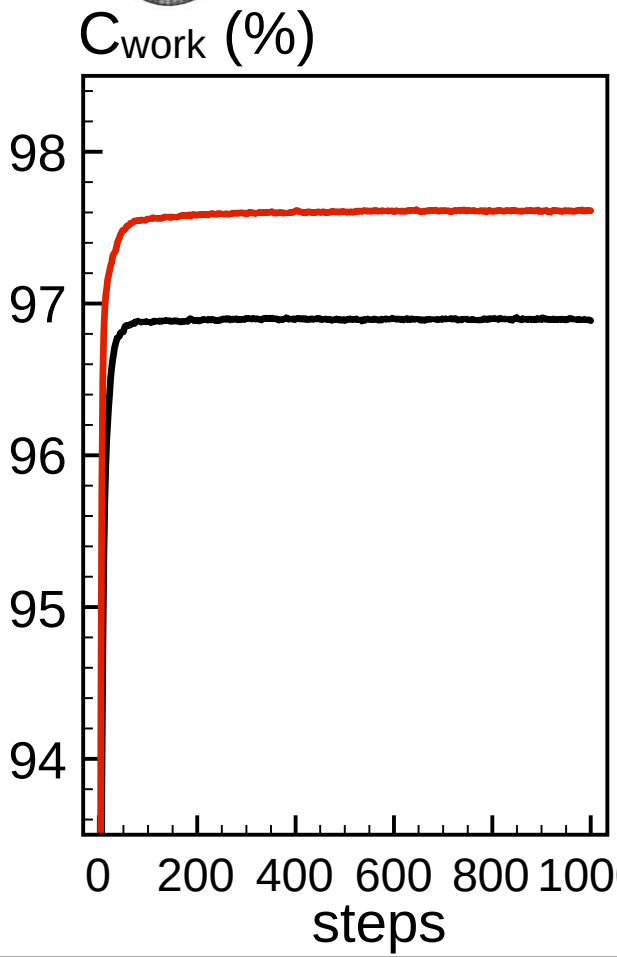
from Benchmark set by Topf et al. (*Structure* 16: 295, 2008)

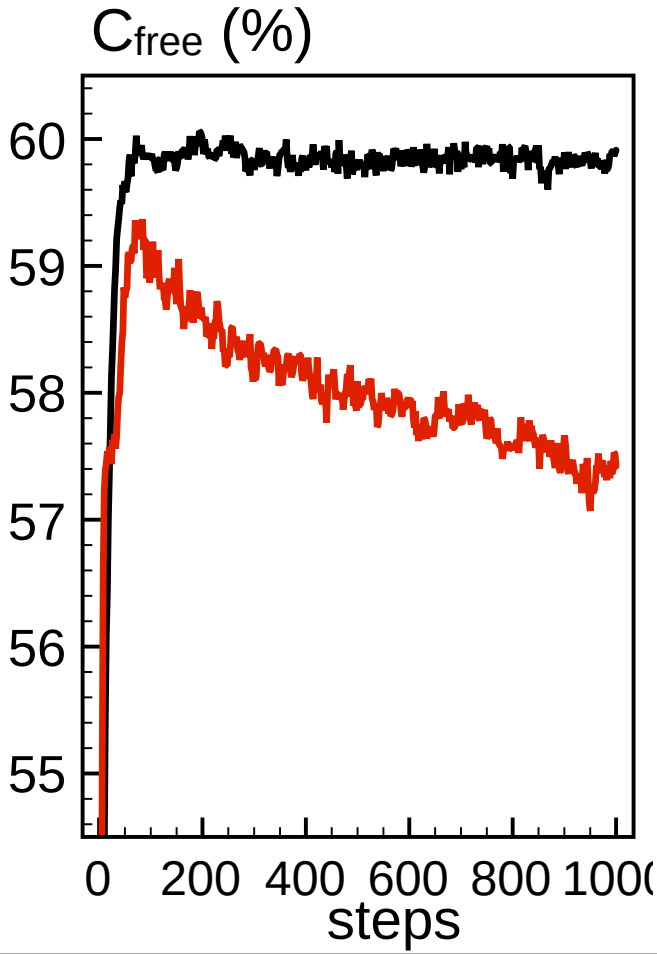
initial RMSD 3.6 Å

refined 1.5 Å

free resolution range 10-13 Å

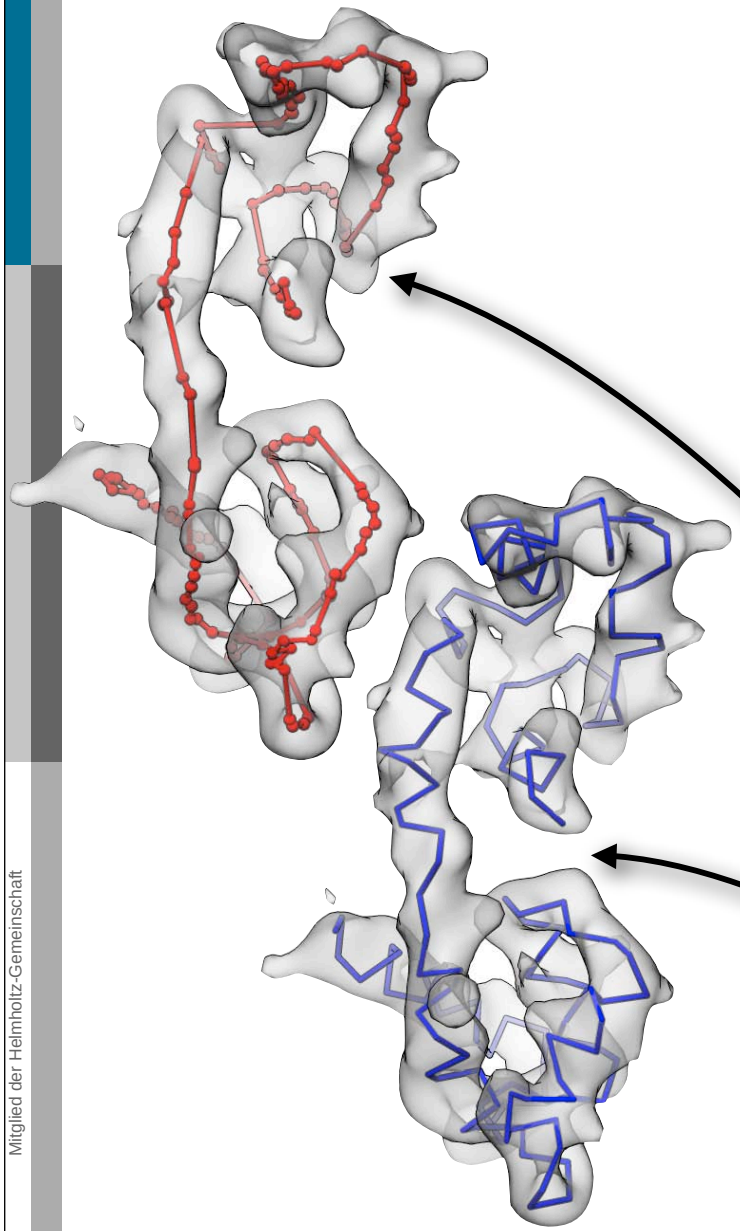






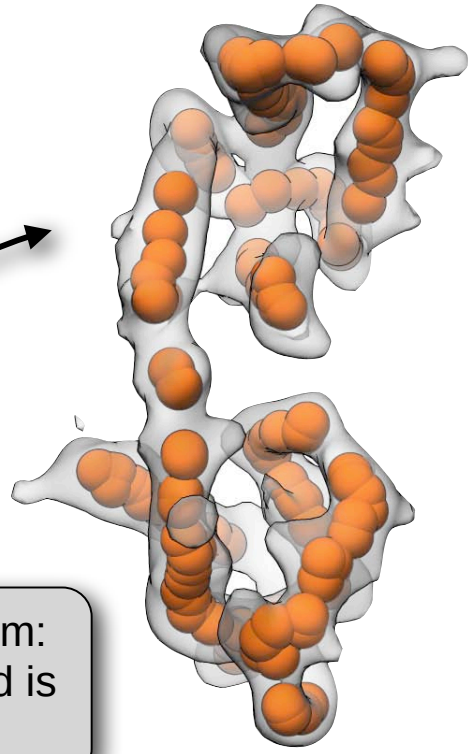
Backbone Tracing at Low Resolution (4 – 6 Å)

Treat tracing as a global optimization problem



Place Beads Randomly into the Density Map

Refine Bead Positions



Solve the Traveling Salesman Problem: find shortest path such that each bead is visited exactly one time

Lin-Kernighan algorithm
+ secondary structure restraints
+ statistical residue pair potential

Refine Backbone Trace (with DireX + Secondary Structure Restraints)

Additional Restraints on the Traveling Salesman Problem

The *Miyazawa-Jernigan* potential is a **statistical potential** which favors contacts of amino acids that are frequently observed to be in contact in the PDB:

$$E_{MJ} = \sum_{i<j} M(a_i, a_j) D_{ij}$$

D_{ij} is the contact matrix between amino acids of types a_i and a_j.
M is the weight according the observed frequency of the a_i,a_j pair.

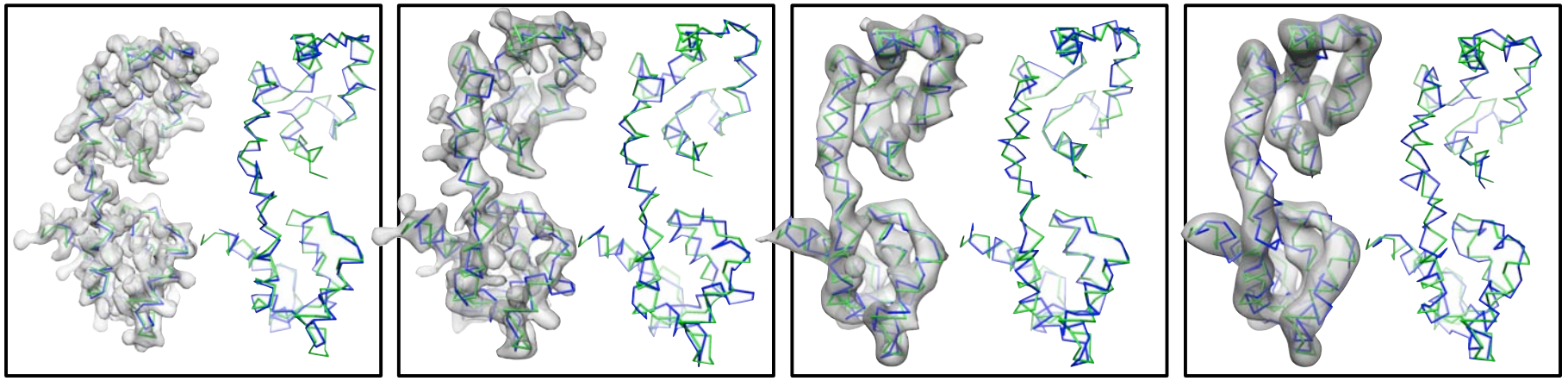
Secondary structure prediction yields restraints on the distances between amino acids that are within the same secondary structure element:

$$E_{SSE} = \sum (d_{ij} - d_{ij}^{seq})^2$$

The tracing algorithm then optimizes the E_{Total}

$$E_{\text{Total}} = E_{\text{Lin-Kernighan}} + E_{\text{MJ}} + E_{\text{SSE}}$$

Calmodulin Backbone Trace at Different Resolutions



4 Å resolution | 5 Å resolution | 6 Å resolution | 7 Å resolution

- Test with synthetic (perfect) density maps at different resolutions
- 10 traces were generated for each resolution
- For all resolution the correct topology was found

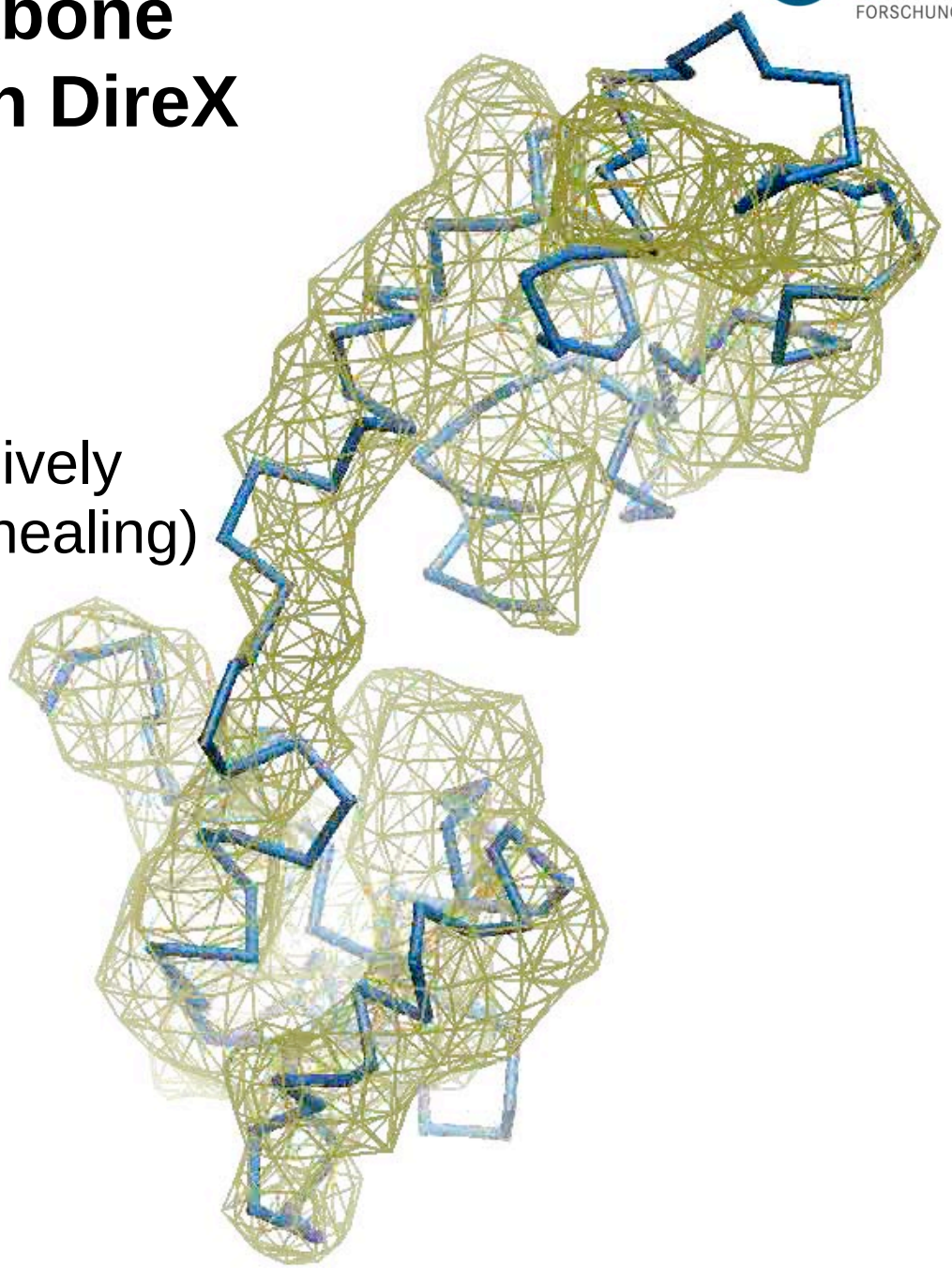
Map Resolution	RMSD
4 Å	6.4 Å
5 Å	7.5 Å
6 Å	4.7 Å
7 Å	8.1 Å

Sampling the backbone conformations with DireX

DireX does not require
a complete input model

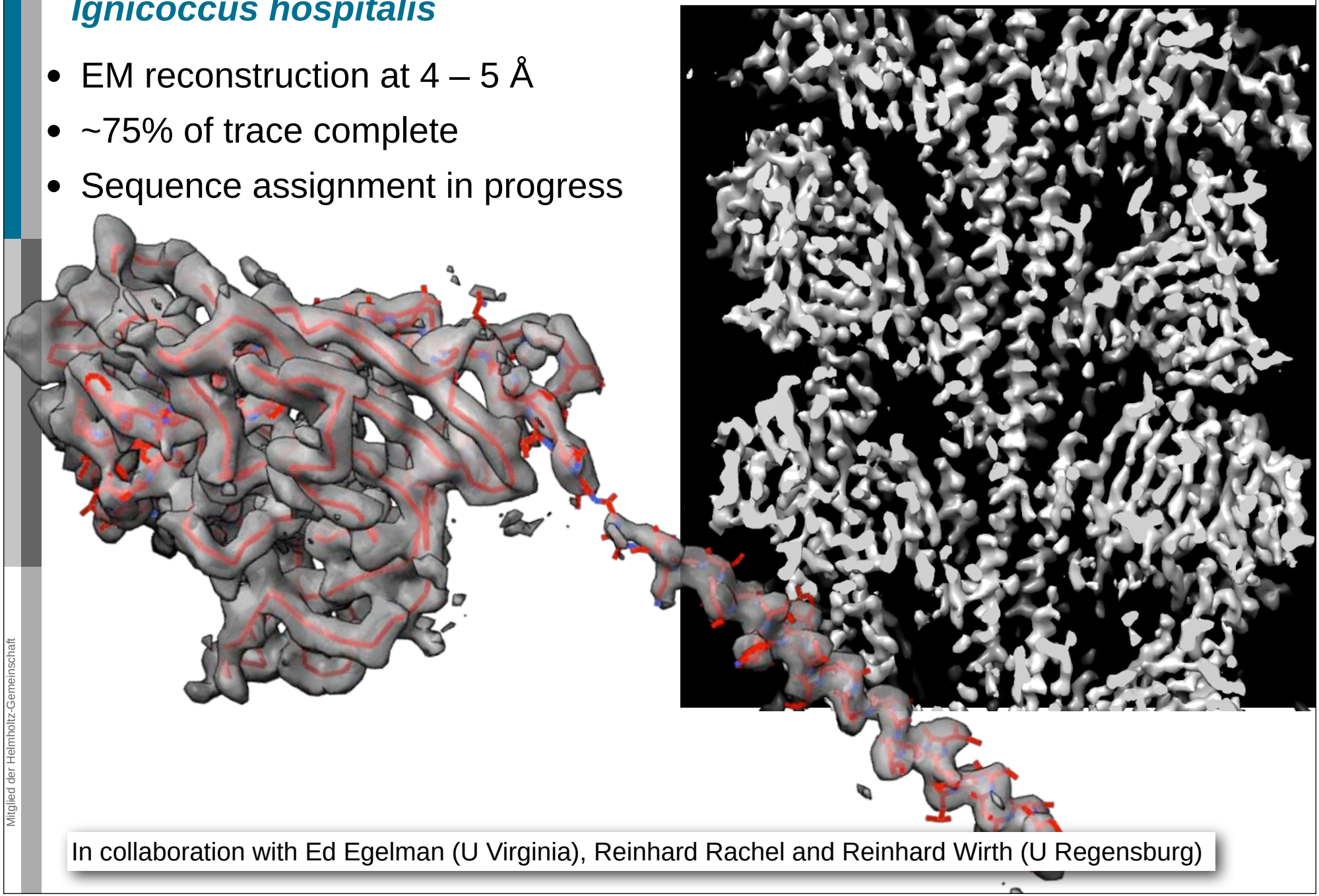
Ca-trace can be extensively
sampled (simulated annealing)

Distance restraints
can impose secondary
structure information



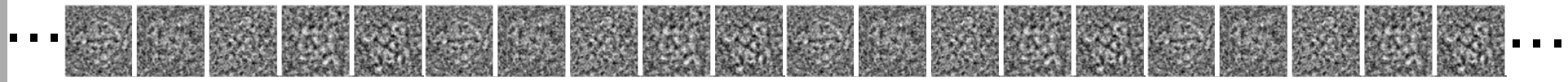
Iho670 adhesion filaments from the *Ignicoccus hospitalis*

- EM reconstruction at 4 – 5 Å
- ~75% of trace complete
- Sequence assignment in progress



In collaboration with Ed Egelman (U Virginia), Reinhard Rachel and Reinhard Wirth (U Regensburg)

Heterogeneity and Flexibility in single-particle Cryo-EM



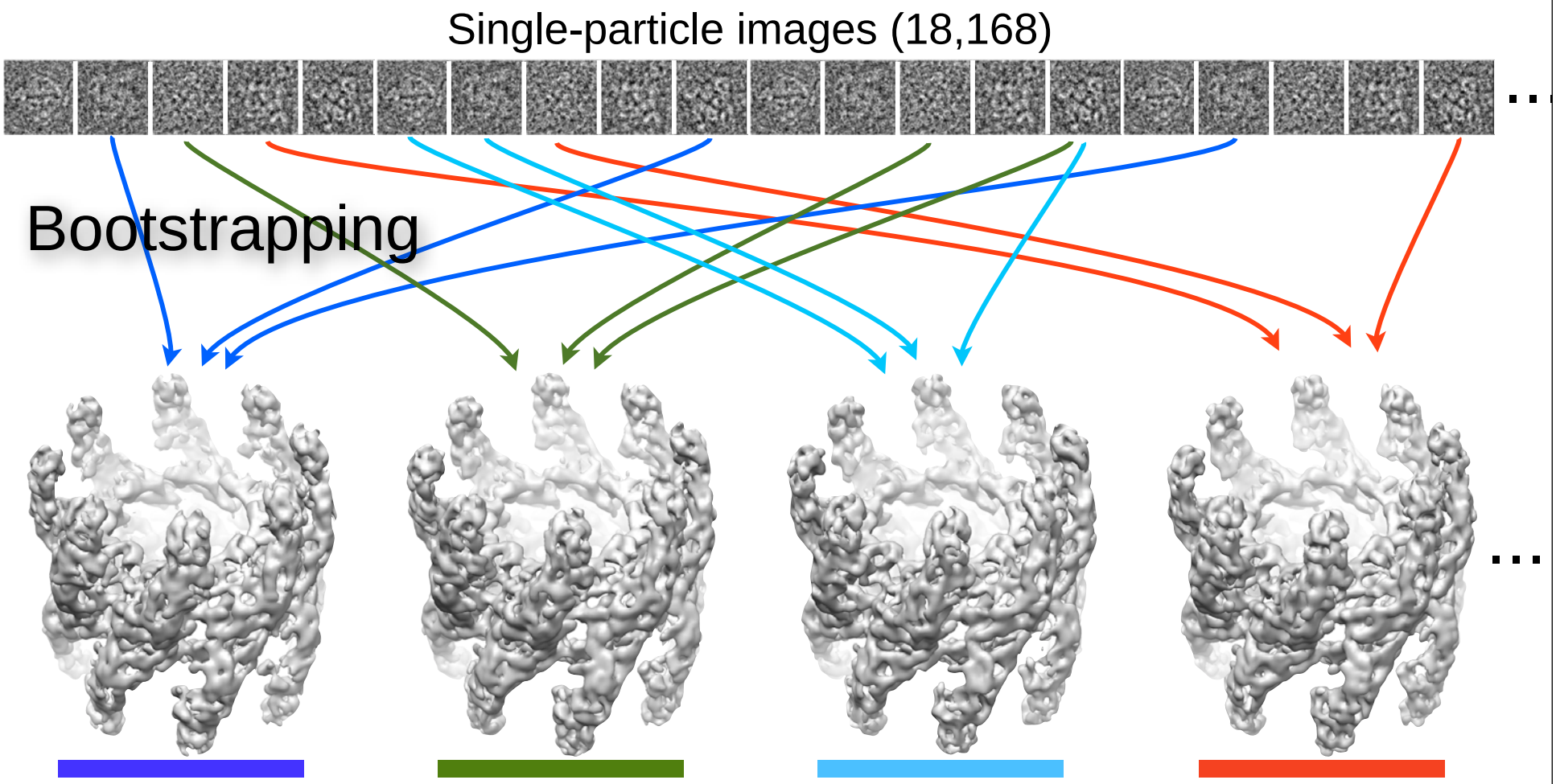
- Heterogeneity severely limits the resolution

- Advantage of Cryo-EM: all information is in the particle images
  (but difficult to extract due to noise).

- <u>Goal:</u> determine conformational variance
  AND improve resolution

Extracting Conformational Dynamics

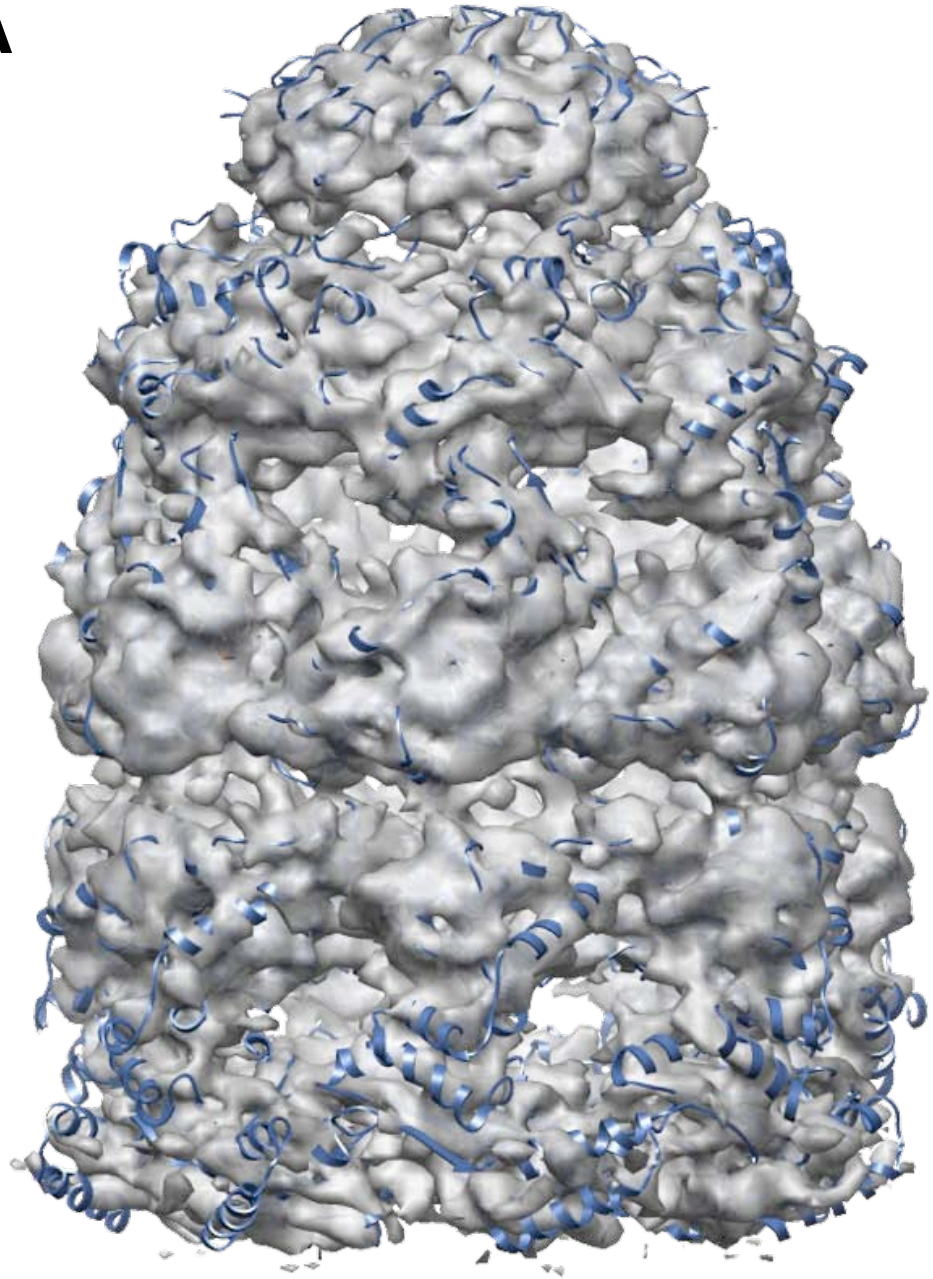
Equilibrium Fluctuations



Penczek, Yang, Frank & Spahn (2006) *J. Struct. Biol.* 154:168-183

GroEL at 8 Å

(GroEL–Aacpn10–ADP)



In collaboration with Wah Chiu (Baylor College)
and Hays Rye (Texas A&M)

Chen, Luke, Zhang, et al. JMB (2008)

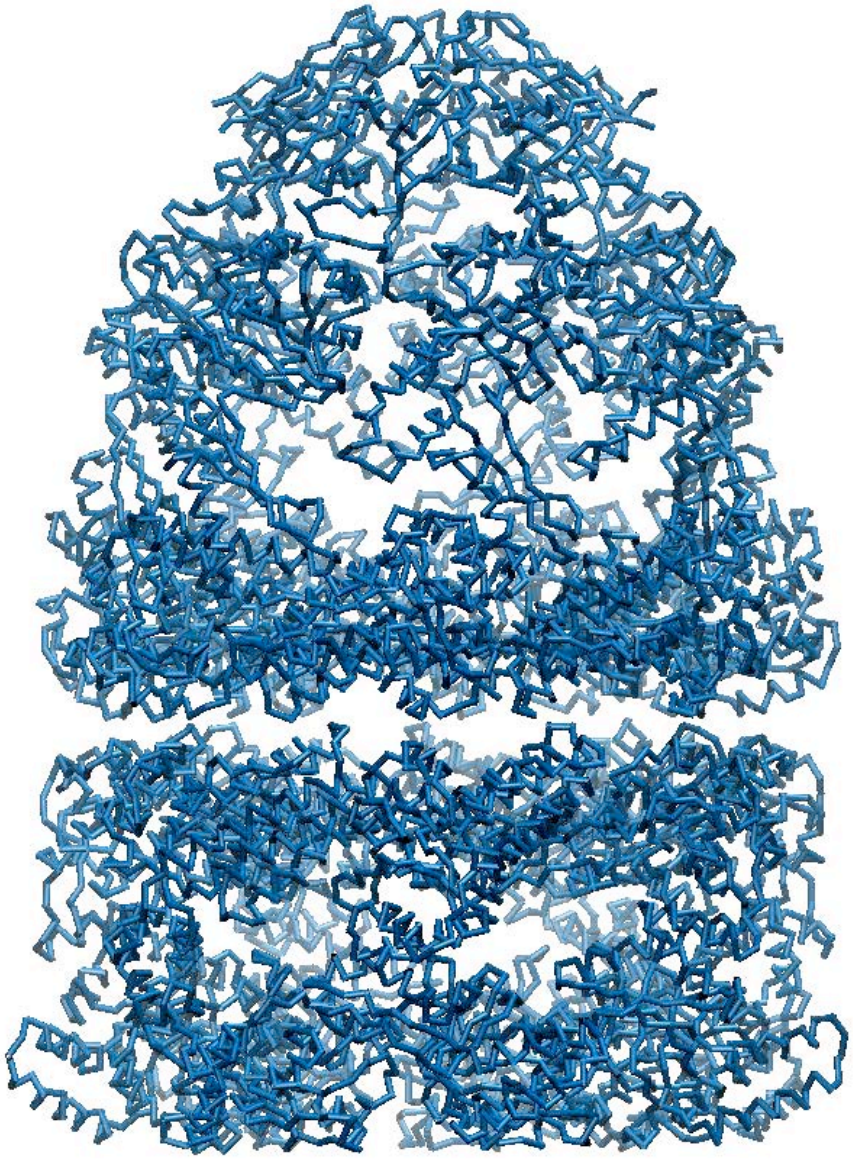
GroEL - Principal Component Analysis of the ensemble fitted models:

1. Eigenvector

- lock-in of GroES

- upward motion of cis-ring apical domains

- rotations of trans-ring apical domains



side view

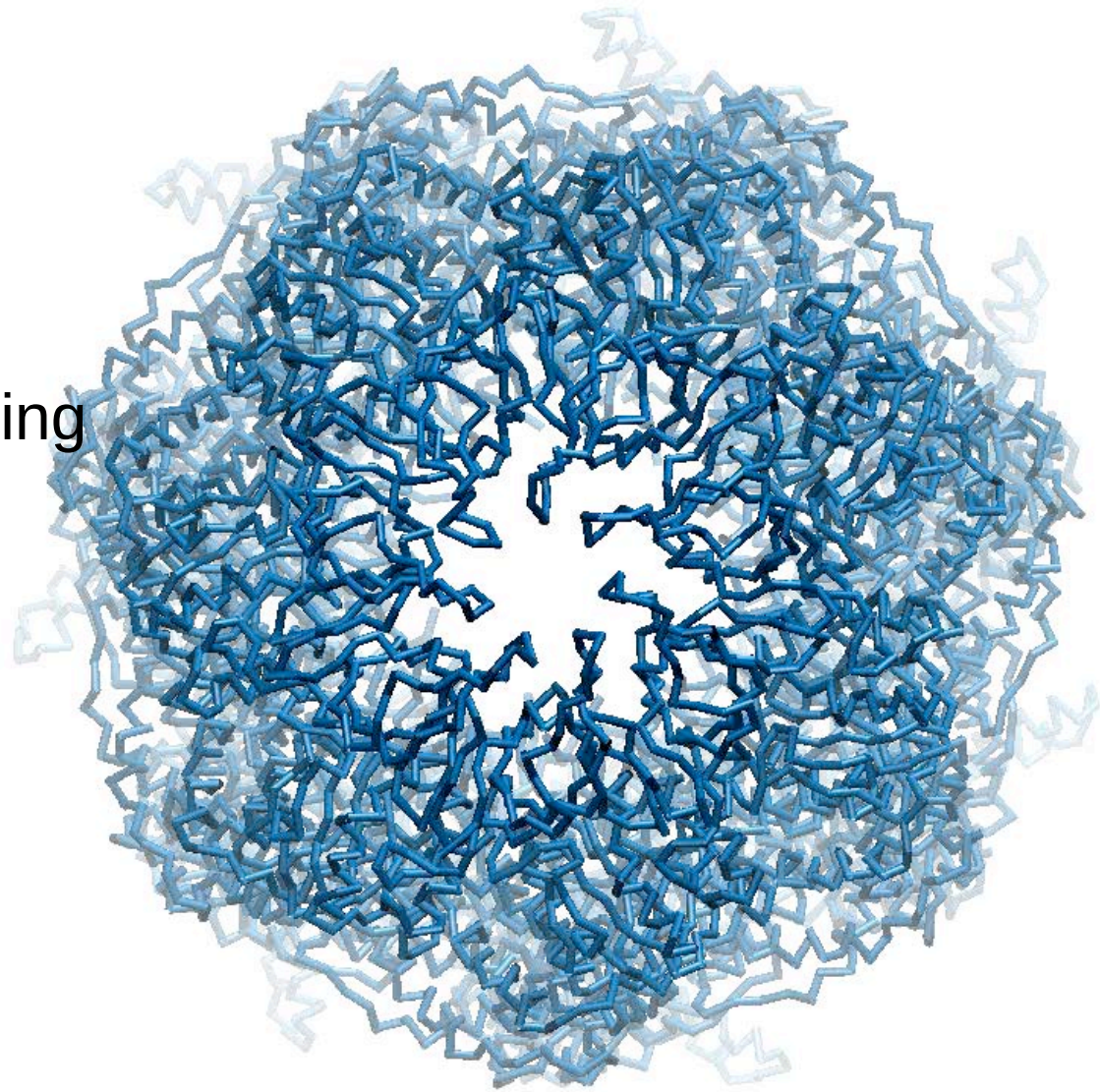
GroEL - Principal Component Analysis of the ensemble fitted models:

1. Eigenvector

- lock-in of GroES

- upward motion of cis-ring apical domains

- rotations of trans-ring apical domains



top view on GroES

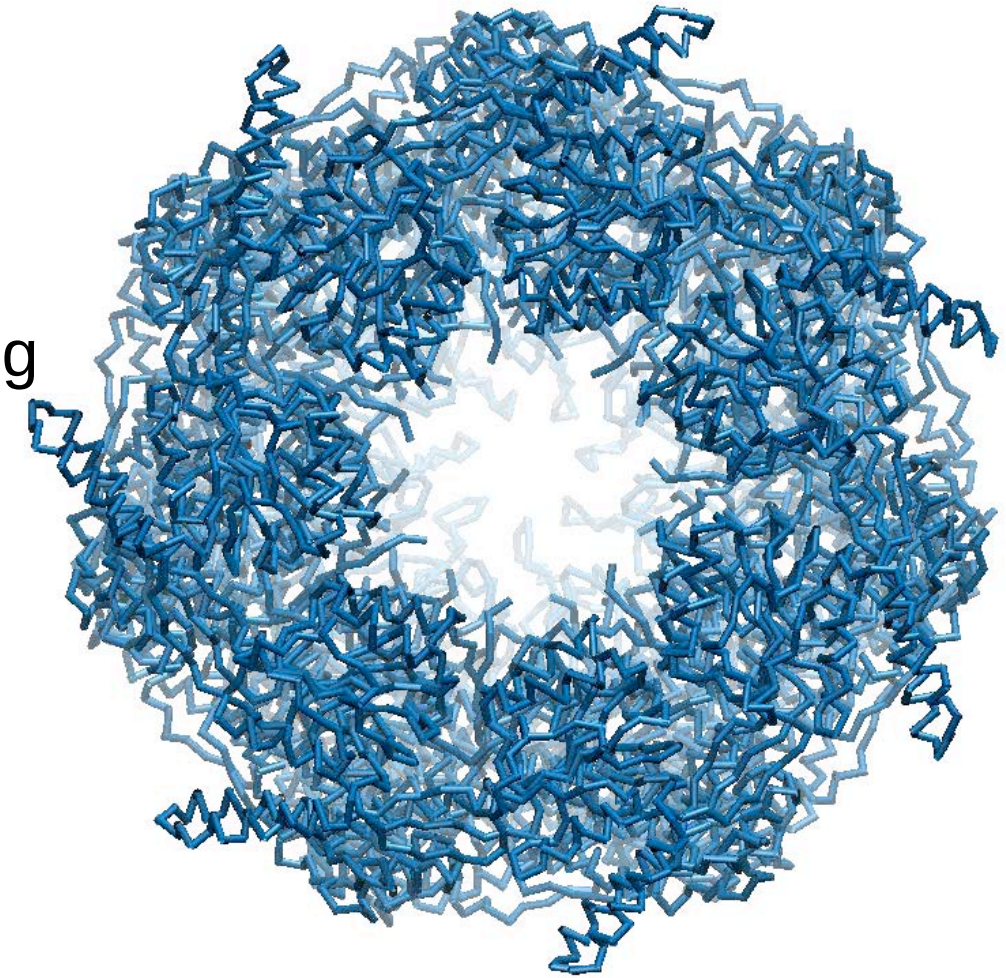
GroEL - Principal Component Analysis of the ensemble fitted models:

1. Eigenvector

- lock-in of GroES

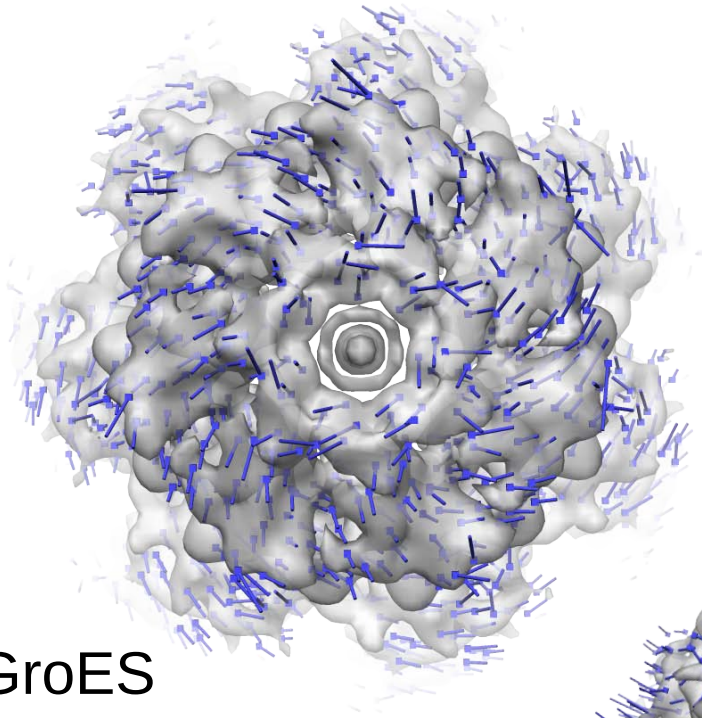
- upward motion of cis-ring apical domains

- rotations of trans-ring apical domains

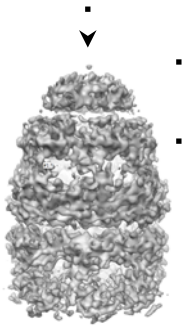


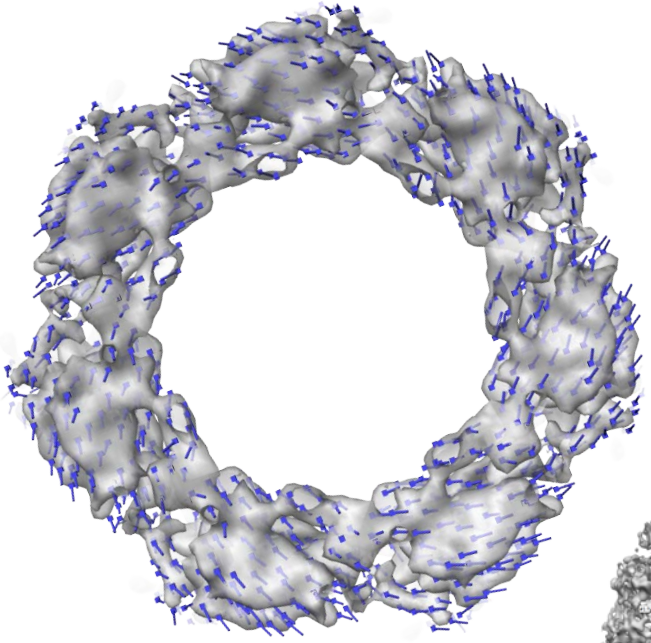
bottom view on trans-apical domain

Comparing GroEL/ES in two nucleotide states $R3_{ATP}$ - $R4_{ADP}$

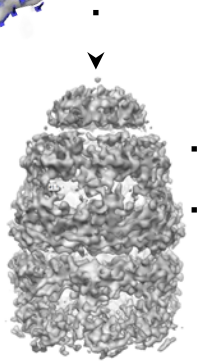


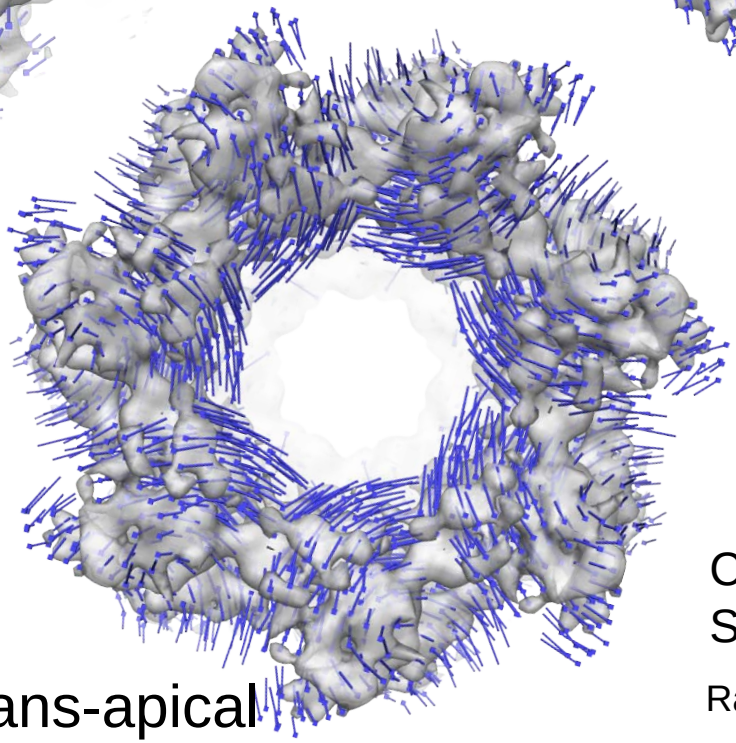
GroES

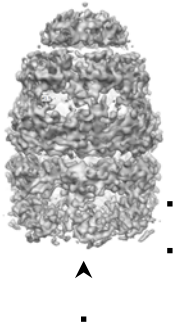




GroEL-cis-apical







GroEL-trans-apical

Chen, Madan, Weaver, Lin, Schröder, Chiu, Rye, *Cell (2013)*

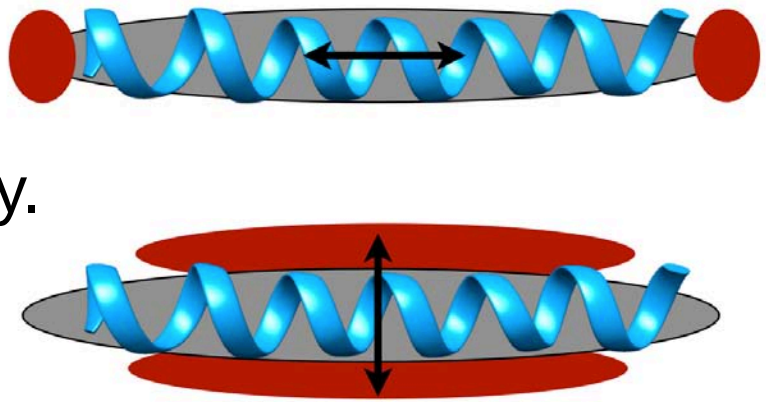
Ranson, et al. *Nat. Struct. Mol. Biol. (2006)*

Image Sorting is key to achieving high resolution (not necessarily a large number of particles)

Images used for reconstruction need to show the molecule in the same conformation!

Image Sorting reveals different conformational states

Standard image classification sorts images according to density similarity.



But: Conformational Variance is not the same as Density Variance

Ongoing Work

- Use principal conformational motions to sort images into classes

- Iteratively determine residual conformational dynamics in subclasses of images for further

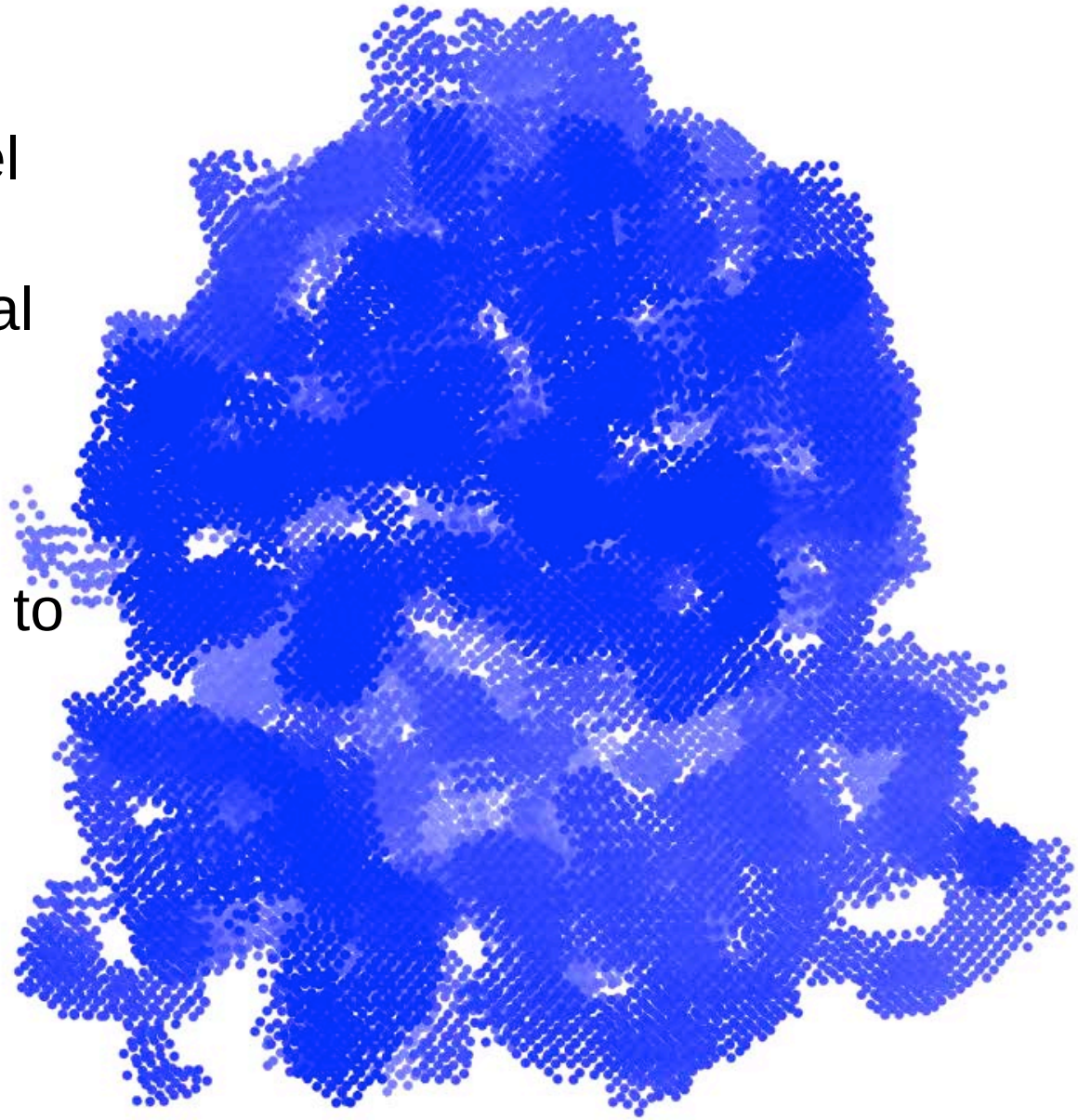
Refinement of generic bead models

DireX can refine any generic geometric model

you do not need a crystal structure to determine principal motions

- Use program *beadgen* to generate a bead model from a density map

- Refine bead model to different density maps



Models refined to 100 bootstrapped maps

Acknowledgements

Group Members

Benjamin Falkner
André Wildberg
Zhe Wang
Kumaran Baskaran
Dennis Della Corte
Amudha Duraisamy
Michaela Spiegel
Lena Möhlenkamp
Robin Pauli

Forschungszentrum Jülich

Dieter Willbold (ICS-6)

Karl-Erich Jaeger (IMET)

Jan Marienhagen (IBG-1)

Stanford University

Michael Levitt
Axel Brunger

Max-Planck-Institute Göttingen

Holger Stark
Niels Fischer

Biozentrum Basel

Henning Stahlberg

University of Virginia

Ed Egelman
Vitold Galkin

Baylor College of Medicine

Junjie Zhang
Dong-Hua Chen
Wah Chiu

Jülich Supercomputing Centre (JSC)