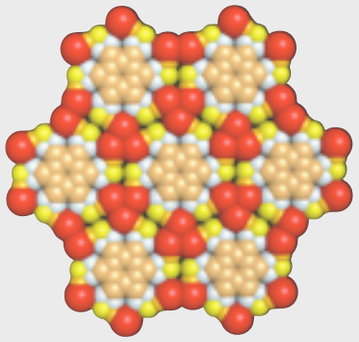
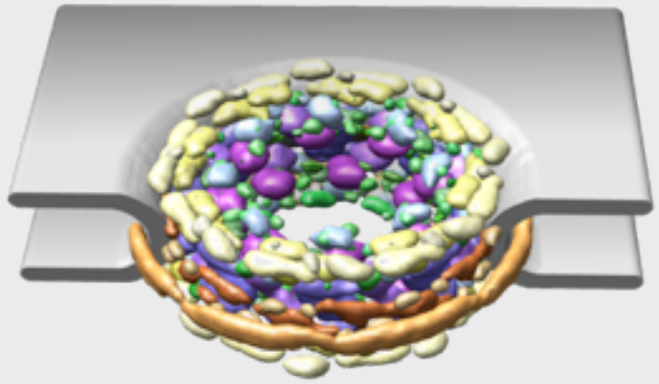
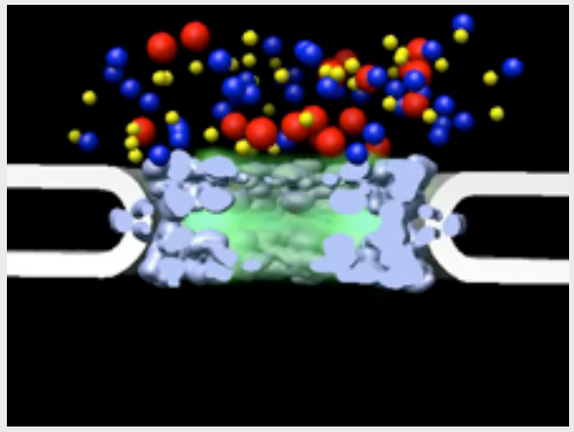
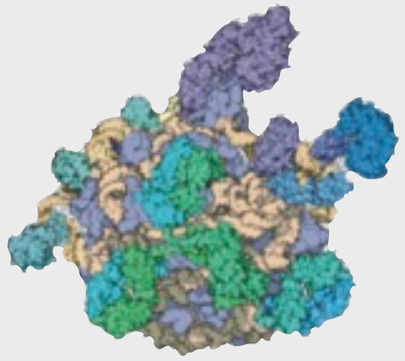
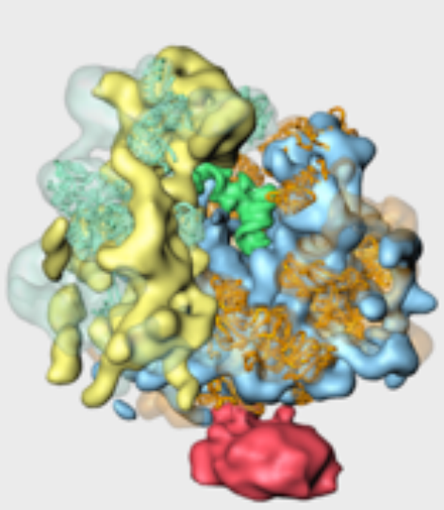
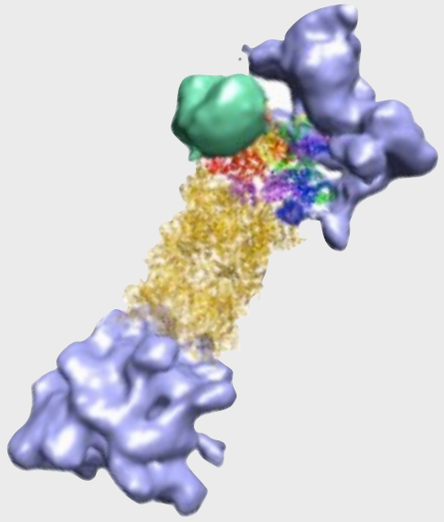
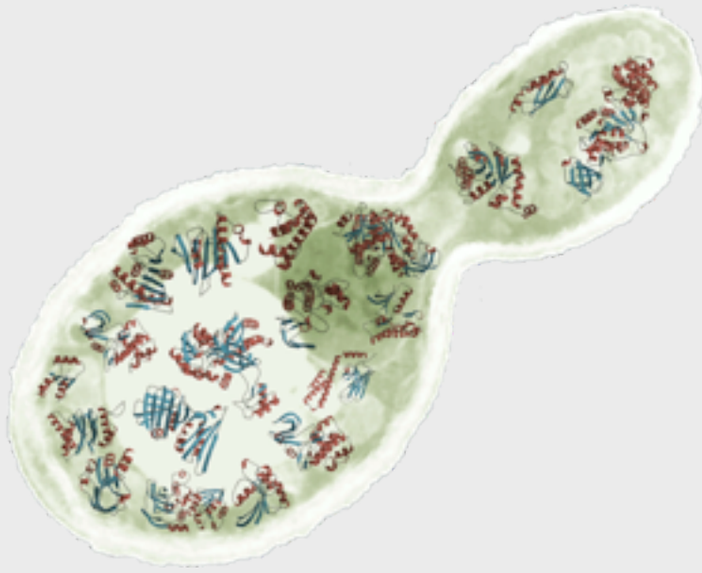
Integrative Structural Biology

Andrej Sali

http://salilab.org/

Department of Bioengineering and Therapeutic Sciences

Department of Pharmaceutical Chemistry

California Institute for Quantitative Biosciences

University of California, San Francisco

Contents

1. Integrative (hybrid) structure determination
2. EM images as a source of spatial restraints

   Application to the Nup84 complex
3. Multiple fitting of subunits into an EM map of the whole assembly

   Application to the 26S proteasome

Contents

1. Integrative (hybrid) structure determination
2. EM images as a source of spatial restraints

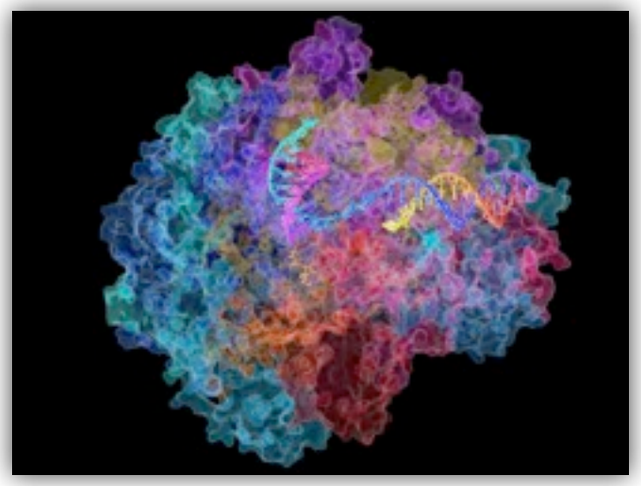
   Application to the Nup84 complex
3. Multiple fitting of subunits into an EM map of the whole assembly

   Application to the 26S proteasome

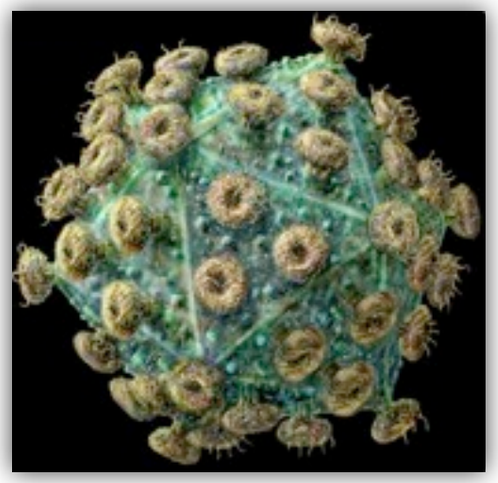
Structural biology:
Maximize accuracy, resolution, completeness, and efficiency of the structural coverage of macromolecular assemblies

Push the envelope: size, dynamics, heterogeneity.

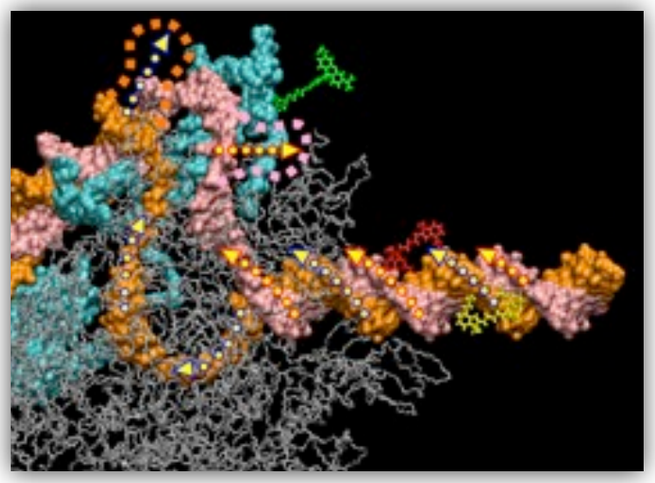
Motivation: Models will allow us to understand how machines work, how they evolved, how they can be controlled, modified, and perhaps even designed.



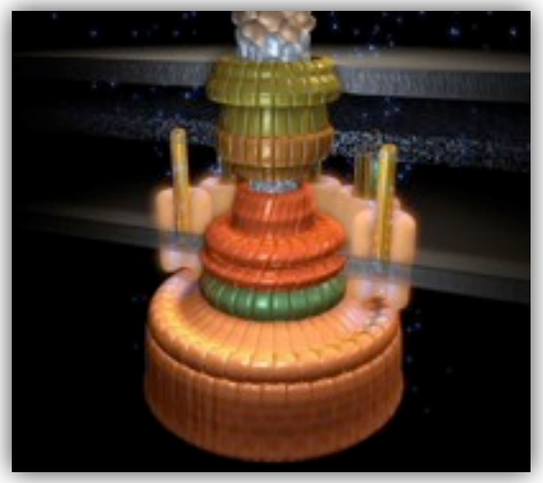
RNA polymerase II



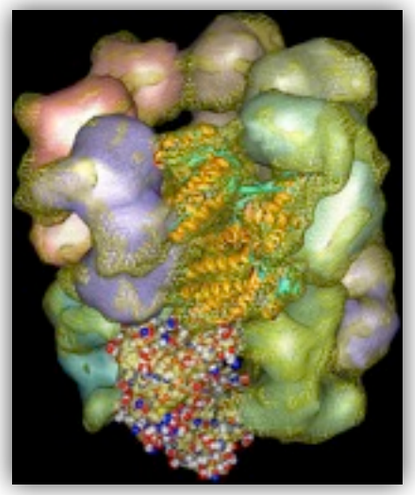
HIV virus



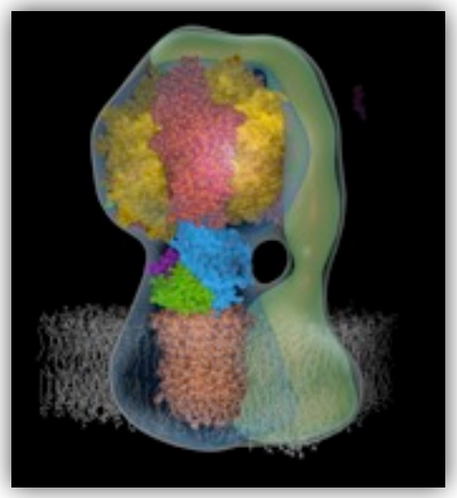
tRNA synthetase



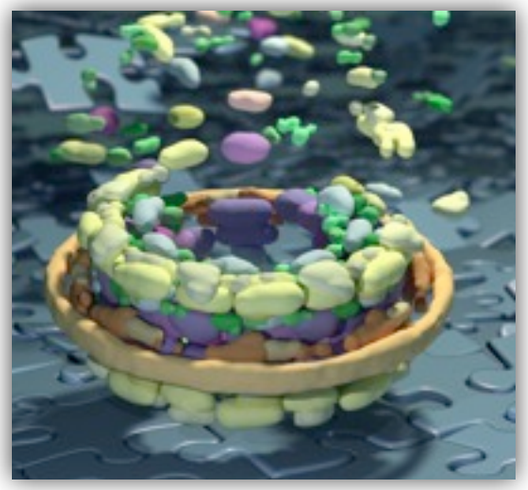
flagellar motor



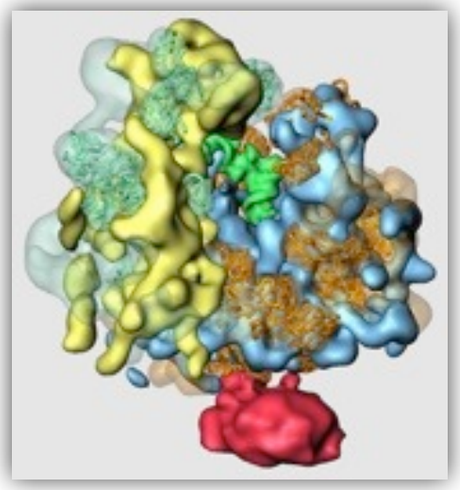
GroEL chaperonin



ATP synthase



nuclear pore complex



ribosome

There may be thousands of biologically relevant macromolecular complexes whose structures are yet to be characterized, involved in a few hundred core biological processes.

Integrative Structural Biology

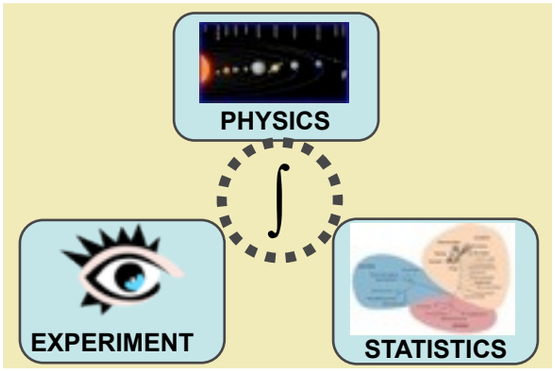
for maximizing accuracy, resolution, completeness, and efficiency of structure determination

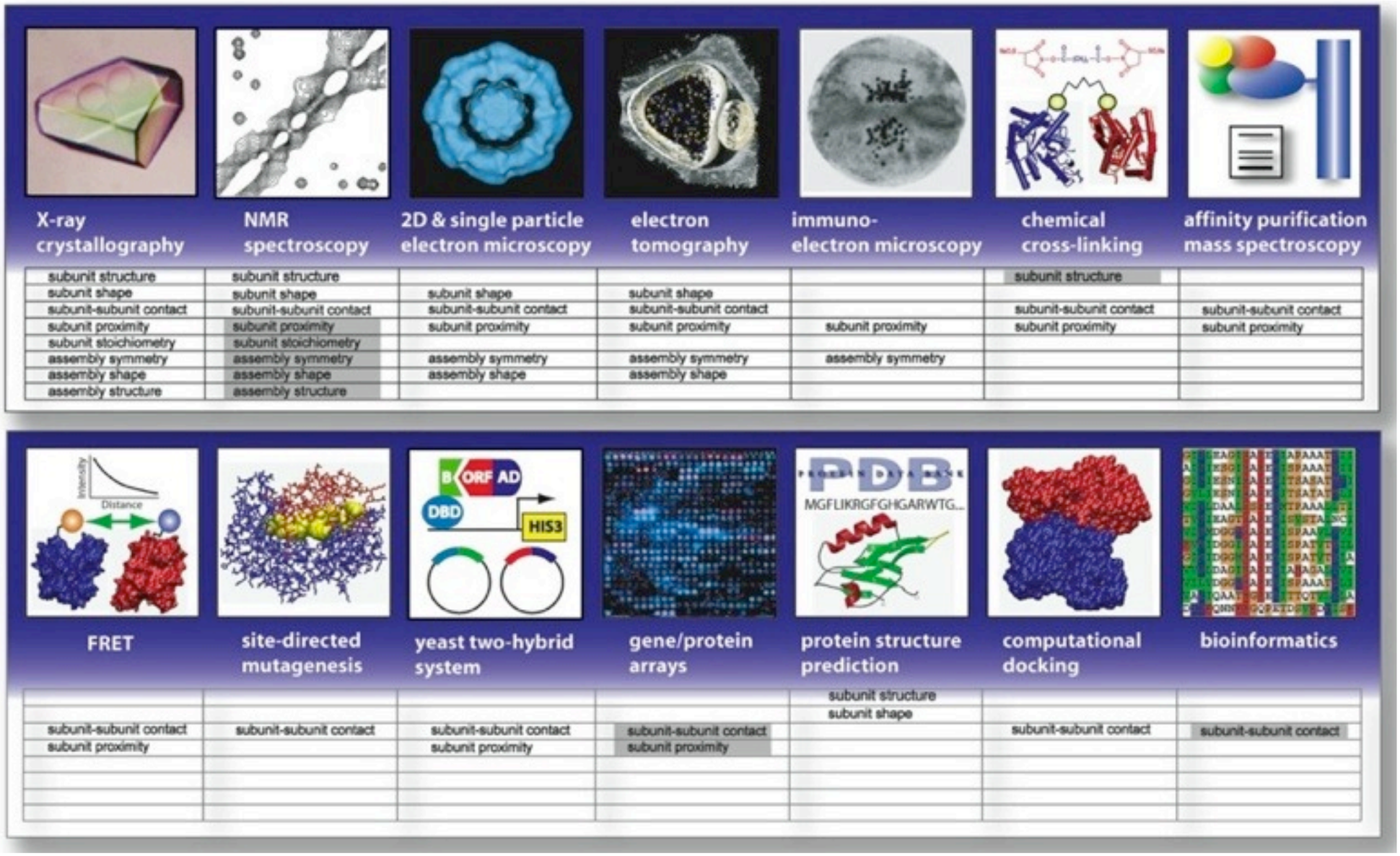
Use structural information from any

source: measurement, first principles, rules;

resolution: low or high resolution

to obtain the set of all models that are consistent with it.





Sali, Earnest, Glaeser, Baumeister. From words to literature in structural proteomics. *Nature* 422, 216-225, 2003.

An approach to integrative structural biology

Alber *et al*. *Nature* **450**, 683-694, 2007

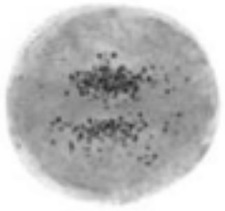
Robinson, Sali, Baumeister. *Nature* **450**, 974-982, 2007

Alber, Foerster, Korkin, Topf, Sali. *Annual Reviews in Biochemistry* **77**, 11.1–11.35, 2008

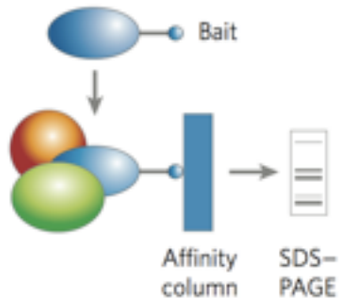
Russel *et al*. *PLoS Biology* **10**, 2012

Gathering information

Cryo-electron microscopy

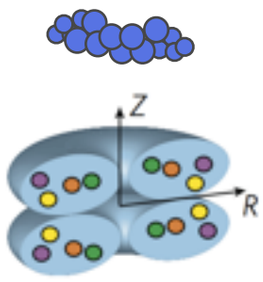


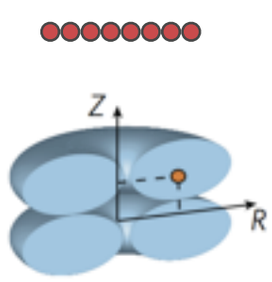
Immuno-electron microscopy

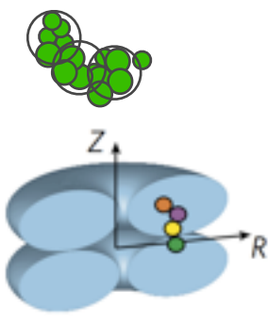


Affinity purification

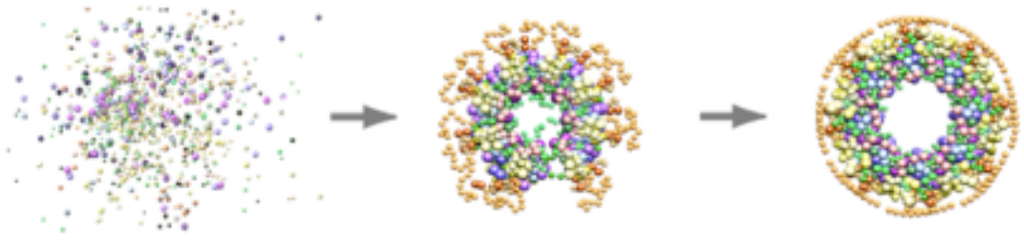
Designing model representation and evaluation



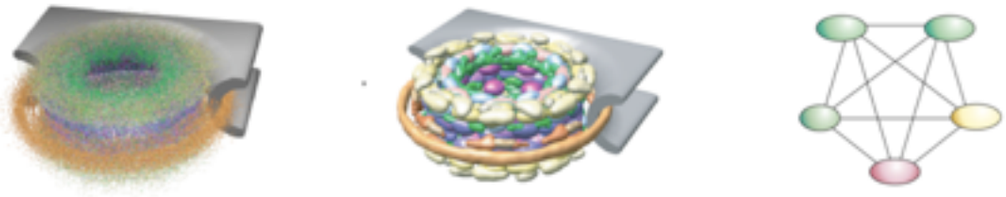




Sampling good models

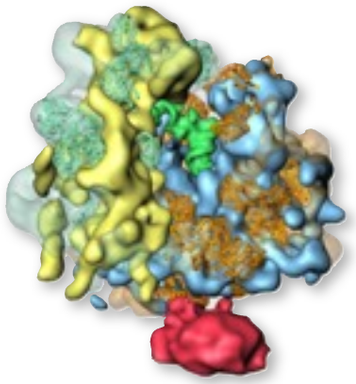


Analyzing models and information

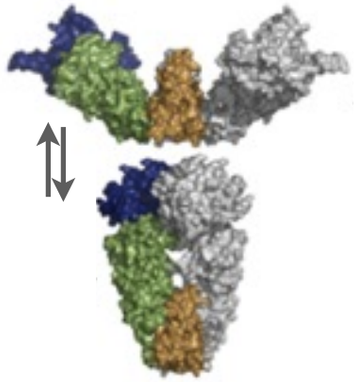


While it may be hard to live with generalization, it is inconceivable to live without it. *Peter Gay*, Schnitzler's Century (2002).

Some IMP applications

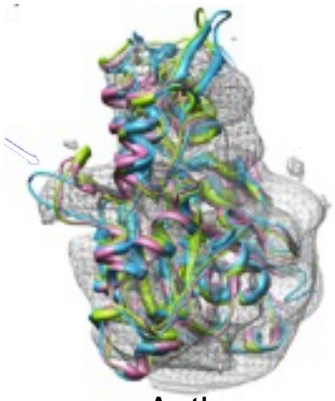


Ribosomes,
w/ Frank, w/ Akey

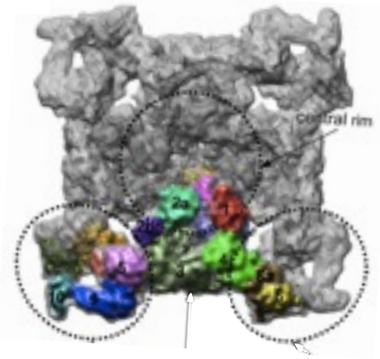


Hsp90 landscape
w/ Agard

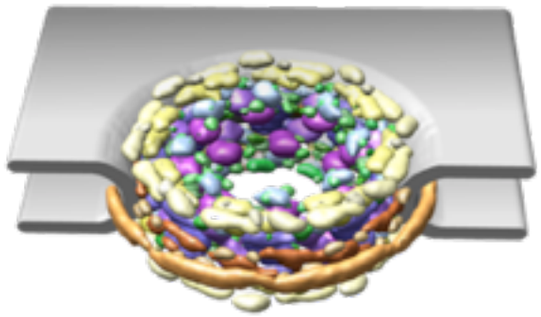
TRiC/CCC
w/ Frydman, Chiu



Actin
w/ Chiu

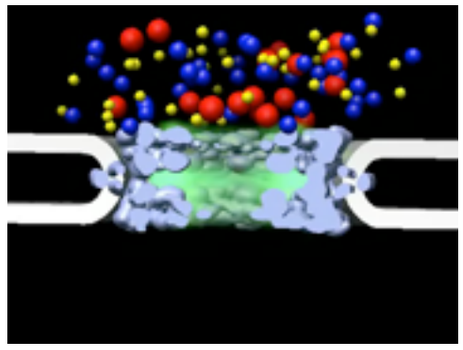


RyR channel
w/ Serysheva, Chiu

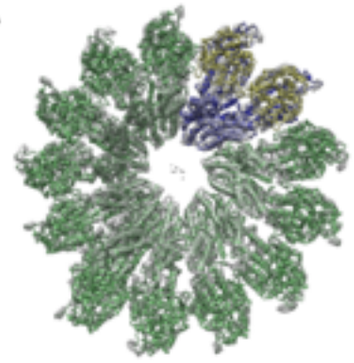


Nuclear Pore Complex,
w/ Rout, Chait

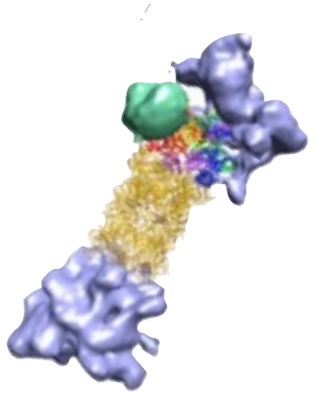
Nup84 complex,
w/ Rout, Chait



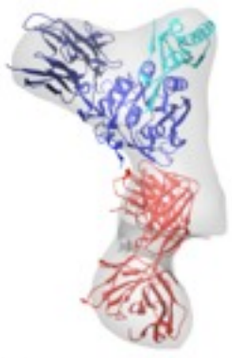
Nuclear Pore Complex transport,
w/ Rout, Chait, Cowburn, Aitchison,
Chook, Liphardt



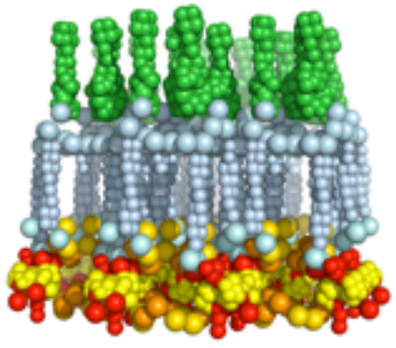
Microtubule nucleation
w/ Agard



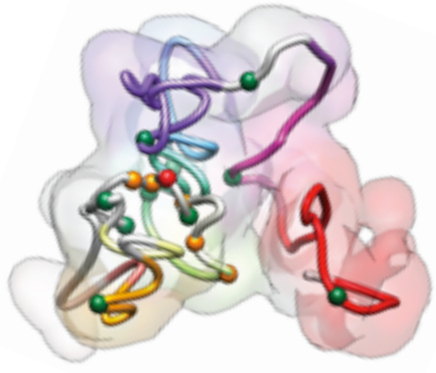
26 Proteasome
w/ Baumeister



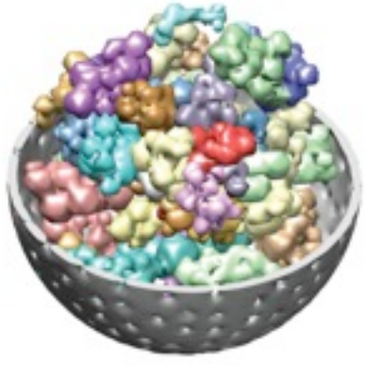
PCS9K-Fab complex
w/ Cheng, Agard, Pons



Spindle Pole Body
w/ Davis, Muller



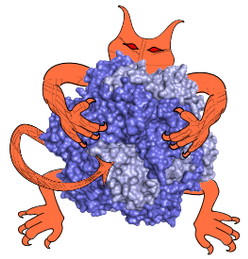
Chromatin globin domain
Marti-Renom



Lymphoblastoid cell genome
Alber, Chen

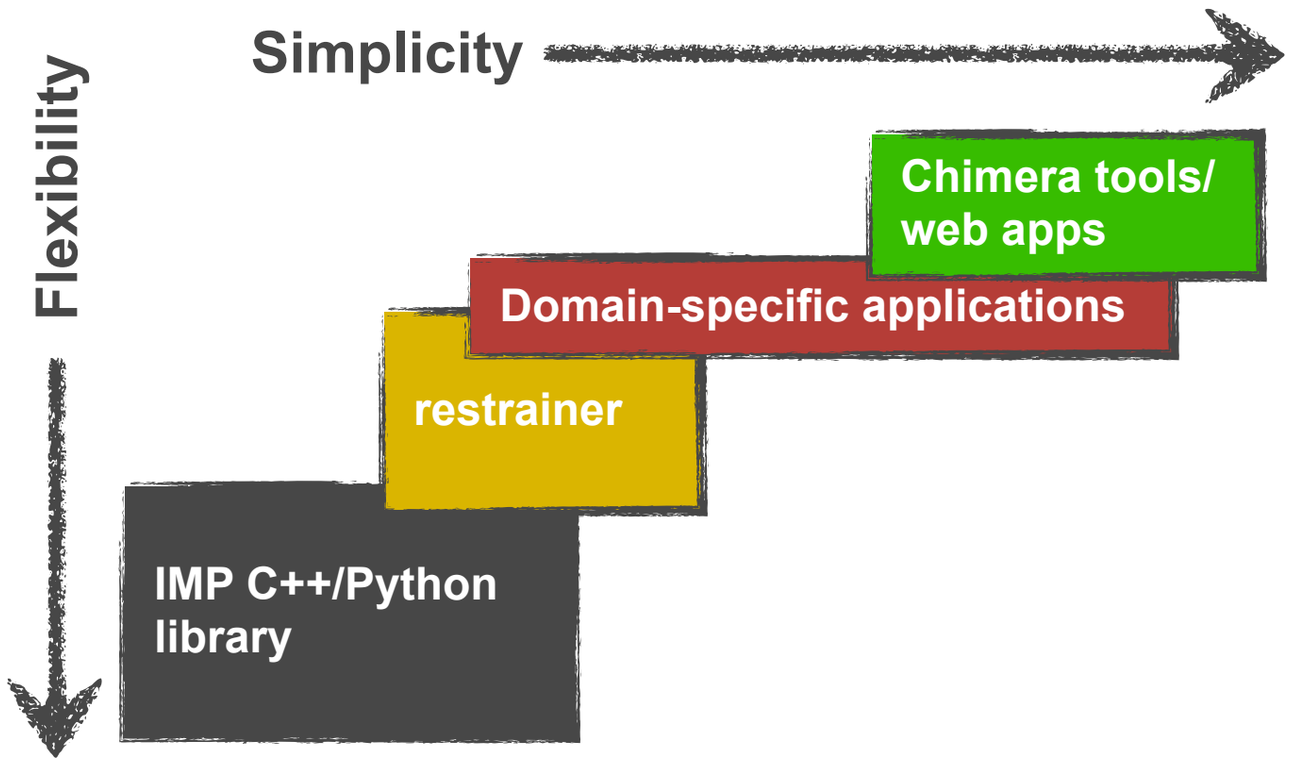
Integrative Modeling Platform (IMP)

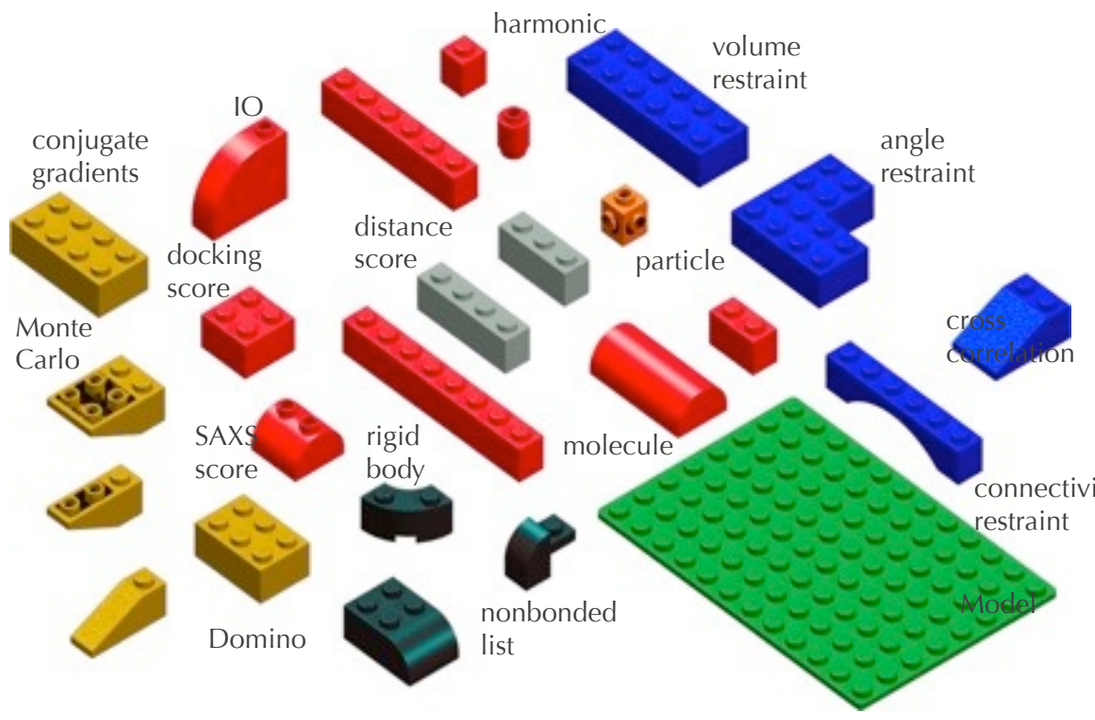
http://integrativemodeling.org



D. Russel, K. Lasker, B. Webb, J. Velazquez-Muriel, E. Tijoe, D. Schneidman, F. Alber, B. Peterson, A. Sali, ***PLoS Biol***, 2012.

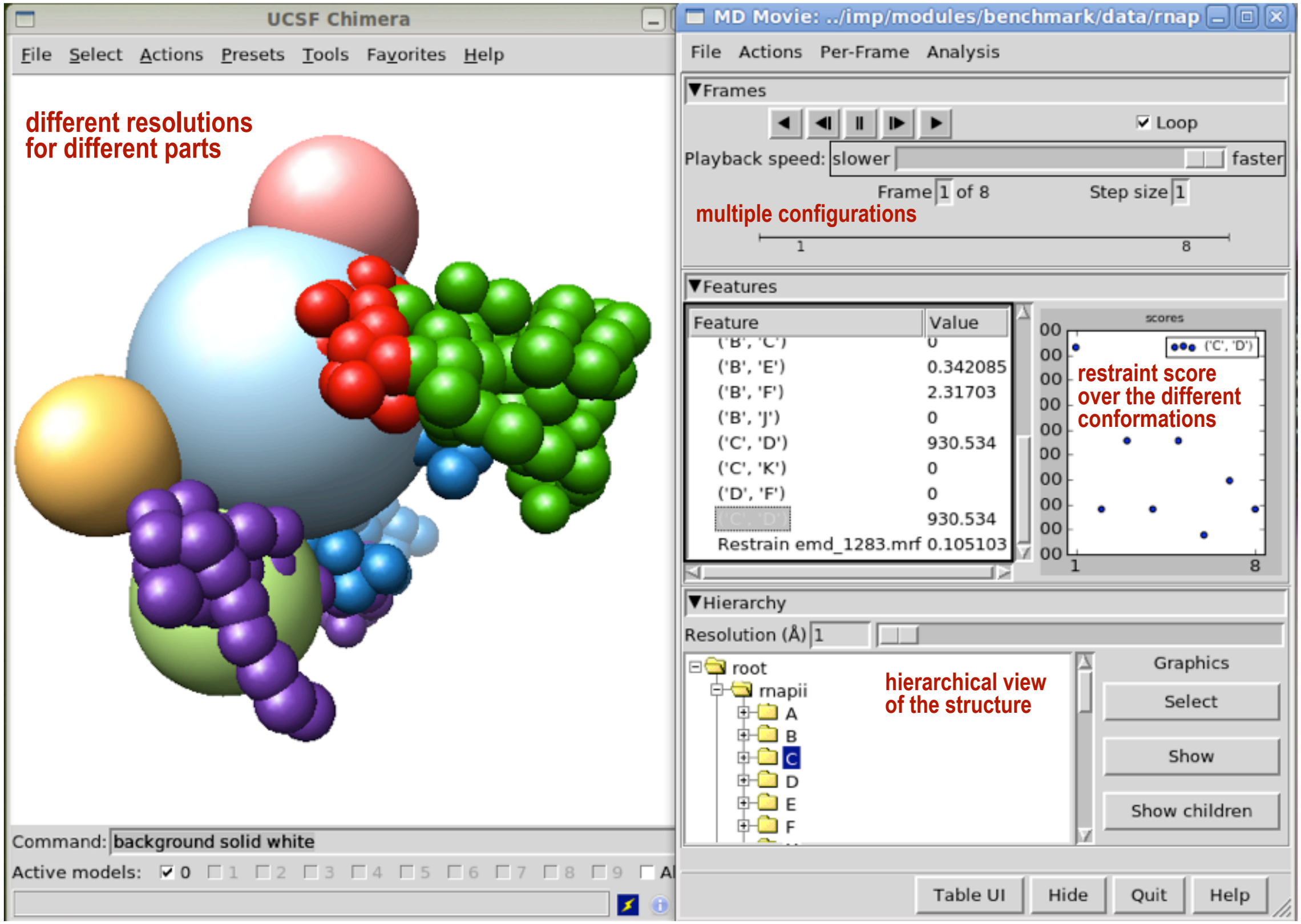
Open source, versions, documentation, wiki, examples, mailing lists, unit testing, bug tracking, ...





Chimera and IMP

Yang Z, Lasker K, Schneidman-Duhovny D, Webb B, Huang C, Pettersen E, Goddard T, Meng E, Sali A, Ferrin T. *J Struct Biol*, 2011



Contents

1. Integrative (hybrid) structure determination
2. EM images as a source of spatial restraints

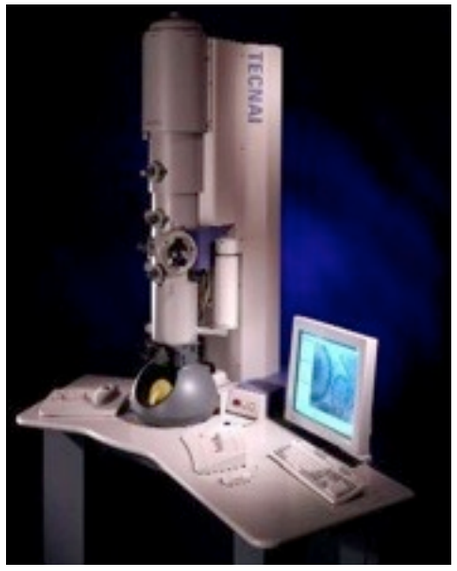
   Application to the Nup84 complex
3. Multiple fitting of subunits into an EM map of the whole assembly

   Application to the 26S proteasome

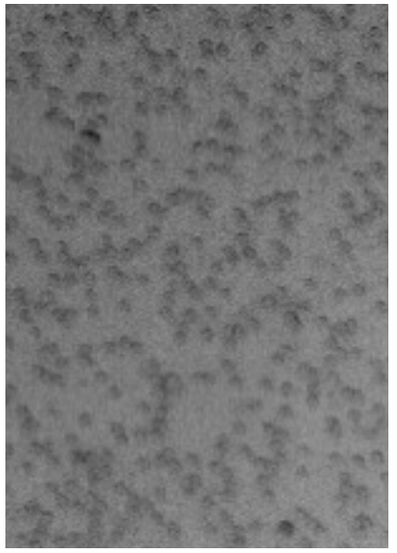
3D-EM process

Specimen

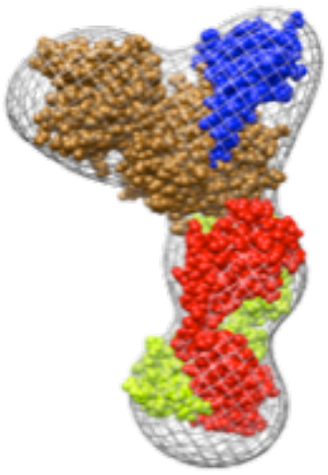
Microscope



Digitized micrograph



Particle selection, alignment, classification, averaging

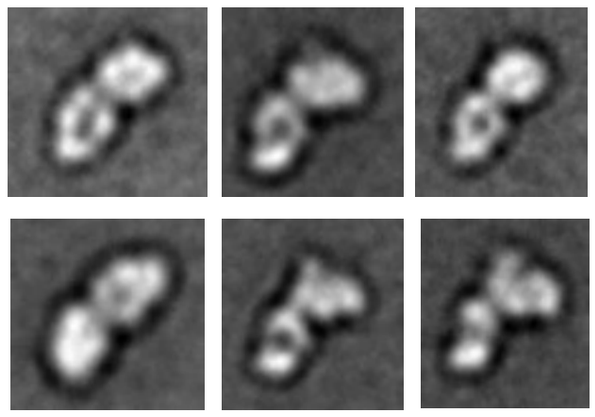


Model

Fitting (& modeling)

3D map

3D reconstruction



2D class averages

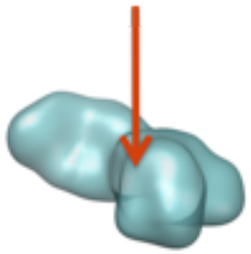
Additional restraints

Velazquez *et al*. PNAS, 2012

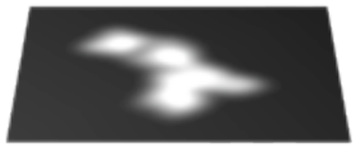
Scoring: Comparison of EM image and model

Velazquez *et al*. PNAS, 2012

model density

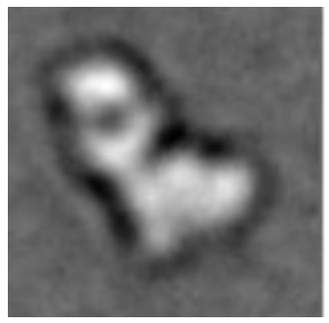


model projection



P(**m**,α)

Class average



d

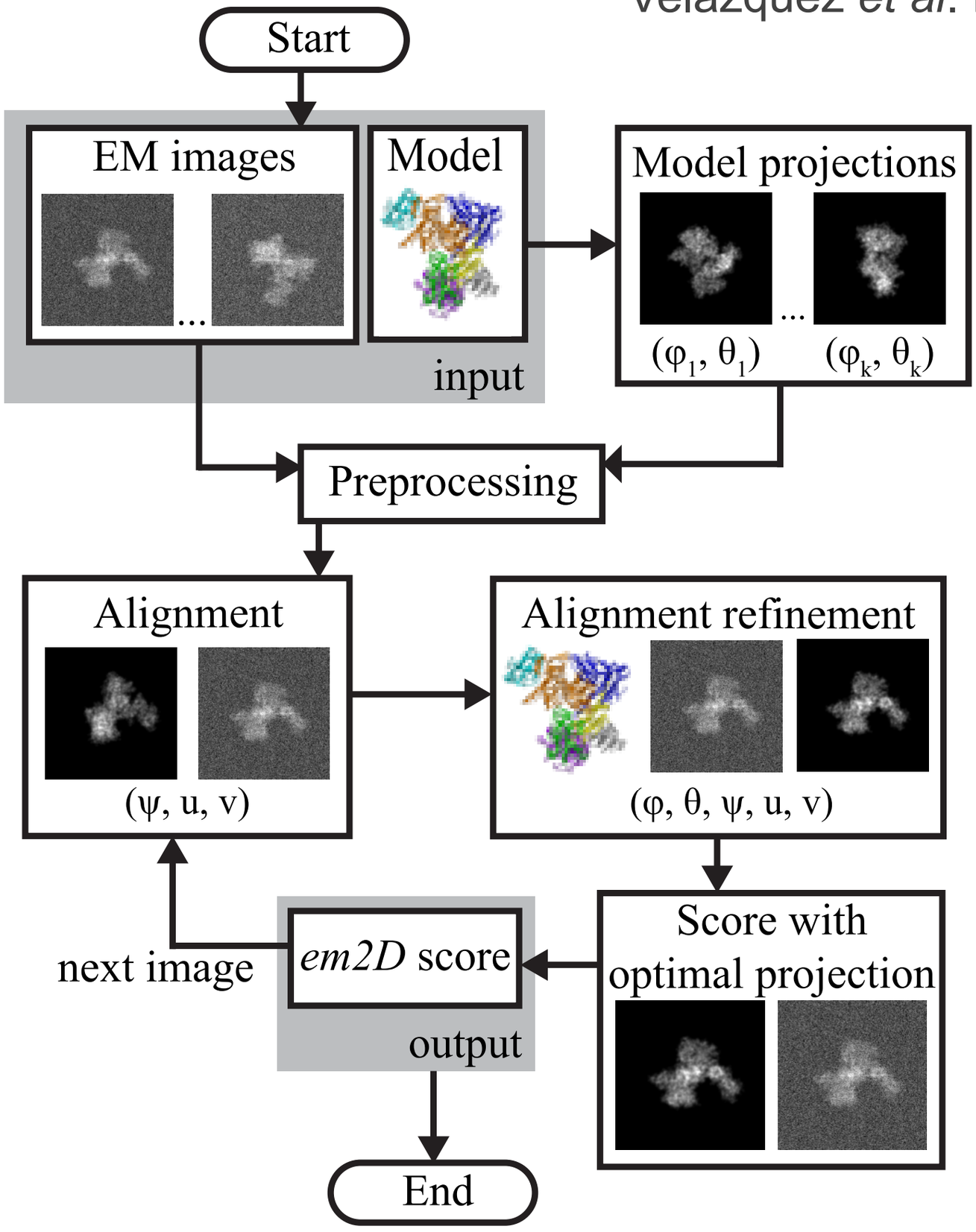
Correlation between an image and closest model projection:

$$em2D = 1 - \max_{\alpha} corr(\mathbf{P}(\mathbf{m},\alpha),\mathbf{d})$$

- An EM image (*eg*, class average) **d** is compared with the most overlapping projection **P**(**m**,α) of a downsampled model **m**.
- This optimal projection is found by optimization over three orientation angles and two translation distances, α.
- Can be easily extended to tilt series of images to improve data-to-parameter ratio.
- It may be possible to address conformational and configurational homogeneity.

Scoring: Comparison of EM image and model

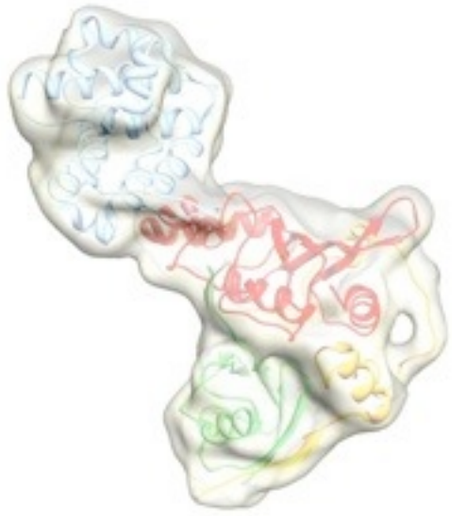
Velazquez *et al*. PNAS, 2012



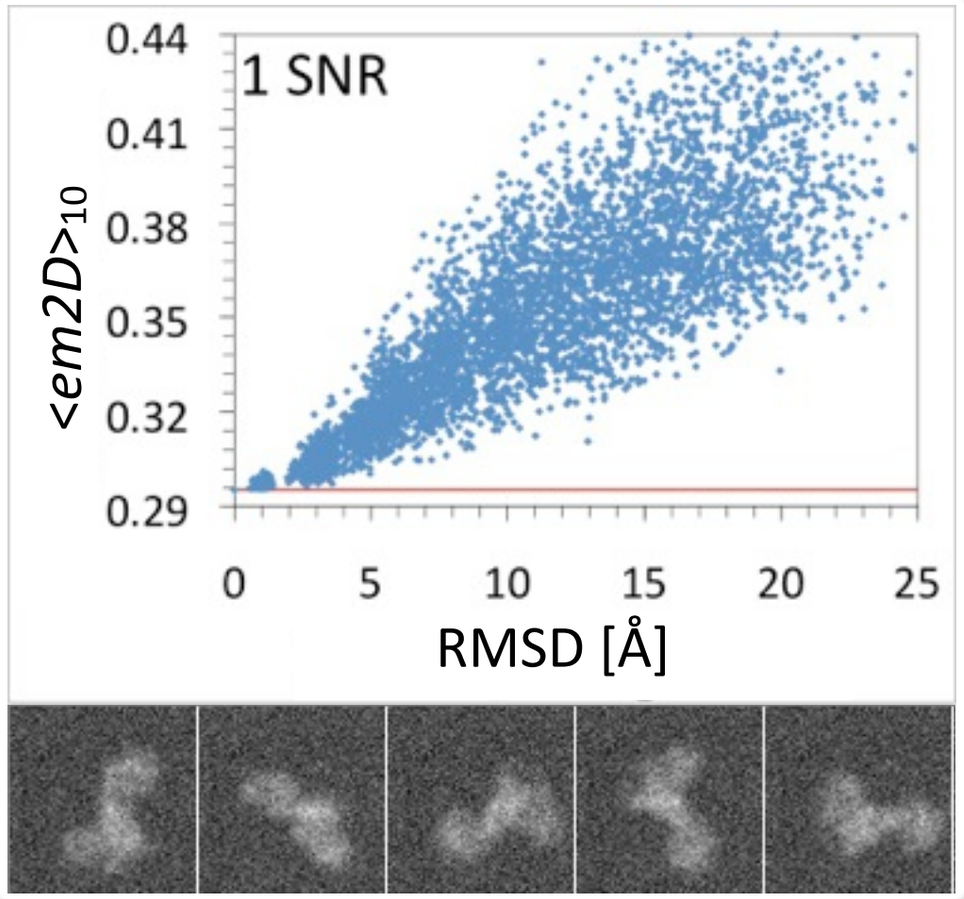
Test with simulated data

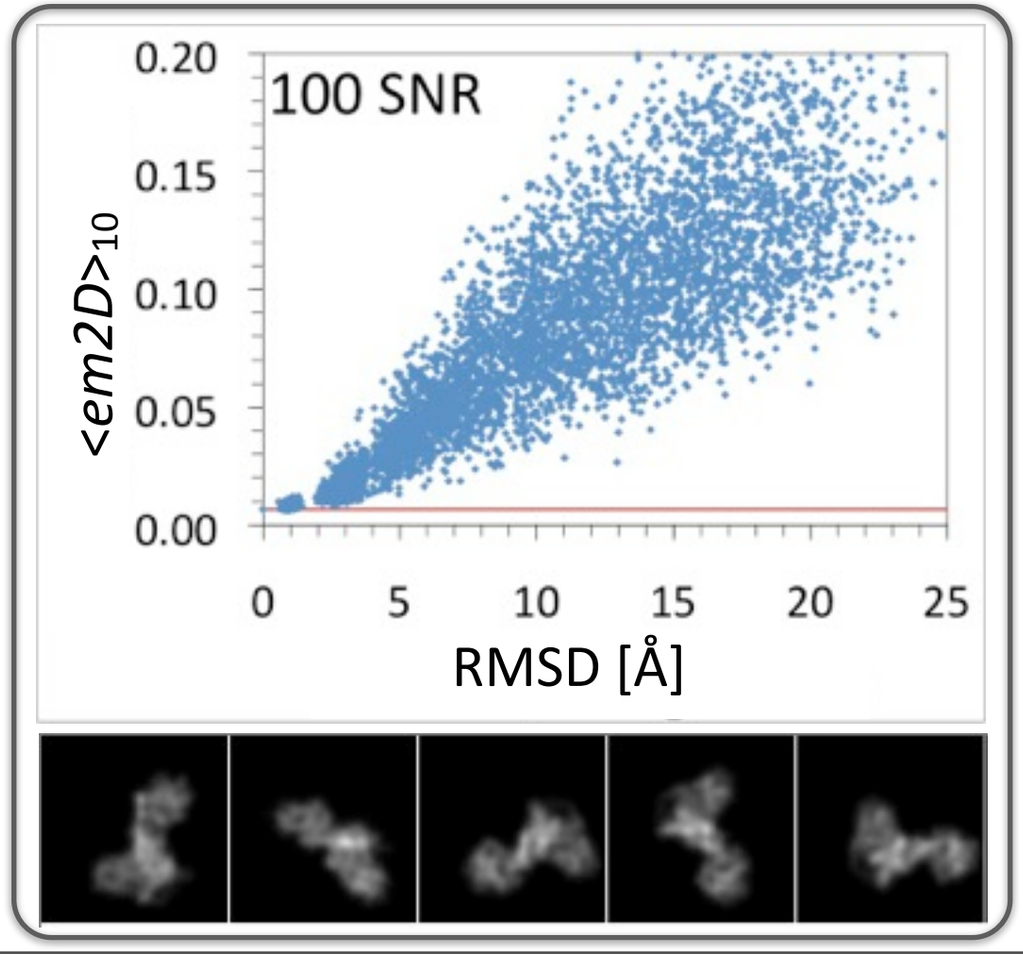
Velazquez *et al*. PNAS, 2012

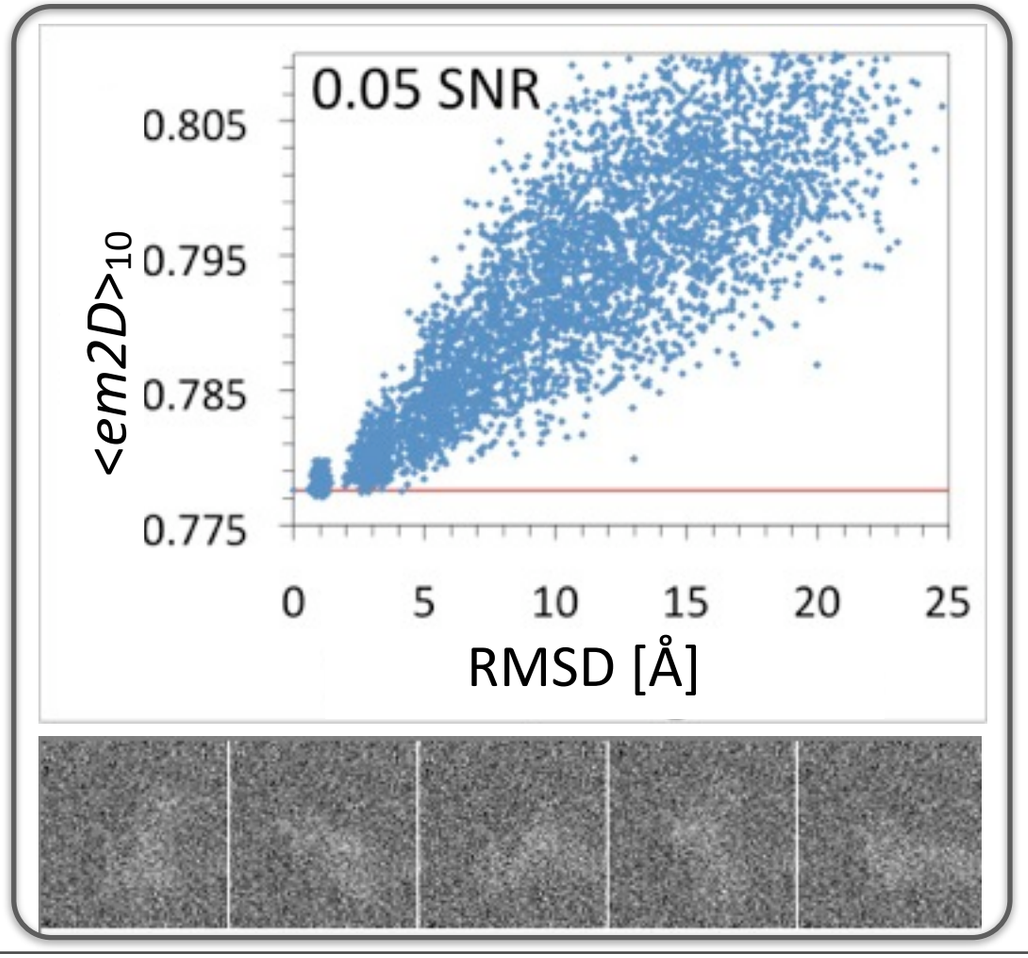
PDB 1z5s:
SUMO - RanGAP1 - Ubc9 - Nup358



5 Å

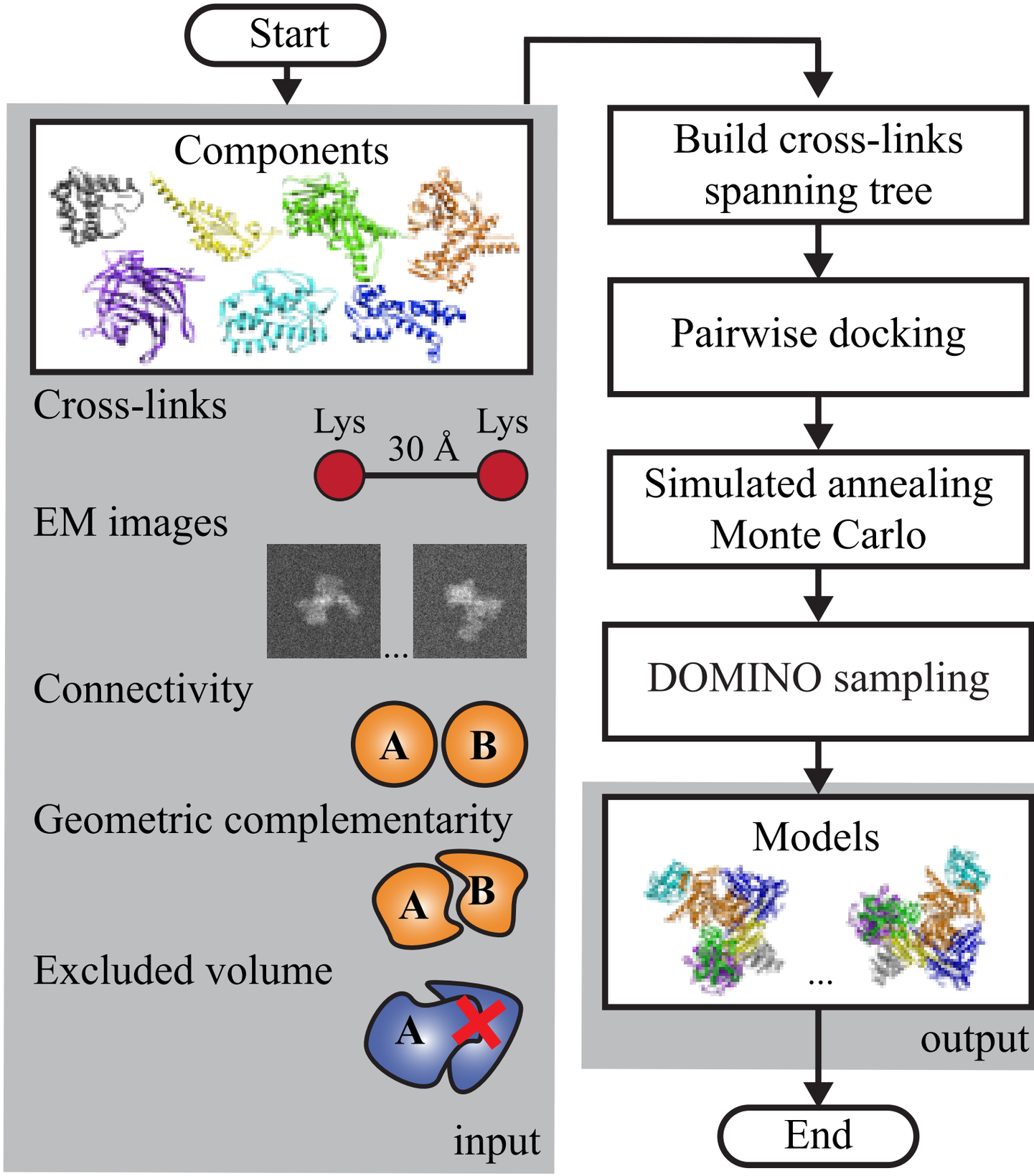






Configurational sampling protocol

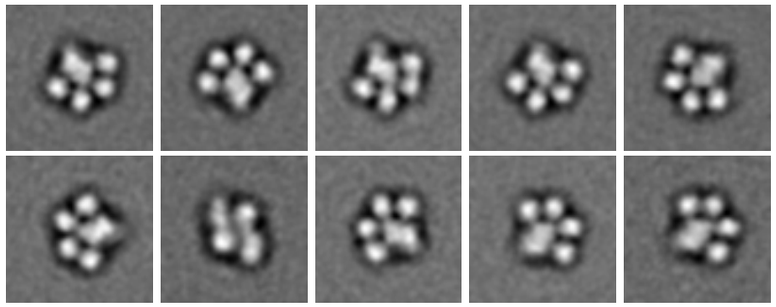
Velazquez *et al*. PNAS, 2012



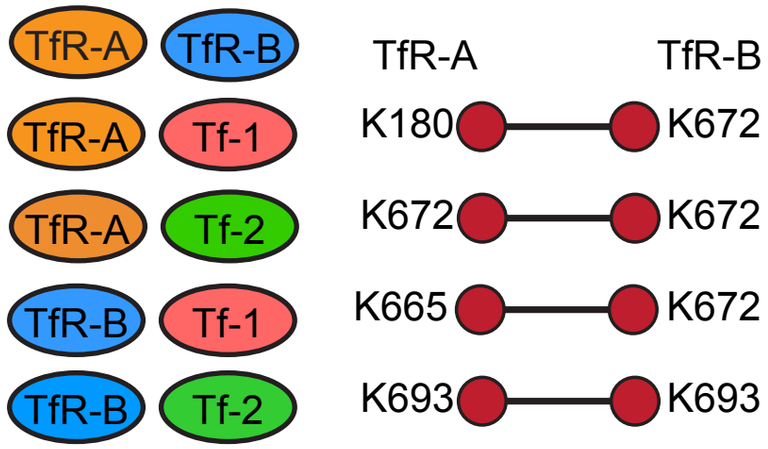
Transferrin-Transferrin Receptor complex

Velazquez *et al*. PNAS, 2012

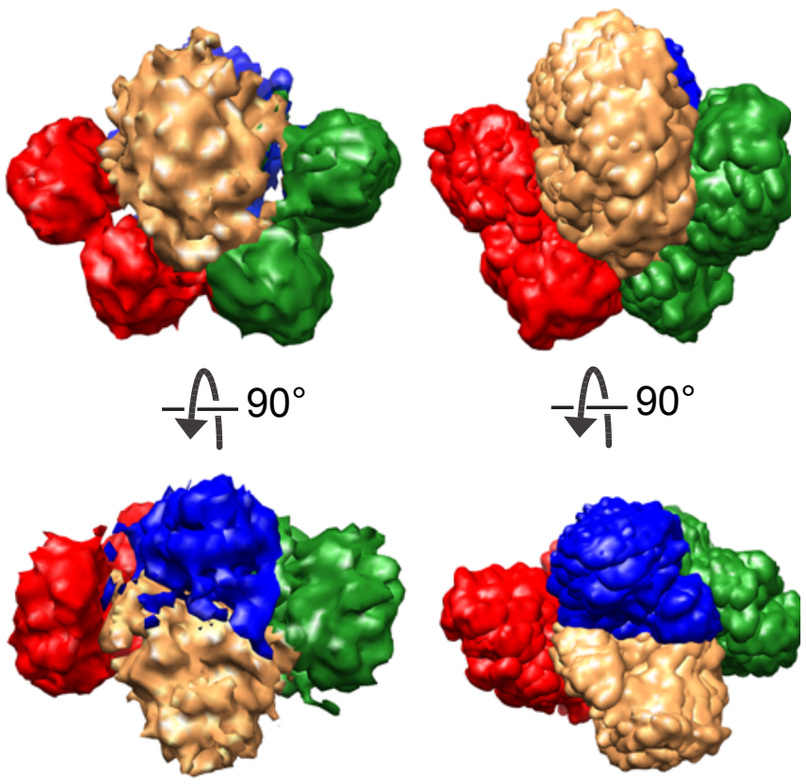
A)



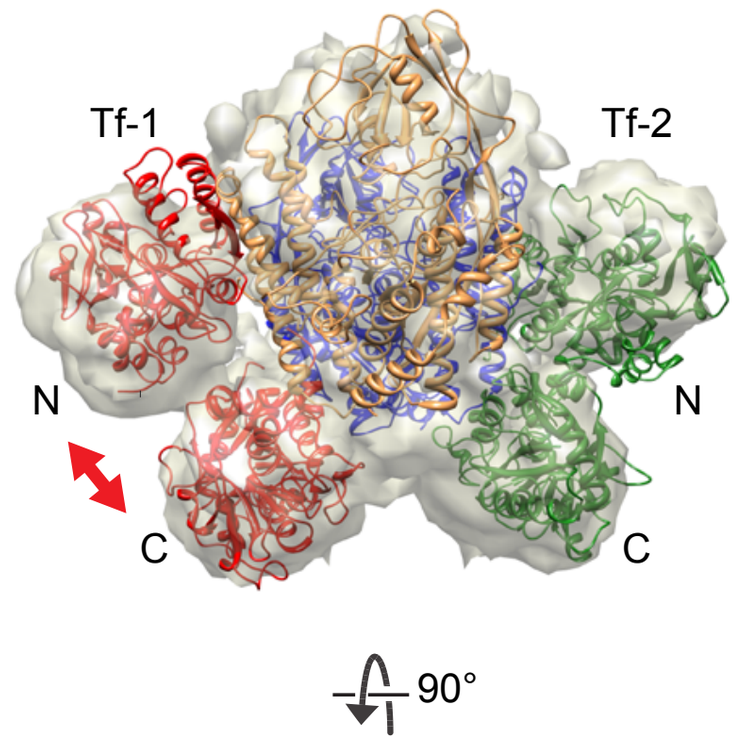
B)

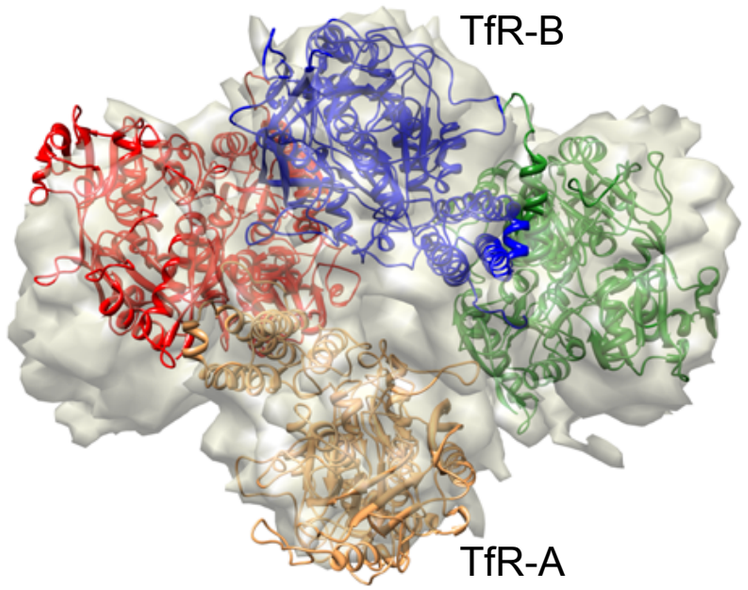


C)



D)

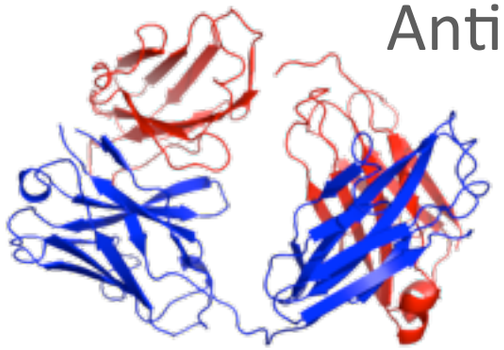




Y. Cheng *et al*, 2004

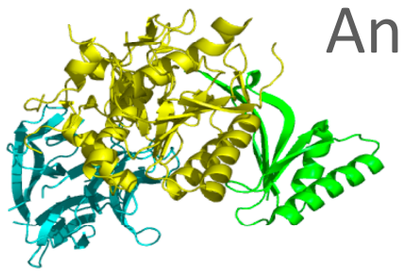
Application to an antigen - antibody complex

Schneidman-Duhovny D, Rossi A, Avila-Sakar A, Kim SJ, Velazquez-Muriel J, Strop P, Liang H, Krukenberg KA, Liao M, Kim HM, Sobhanifar S, Dotsch V, Raipal A, Pons J, Agard DA, Cheng Y, Sali A. A Method for Integrative Structure Determination of Protein-Protein Complexes. Bioinformatics, 2012.



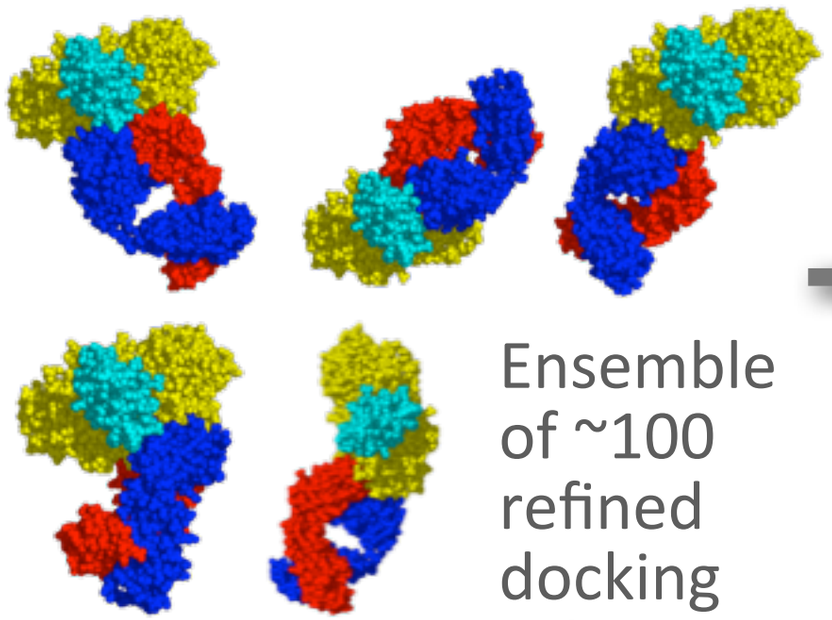
Antibody

+



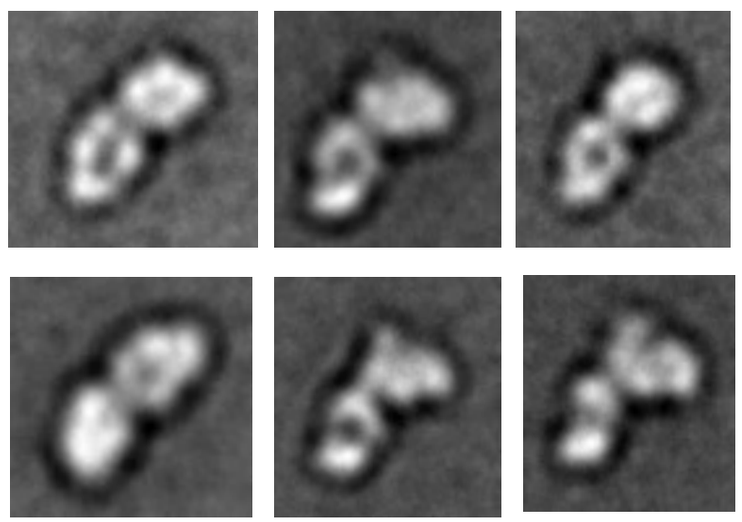
Antigen

Docking by PatchDock, FireDock

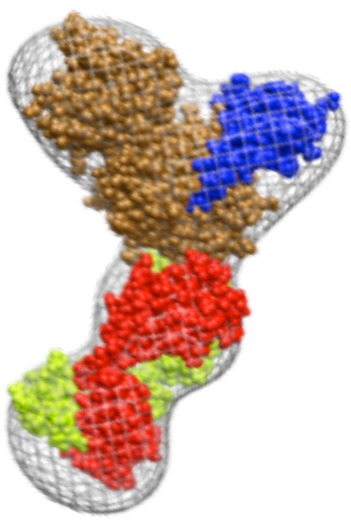


Ensemble of ~100 refined docking solutions

2D class averages from negative stain EM



Filtering of docking models



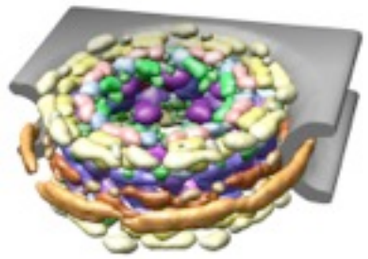
Final model; confirmed by 3D reconstruction and X-ray crystallography

Contents

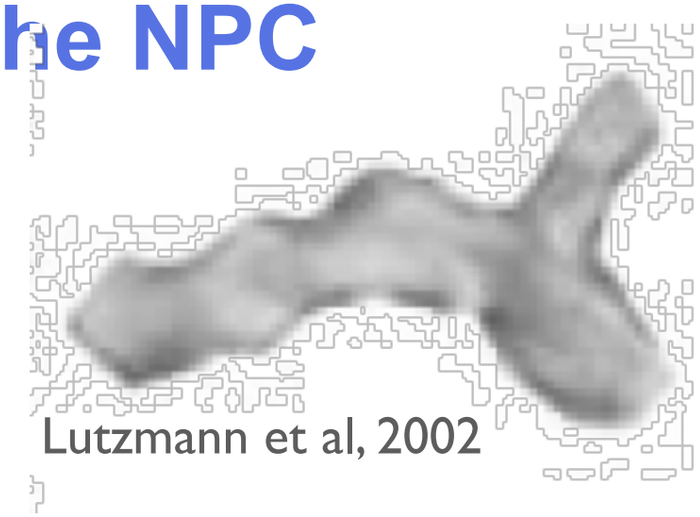
1. Integrative (hybrid) structure determination
2. EM images as a source of spatial restraints

   Application to the Nup84 complex
3. Multiple fitting of subunits into an EM map of the whole assembly

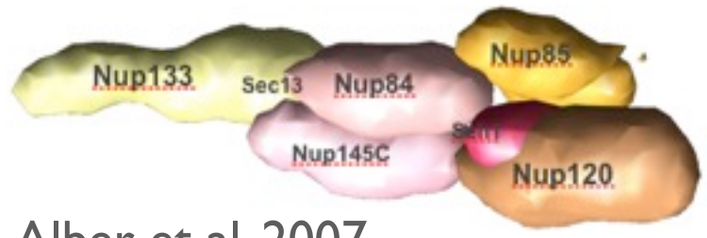
   Application to the 26S proteasome



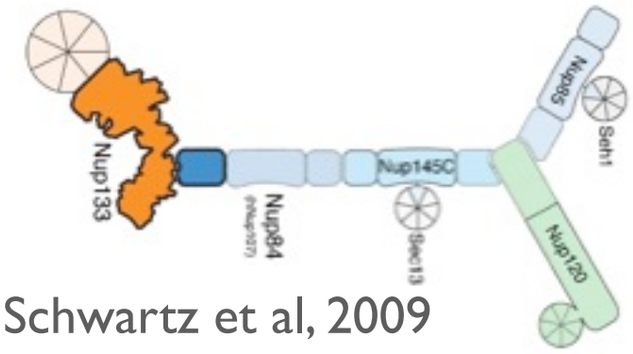
The Nup84 complex in the NPC



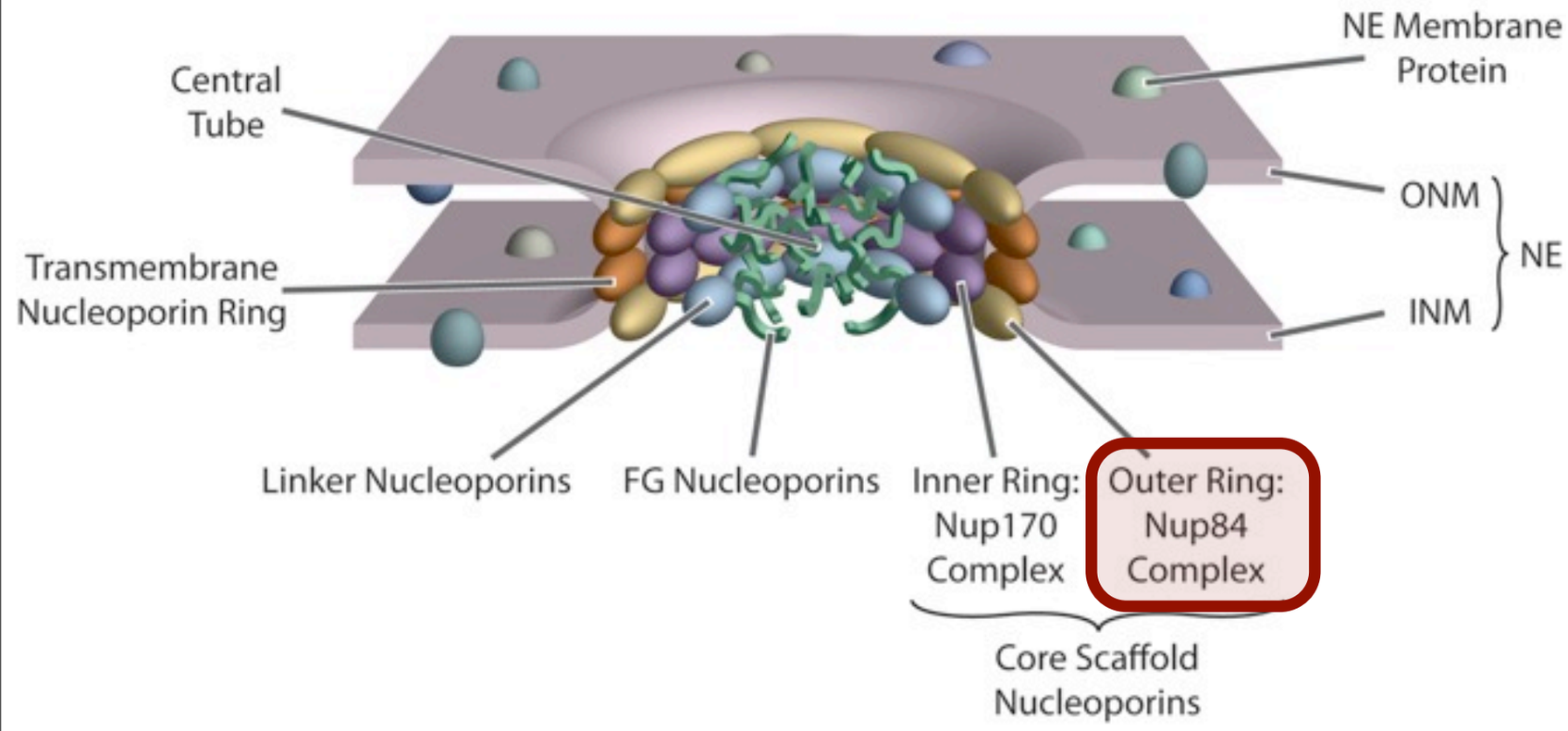
Lutzmann et al, 2002



Alber et al, 2007



Schwartz et al, 2009

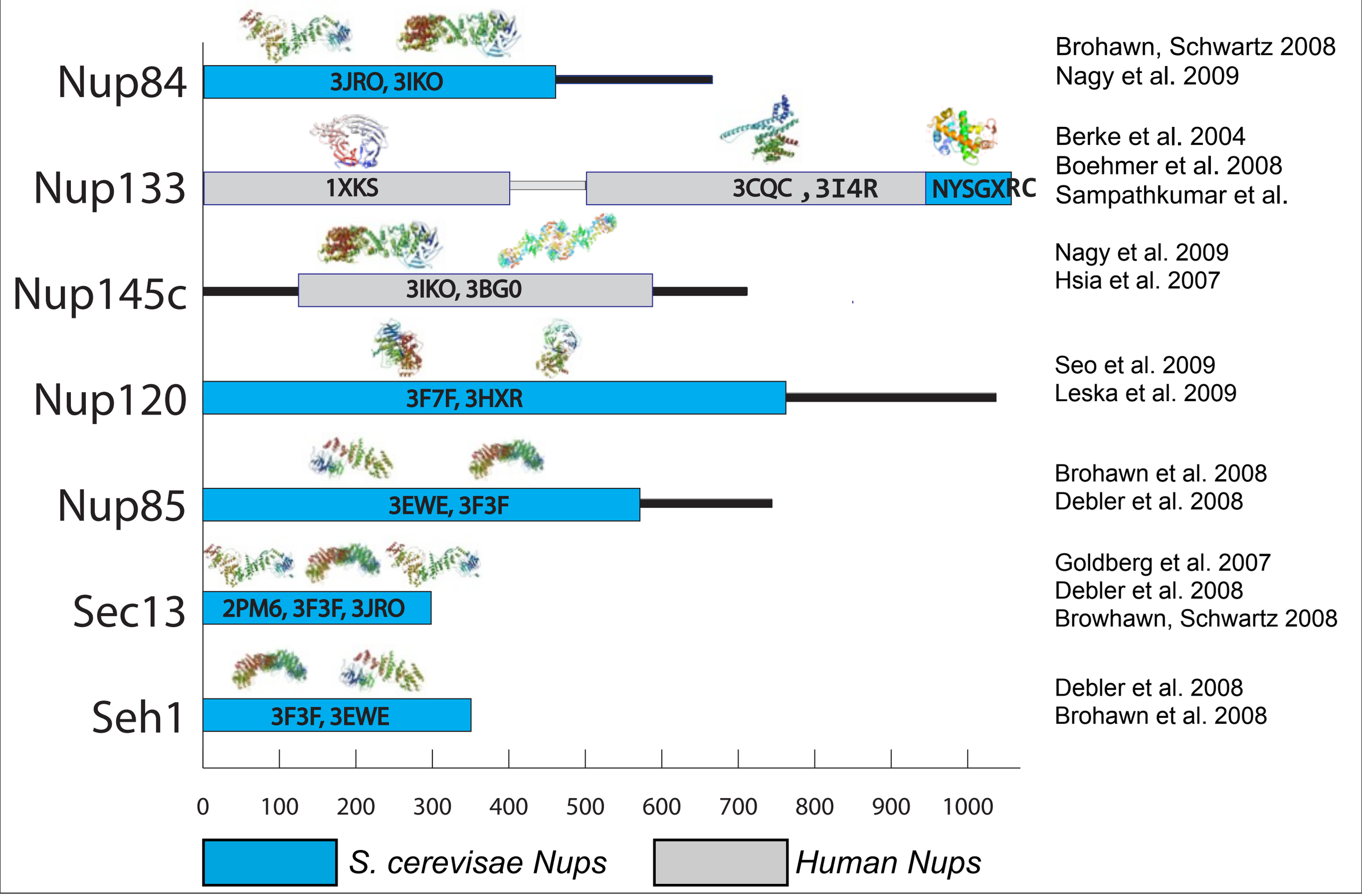


Kampmann et al, 2009

- 7-protein complex
- Forms the two outer rings of the NPC
- Present in 16 copies in the NPC
- Proteins share a common ancestor with vesicle coating complexes

Nup84 complex: Representation

Fernandez, Phillips, ... Stokes, Chait, Rout. JCB, 2012



Brohawn, Schwartz 2008
Nagy et al. 2009

Berke et al. 2004
Boehmer et al. 2008
Sampathkumar et al.

Nagy et al. 2009
Hsia et al. 2007

Seo et al. 2009
Leska et al. 2009

Brohawn et al. 2008
Debler et al. 2008

Goldberg et al. 2007
Debler et al. 2008
Browhawn, Schwartz 2008

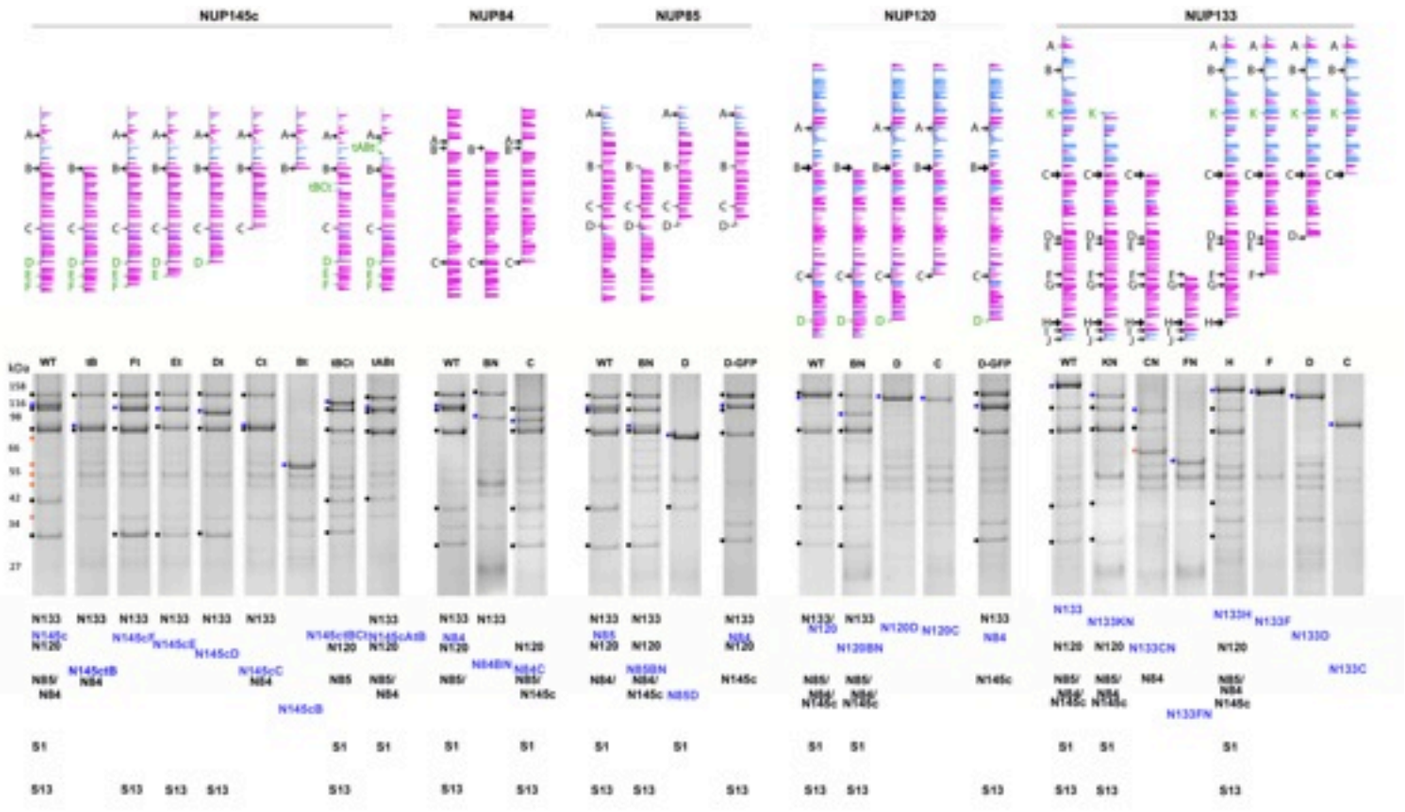
Debler et al. 2008
Brohawn et al. 2008

Nup84 complex: Data

Subunit positions & orientations

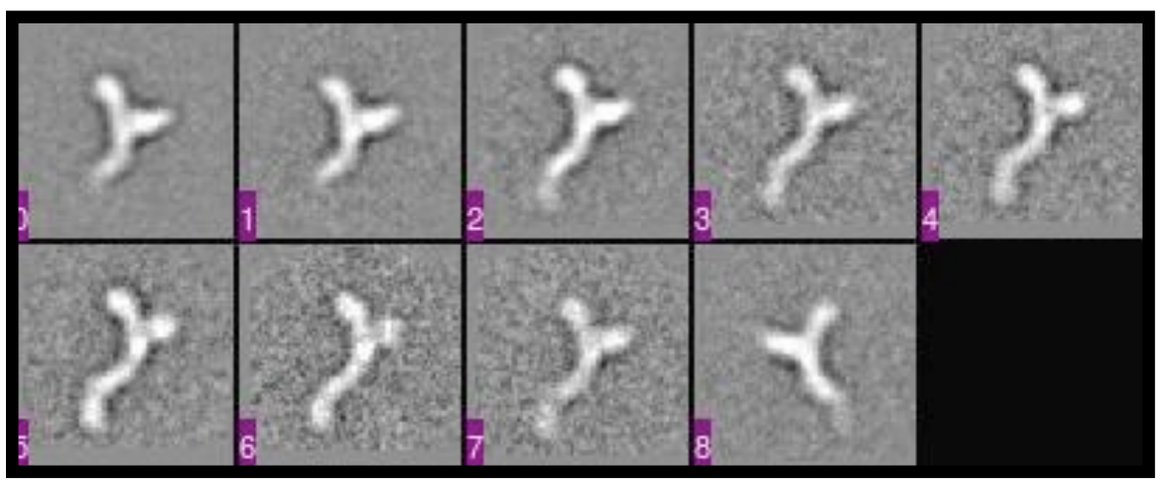
Affinity purifications with domain truncations

J. Fernandez, J. Franke, B. Chait, M. Rout



Negative stain EM particle averages at ~3nm resolution

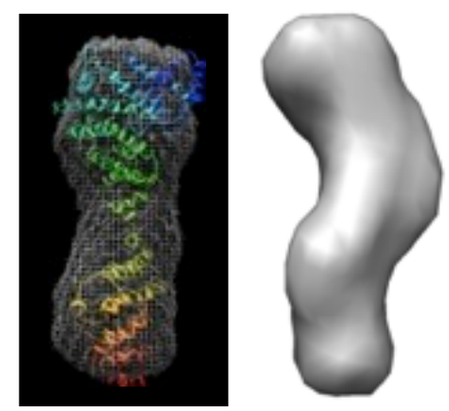
R. Diaz, D. Stokes, J. Velazquez

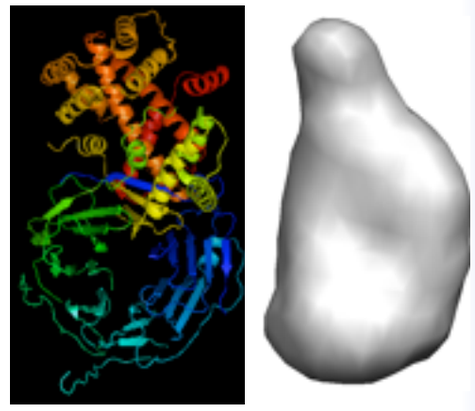


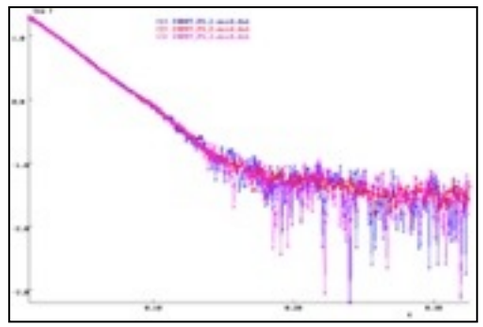
Subunit conformations

Small angle X-ray scattering

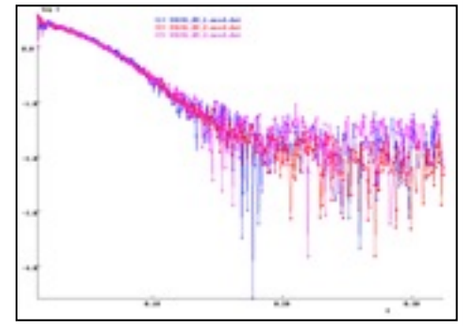
S.J. Kim, A. Martel, H. Tsuruta, NYSGXRC, J. Tainer







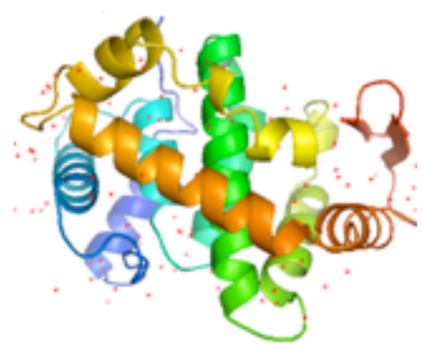
Nup133



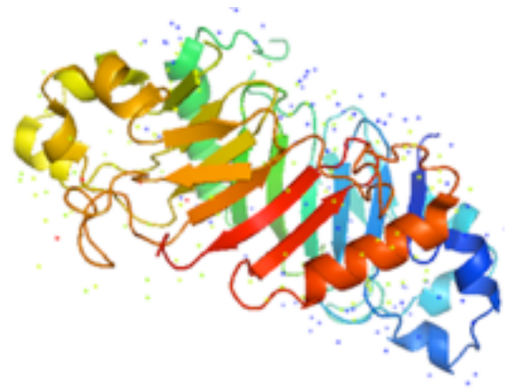
Nup120

High-throughput crystallography

NYSGXRC, P. Sampathkumar, M. Sauder, S. Burley



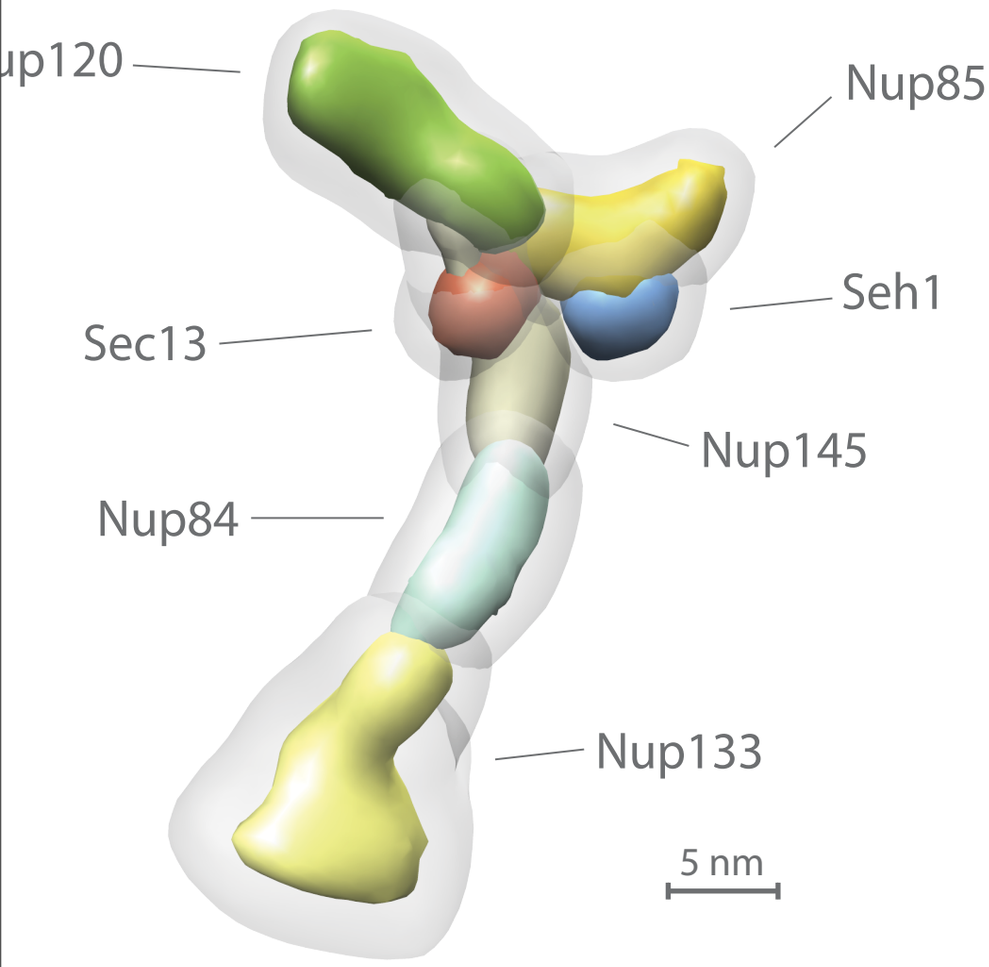
Yeast Nup133

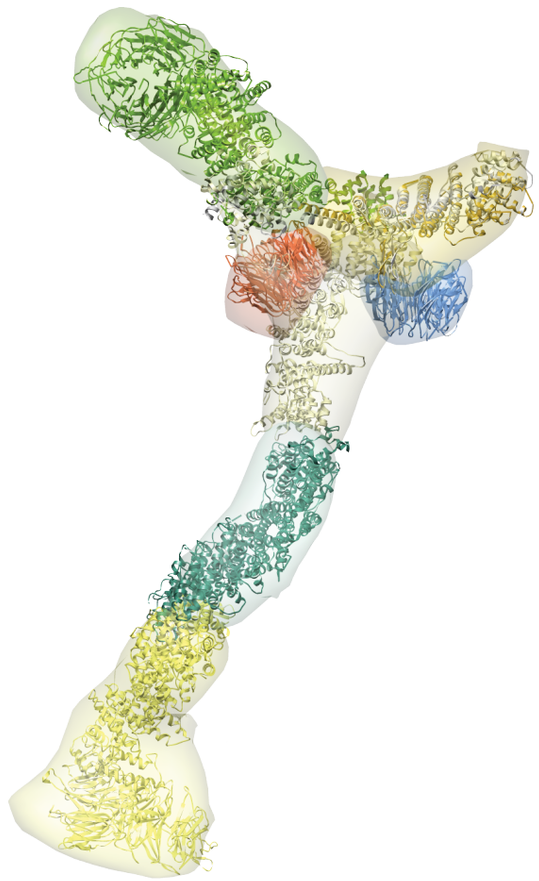


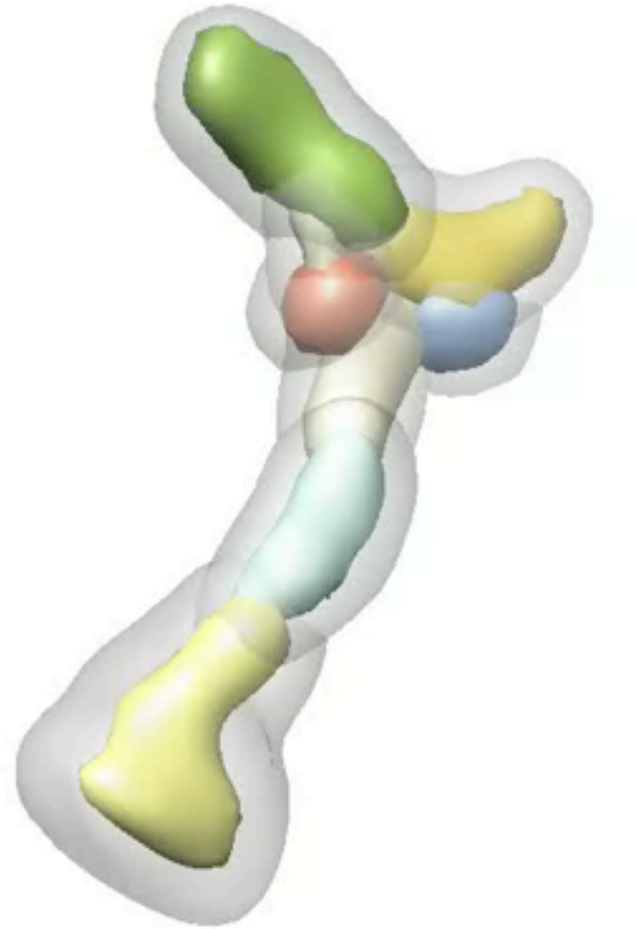
Yeast Nup145

Nup84 complex:
Ensemble of good scoring solutions

Fernandez, Phillips, ... Stokes, Chait, Rout. JCB, 2012

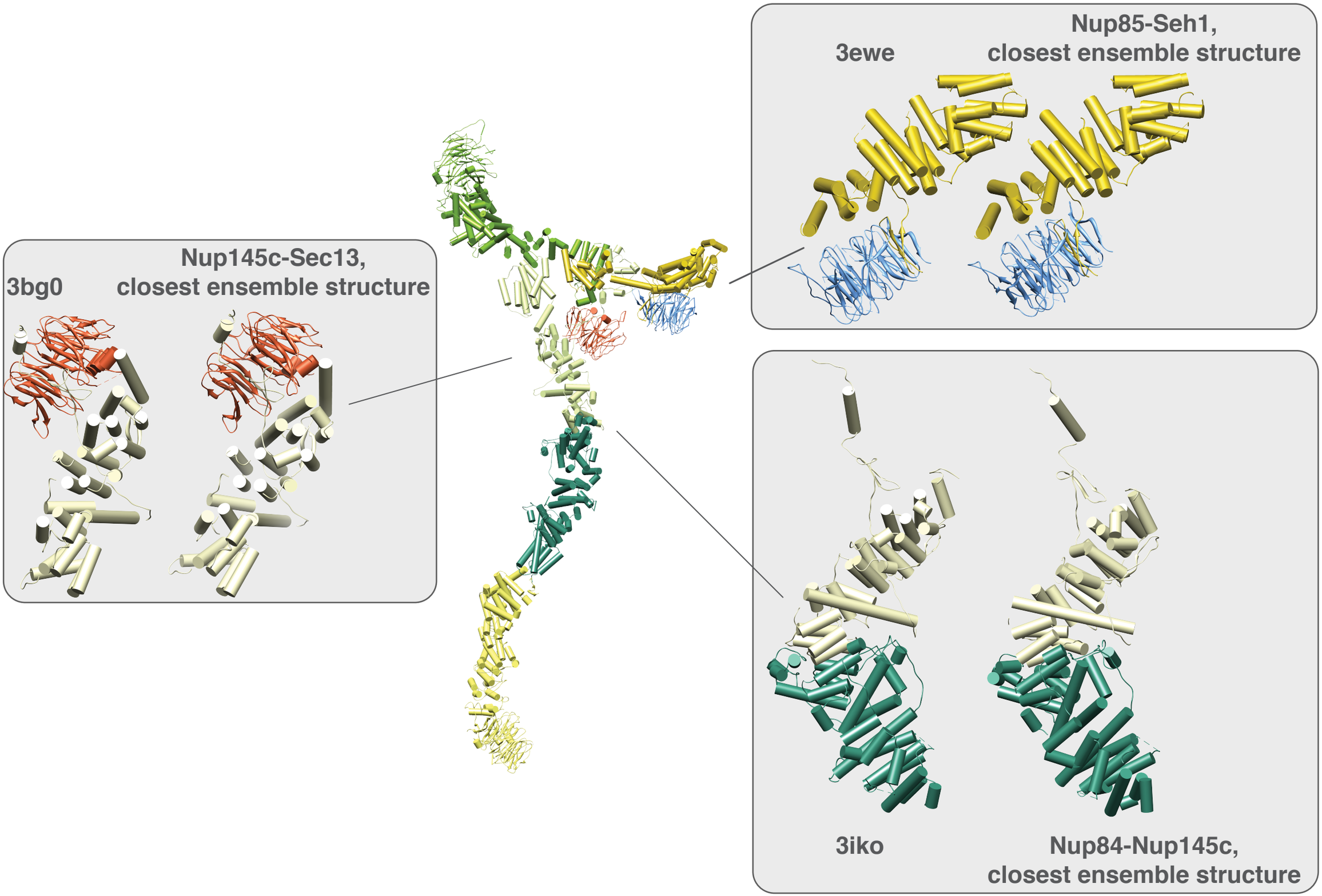






- 10,000 good scoring structures
- All restraints are satisfied (2D-EM, domain deletion, ...)
- Domain-domain orientations are resolved uniquely.
- Full ensemble precision is ~1 nm

Assessment: Agreement with heterodimeric crystallographic structures



Contents

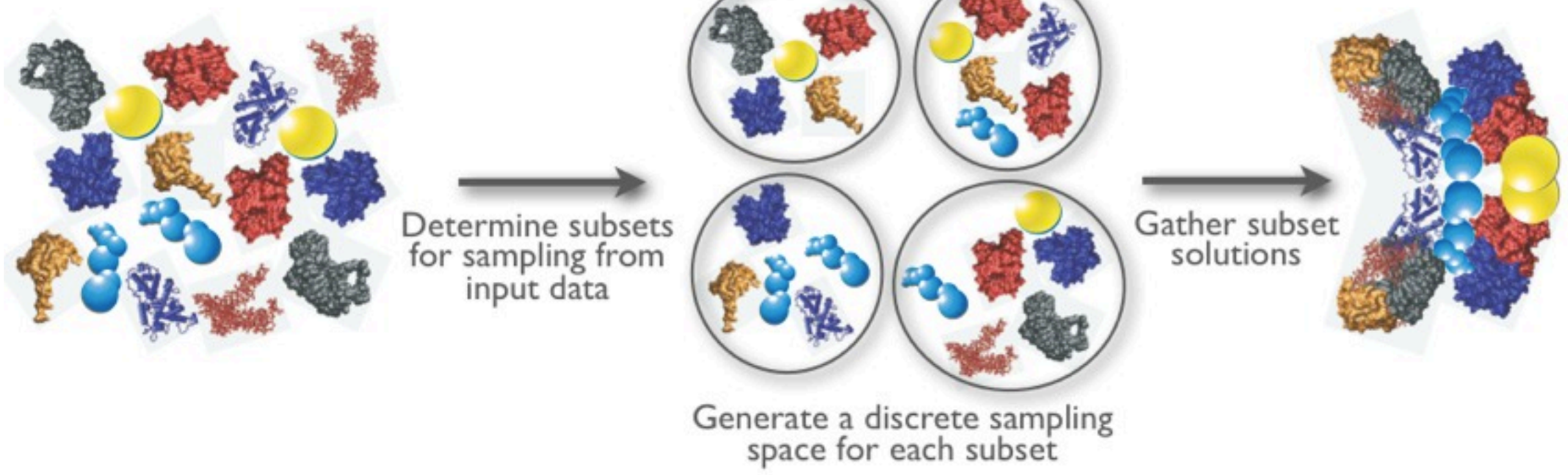
1. Integrative (hybrid) structure determination
2. EM images as a source of spatial restraints

   Application to the Nup84 complex
3. Multiple fitting of subunits into an EM map of the whole assembly

   Application to the 26S proteasome

Sampling

Divide-and-conquer sampling



M.I. Jordan, Graphical models. *Stat. Sci*. **19**, 140–155, 2004.
K. Lasker, M. Topf, A. Sali, **H. Wolfson**, J. Mol. Biol. 388, 180-194, 2009.

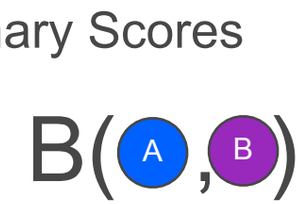
Divide-and-conquer sampling in IMP

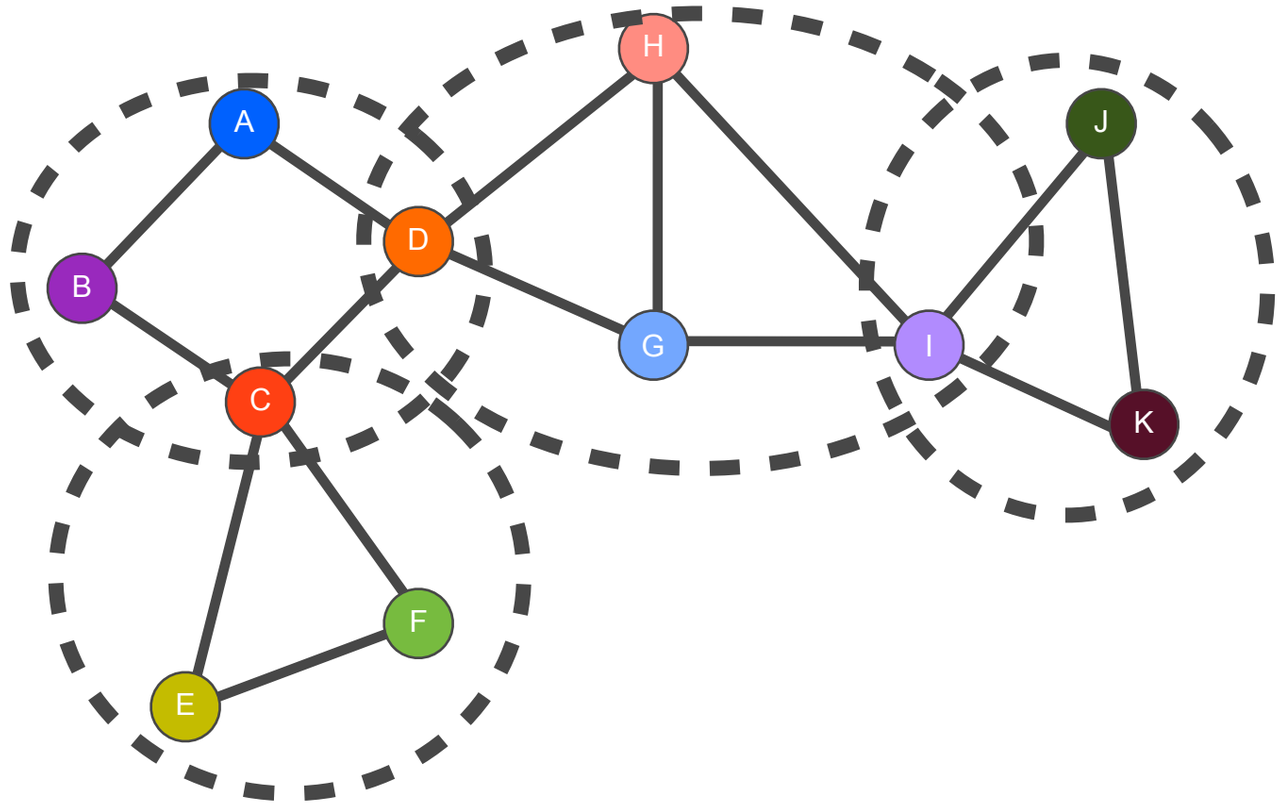
K. Lasker, D. Russel, B. Webb

System components

Unary Scores

Binary Scores

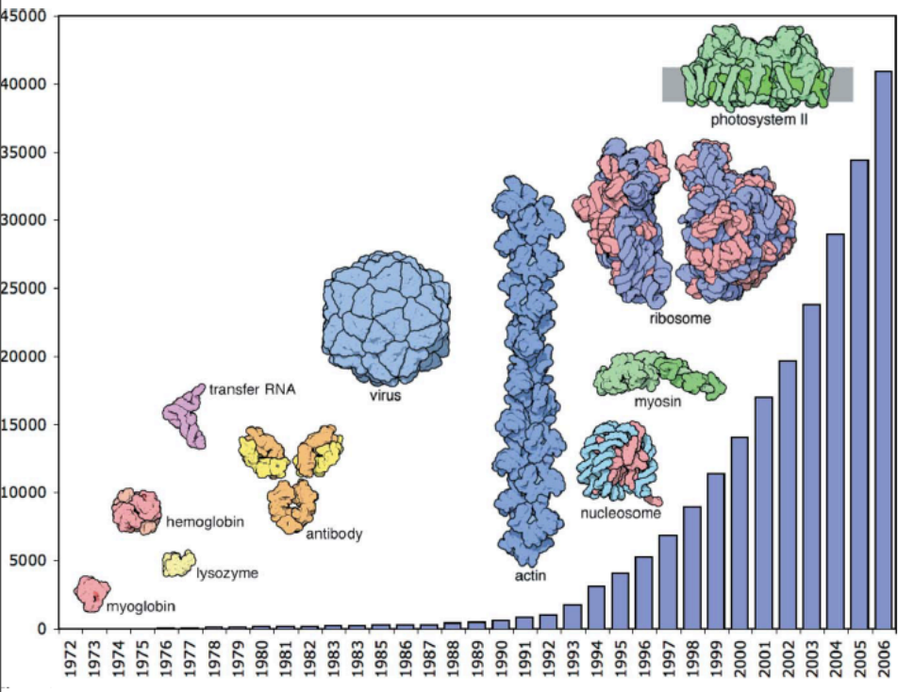




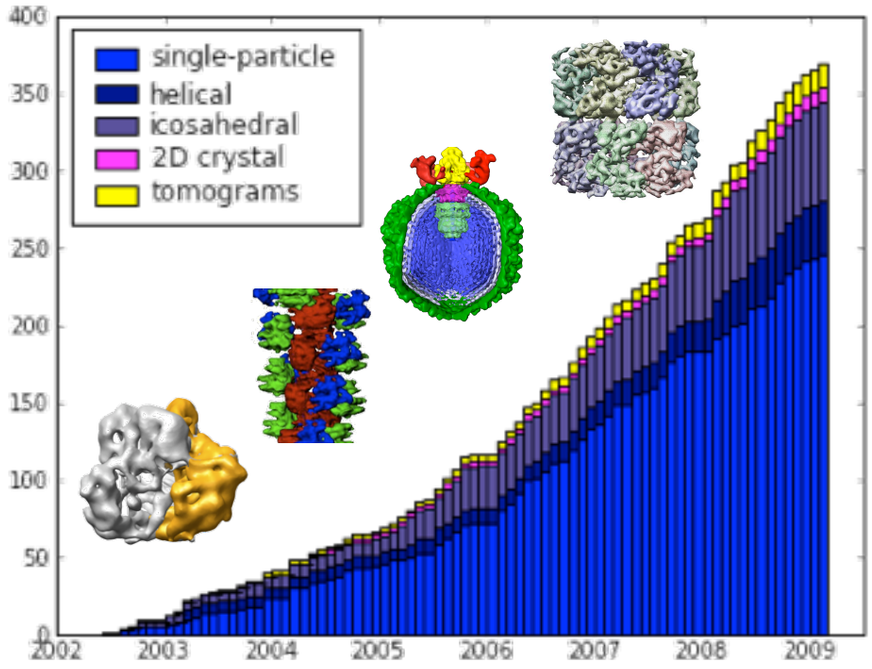
1. discretize the sampling space,
2. break the system into overlapping subsets,
3. find acceptable local solutions,
4. then merge them together self-consistenly into increasingly larger subsets.

Assembly architecture from atomic structures of subunits, EM density map of assembly, and proteomics

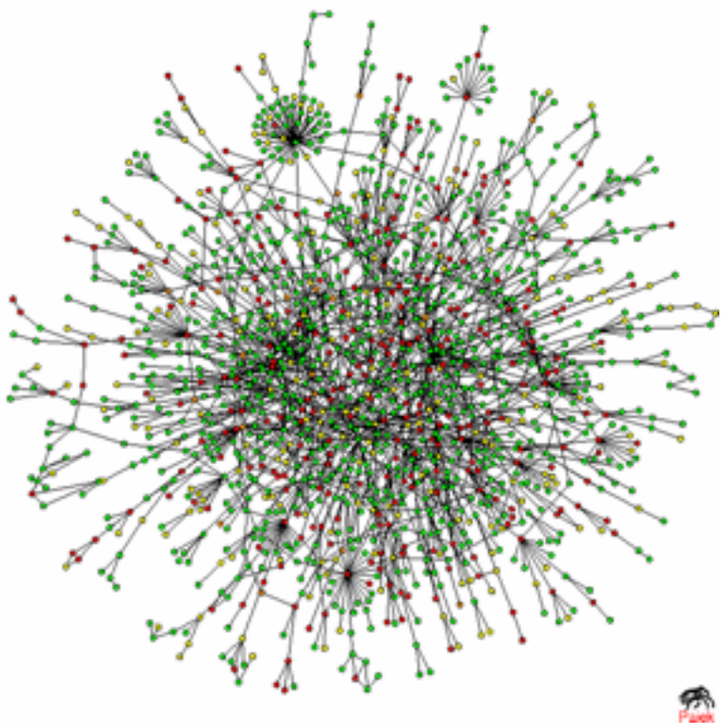
Protein Data Bank



EM Data Bank



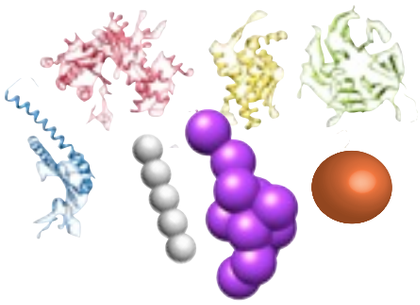
BioGrid, ...



Aligning proteomics networks to EM density maps

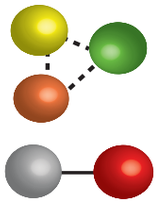
Lasker et al. *J. Mol. Biol*. 388, 180-194, 2009
Lasker et al. *in preparation*

Input: components, map, interactions



component sequences

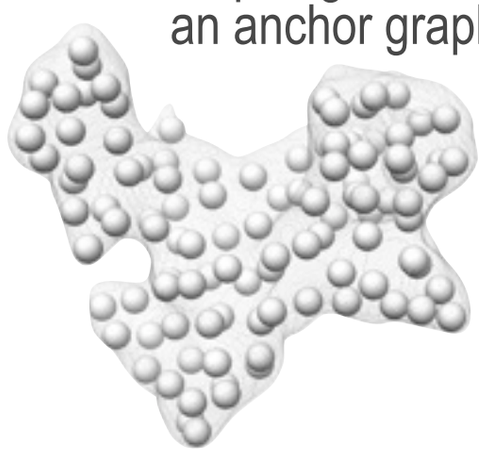
assembly map



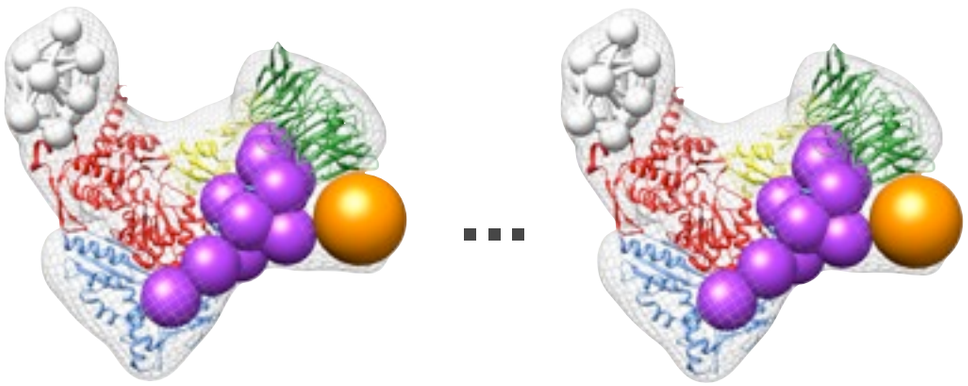
contacts, proximities

Discretize map

Map segmented into an anchor graph

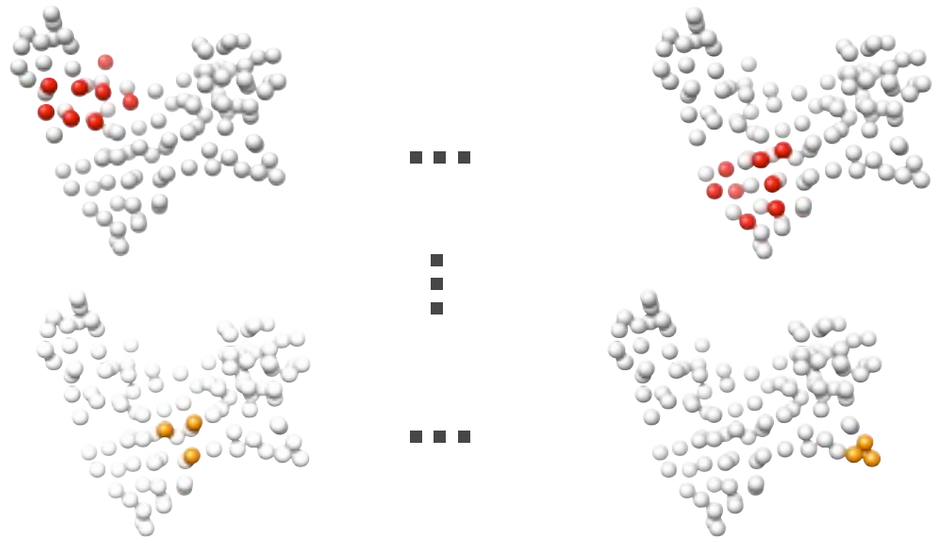


Build a discrete sampling space



Output: Enumerated best-scoring configurations

Assemble partial solutions



Positions of a component with known atomic structures are searched by rigid body fitting; Positions of a component without a known atomic structure are found by path search

Contents

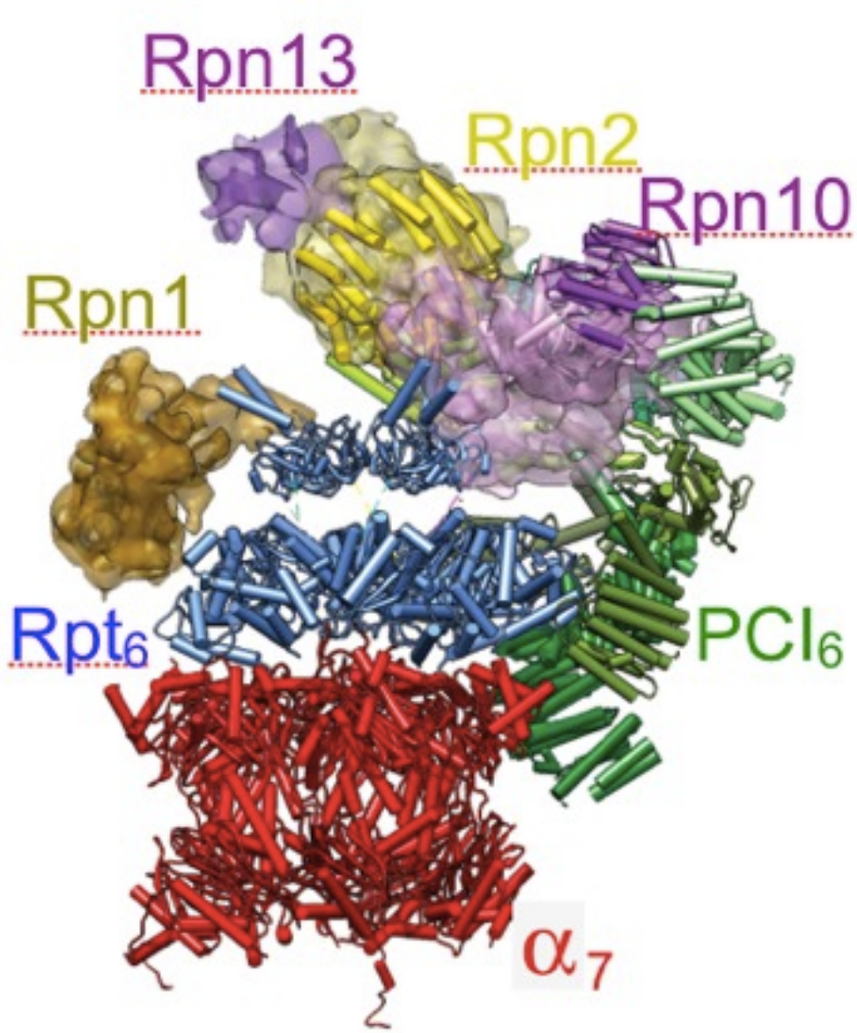
1. Integrative (hybrid) structure determination
2. EM images as a source of spatial restraints

   Application to the Nup84 complex
3. Multiple fitting of subunits into an EM map of the whole assembly

   Application to the 26S proteasome

Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach

Keren Lasker, Friedrich Förster, Stefan Bohn, Thomas Walzthoeni, Elizabeth Villa, Pia Unverdorben, Florian Beck, **Ruedi Aebersold**, Andrej Sali, and **Wolfgang Baumeister**. ***Proc. Natl. Acad. Sci. USA***, 2012.

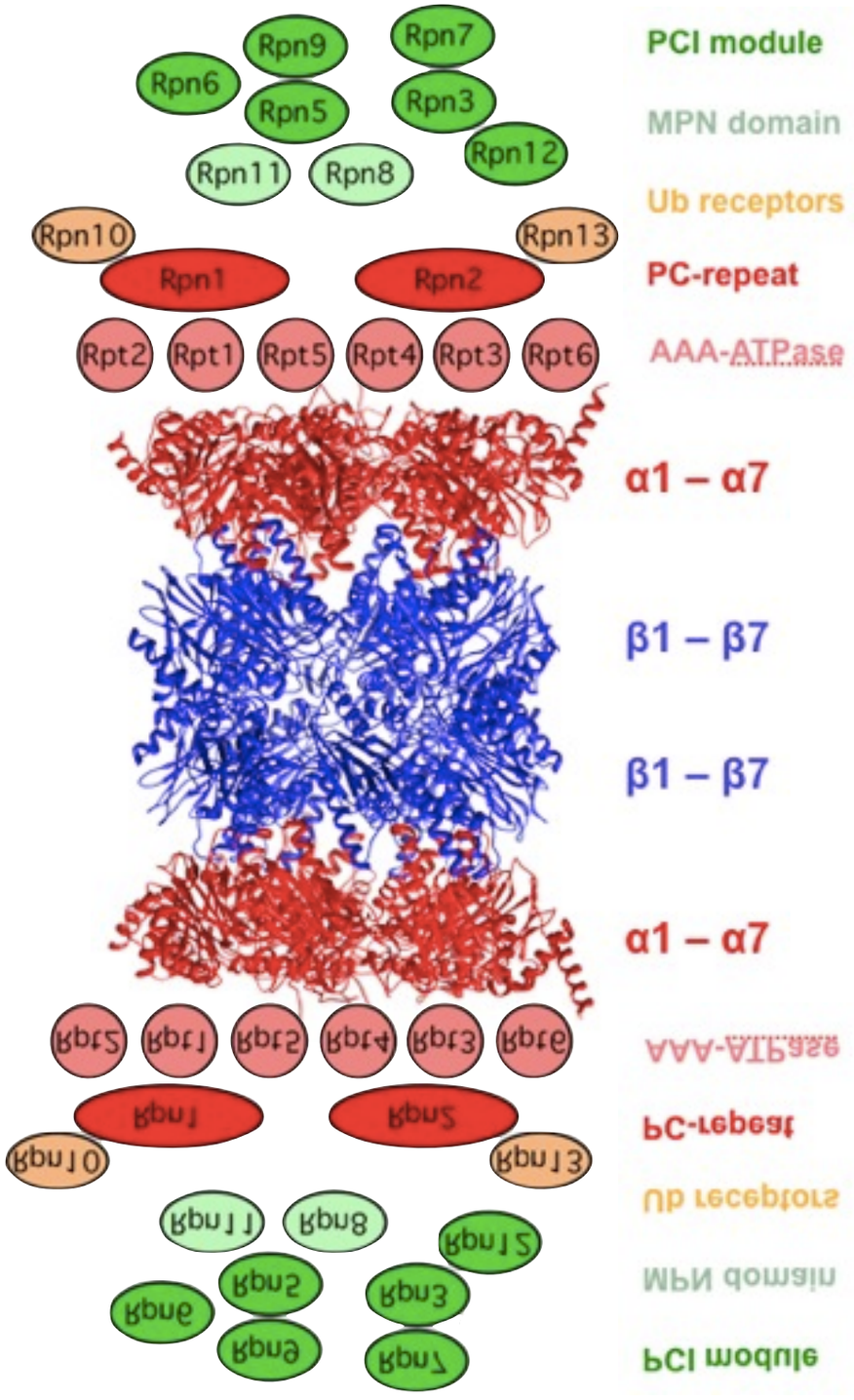


The 26S proteasome architecture

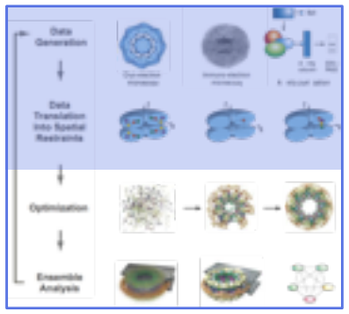
19S Regulatory Particle

20S Core Particle

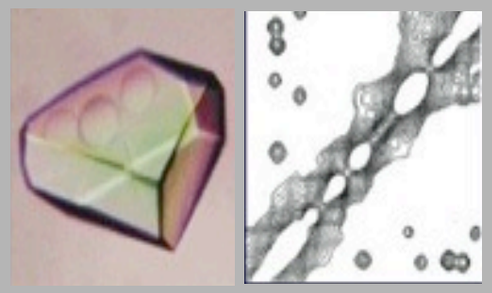
19S Regulatory Particle



Bohn S. and Förster F. *Handbook of Proteolytic Enzymes,* 2012

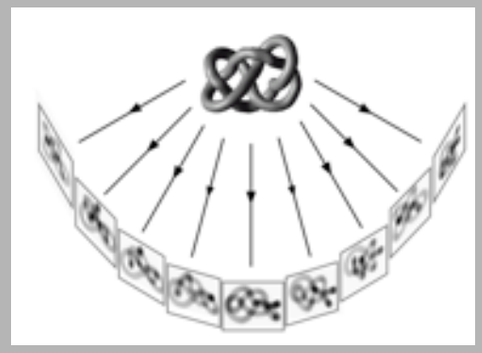


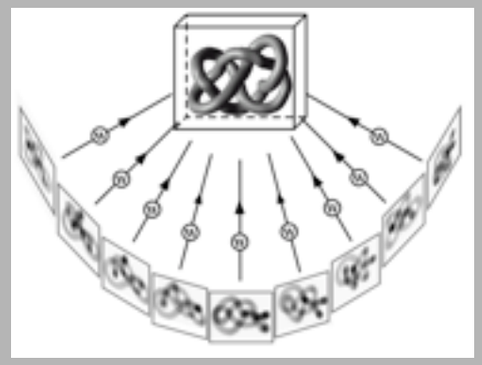
Gathering information and translation into spatial restraints



X-ray, NMR, homology modeling

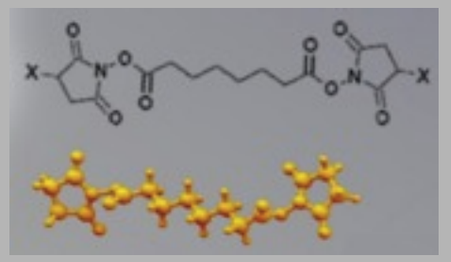
Component atomic models

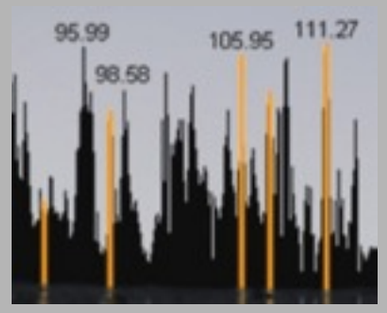




Electron microscopy

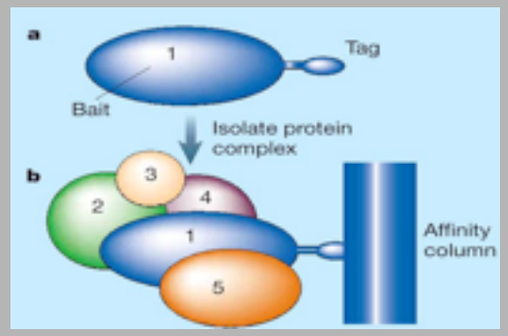
Overall shape, component positions

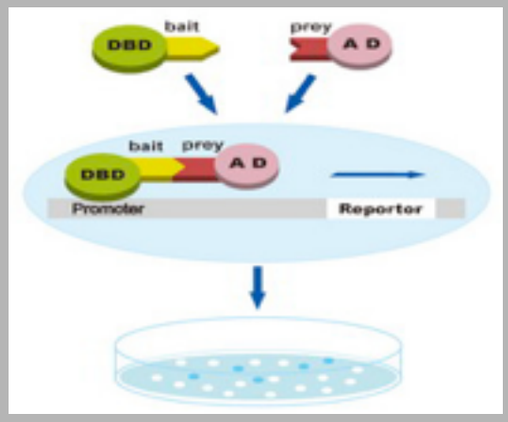




Chemical cross-linking

Protein-protein contacts





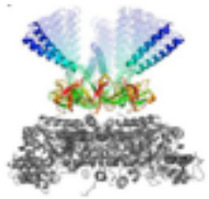
Proteomics

Protein-protein proximities

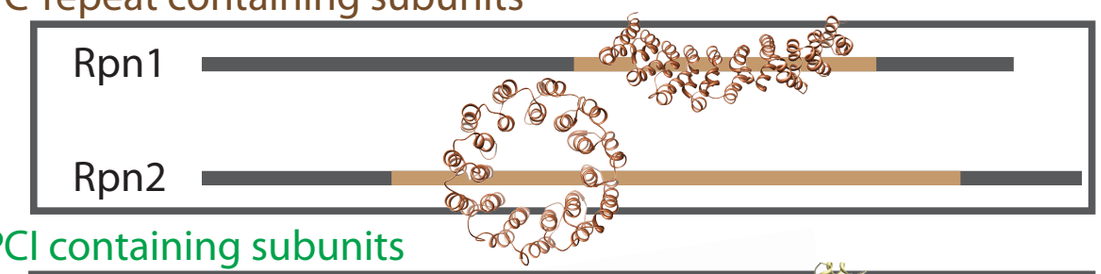
RP components and their representation

AAA-ATPase hexamer ring

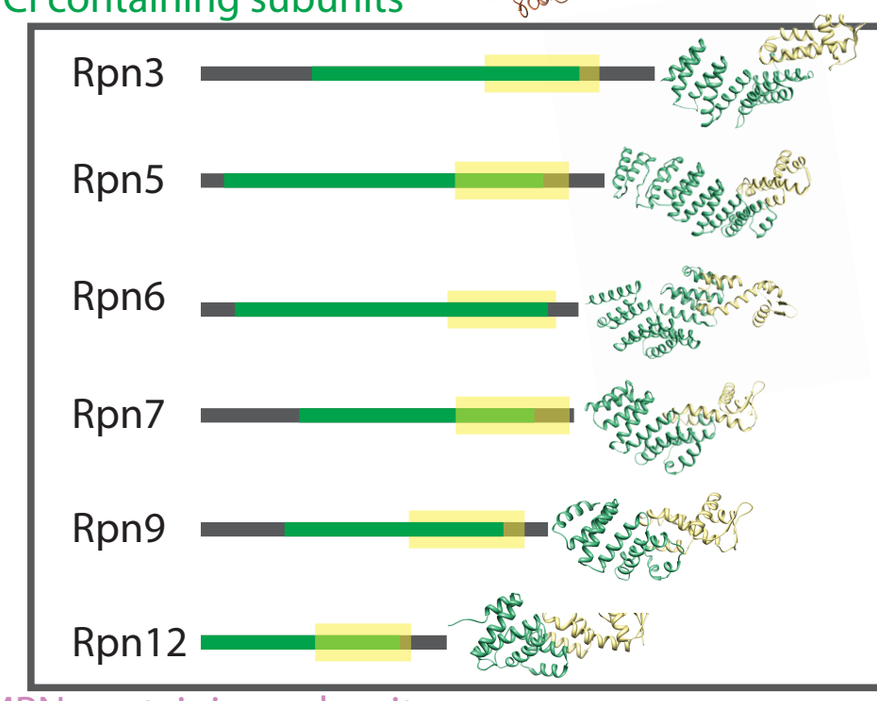
homology model based on PAN structure (Bohn *et al*, ***PNAS***, 2010).



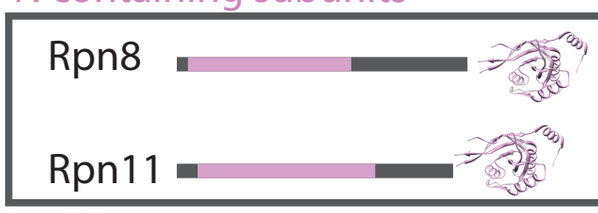
PC-repeat containing subunits



PCI containing subunits

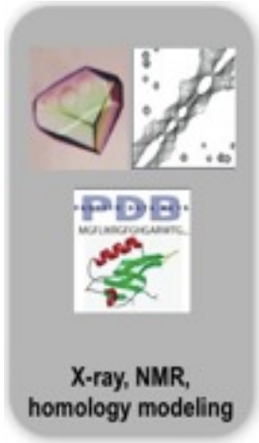


MPN containing subunits



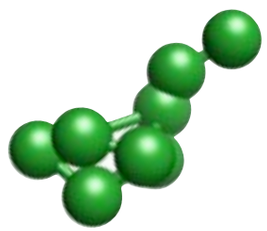
Ubiquitin receptors

precision, efficiency, availability



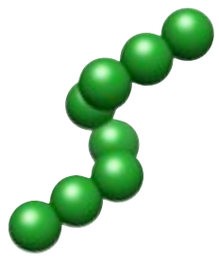
Atomic

Restraints: Geometric complementarity
Excluded volume



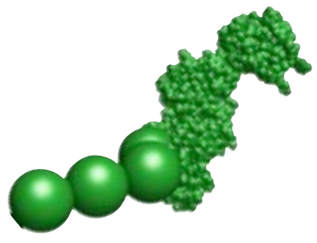
Fixed coarse

Restraints: Excluded volume



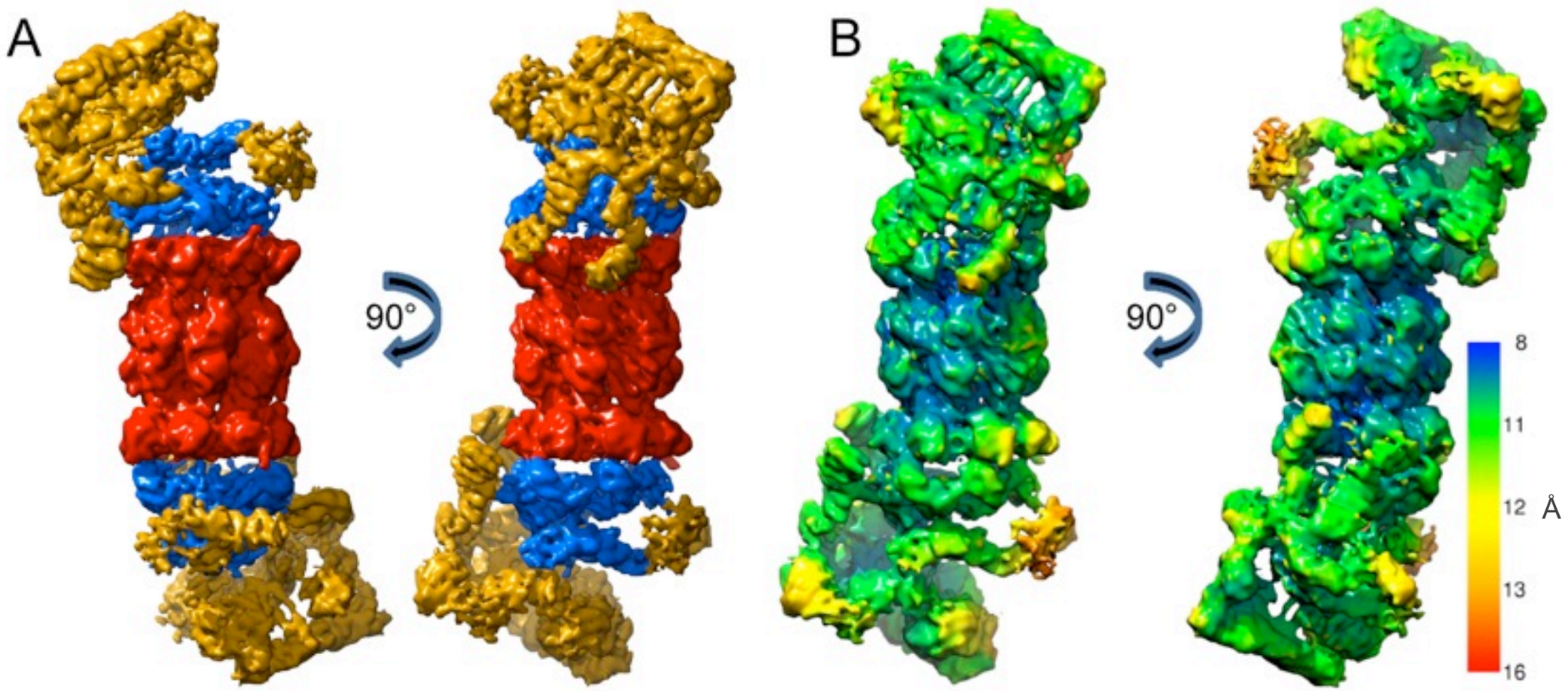
Flexible coarse

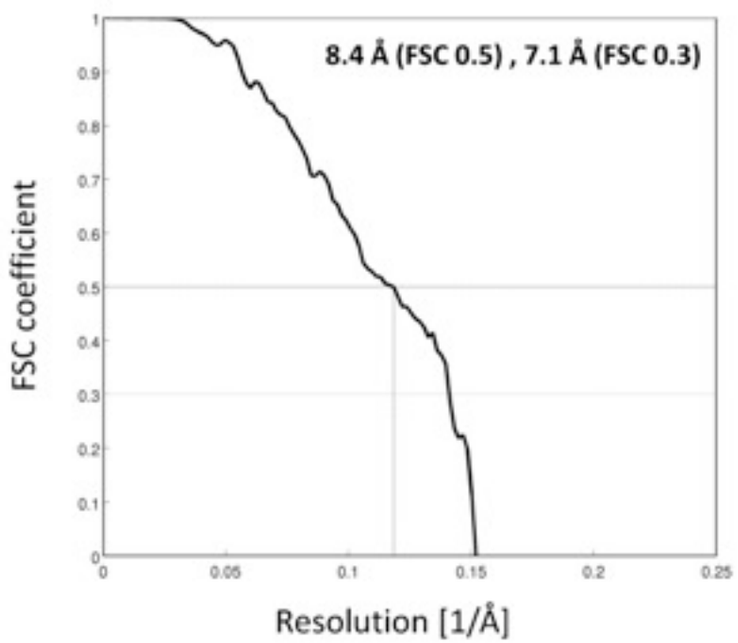
Restraints: Chain connectivity
Radius of gyration
Excluded volume



Hybrid

Cryo-EM map of the *S. pombe* 26S proteasome



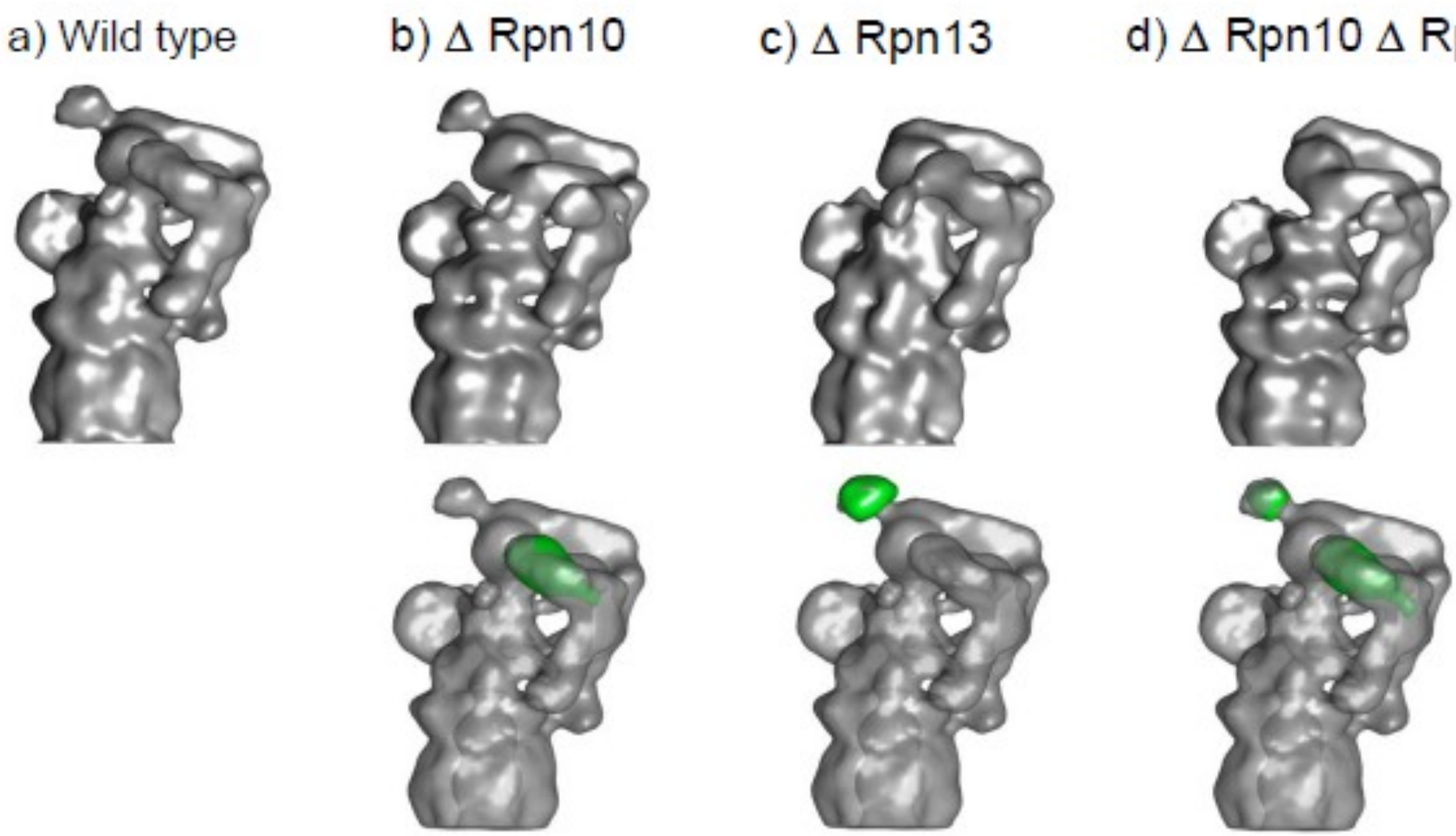


Particles	375,000
Symmetry	C2
Increment	0.5°
FSC @ 0.5	8.4 Å
FSC @ 0.3	7.1 Å

Restraints: Cross-correlation between a model and the map

F. Foerster, S. Bohn, W. Baumeister

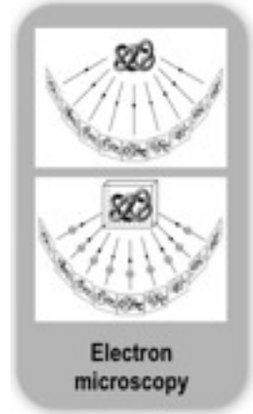
Cryo-EM of knockout mutants localizes Rpn10 and Rpn13

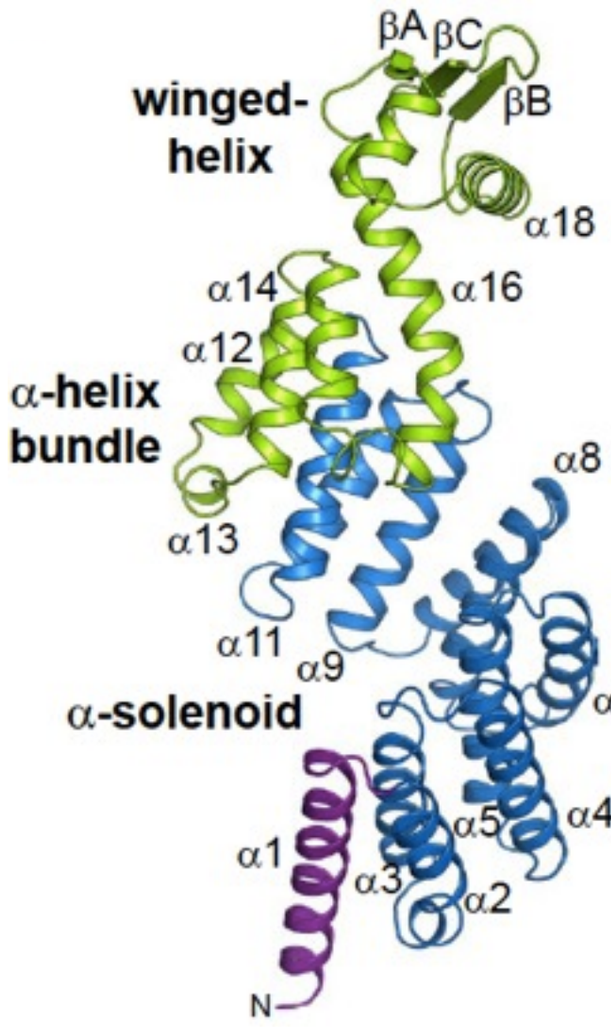


Restraints: Positions of Rpn10 and Rpn13 are fixed while sampling other subunits.

Sakata S, Bohn S, Mihalache O, Kiss P, Beck F, Nagy I, Nickell S, Tanaka K, Saeki Y, Förster F, Baumeister W, ***PNAS***, 2012.

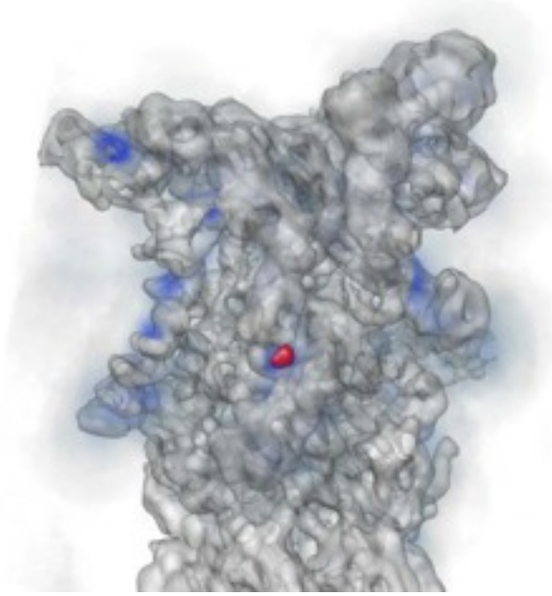
Fitting of *D. melanogaster* Rpn6 X-ray structure into the cryo-EM map localizes Rpn6

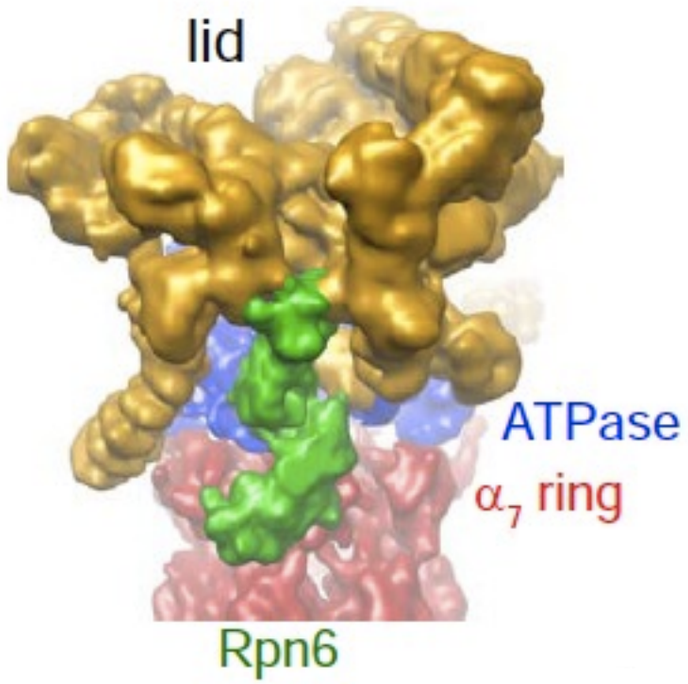




Pathare GR, Nagy I, Bohn S, Unverdorben P, Hubert A, Körner R, Nickell S, Lasker K, Sali A, Tamura T, Nishioka T, Förster F, Baumeister W & Bracher A., ***PNAS***, 2012.

Structure - map cross-correlation

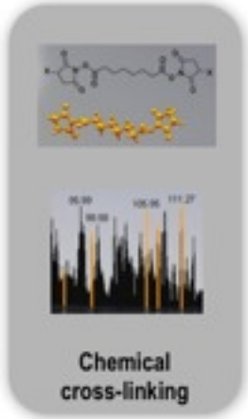




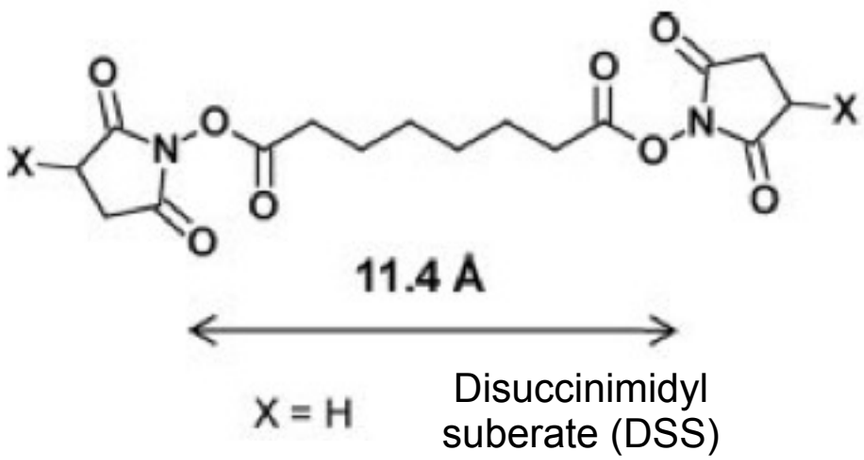
Restraints: Position of Rpn6 is fixed while sampling other subunits.

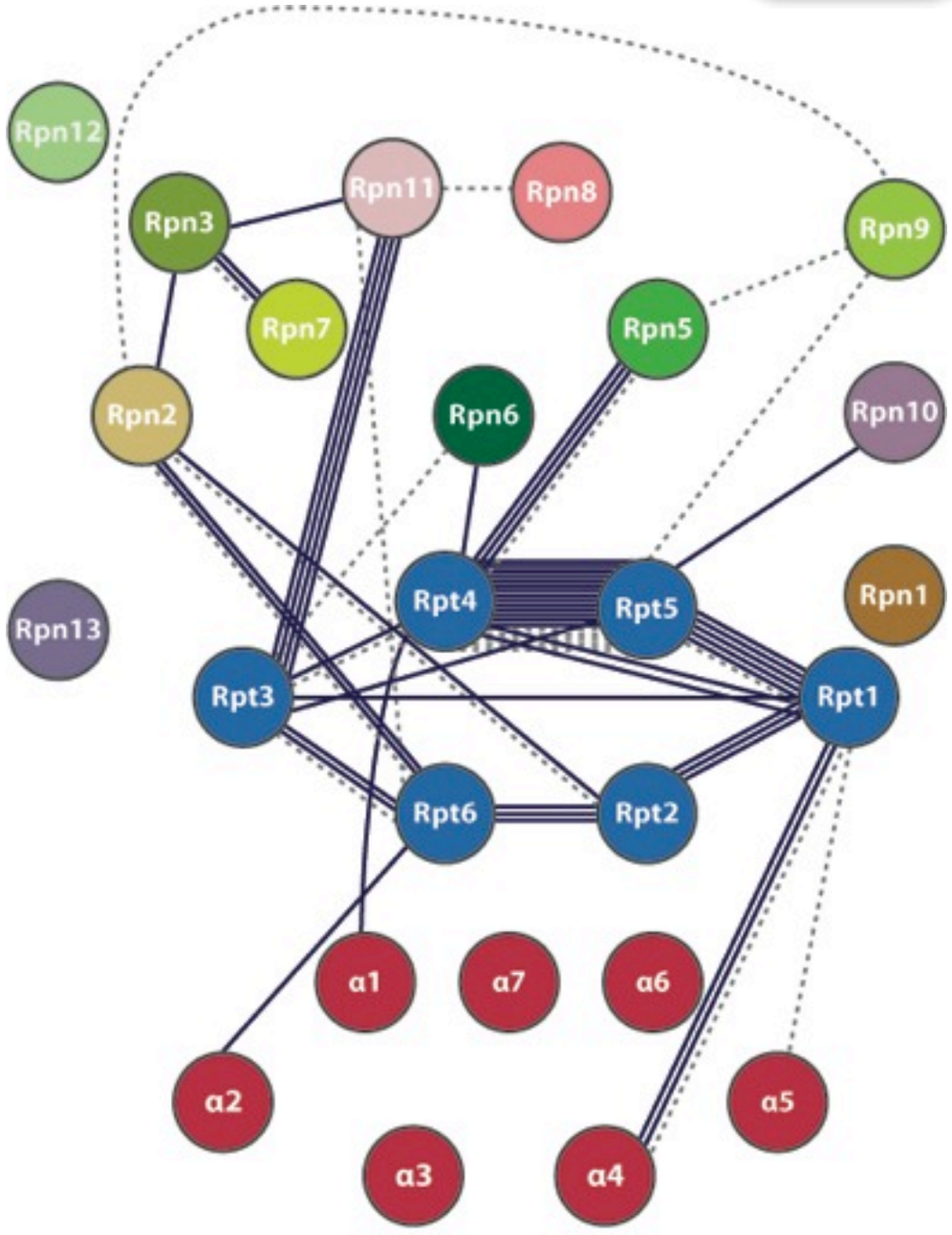
Similarly, for the AAA-ATPase Rtp1-6 heteromeric ring (Bohn *et al*, ***PNAS***, 2010)**.**

Cross-linking / mass spectrometry data



Leitner, Walzthoeni, Kahraman, Herzog,
Rinner, Beck, Aebersold. ***MCP***, 2010





Inter-molecular cross-linking of exposed Lys residues:

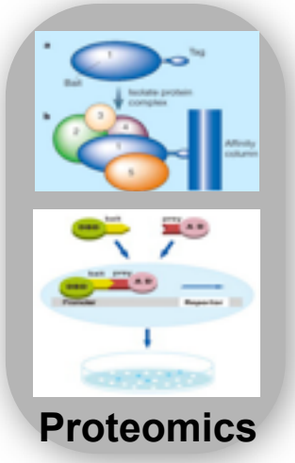
- 12 Rpt-Rpn residue-specific crosslinks (*S.p.*)
- 3 Rpn-Rpn residue-specific crosslinks (*S.p.*)

Restraints: upper distance bounds on cross-linked atoms or beads.

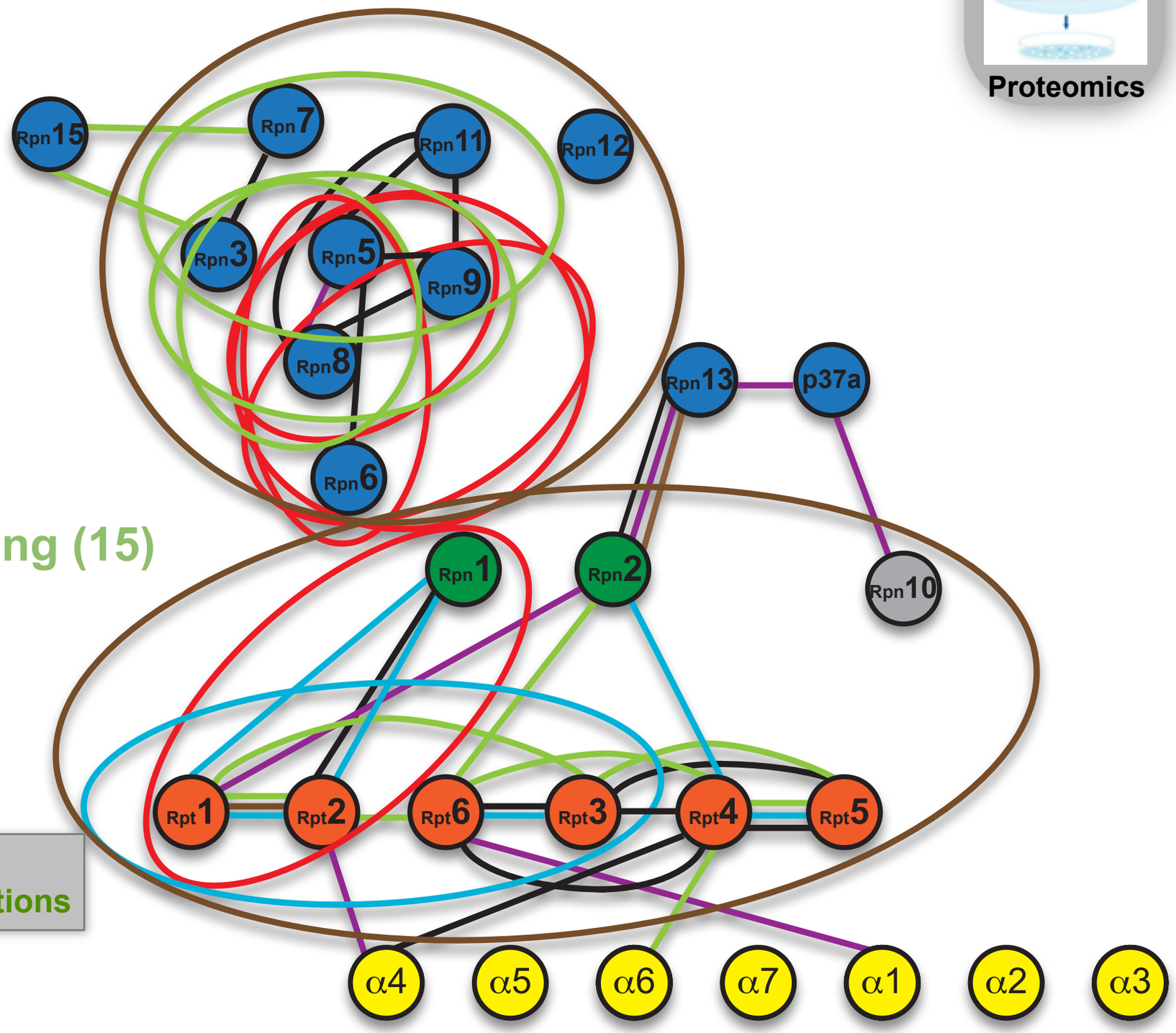
***S.p.* cross-links**
S.c. cross-links

T. Walzthoni, A. Leitner,
M. Beck, R. Aebersold

Public proteomics data

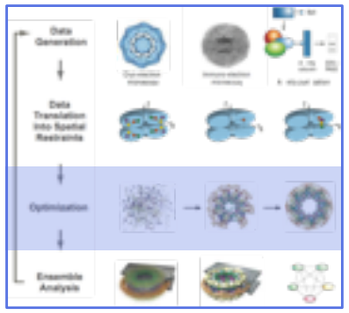


- Two Hybrid (16)
- Pulldown (10)
- Co-Purification (4)
- MS-MS (4)
- Chemical Cross-Linking (15)
- Filter Binding (14)



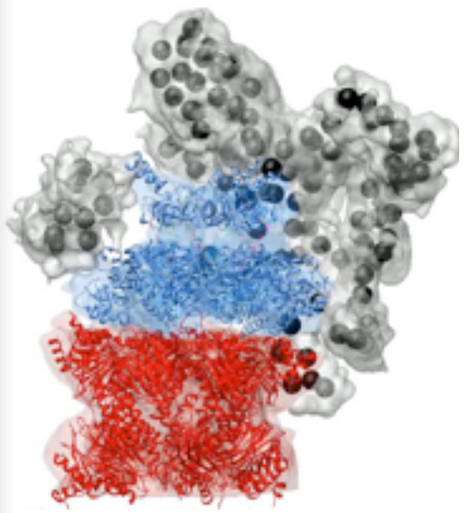
Restraints: connectivity restraints for coarse subunit representations

Förster F, Lasker K, Nickell S, Sali A, and Baumeister W, ***Mol. Cell. Proteomics,*** 2010
Stengel F, Robinson, C.



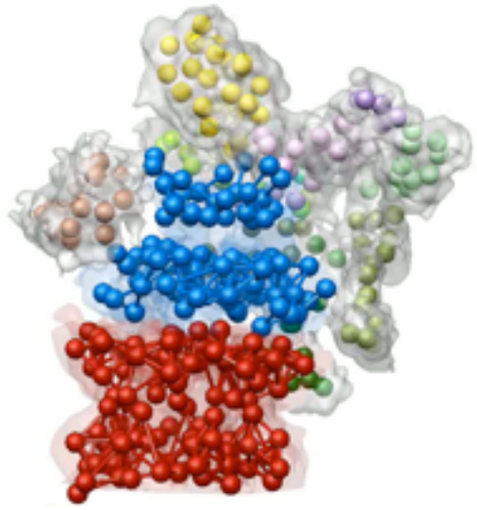
Sampling good-scoring 19S structures

Discretization



discretization of the map into 238 anchor points

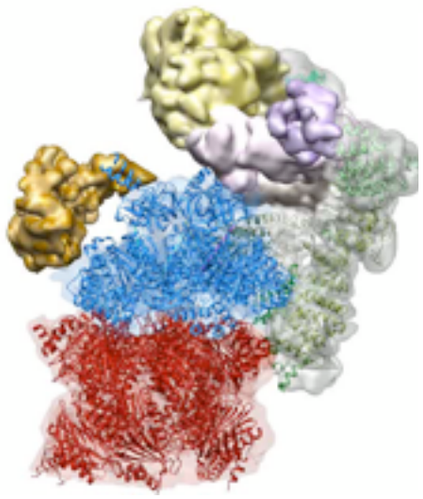
Localization



localization of coarse subunit models, subject to proteomics data

enumeration of all configurations with at most 5 violations

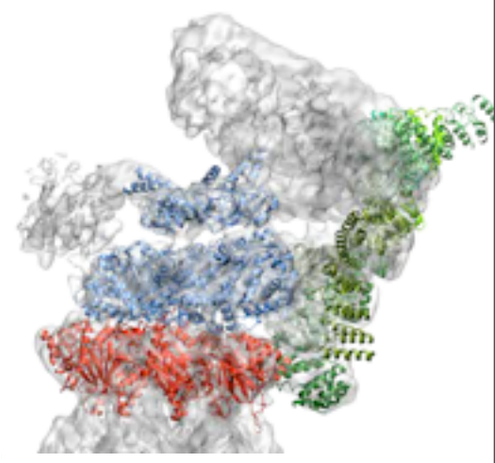
Fitting



local rigid body fitting of alternative atomic subunit models

selection of best subunit models by fitting quality

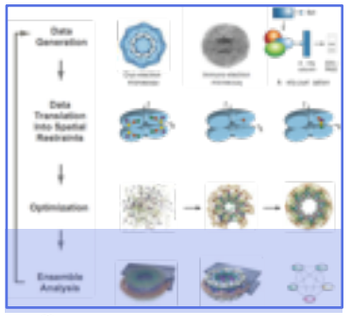
Refinement



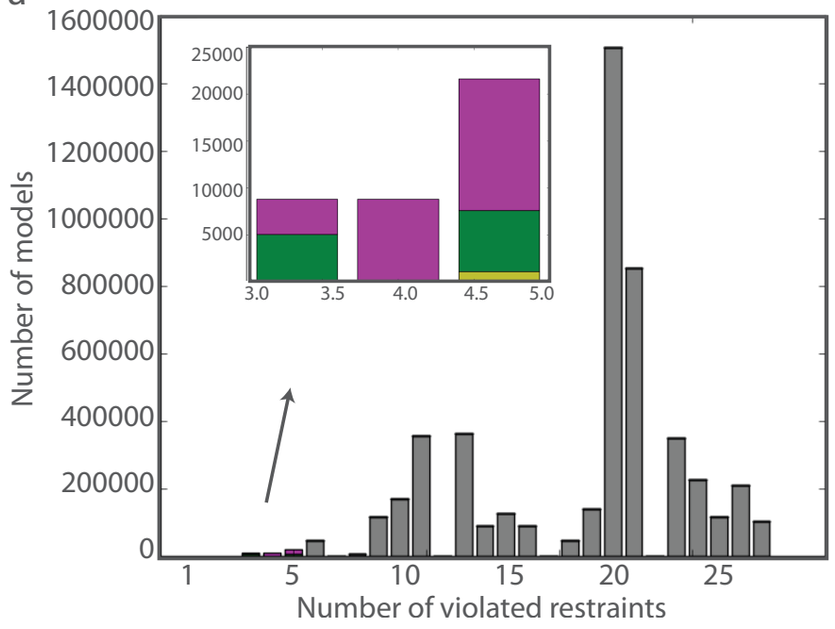
atomic model refinement subject to cross-linking and position restraints

Elizabeth Villa

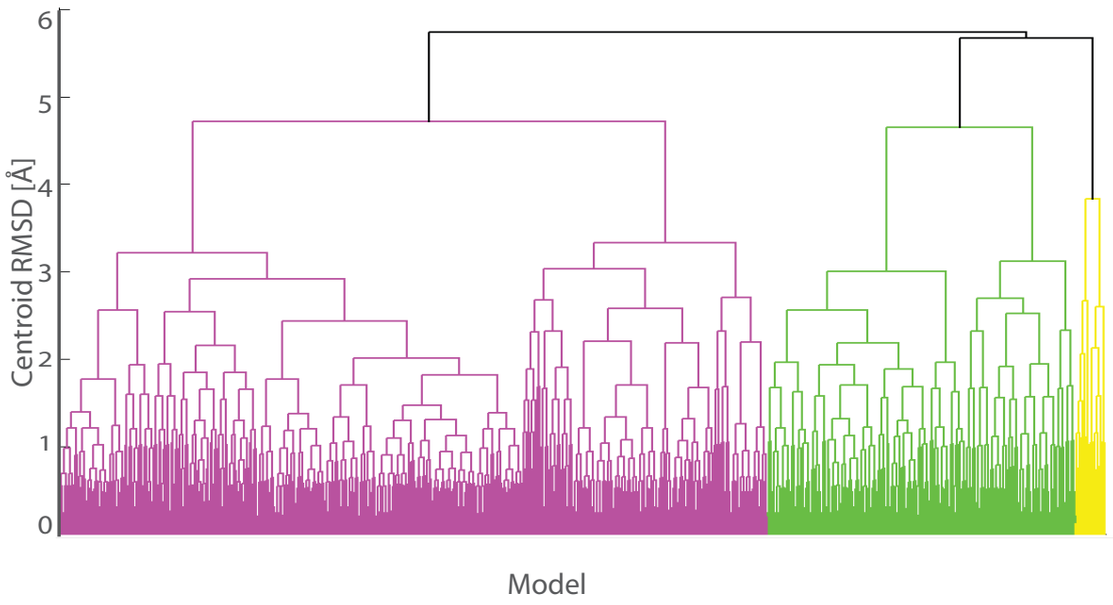
Ensemble of ~0.5 million best-scoring models



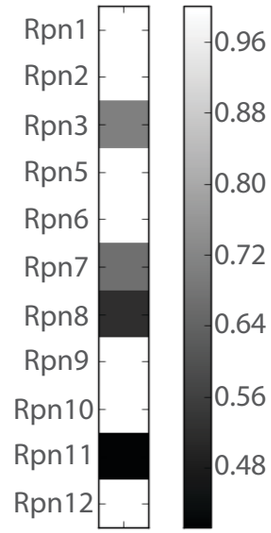
a



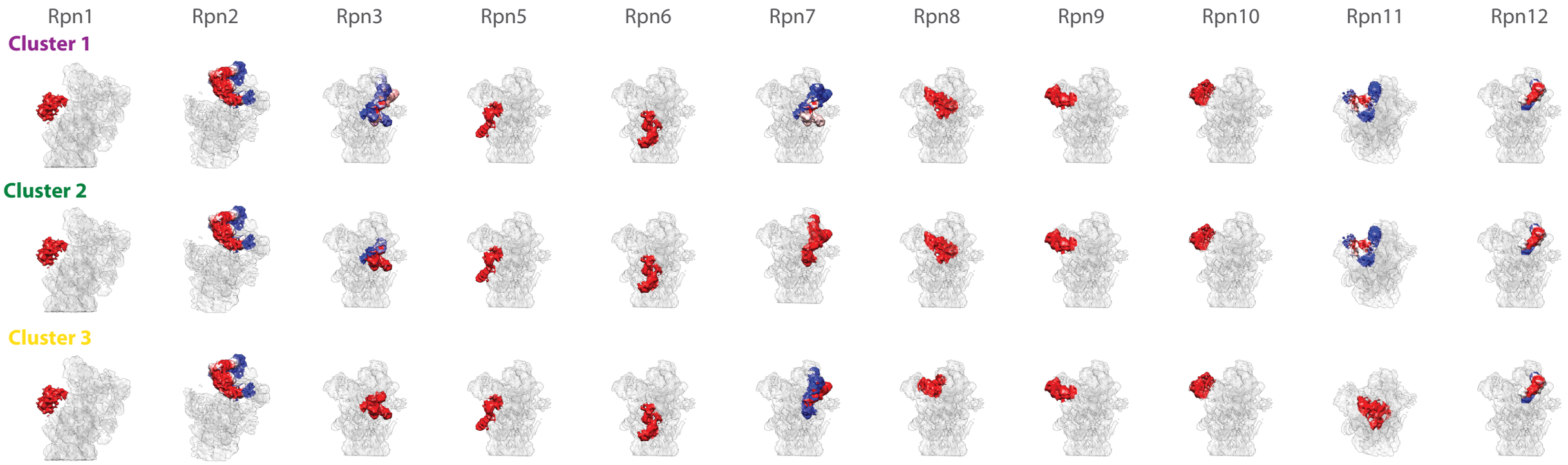
b



c



d



Assessing the well-scoring models

(in the absence of Bayesian inference)

1. Existence of a good-scoring model.

2. Precision of the ensemble of good-scoring models.

3. Check model against unused data (cross-validation).

4. Known precision / accuracy for “similar” cases.

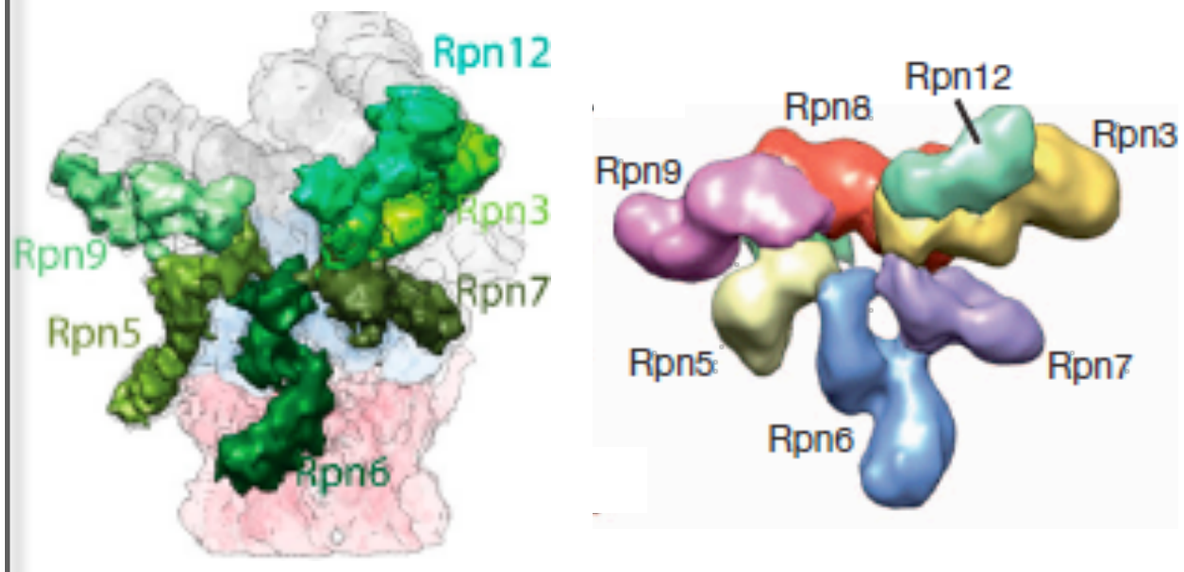
5. Non-random patterns in the model.

Modeling facilitates assessing the data as well as models in terms of precision and accuracy.

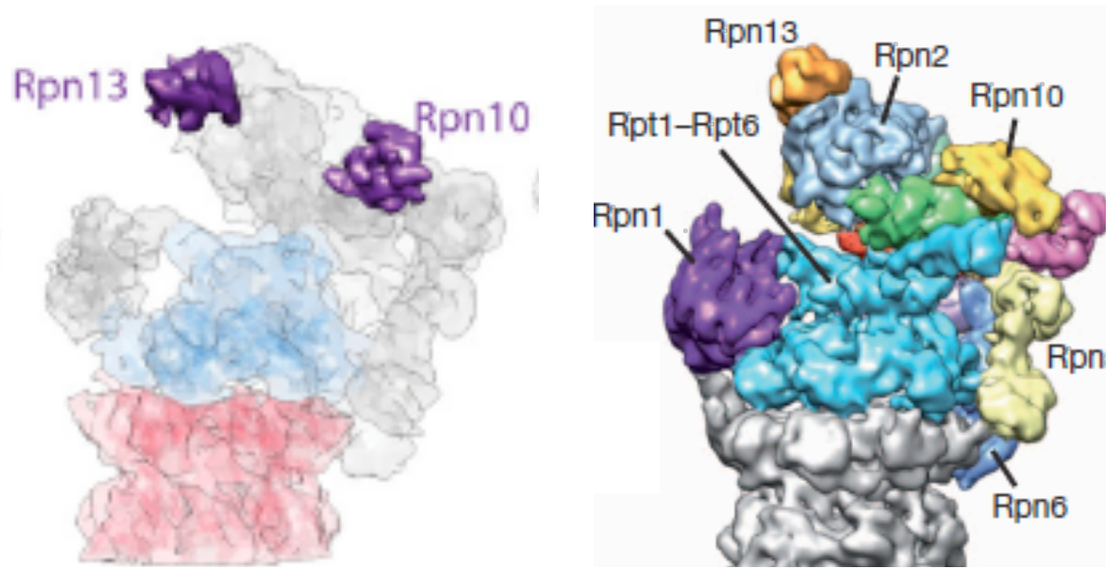
Comparison with an independently determined model

Left: Lasker, Förster, Bohn, Walzthoeni, Villa, Unverdorben, Beck, Aebersold, Sali, Baumeister. ***PNAS***, 2012.
Right: Lander, Estrin, Matyskiela, Bashore, Nogales, Martin. ***Nature***, 2012.

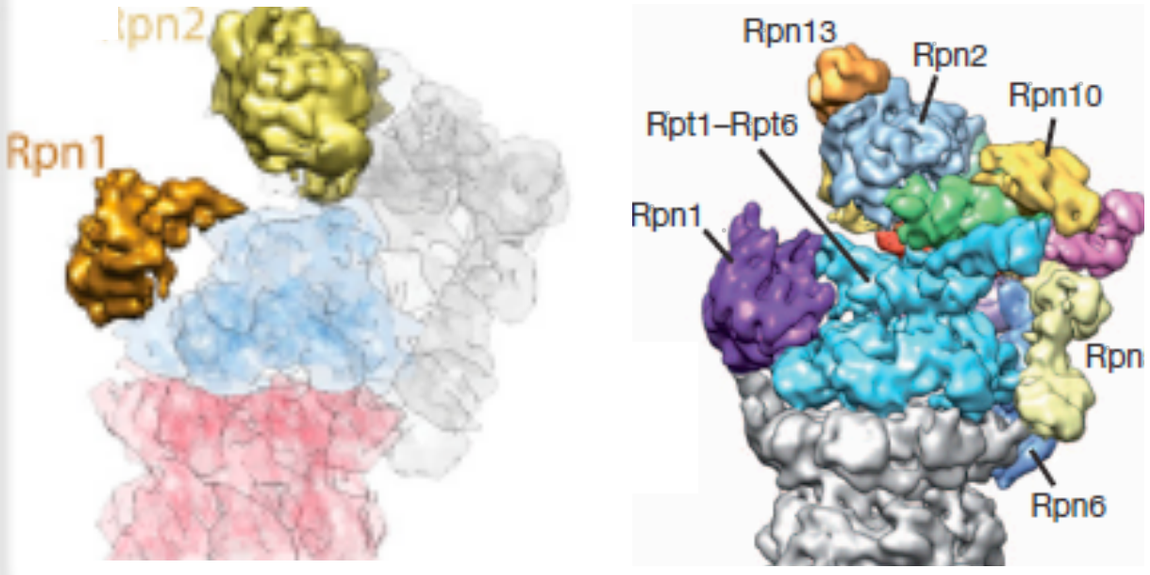
PCI containing proteins



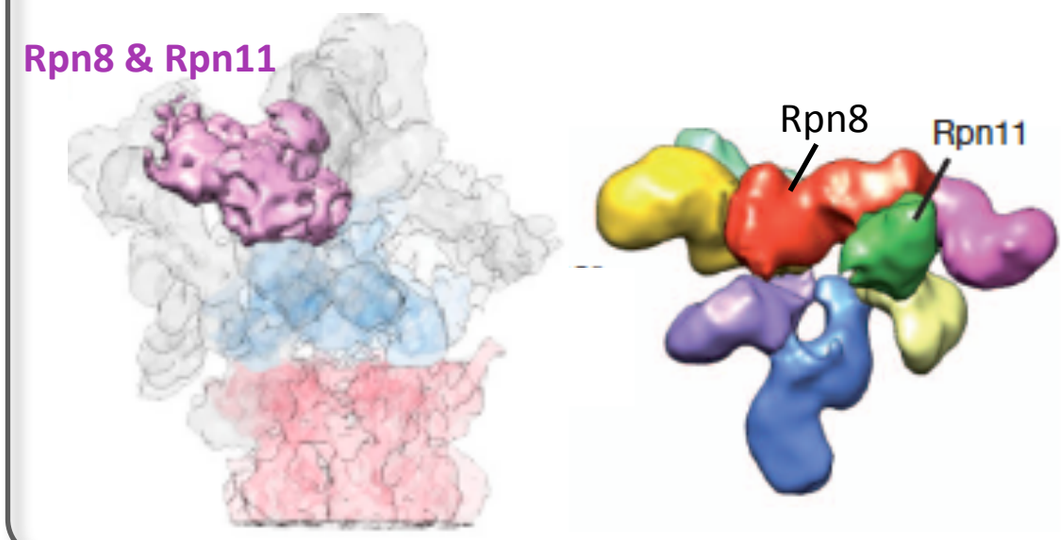
Ubiquitin receptors



PC-repeat containing proteins



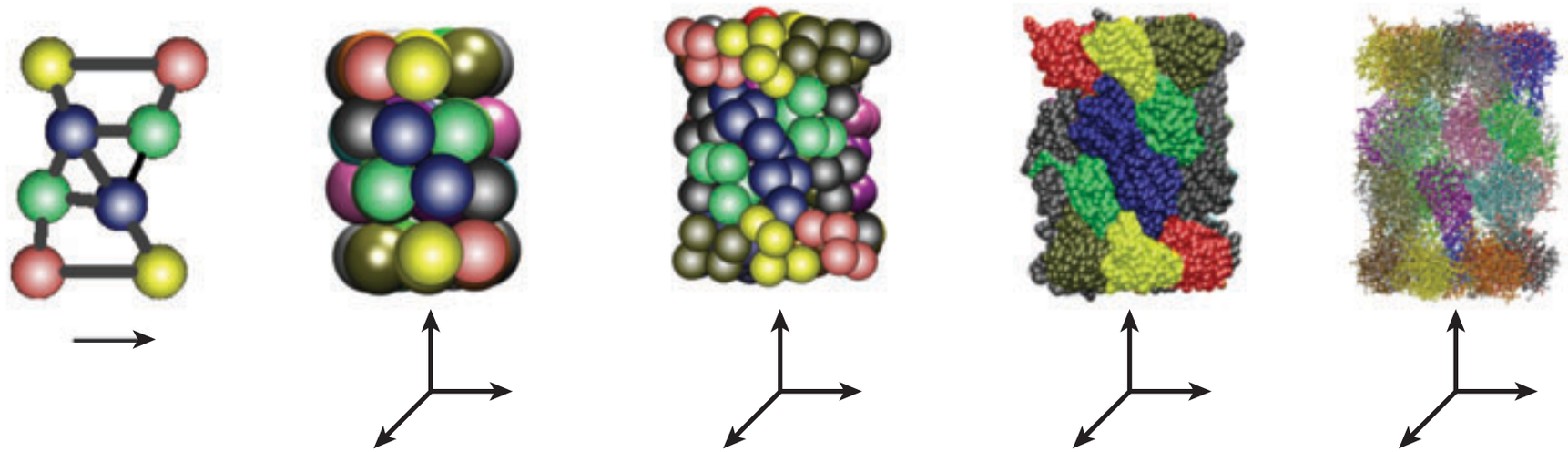
MPN containing proteins



Need for multi-scale (hierarchical) sampling

Component interactions	Component configuration	Molecular architecture	Pseudo-atomic structure	Atomic structure

Component instances



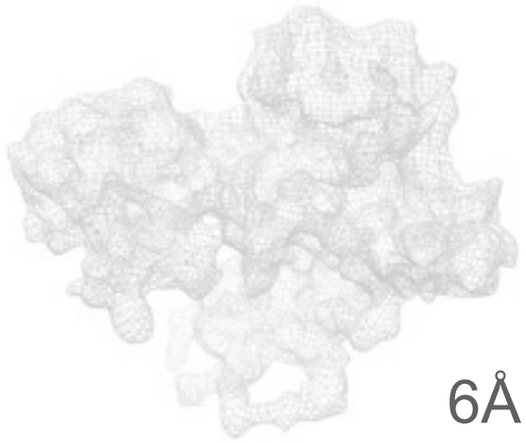
Interaction instances	Component position	Component position and orientation	Residue positions	Atomic positions

Spatial information

Immuno-EM	Comparative modeling	Cryo-EM	X-ray crystallography
FRET spectroscopy	Ab initio prediction	Cryo-ET	NMR spectroscopy
Electron microscopy	Comparative patch analysis	SAXS	
Symmetry information	Computational docking	H/D exchange	
Mass spectrometry	Electron spin resonance	Footprinting	
	Bioinformatics		

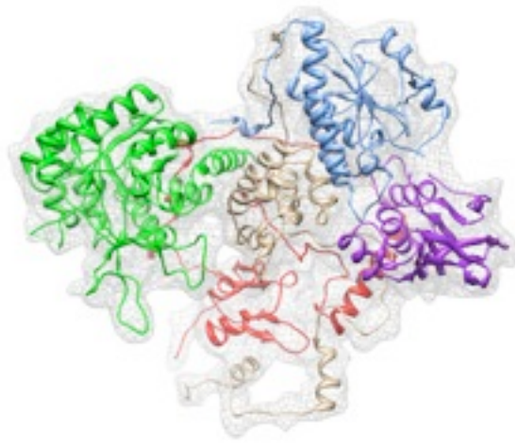
Example

Simulated map:

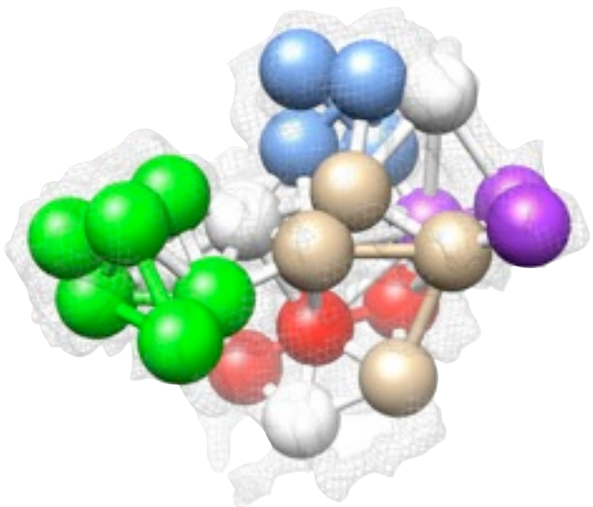


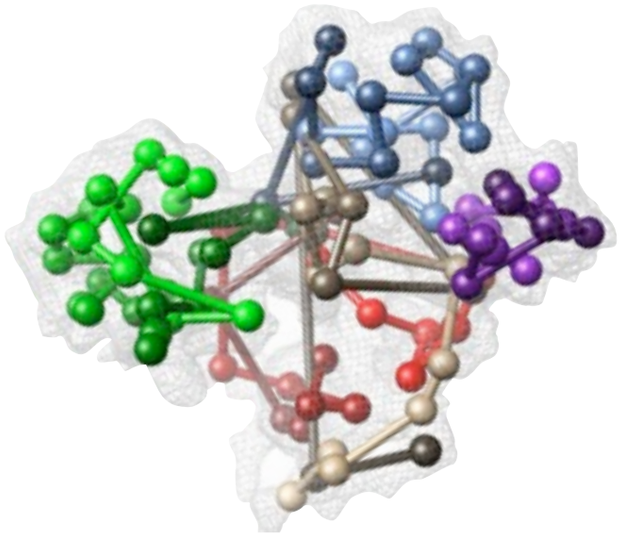
6Å

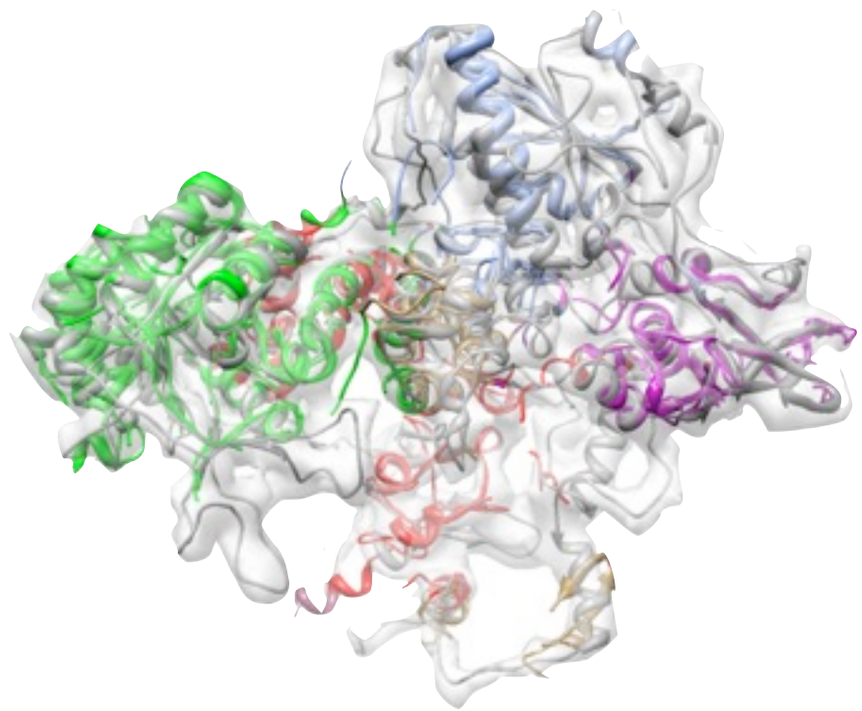
Native X-ray:
DIHYDROPYRIMIDINE
DEHYDROGENASE (1gte)



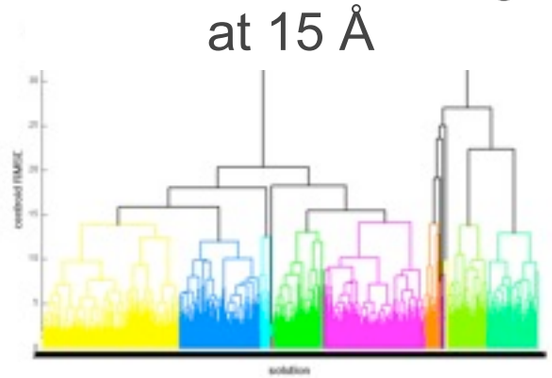
50 residue per bead → 10 residue per bead → 1 residue per bead



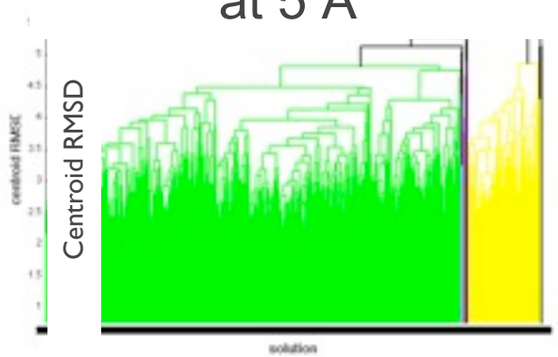




Solutions clustering
at 15 Å

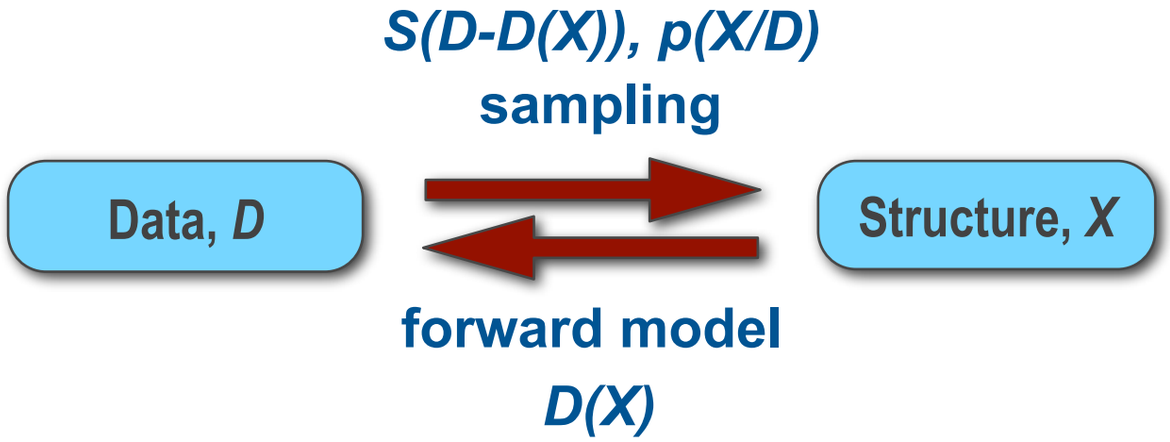


Solutions clustering
at 5 Å



Overlay of fragments
on bead model

Bayesian inference of structures



Structures from NMR data and a molecular mechanics force field

W. Rieping, M. Habeck, M. Nilges. Inferential Structure Determination. *Science* **309**, 2005.

Least-squares scoring function:

$$S(D-D(X))=E_{MM}+w\cdot E_{NMR}$$

Bayesian scoring function:

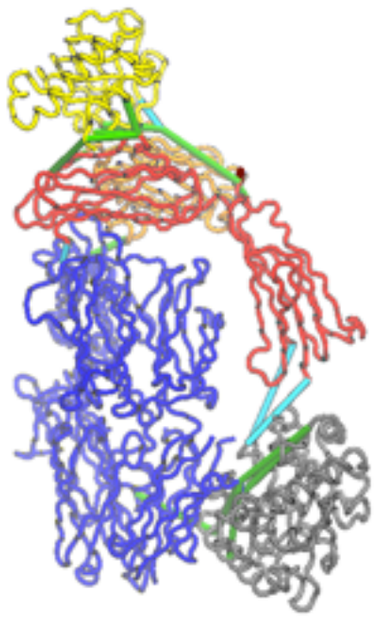
$$p(X,\sigma\,|\,D,E_{MM})\propto p(D\,|\,X,\sigma)\cdot p(X\,|\,E_{MM})\cdot p(\sigma)$$

Single structure from inconsistent cross-links

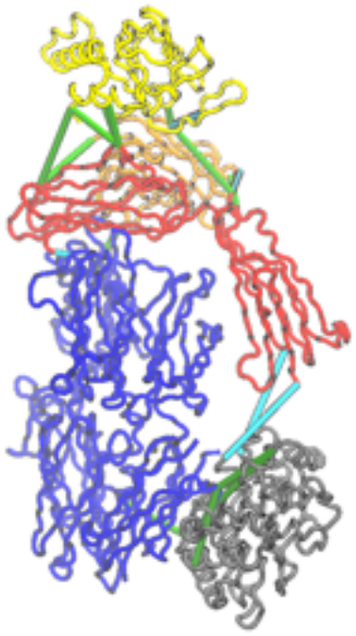
C3 and C3b forms of human Complement factor 3

self-consistent data:

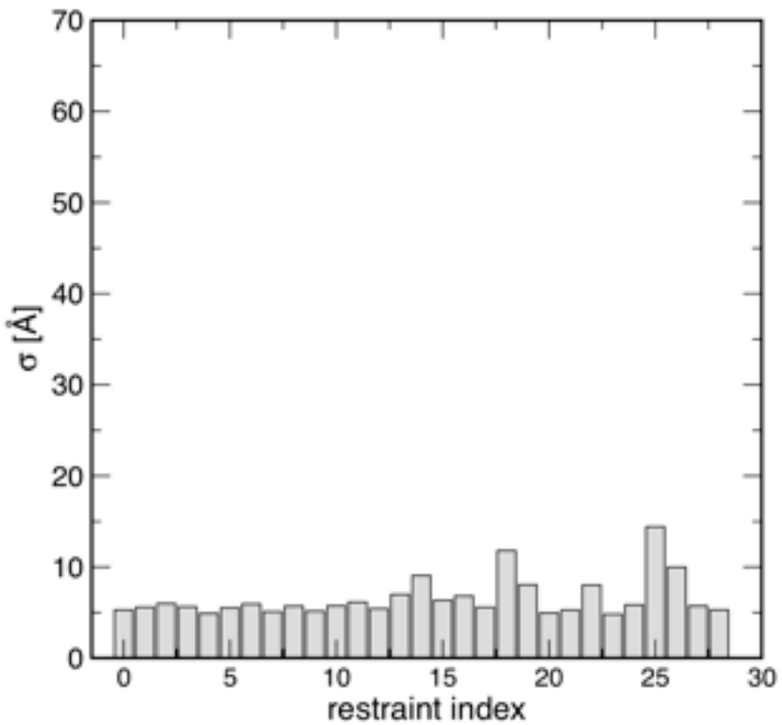
C3b X-ray structure



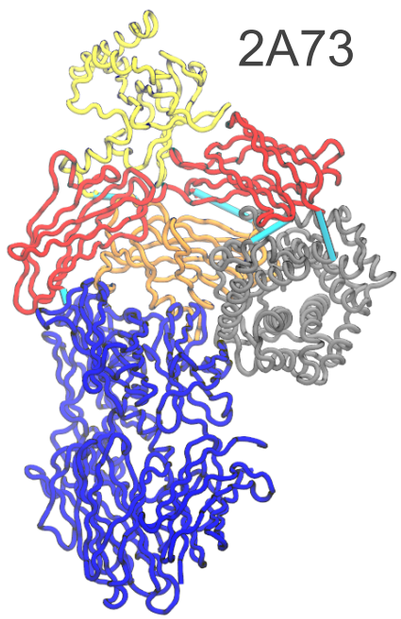
Bayesian modeling



least-squares modeling

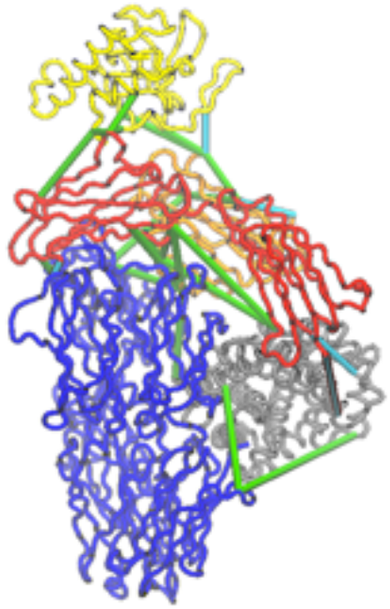


self-inconsistent data:

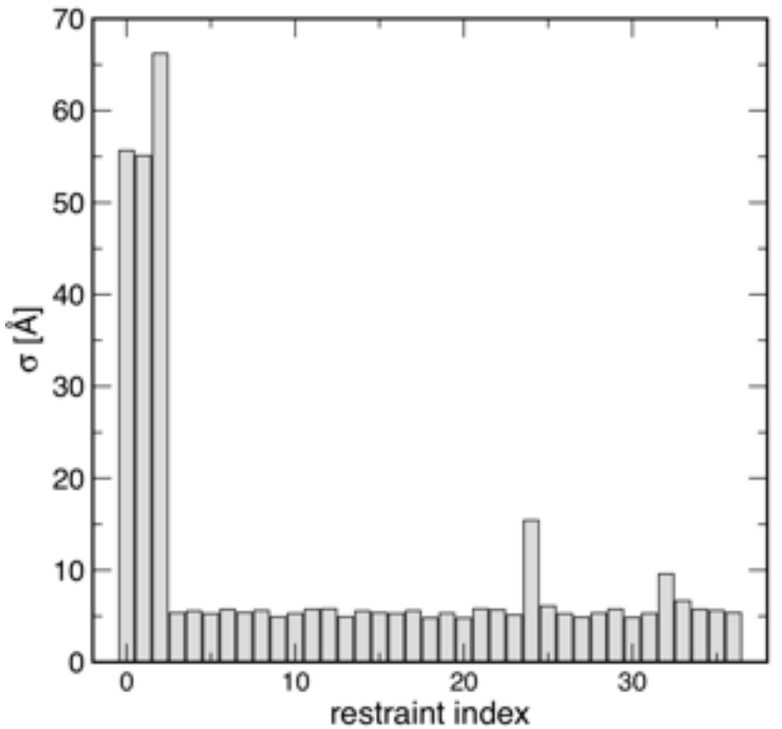


C3 X-ray structure

Bayesian modeling



least-squares modeling



Data from ZA Chen and J Rappsilber

R. Pellarin, w/ Y. Spill, M. Nilges

Disseminating structural models

Publishing models in a **printed paper**

Depositing models in a **computer database**

RCSB PDB PROTEIN DATA BANK

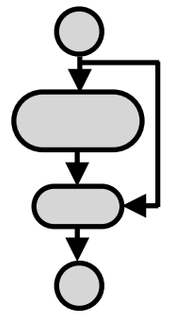
BioGRID

Depositing **input data** in a computer database

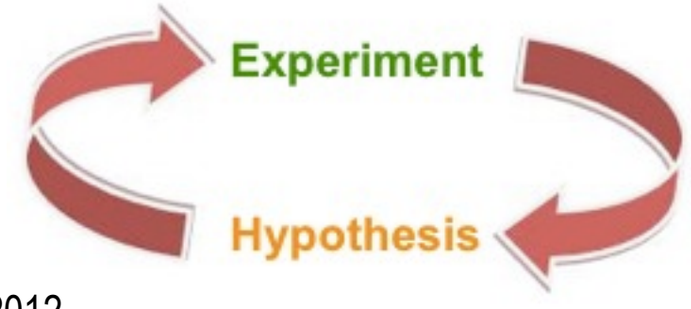
EMDB EM DATA BANK

PubMed.gov U.S. National Library of Medicine National Institutes of Health

Depositing modeling **protocols** for converting data to models



Enable others to interact with data and models:
test, improve, use data and models



Russel D, Lasker K, Webb B, Velazquez-Muriel A, Tijoe E, Schneidman D, Alber F, Peterson B, Sali A, *PLoS Biol* **10**, 2012.
Morin A, Urban J, Adams PD, Foster I, Sali A, Baker D, Sliz P. Shining Light into Black Boxes. *Science* **336**, 159-160, 2012.

Acknowledgments

QB3 @ UCSF:

Javier Velazquez (2D EM)
Keren Lasker (26S, MultiFit)
Jeremy Phillips (Nup84)
Charles Greenberg (EM)
Daniel Russel (IMP)
Ben Webb (IMP)
Elina Tjioe (IMP)
Riccardo Pellarin (Bayesian,XL)
Massimiliano Bonomi (Bayesian,SPB)
GQ Dong (Bayesian)
Seung Joong Kim (NPC)
Dina Schneidman (SAXS)
Peter Cimermancic (HPC)
Natalia Khuri (BD)
Barak Raveh (NPC transport)

Former members:

Frank Alber (USC)
Friederich Förster (MPI)
Damien Devos (EMBL)
Maya Topf (Birkbeck College)
Narayanan Eswar (Du Pont)
Marc Marti-Renom (CNAG)
Mike Kim (Google)
Dmitry Korkin (UM, Columbia)
Fred Davis (HHMI)
M. Madhusudhan (Singapore)
D. Eramian (UCSF)
Min-Yi Shen (Applied Biosys)
Bret Peterson (Google)

Mike Rout (Rockefeller U)
Javier Fernandez
Brian Chait (Rockefeller U)
David Stokes (NYSBC)
Steven Burley (Lilly)
David Cowburn (AECOM)
Bo Huang (UCSF)
Haim Wolfson (TAU)
Wolfgang Baumeister (MPI)
Friedrich Foerster
Elizabeth Villa
Stefan Bohn
Stefan Nickell
Ruedi Aebersold (ETH)
Michael Nilges (Pasteur)
Yannick Spill
Juri Rappsilber (U Edinborough)
Tom Ferrin (UCSF)
Tom Goddard
Trisha Davis (Univ of Wash)
David Agard (UCSF)
Wah Chiu (Baylor)
Joachim Frank (Columbia)
Nevan Krogan (UCSF)
Al Burlingame (UCSF)
Robert Stroud (UC

Funding

NIH
NSF

Keystone International Conference on

Structural Analysis of Supramolecular Assemblies by Hybrid Methods

March 3-7, 2013
Granlibakken Conference Center Lodge
Lake Tahoe, CA, USA

SESSION TOPICS

SESSION 1: Computation for Hybrid Approaches
SESSION 2: Hybrid Approaches to Studying Dynamic Systems
SESSION 3: Hybrid Approaches to Studying Macromolecular Structures
SESSION 4: Hybrid Approaches to Studying Cellular Organization
SESSION 5: Single Molecule Methods
SESSION 6: Latest Advances in Hybrid Methods

ORGANIZERS
Chair: Andrej Sali, USA
Co-Chairs: David Baker, USA
Brian Chait, USA

Figures:
Dynein cargo transport, Graham Johnson, www.grahamj.com; Adenovirus vertex, Stewart Laboratory, Case Western Reserve University