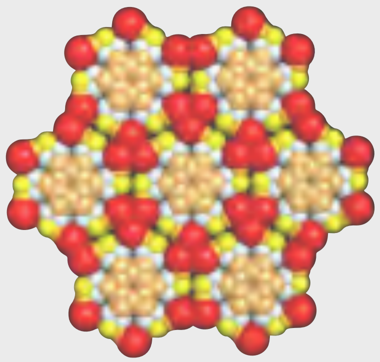
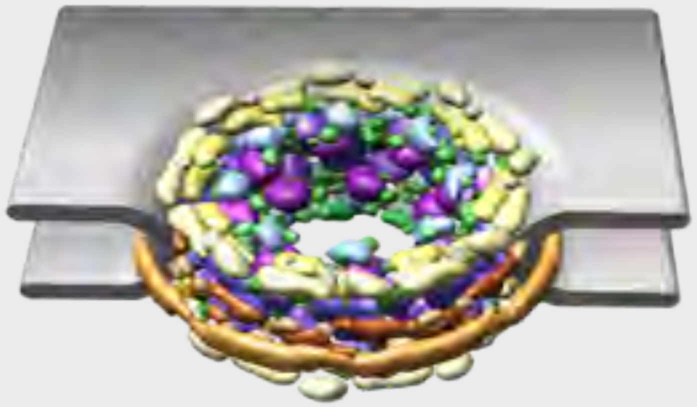
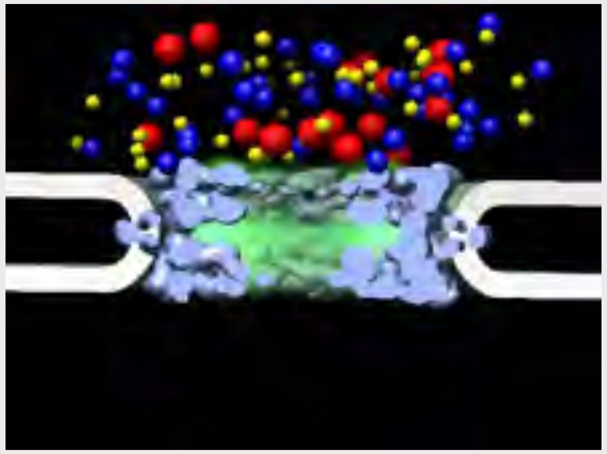
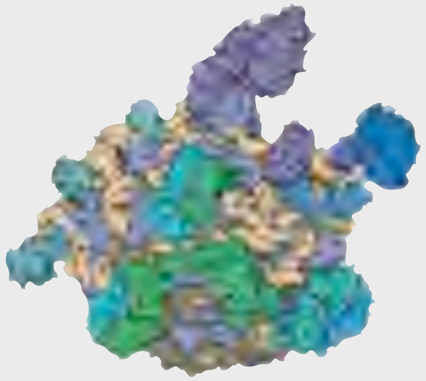
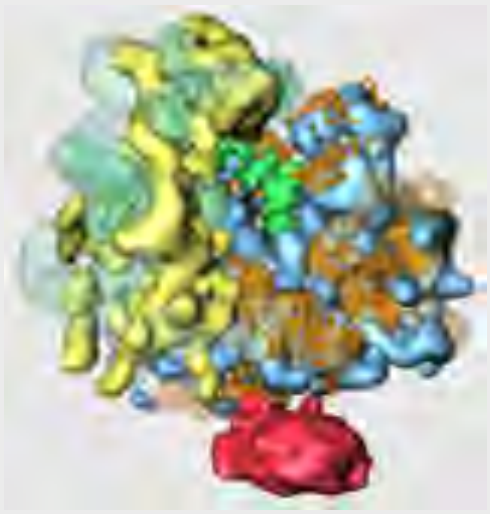
Integrative structure determination of macromolecular assemblies

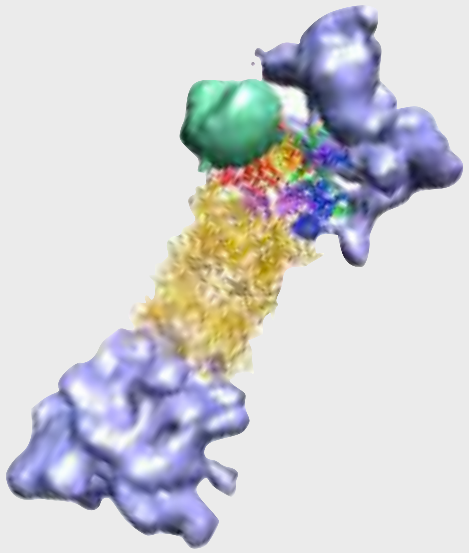












Andrej Sali
http://salilab.org/

Department of Bioengineering and Therapeutic Sciences
Department of Pharmaceutical Chemistry
California Institute for Quantitative Biosciences
University of California, San Francisco

Disseminating structural models

Publishing models in a **printed paper**

Depositing models in a **computer database**

RCSB PDB PROTEIN DATA BANK

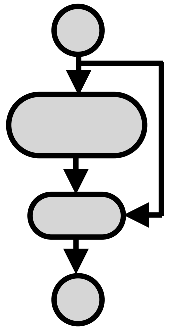
BioGRID

Depositing **input data** in a computer database

EMDB EM DATA BANK

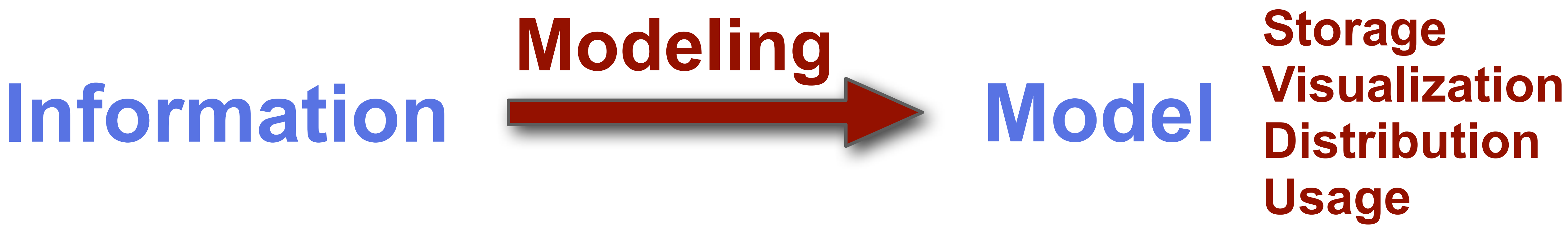
PubMed.gov U.S. National Library of Medicine National Institutes of Health

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



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

Experiment
Hypothesis



Information

Modeling

Model

Storage
Visualization
Distribution
Usage

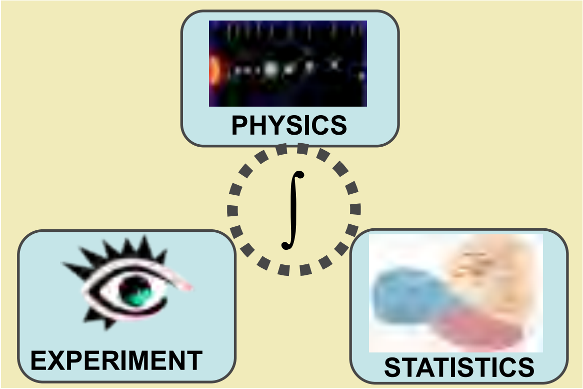
- **Types** of structural models (static and dynamic):
  - **information**: X-ray, NMR, EM, and SAXS structures; “theoretical” models; hybrid models
  - **representation**: atomic, coarse-grained, multi-scale models
- **PDB** is a natural facilitator of establishing conventions, standards, interfaces, assessment criteria, publication criteria, *etc*, thus catalyzing a collaborative community

Contents

1. Integrative (hybrid) structure determination
2. Fitting multiple subunits into an EM map subject to restraints from proteomics
3. Structure of the yeast Nup84 complex

Integrative determination of macromolecular structures

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

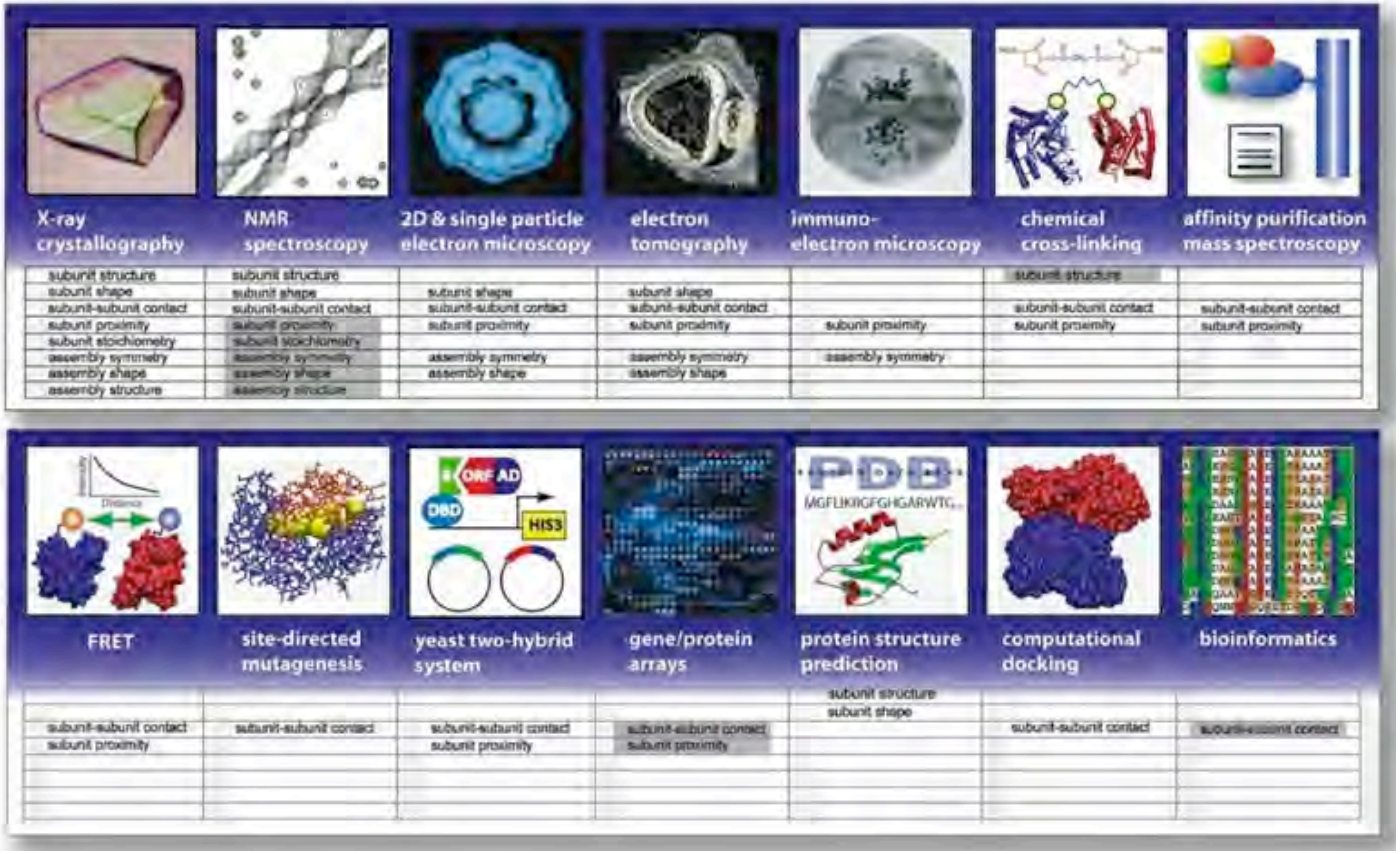


Use structural information from any

source: measurement, first principles, rules;

resolution: low or high resolution

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



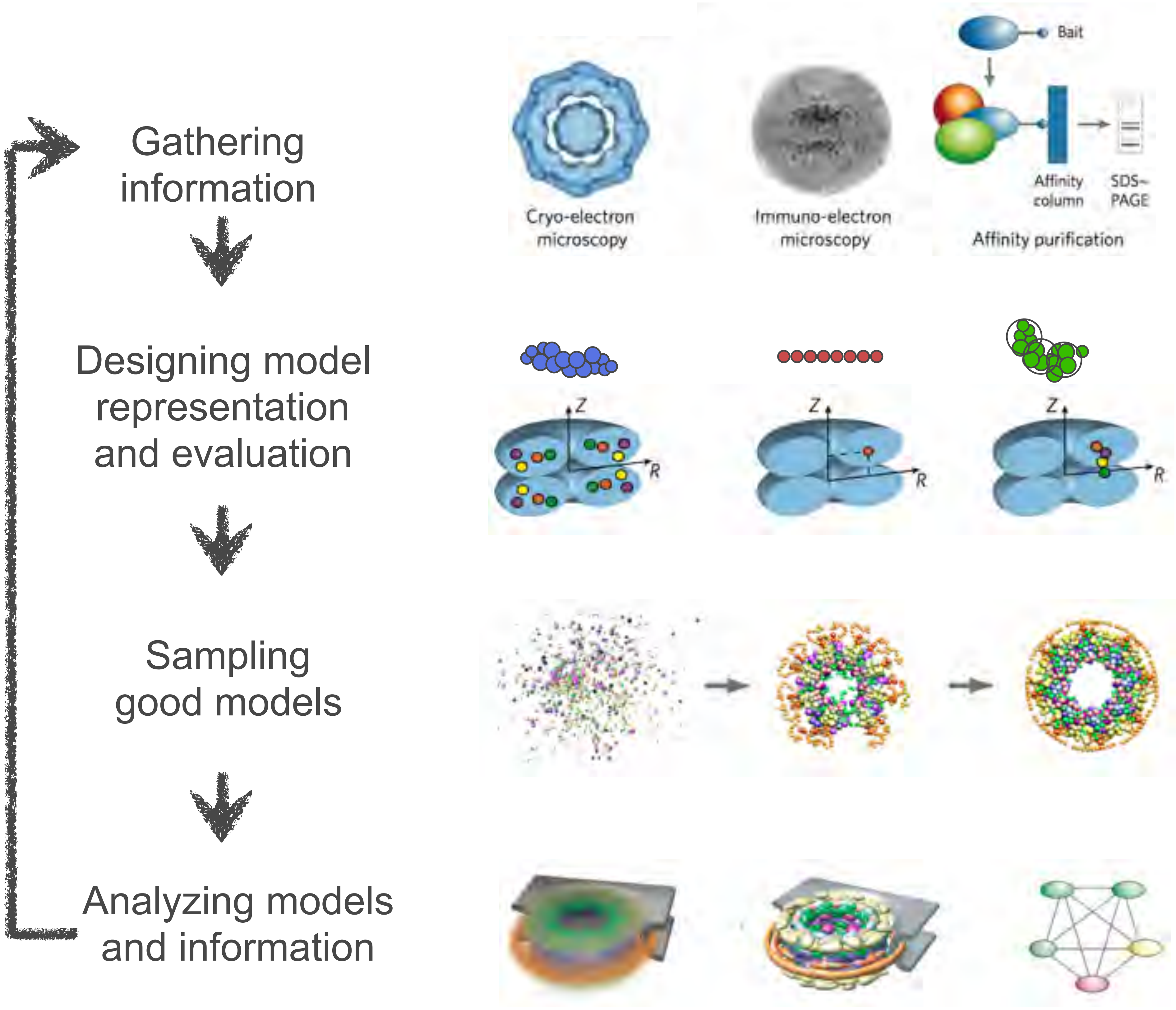
X-ray crystallography	NMR spectroscopy	2D & single particle electron microscopy	electron tomography	immuno-electron microscopy	chemical cross-linking	affinity purification mass spectroscopy
subunit structure	subunit structure				subunit structure	
subunit shape	subunit shape	subunit shape	subunit shape			
subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit-subunit contact		subunit-subunit contact	subunit-subunit contact
subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity
subunit stoichiometry	subunit stoichiometry					
assembly symmetry	assembly symmetry	assembly symmetry	assembly symmetry	assembly symmetry		
assembly shape	assembly shape	assembly shape	assembly shape			
assembly structure	assembly structure					

FRET	site-directed mutagenesis	yeast two-hybrid system	gene/protein arrays	protein structure prediction	computational docking	bioinformatics
				subunit structure		
				subunit shape		
subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit-subunit contact		subunit-subunit contact	subunit-subunit contact
subunit proximity		subunit proximity	subunit proximity			

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

An approach to integrative structure determination

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.



Gathering information

Designing model representation and evaluation

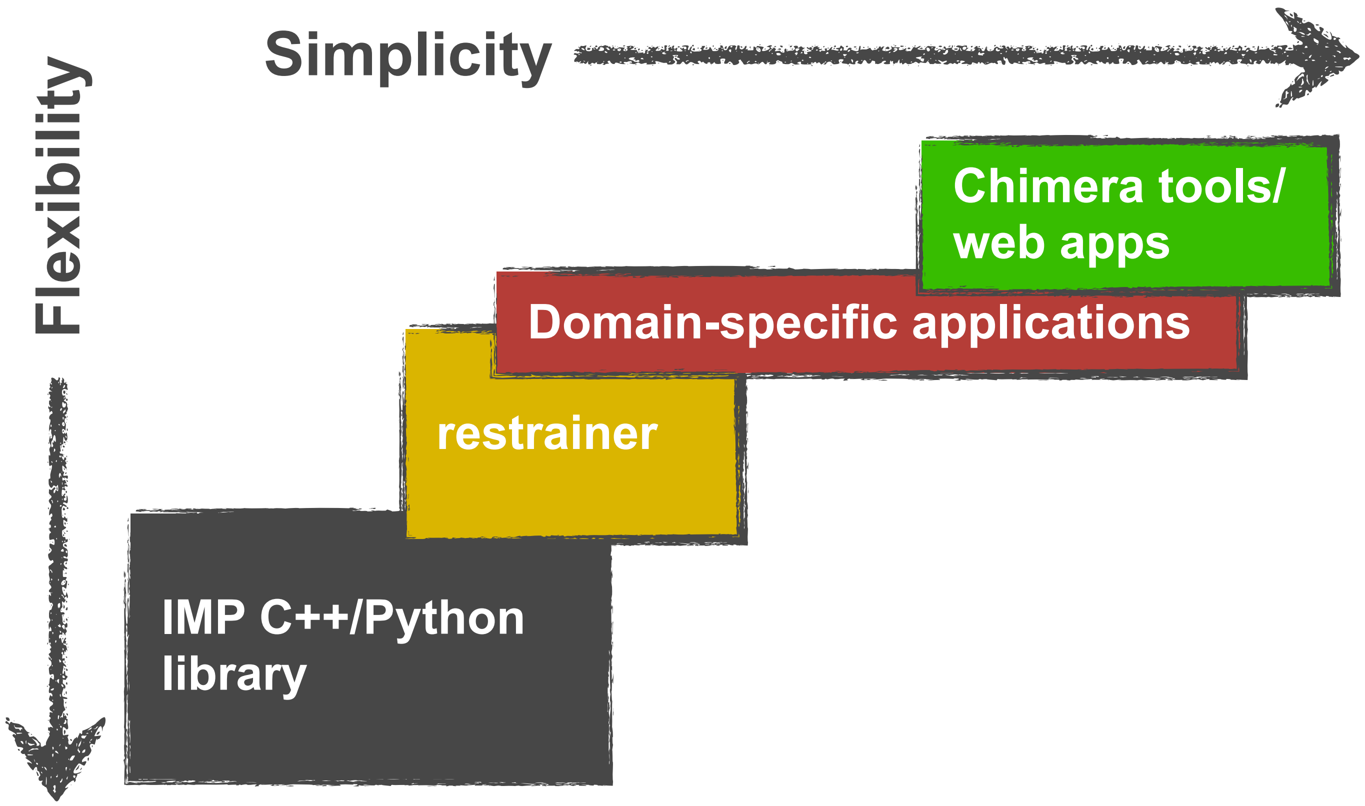
Sampling good models

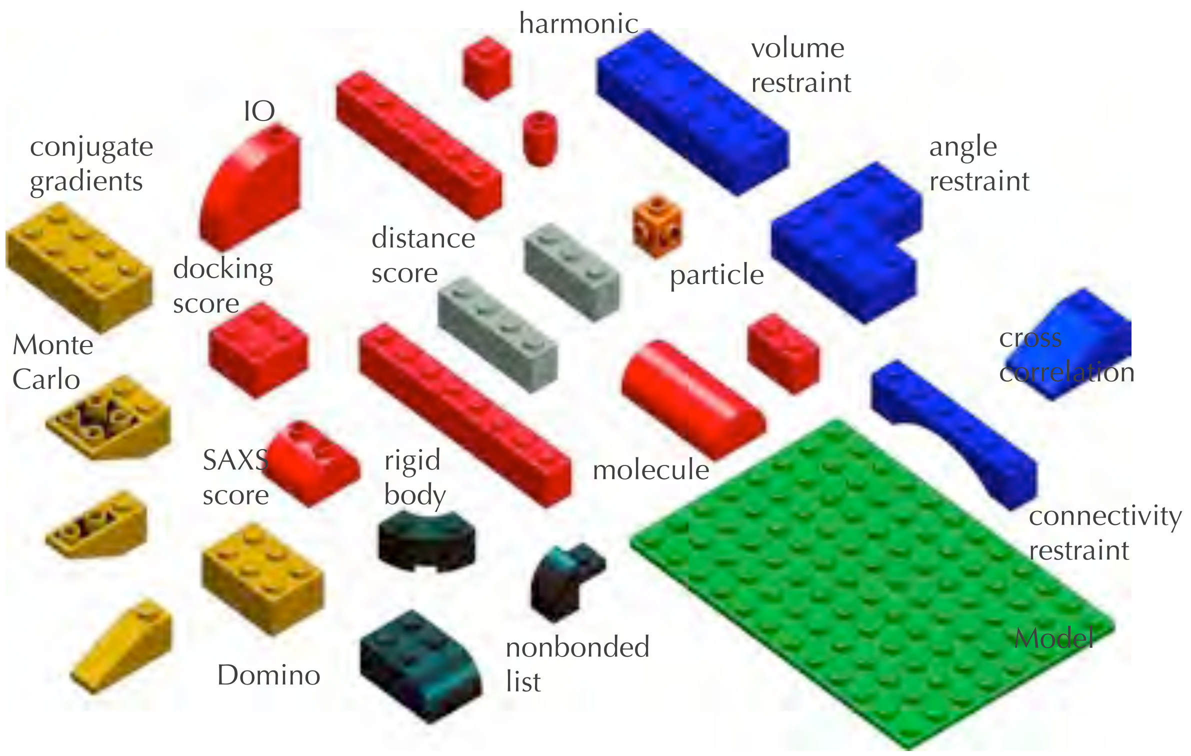
Analyzing models and information

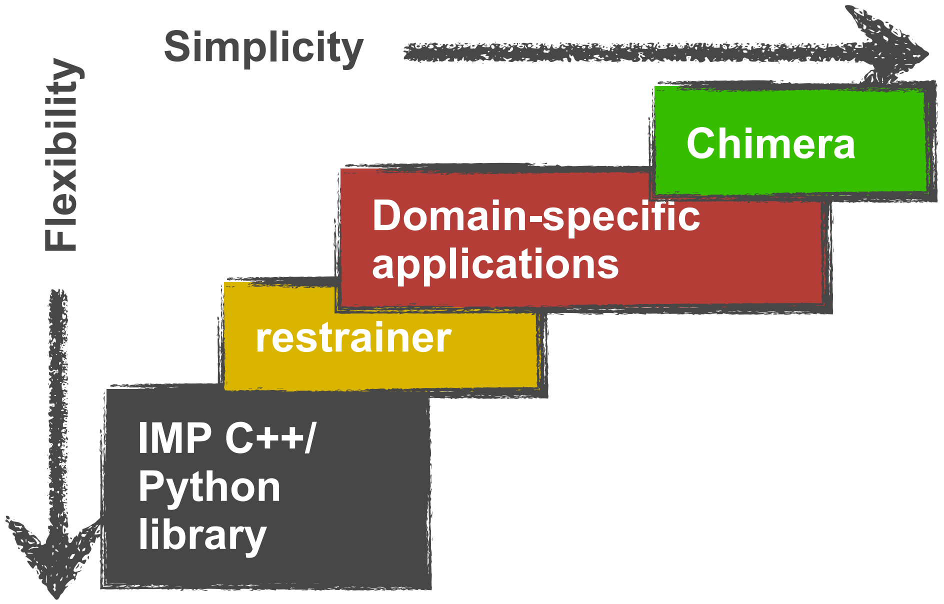
Integrative Modeling Platform (IMP)

D. Russel, K. Lasker, B. Webb, J. Velazquez-Muriel, E. Tijoe, D. Schneidman,
F. Alber, B. Peterson, A. Sali, PLoS Biol, 2011.

- IMP-1.0 available at http://salilab.org/imp/   (3/10/10)
- Open source, SVN, documentation, wiki, examples, mailing lists, unit testing, bug tracking, ...







Simplicity

Flexibility

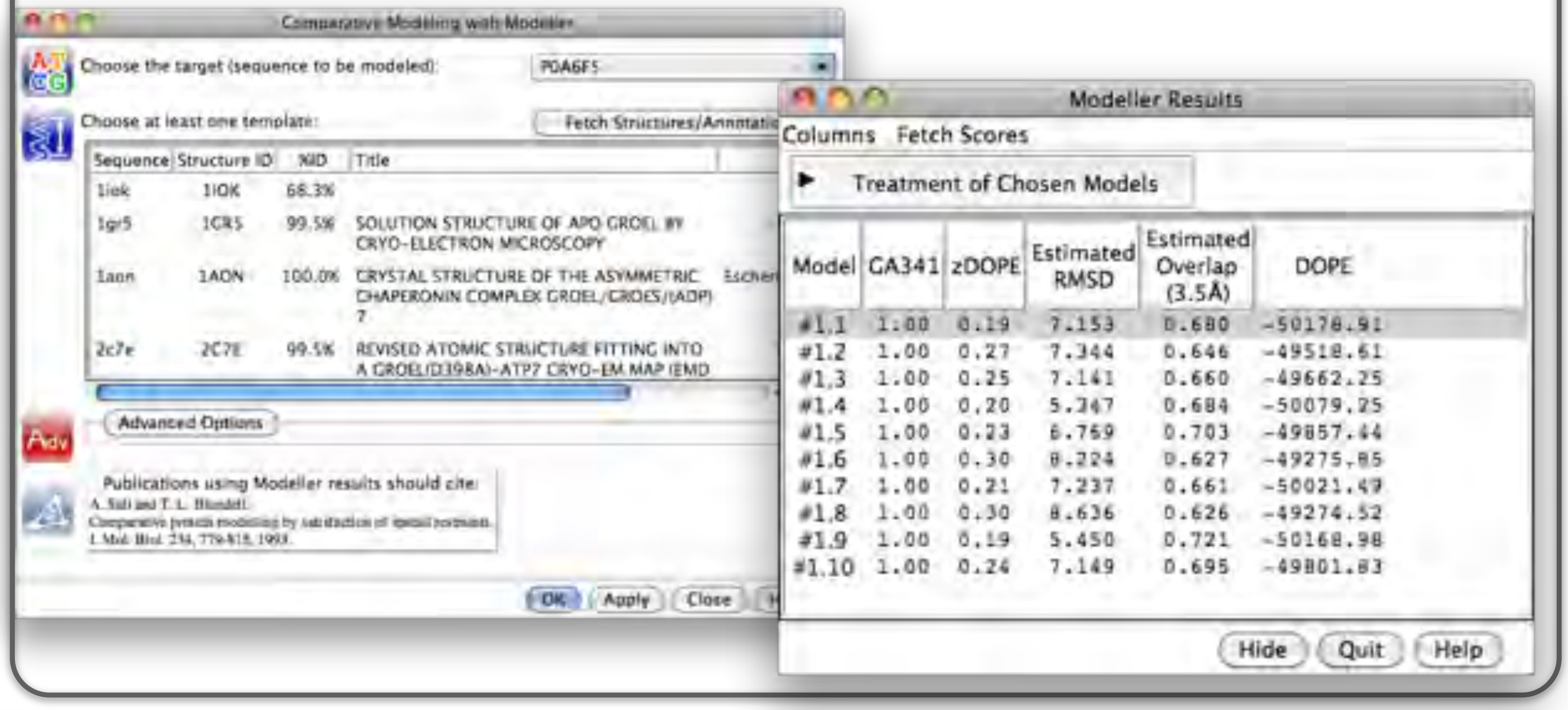
Chimera

Domain-specific applications

restrainer

IMP C++/ Python library

Modeller interface



Comparative Modeling with Modeller

Choose the target (sequence to be modeled): P0A6F5

Choose at least one template:

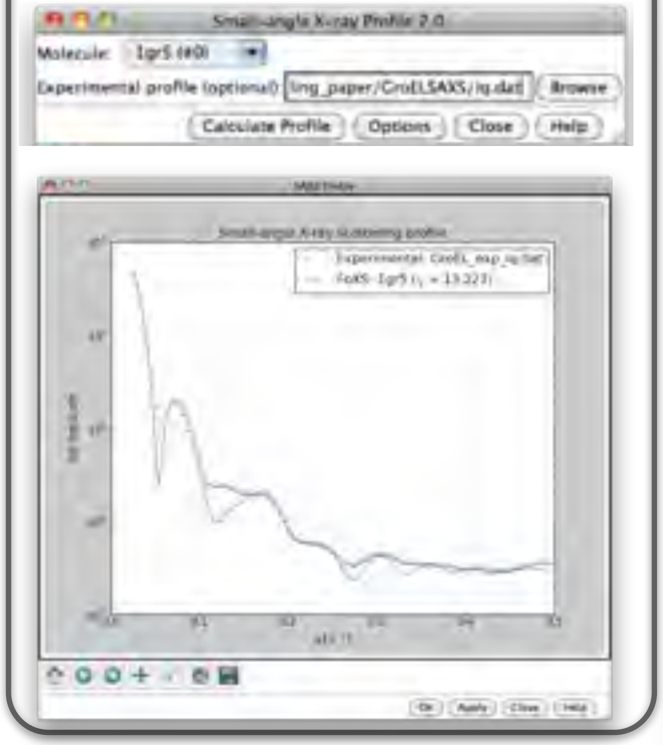
Fetch Structures/Annotatio

Sequence	Structure ID	%ID	Title	
1iok	1IOK	68.3%		
1gr5	1GR5	99.5%	SOLUTION STRUCTURE OF APO GROEL BY CRYO-ELECTRON MICROSCOPY	
1aon	1AON	100.0%	CRYSTAL STRUCTURE OF THE ASYMMETRIC CHAPERONIN COMPLEX GROEL/GROES/(ADP) 7	Escher
2c7e	2C7E	99.5%	REVISED ATOMIC STRUCTURE FITTING INTO A GROEL(D398A)-ATP7 CRYO-EM MAP (EMD	

Publications using Modeller results should cite:
A. Sali and T. L. Blundell.
Comparative protein modelling by satisfaction of spatial restraints.
J. Mol. Biol. 234, 779-815, 1993.

Model	GA341	zDOPE	Estimated RMSD	Estimated Overlap (3.5Å)	DOPE
#1.1	1.00	0.19	7.153	0.680	-50178.91
#1.2	1.00	0.27	7.344	0.646	-49518.61
#1.3	1.00	0.25	7.141	0.660	-49662.25
#1.4	1.00	0.20	5.347	0.684	-50079.25
#1.5	1.00	0.23	6.769	0.703	-49857.44
#1.6	1.00	0.30	8.224	0.627	-49275.85
#1.7	1.00	0.21	7.237	0.661	-50021.49
#1.8	1.00	0.30	8.636	0.626	-49274.52
#1.9	1.00	0.19	5.450	0.721	-50168.98
#1.10	1.00	0.24	7.149	0.695	-49801.83

FoXS interface



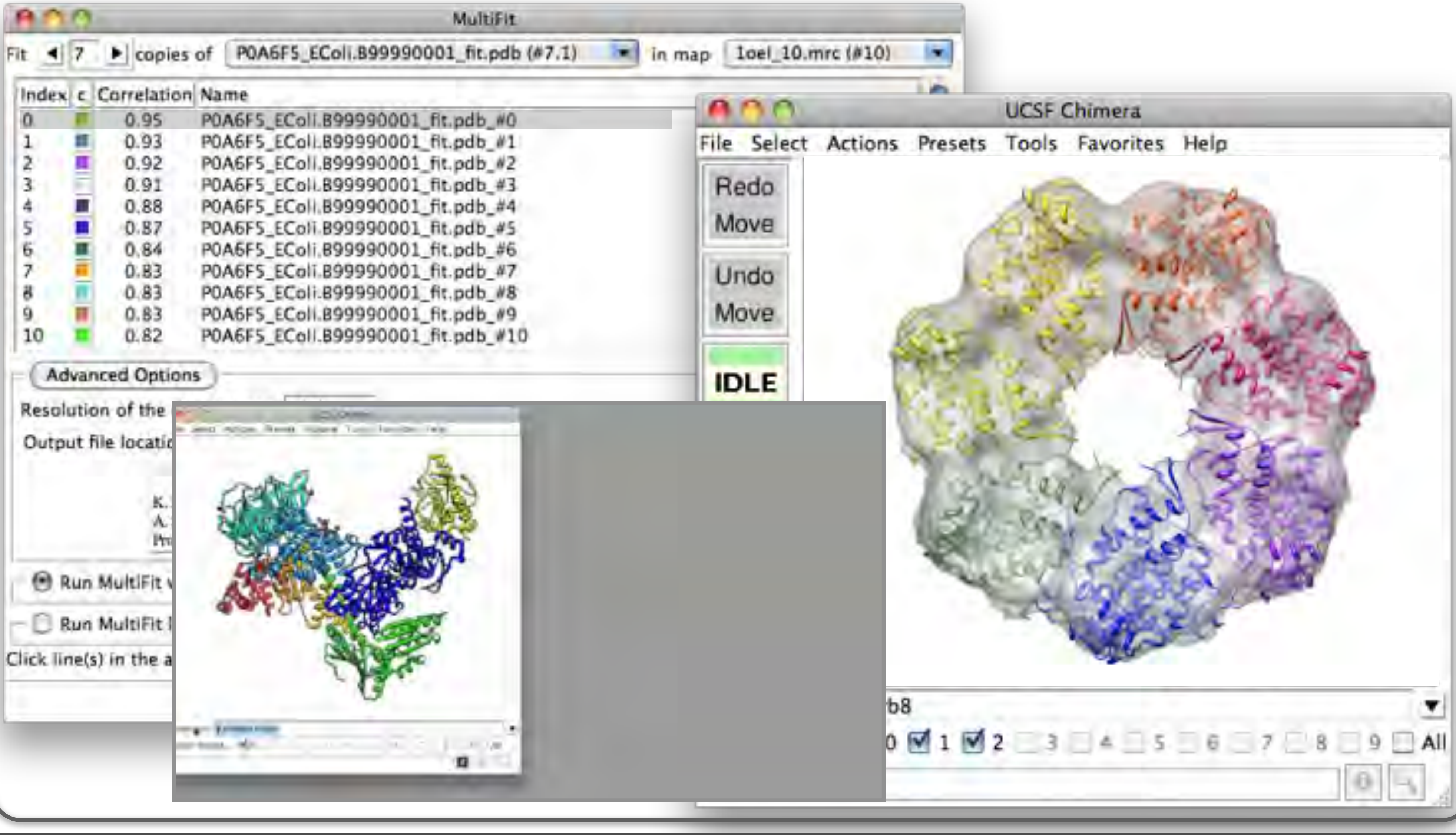
Small-angle X-ray Profile 2.0

Molecule: 1gr5 (#0)

Experimental profile (optional): ing_paper/GroELSAXS/iq.dat

Browse
Calculate Profile
Options
Close
Help
Small-angle X-ray scattering profile

MultiFit interface



Fit 7 copies of P0A6F5_EColi.B99990001_fit.pdb (#7.1) in map 1oel_10.mrc (#10)

Index	c	Correlation	Name
0	■	0.95	P0A6F5_EColi.B99990001_fit.pdb_#0
1	■	0.93	P0A6F5_EColi.B99990001_fit.pdb_#1
2	■	0.92	P0A6F5_EColi.B99990001_fit.pdb_#2
3	■	0.91	P0A6F5_EColi.B99990001_fit.pdb_#3
4	■	0.88	P0A6F5_EColi.B99990001_fit.pdb_#4
5	■	0.87	P0A6F5_EColi.B99990001_fit.pdb_#5
6	■	0.84	P0A6F5_EColi.B99990001_fit.pdb_#6
7	■	0.83	P0A6F5_EColi.B99990001_fit.pdb_#7
8	■	0.83	P0A6F5_EColi.B99990001_fit.pdb_#8
9	■	0.83	P0A6F5_EColi.B99990001_fit.pdb_#9
10	■	0.82	P0A6F5_EColi.B99990001_fit.pdb_#10

Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C. Huang, E. Pettersen, T. Goddard, E. Meng, A. Sali, T. Ferrin. UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J Struct Biol*, in press.

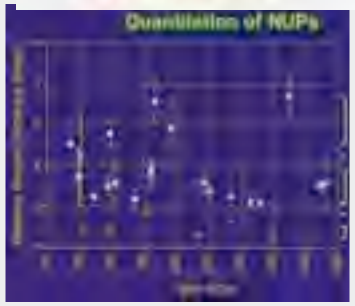
Configuration of 456 proteins in the Nuclear Pore Complex

with M. Rout & B. Chait

Alber *et al*. *Nature* 450, 684-694, 2007.
Alber *et al*. *Nature* 450, 695-702, 2007.

Quantitative Immunoblotting

30 relative abundances



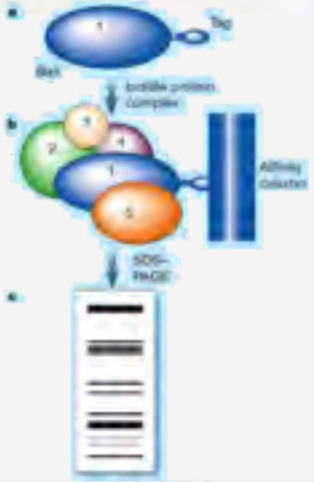
Protein Stoichiometry →

Affinity Purification

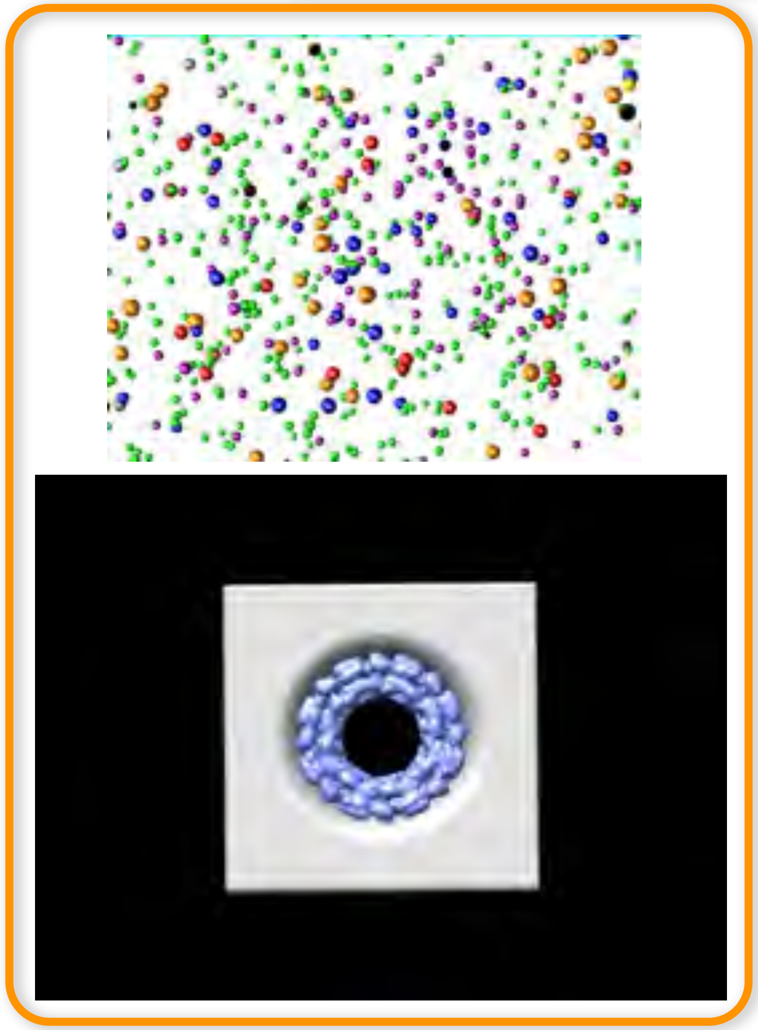
75 composites

Overlay Assay

7 contacts



Protein-protein Proximities →



Immuno-Electron Microscopy

10,615 gold particles

← Protein Localization

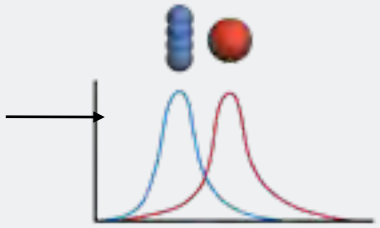
Electron Microscopy

electron microscopy map

← Symmetry

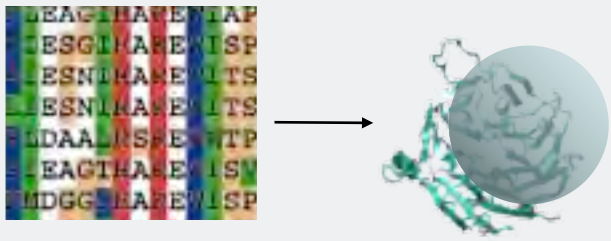
Ultracentrifugation

30 S-values 1 S-value



Bioinformatics and Membrane Fractionation

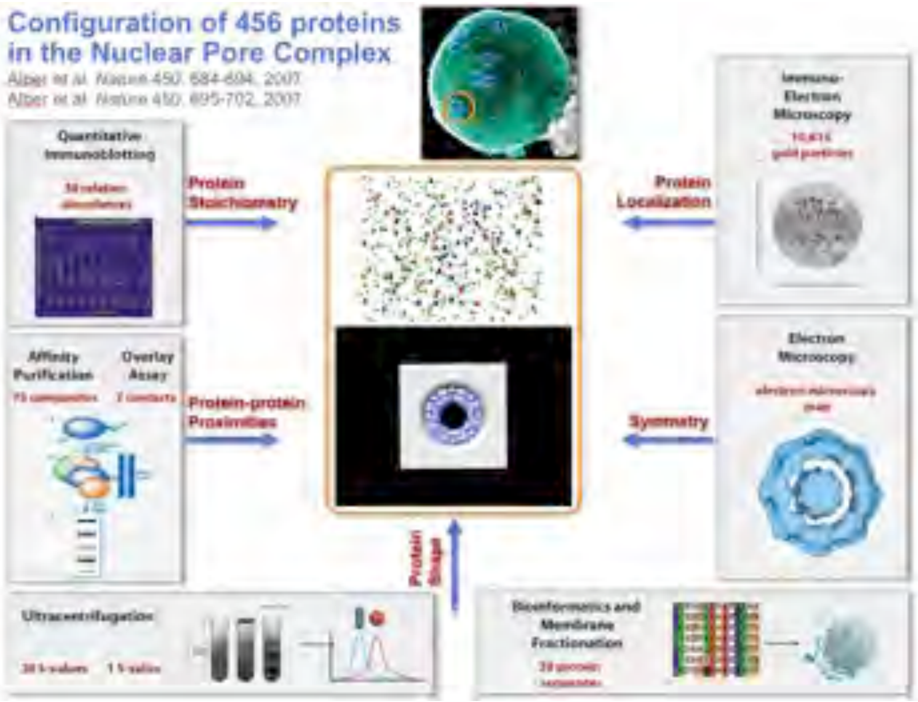
30 protein sequences



↑ Protein Shape

Determination by experiment *versus* prediction by modeling

Integrative structure determination

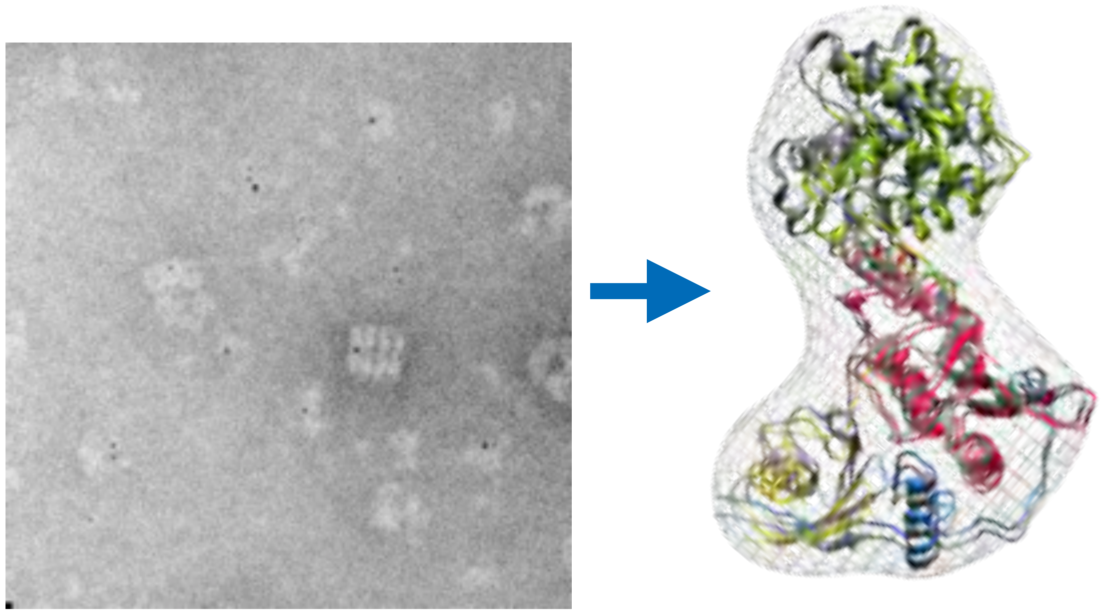


Configuration of 456 proteins in the Nuclear Pore Complex

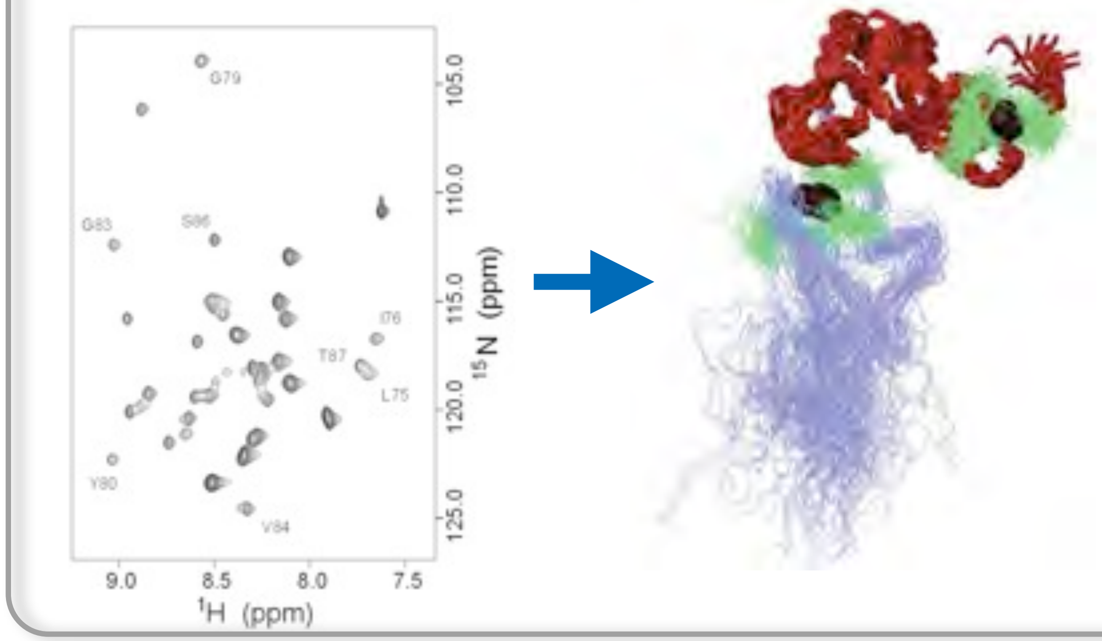
Alber et al. Nature 450: 684-694, 2007

Alber et al. Nature 450: 695-702, 2007

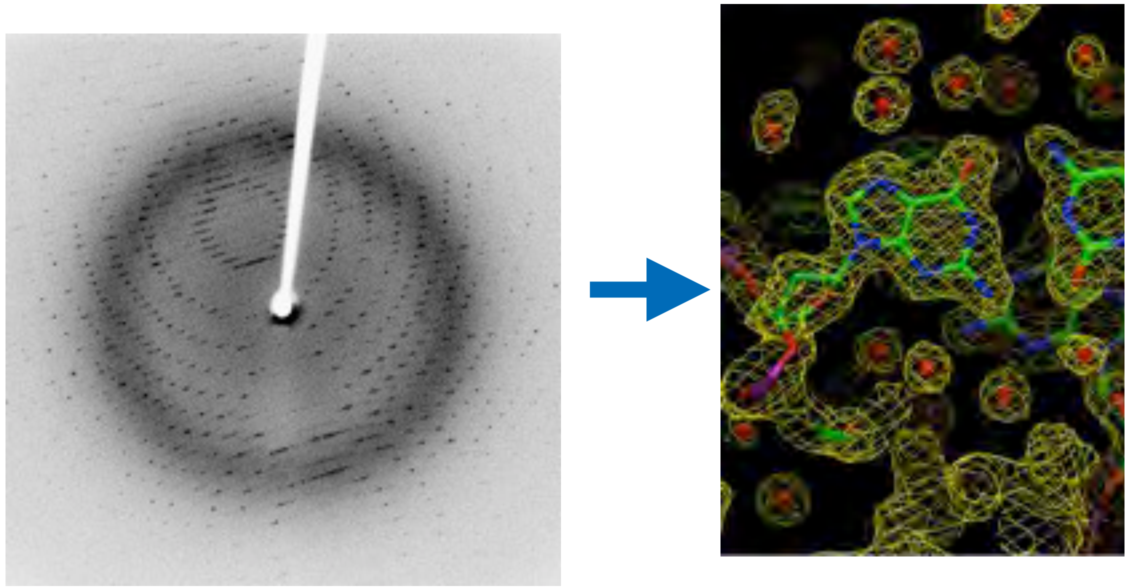
EM microscopy



NMR spectroscopy



X-ray crystallography

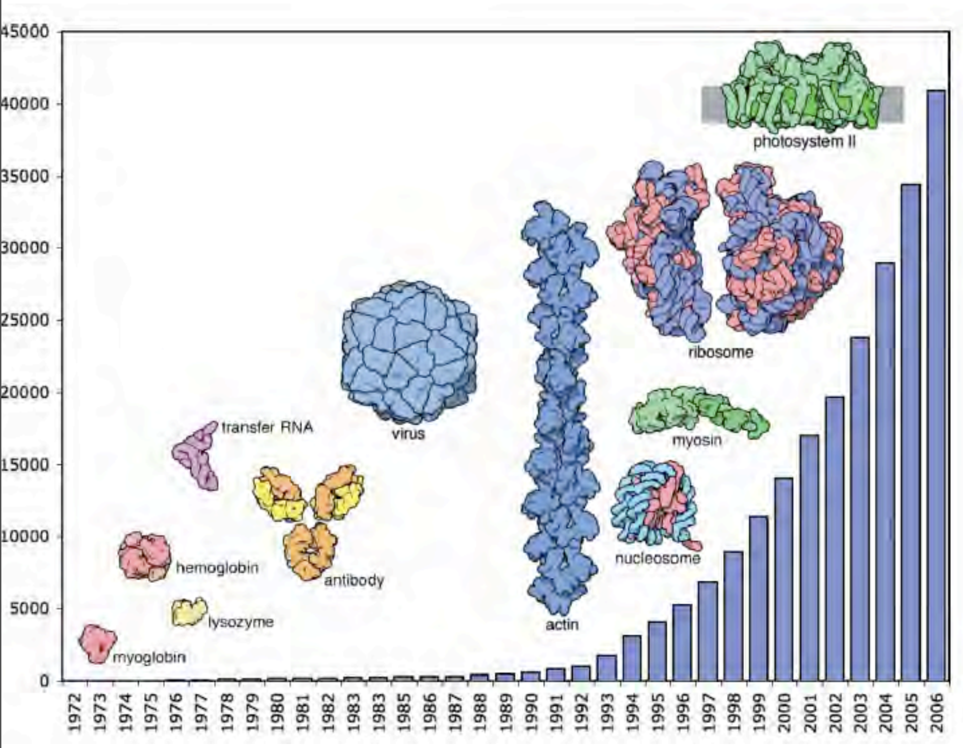


Contents

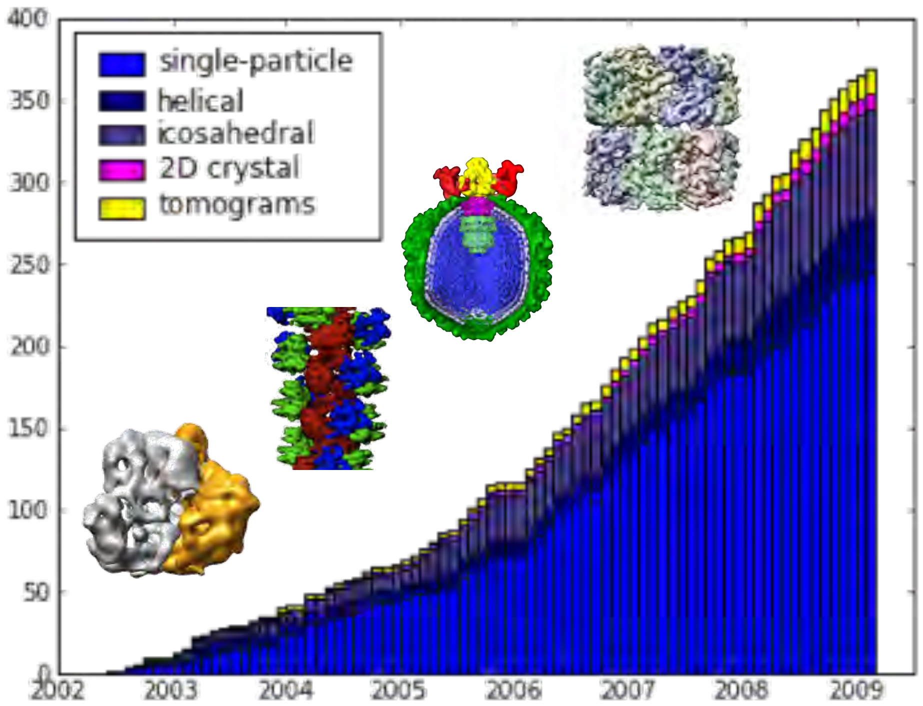
1. Integrative (hybrid) structure determination
2. Fitting multiple subunits into an EM map subject to restraints from proteomics
3. Structure of the yeast Nup84 complex

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

Protein Data Bank



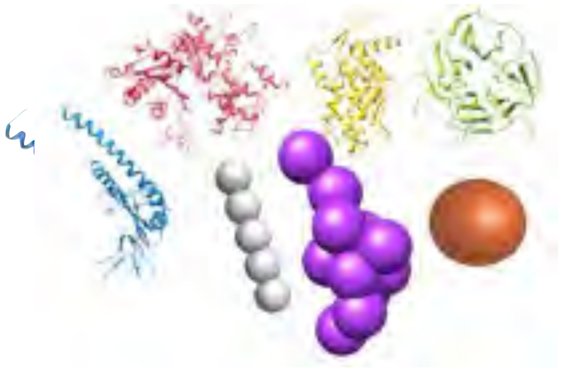
EM Data Bank

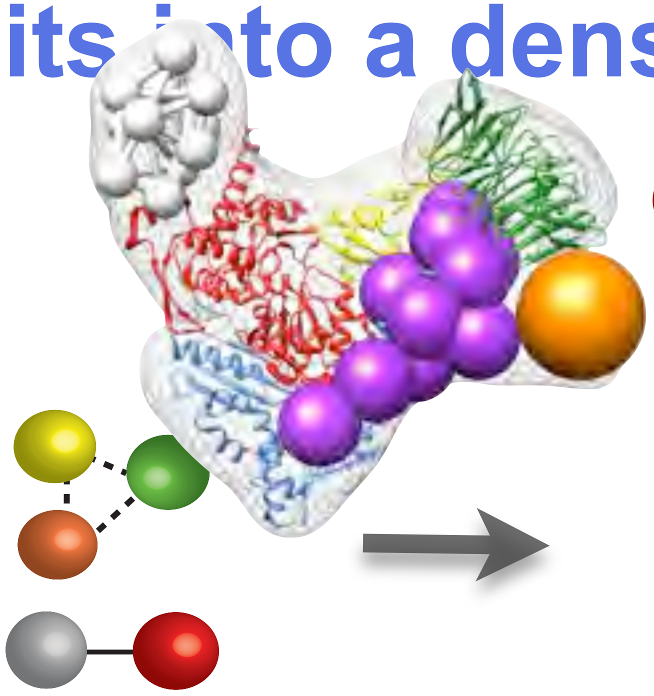


BioGrid, ...

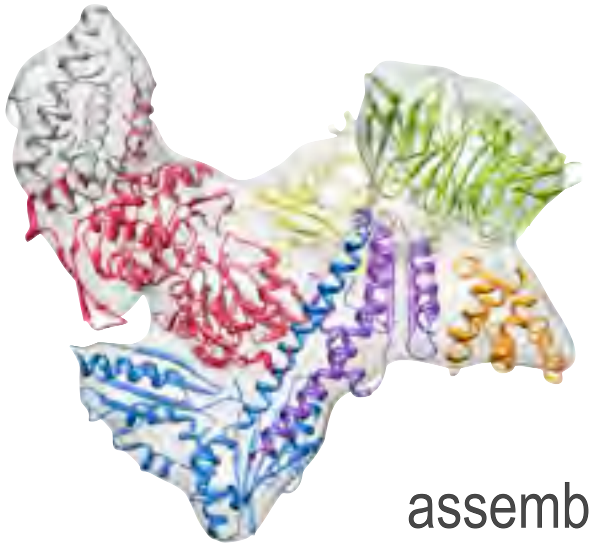
Fitting multiple subunits into a density map: Scoring

Input:





Output:



atomic, coarse components

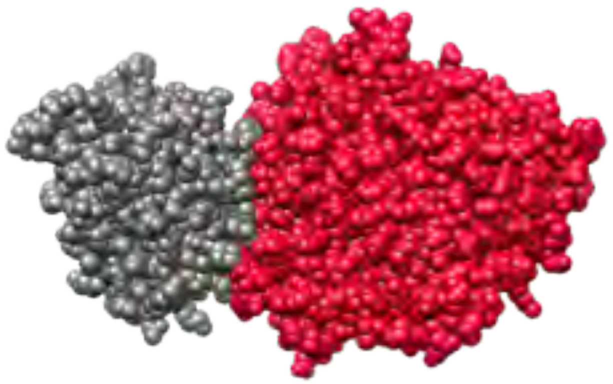
low resolution density map of the assembly

proteomics data

assembly configuration

Find assembly configurations that satisfy:

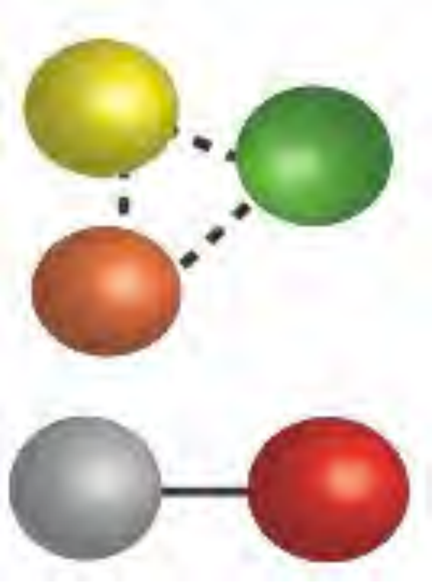
Shape complementarity



Quality-of-fit

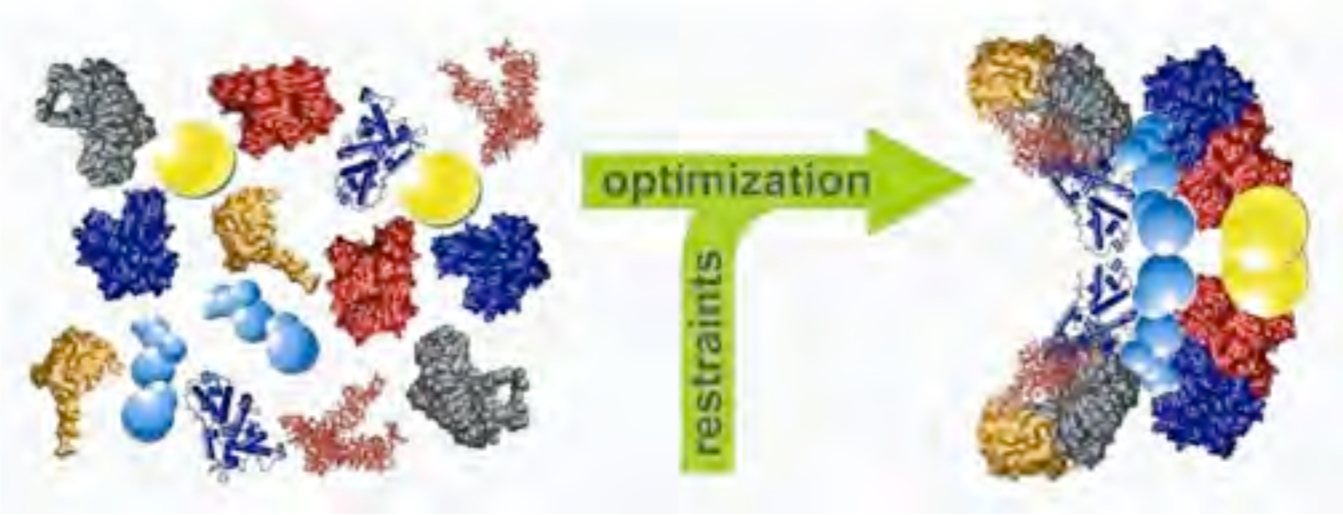
Envelope protrusion

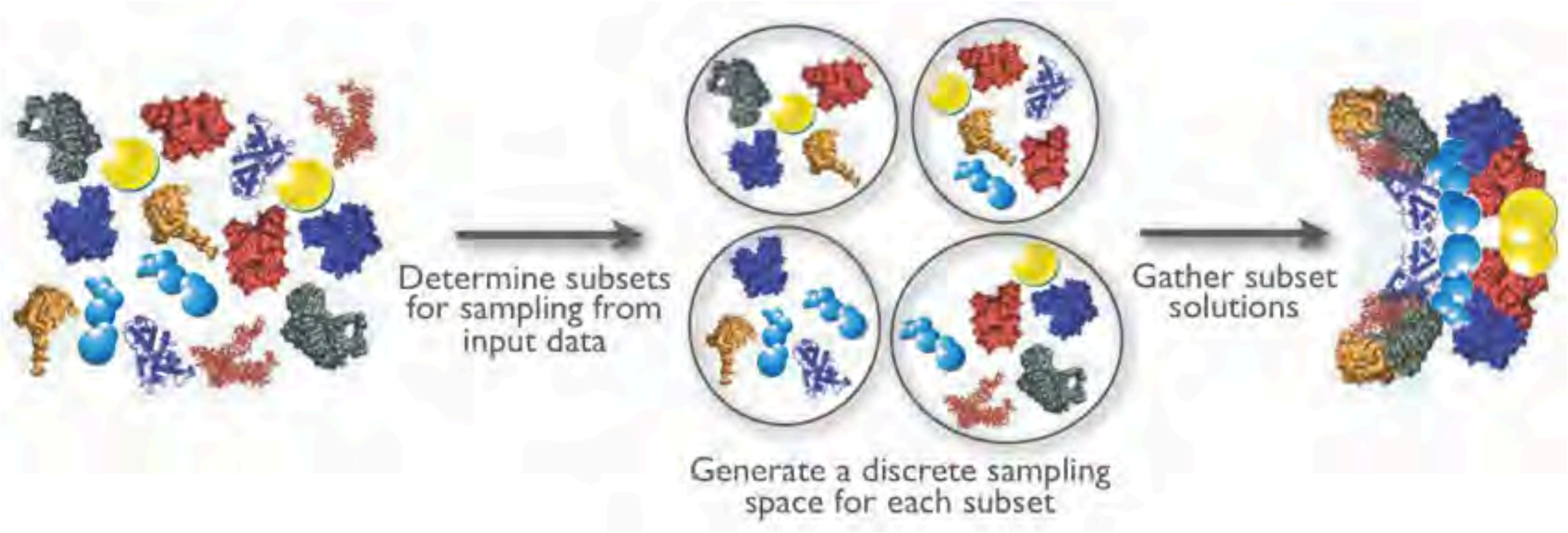
Connectivity



K. Lasker, M. Topf, A. Sali, H. Wolfson, *J. Mol. Biol.* 388, 180-194, 2009.
K. Lasker et al, *Proteins*, 2010.
K. Lasker et al, *Mol Cel Prot*, 2010.

Optimization / sampling



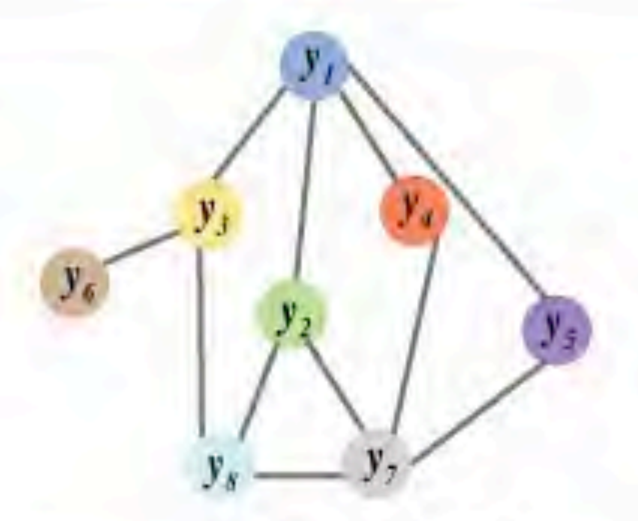

Determine subsets for sampling from input data
Generate a discrete sampling space for each subset
Gather subset solutions

K. Lasker, M. Topf, A. Sali, **H. Wolfson**, J. Mol. Biol. 388, 180-194, 2009.

Divide-and-Conquer (DOMINO)

1.**Represent** the scoring function as a graph.

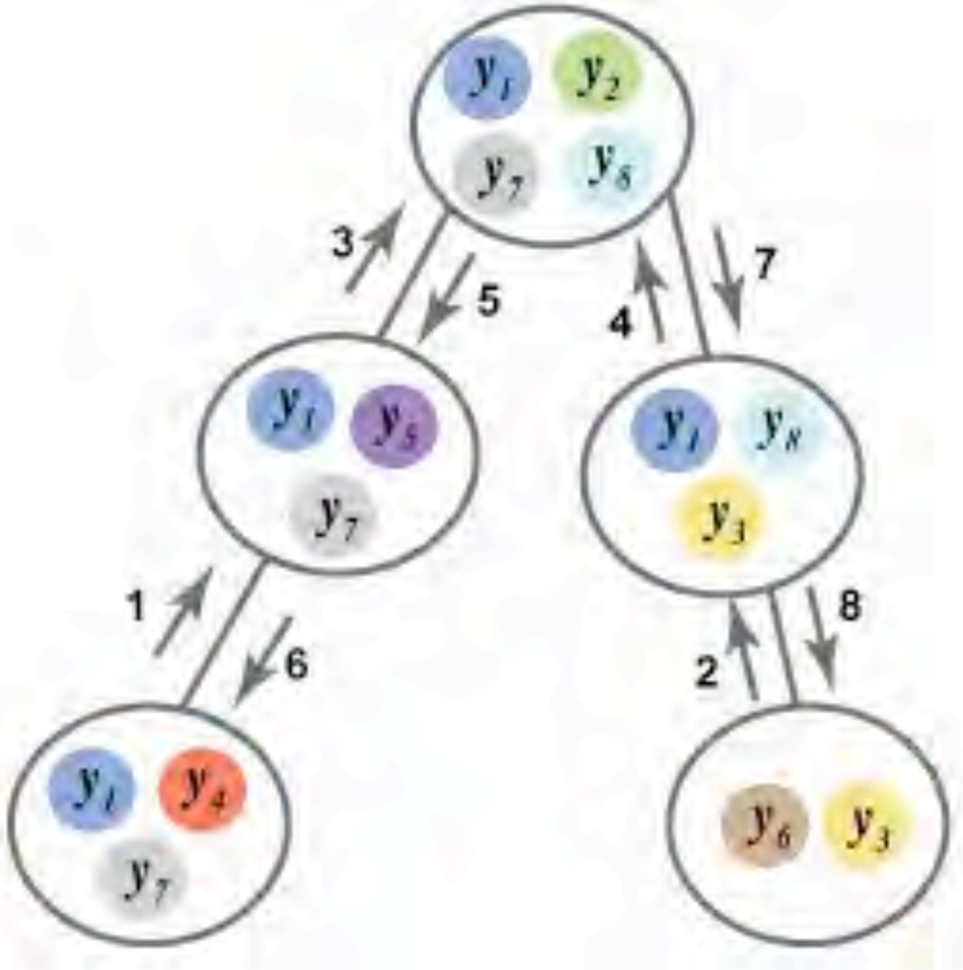
$$\begin{aligned}F(y_1,\dots,y_8) = {} & \alpha_2(y_2)+\alpha_6(y_6)+\alpha_7(y_7)\\ & +\beta_{1,2}(y_1,y_2)+\beta_{1,3}(y_1,y_3)+\beta_{1,4}(y_1,y_4)+\beta_{1,5}(y_1,y_5)\\ & +\beta_{2,7}(y_2,y_7)+\beta_{2,8}(y_2,y_8)+\beta_{3,6}(y_3,y_6)+\beta_{3,8}(y_3,y_8)\\ & +\beta_{4,7}(y_4,y_7)+\beta_{5,7}(y_5,y_7)+\beta_{7,8}(y_7,y_8)\end{aligned}$$



2. **Decompose** the set of variables into relatively decoupled subsets (a junction tree algorithm).

3. **Optimize** each subset independently by a traditional optimizer, to get the optimal and a number of suboptimal solutions.

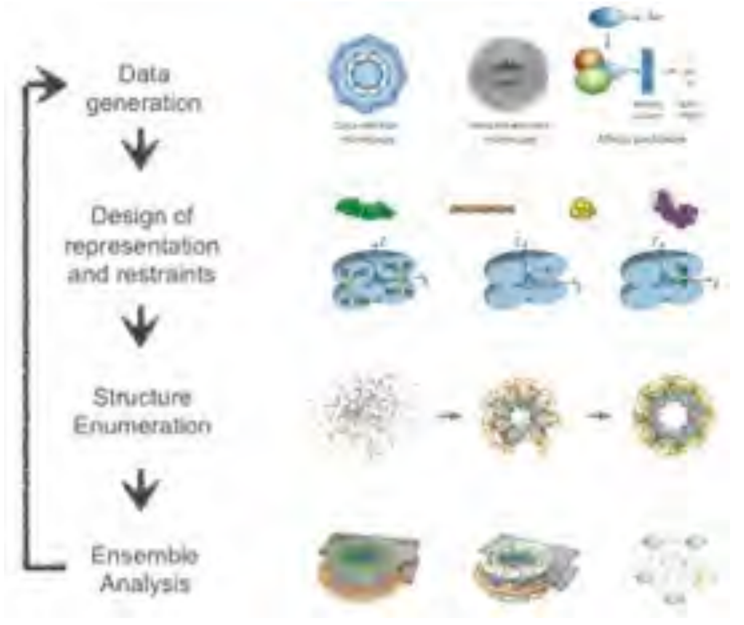
4. **Gather** subset solutions into the best possible global solutions (message passing algorithms; *eg*, belief-propagation).



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

Proof-of-principle: Integrative structure determination of human RNAPII

Lasker *et al*, MCP 2010



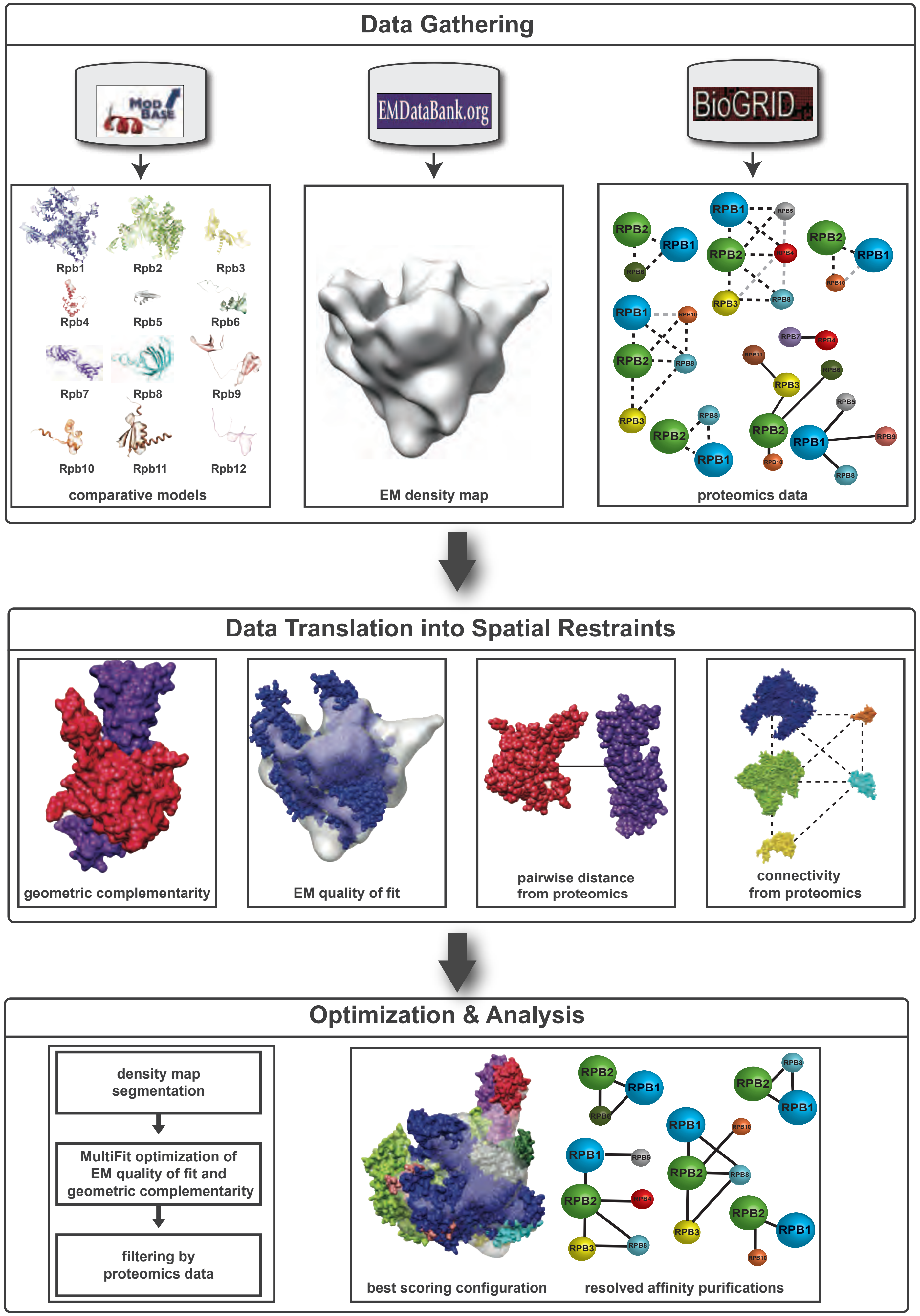
Data generation

Design of representation and restraints

Structure Enumeration

Ensemble Analysis

Cramer *et al*, *Science*, 2000 (X-ray)
Kostek *et al*, *Structure*, 2006 (EM)
Gavin *et al*, *Nature* 2006 (proteomics)
Krogan *et al*, *Nature*, 2006 (proteomics)



Data Gathering

comparative models

EM density map

proteomics data

Data Translation into Spatial Restraints

geometric complementarity

EM quality of fit

pairwise distance from proteomics

connectivity from proteomics

Optimization & Analysis

density map segmentation

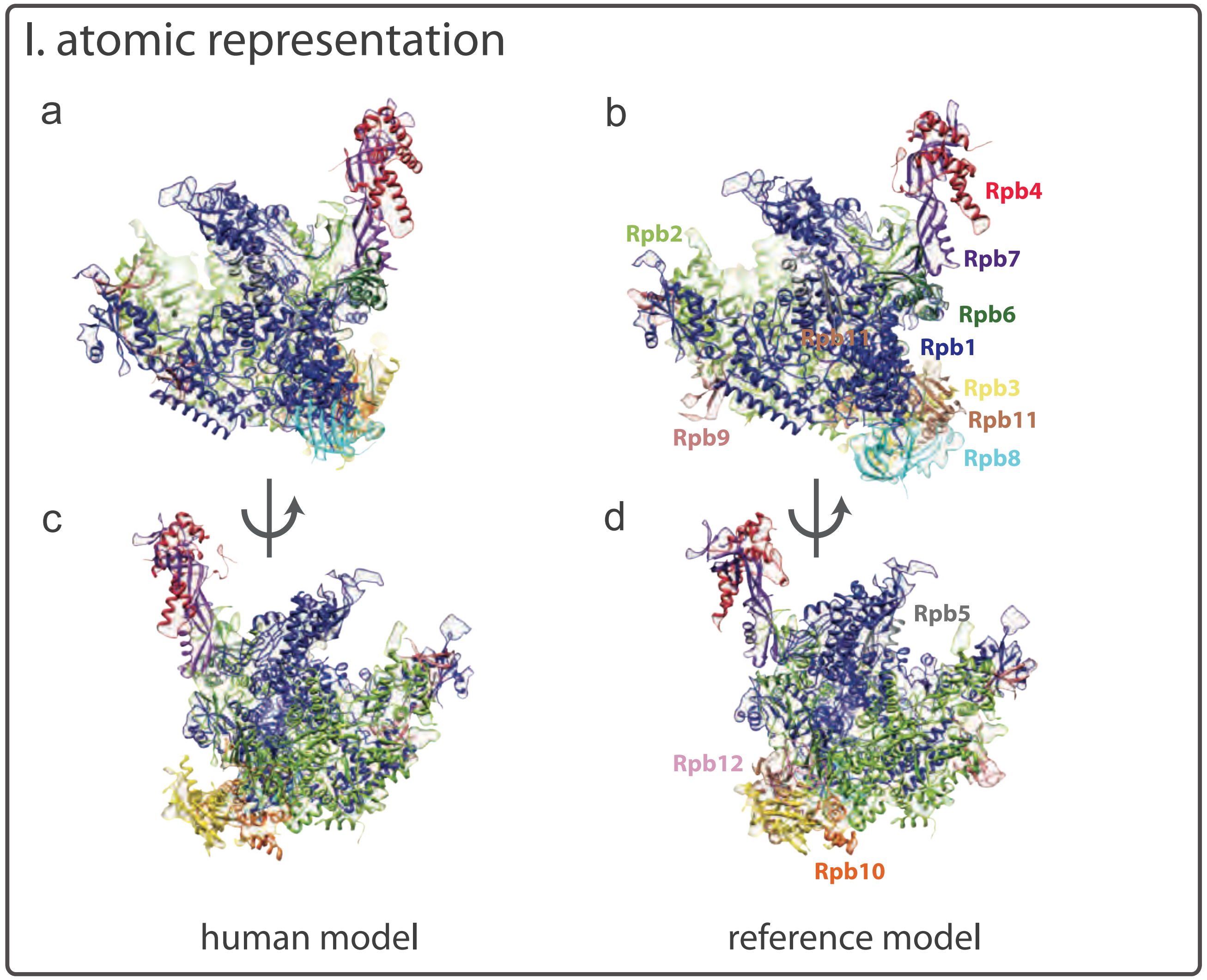
MultiFit optimization of EM quality of fit and geometric complementarity

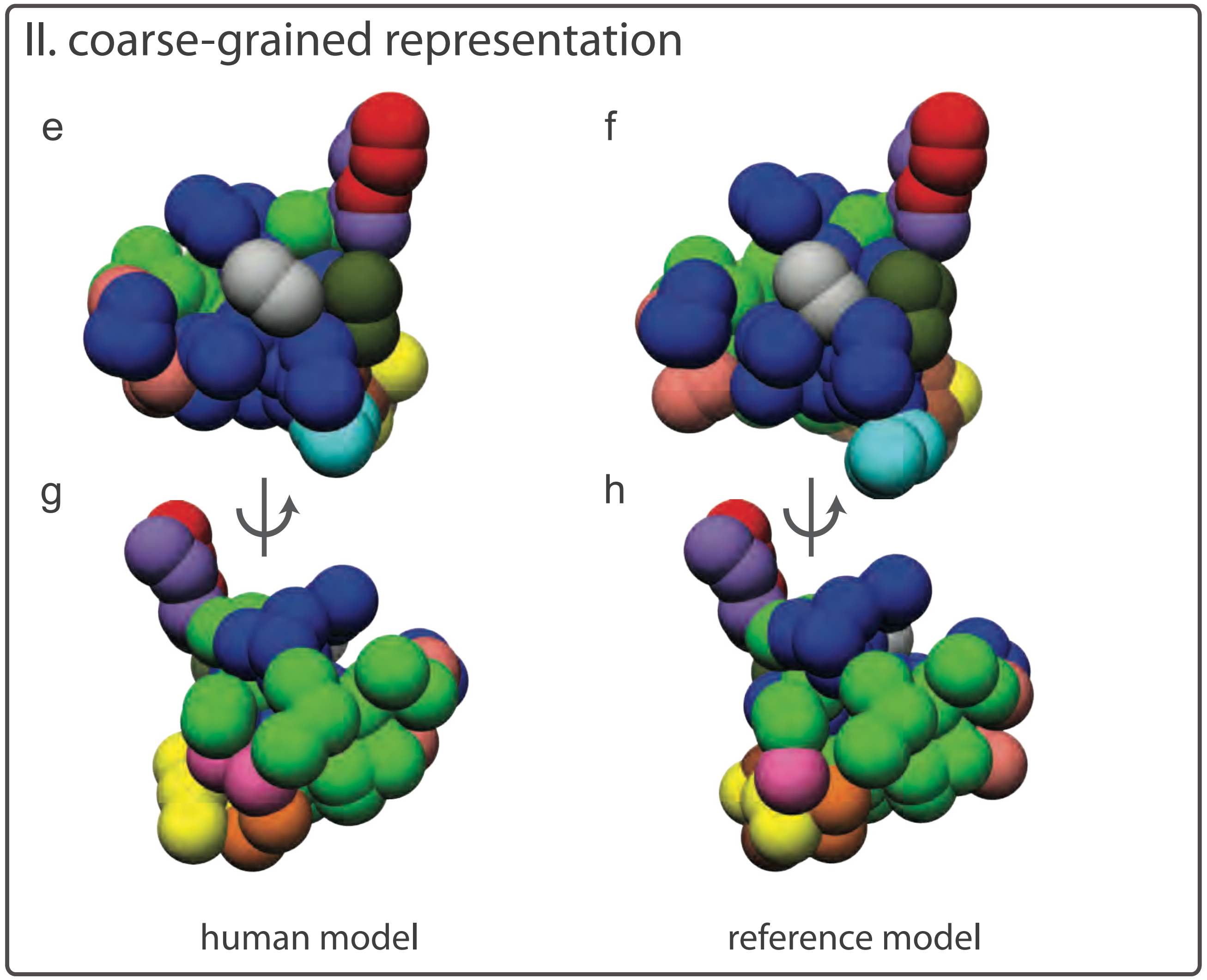
filtering by proteomics data

best scoring configuration

resolved affinity purifications

Assessment of an integrative model of human RNAPII


I. atomic representation


II. coarse-grained representation

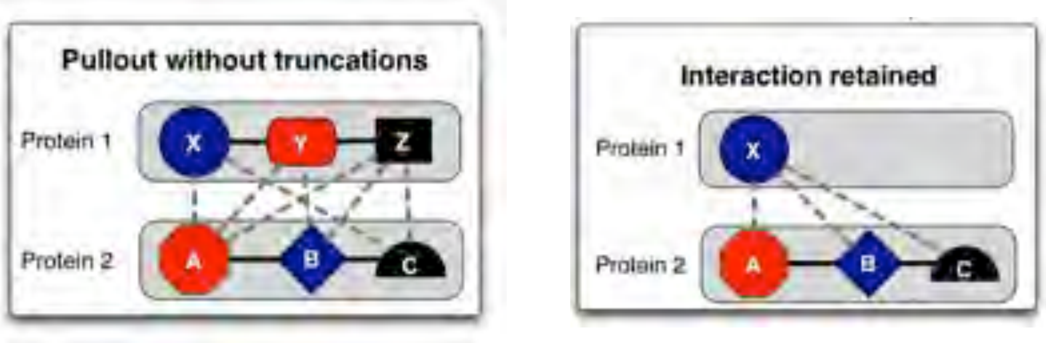
reference model - human subunit models fit on the corresponding subunits in the crystallographic yeast RNAPII structure

Additional configurational restraints

1. Affinity purification with domain deletion constructs

Orienting subunits by identification of interacting domains

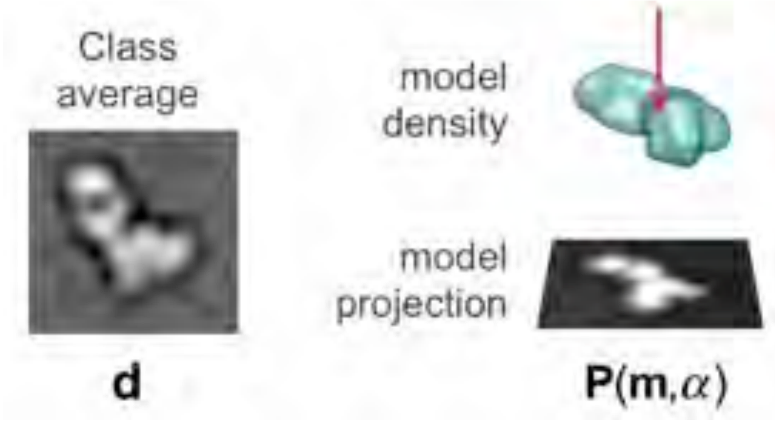
J. Phillips; with J. Fernandez, M. Rout:



2. 2D EM class averages

Filtering models by matching their optimal projections to images

J. Velazquez, D. Schneidman



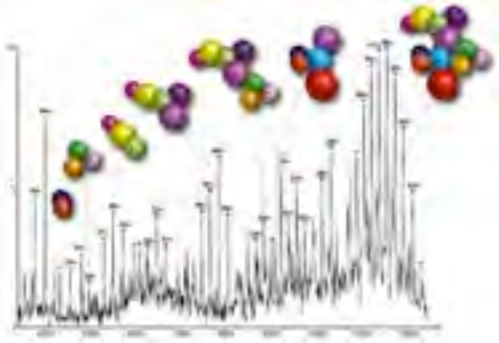
Correlation between an image and closest model projection:

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

3. Assembly subcomplex stoichiometry by native mass spectrometry

Ambiguous network of protein proximities

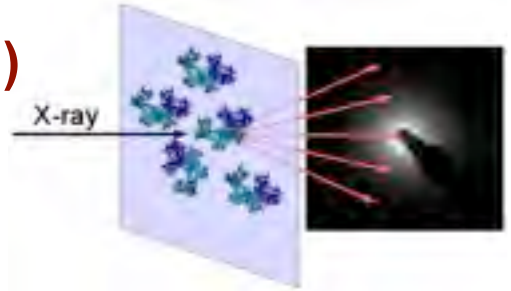
D. Russel, J. Phillips; with A. Politis, C. Robinson:

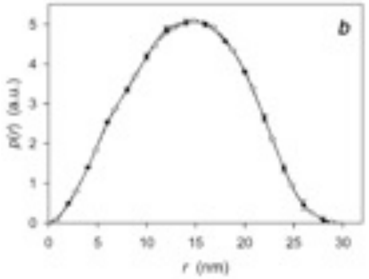


4. Small Angle X-ray Scattering (SAXS)

Filtering models by their shape

D. Schneidman, S.-J. Kim





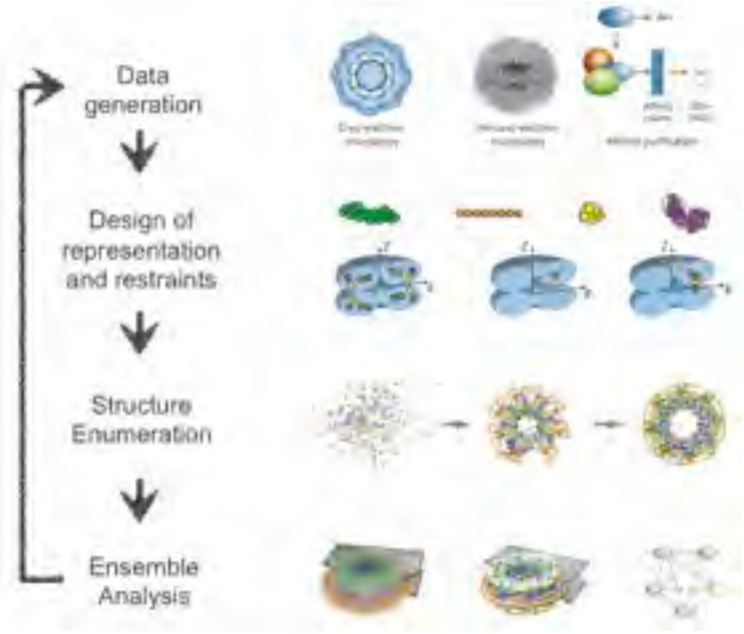
$$\chi^2 = \frac{1}{Q}\sum_{k=1}^{Q}\frac{1}{\sigma_{\mathrm{exp}}^2(q_k)}\cdot\left(I_{\mathrm{exp}}(q_k) - c\cdot I_{\mathrm{m}}(q_k)\right)^2$$

Contents

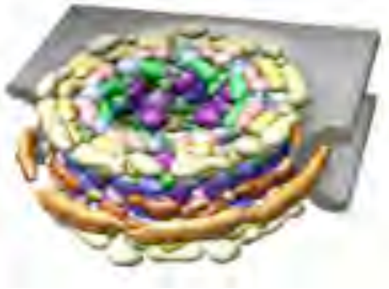
1. Integrative (hybrid) structure determination
2. Fitting multiple subunits into an EM map subject to restraints from proteomics
3. Structure of the yeast Nup84 complex

Towards a higher resolution structure of the NPC

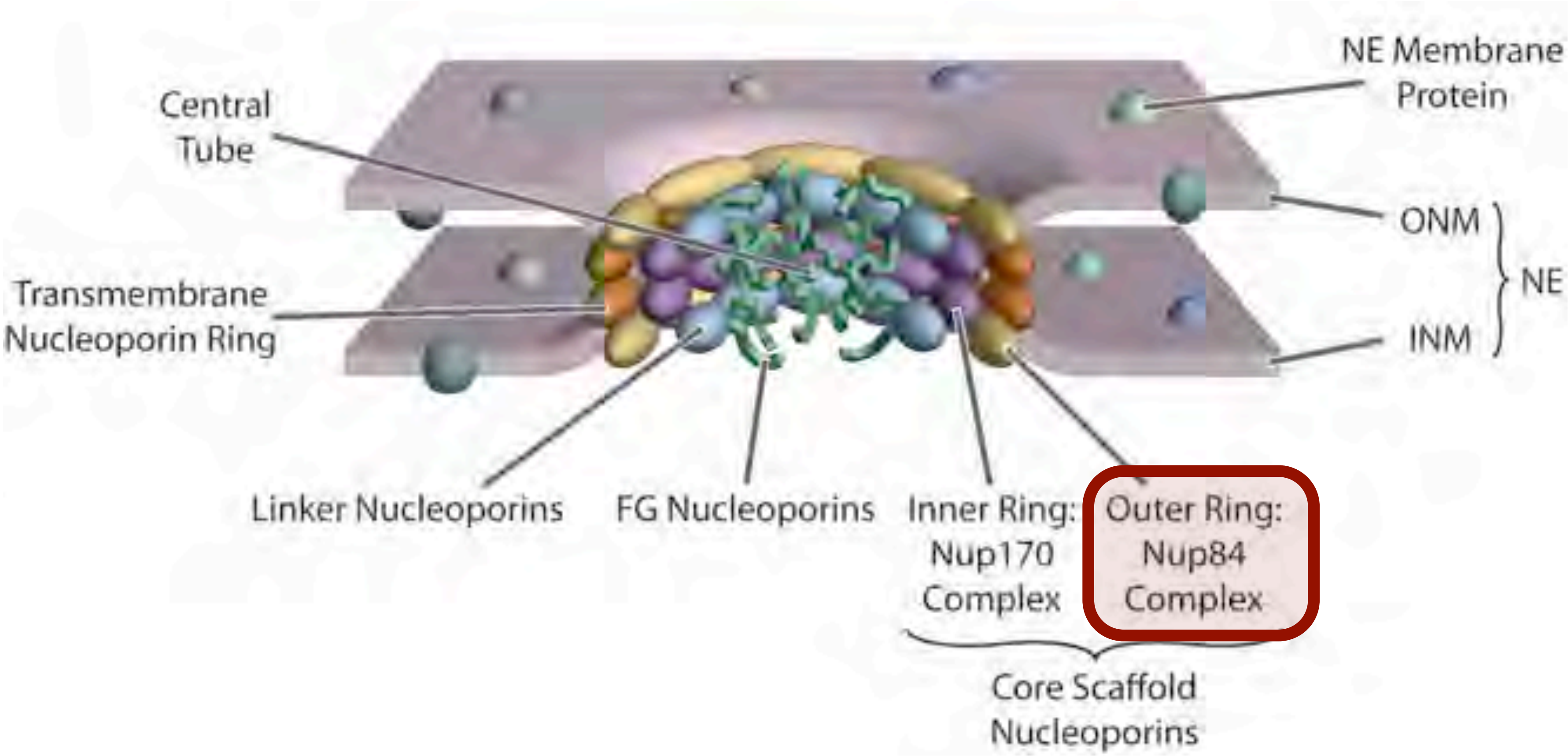
Characterize structures of the individual subunits, then fit them into the current low-resolution structure, aided by additional experimental information.



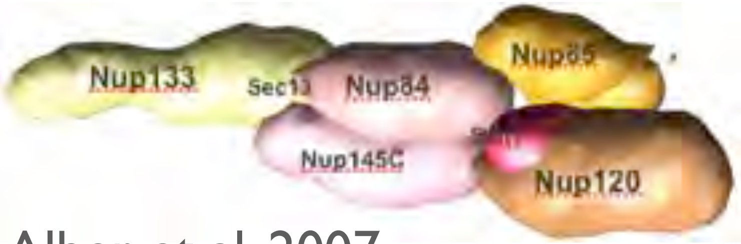
Alber *et al*. *Nature* 450, 684-694, 2007.
Alber *et al*. *Nature* 450, 695-702, 2007.



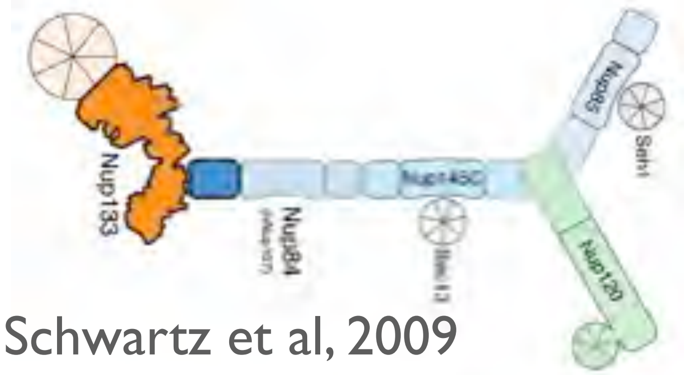
The Nup84 complex in the NPC


Central Tube
NE Membrane Protein
ONM
NE
INM
Transmembrane Nucleoporin Ring
Linker Nucleoporins
FG Nucleoporins
Inner Ring: Nup170 Complex
Outer Ring: Nup84 Complex
Core Scaffold Nucleoporins

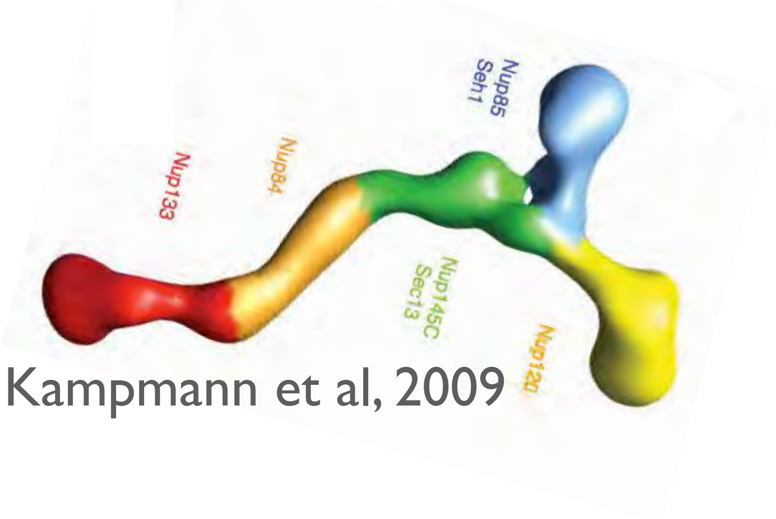
Lutzmann et al, 2002



Alber et al, 2007



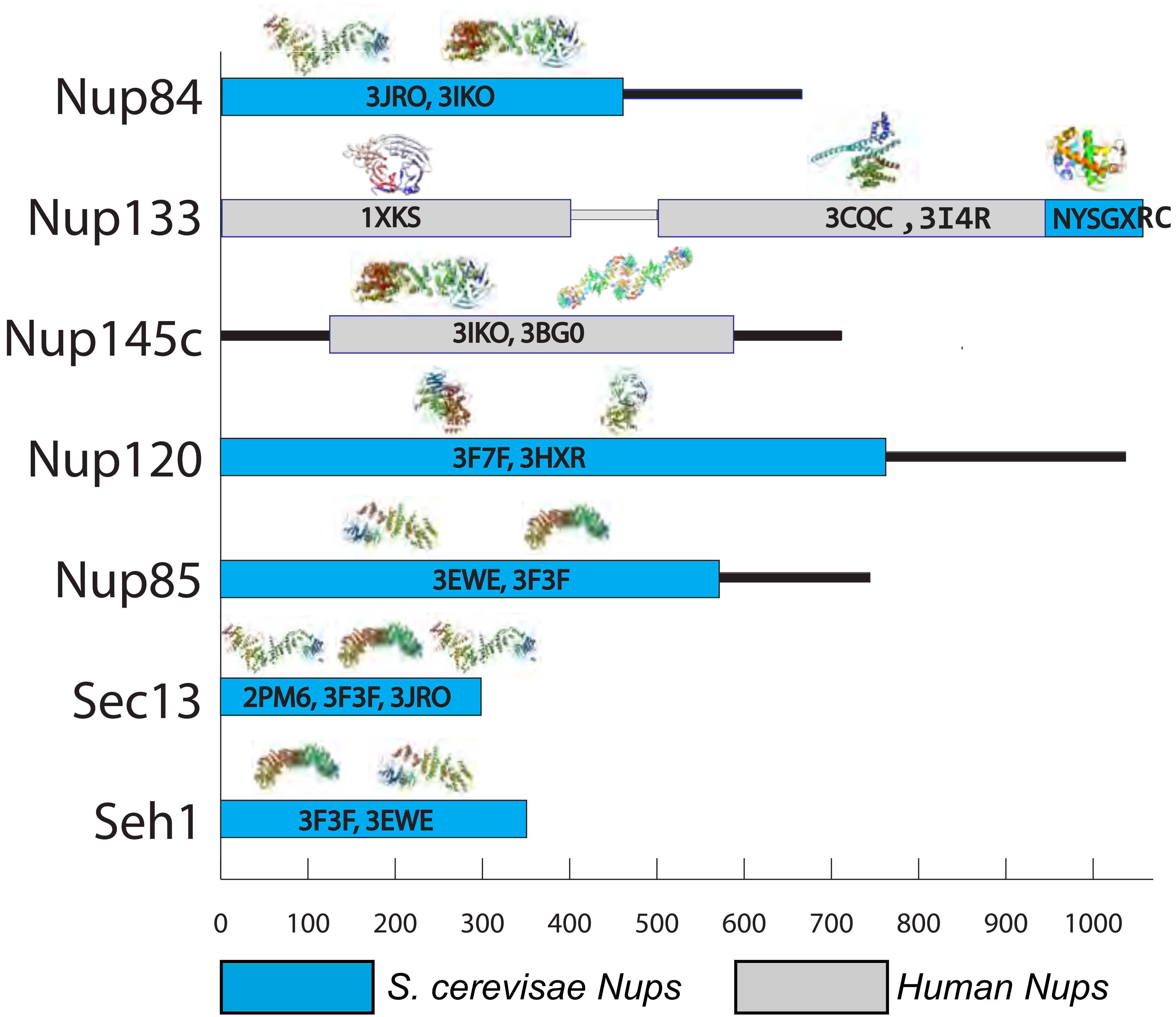
Schwartz et al, 2009



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



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

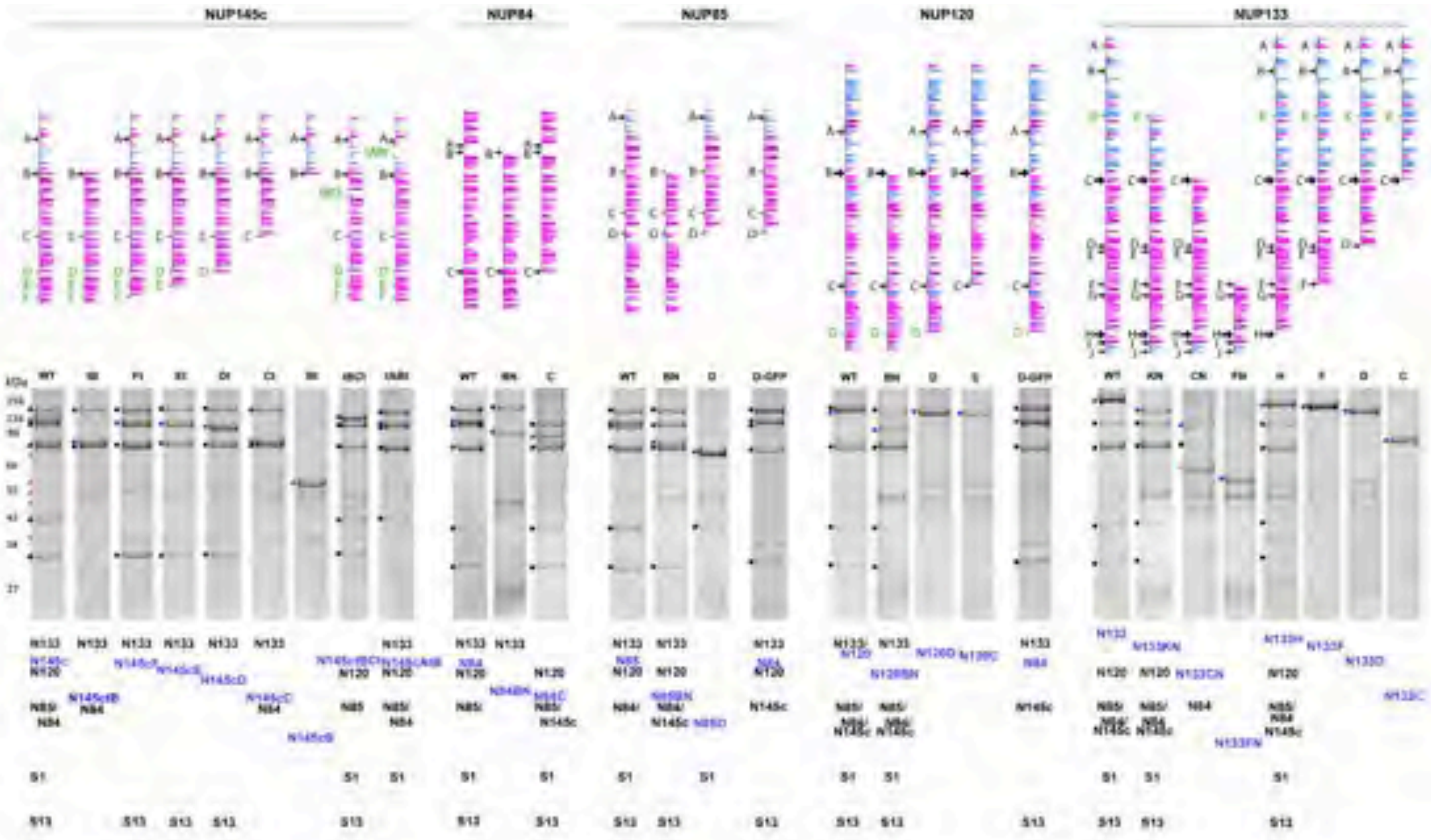
Debler et al. 2008
Brohawn et al. 2008

Nup84 complex: Data

Subunit positions & orientations

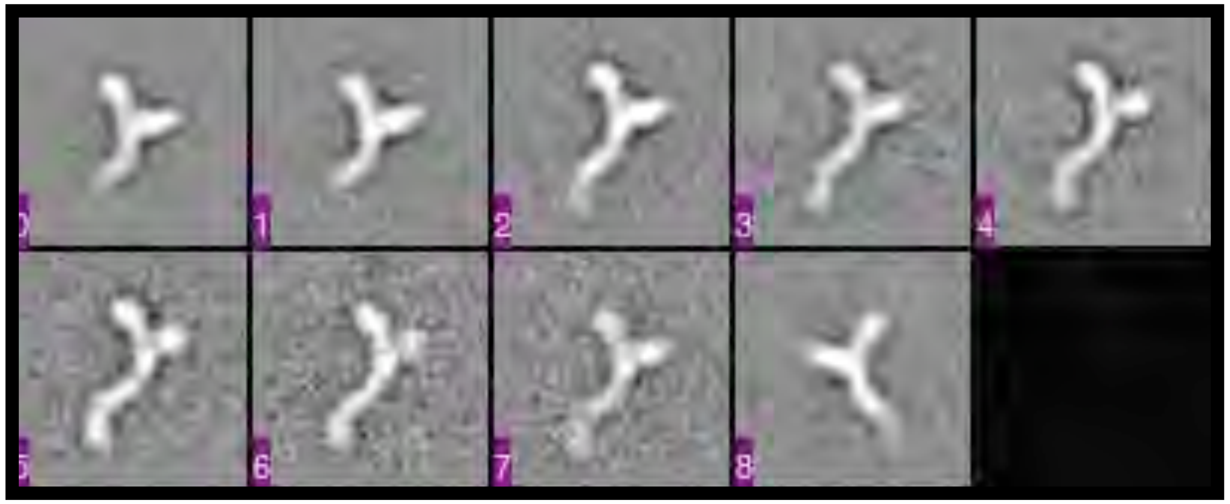
Affinity purifications with domain truncations

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



Negative stain EM particle averages at ~3nm resolution

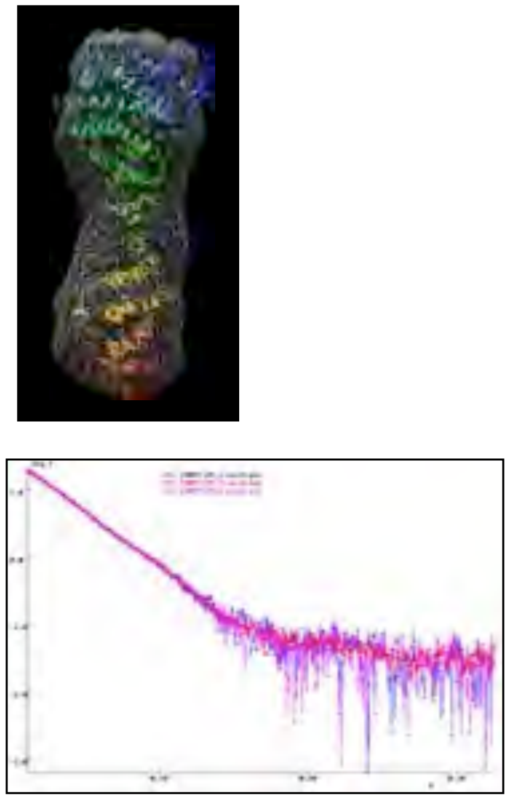
R. Diaz, D. Stokes, J. Velazquez



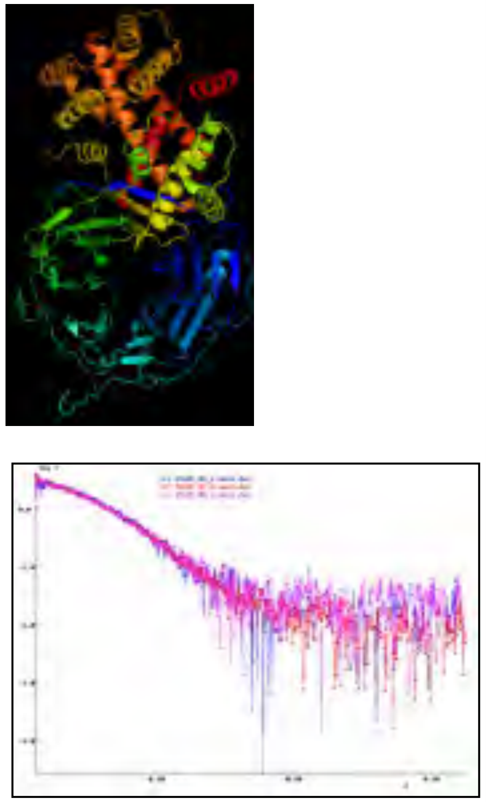
Subunit conformations

Small angle X-ray scattering

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



Nup133



Nup120

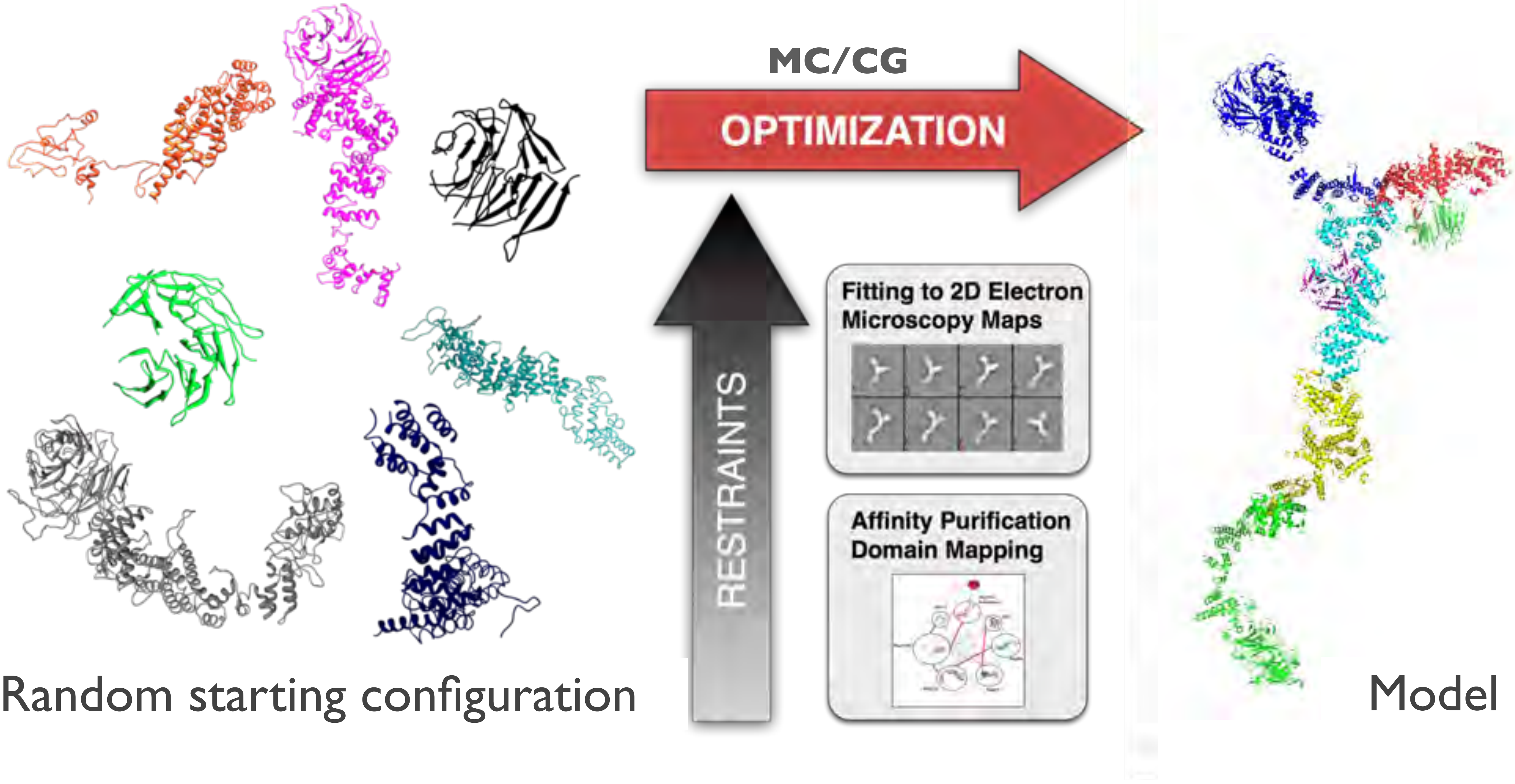
High-throughput crystallography

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

Yeast Nup133

Yeast Nup145

Nup84 complex: Optimization



MC/CG

OPTIMIZATION

RESTRAINTS

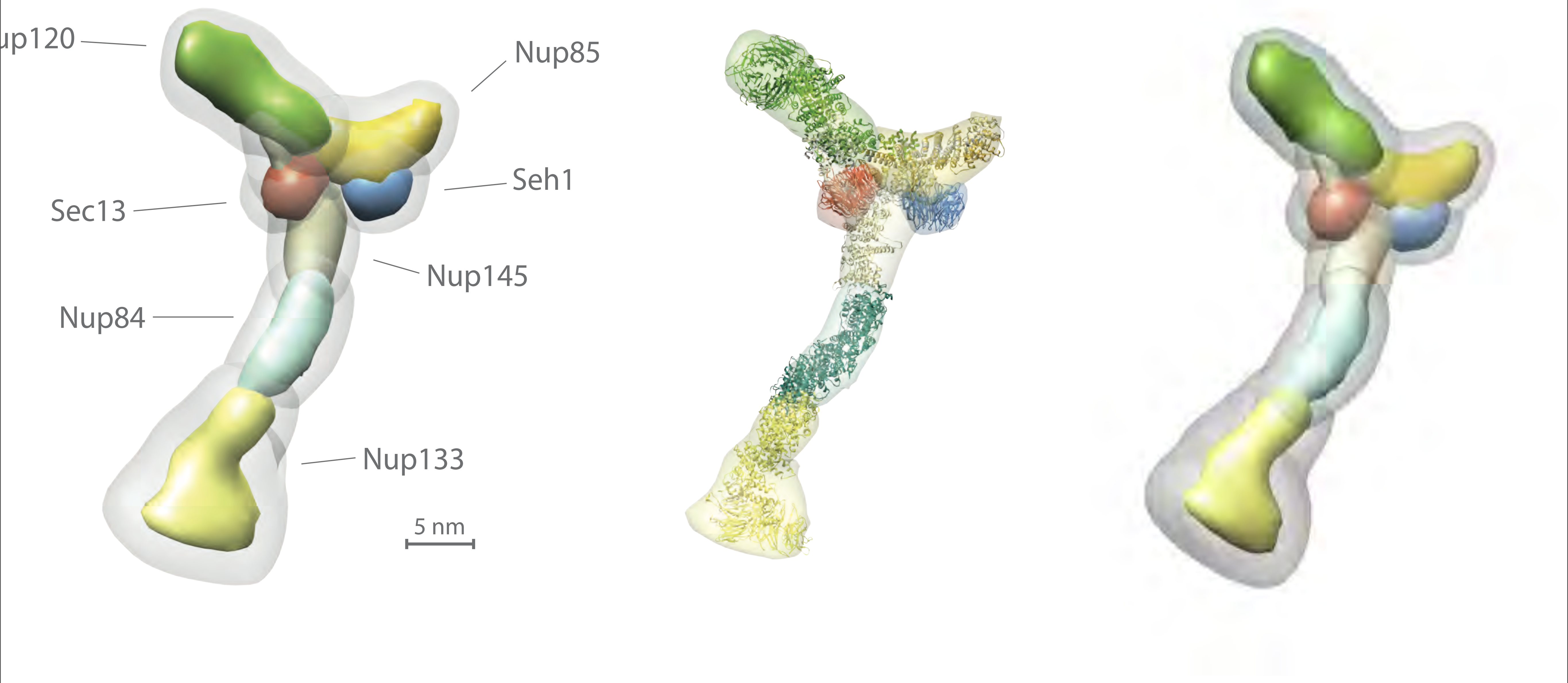
Fitting to 2D Electron Microscopy Maps

Affinity Purification Domain Mapping

Random starting configuration

Model

Nup84 complex:
Ensemble of good scoring solutions



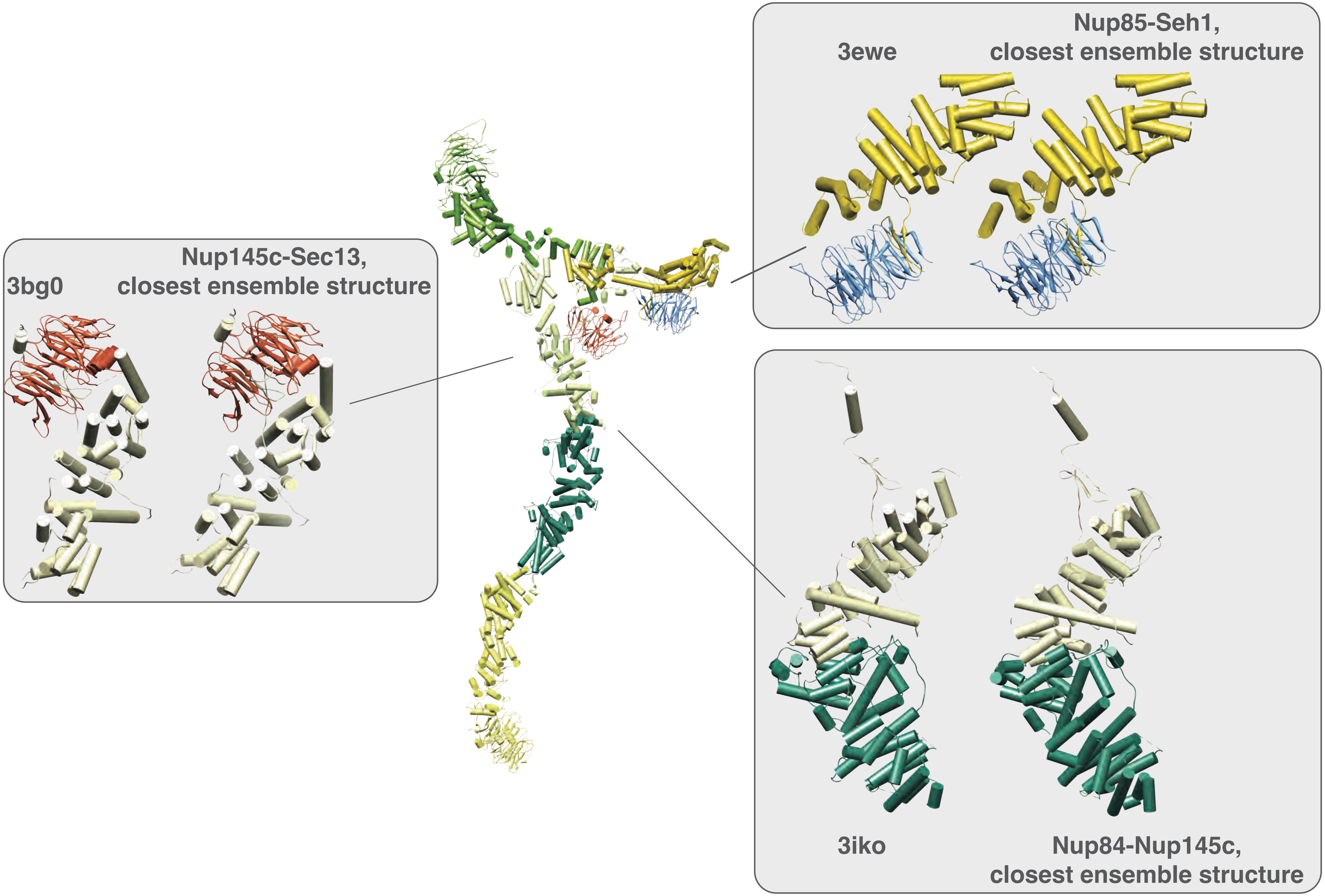
- 10,000 good scoring structures
- All restraints are satisfied (2D-EM, domain deletion, ...)
- Domain-domain orientations are resolved uniquely.
- Full ensemble precision is ~1 nm

Assessing the well-scoring models

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.

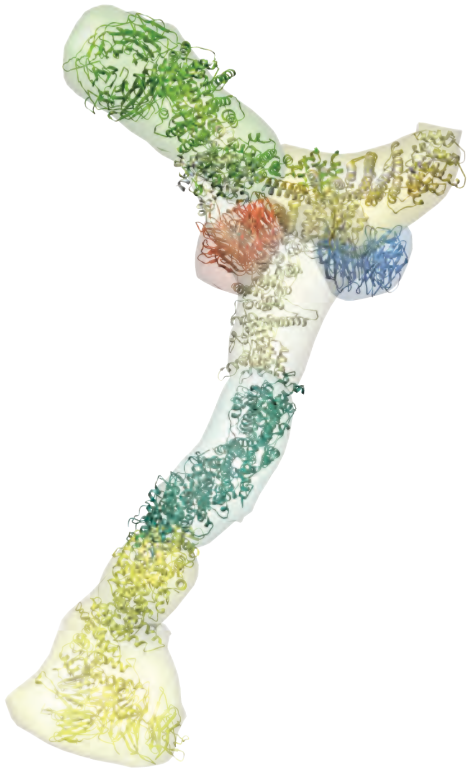
Modeling facilitates assessing the data as well as models in terms of precision and accuracy.

Assessment: Agreement with heterodimeric crystallographic structures


3ewe
Nup85-Seh1,
closest ensemble structure
3bg0
Nup145c-Sec13,
closest ensemble structure
3iko
Nup84-Nup145c,
closest ensemble structure

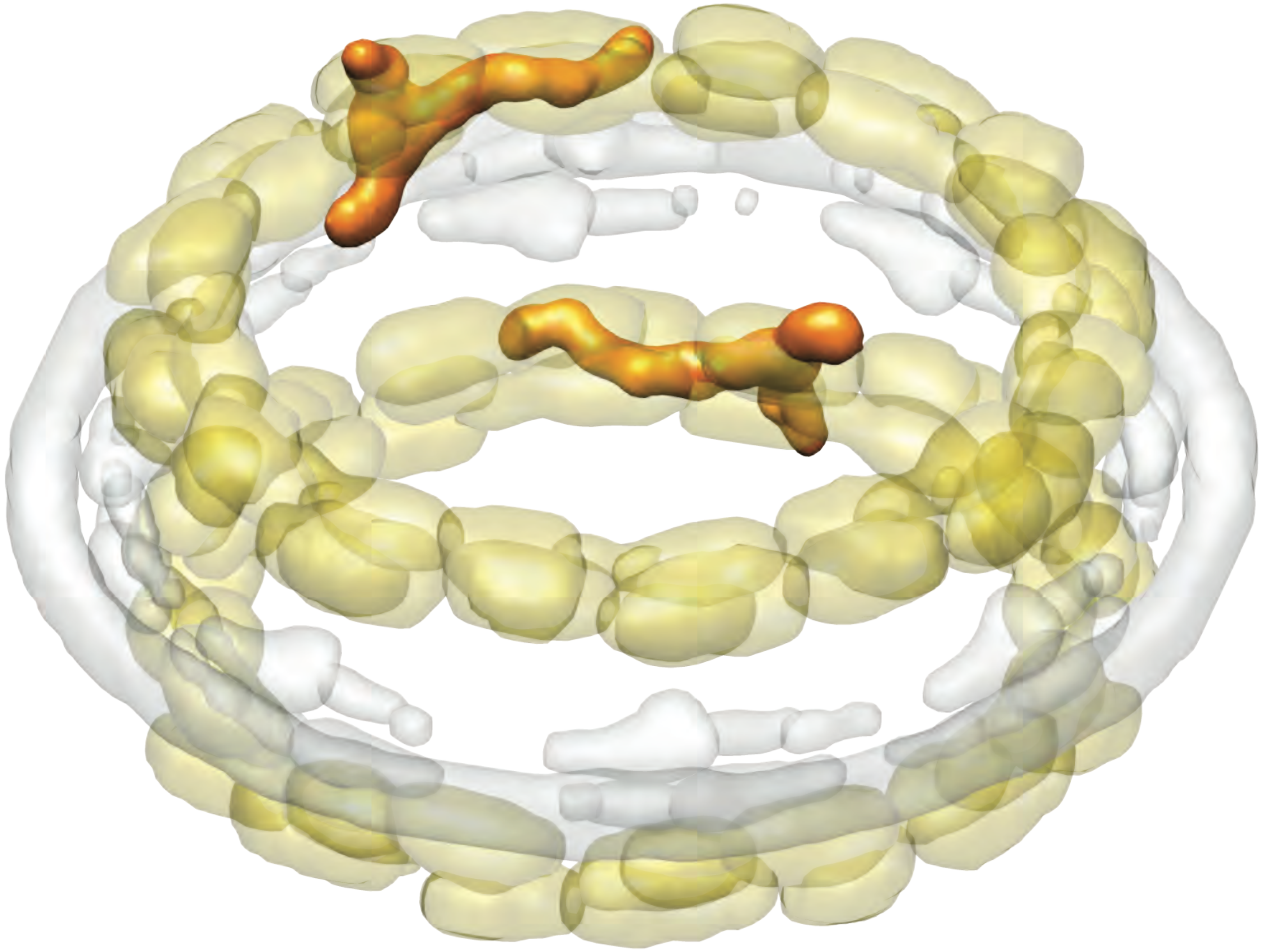
Towards a near-atomic structure of the NPC

Nup84 complex

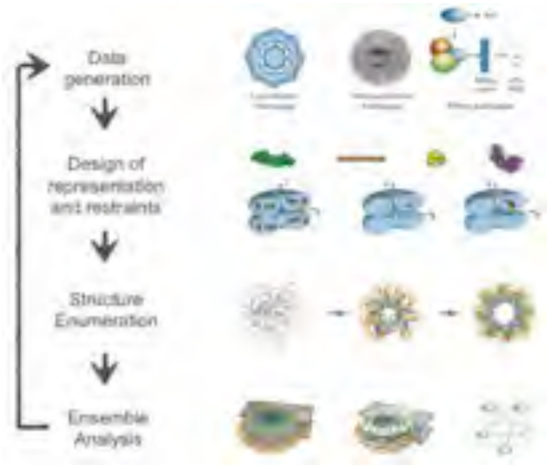


16 x

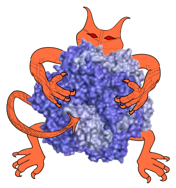
NPC



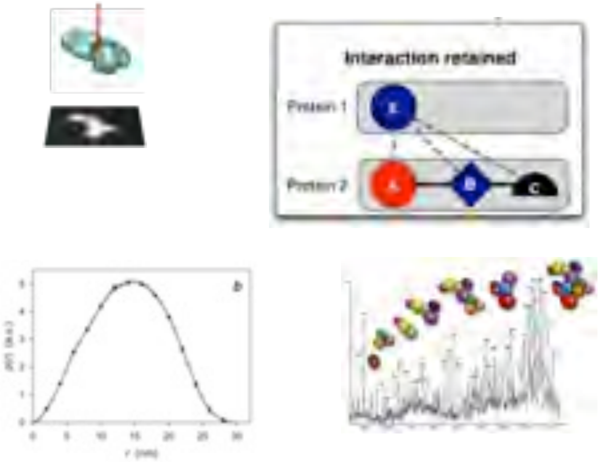
Conclusions



1. Assembly structure determination benefits greatly from the inclusion of all available information, including heterogeneous data sources.



2. Open source *Integrative Modeling Platform* (IMP). Developers and users of IMP are most welcome.



3. General and efficient assembly of subunit models based on domain deletion pullouts, 2D EM projections, 3D EM maps, SAXS profiles, and native MS.

4. Near atomic model of the Nup84 complex.

Acknowledgments

QB3 @ UCSF

Keren Lasker (DOMINO)
Jeremy Phillips (NPC)
Seung Joong Kim (NPC)
Daniel Russel (IMP)
Javier Velazquez (2D EM)
Ben Webb (IMP)
Massimiliano Bonomi (SPB)
Charles Greenberg (EM)
Riccardo Pellarin (proteomics)
Elina Tjioe (IMP)
Dina Schneidman (SAXS)
Peter Cimermancic
Natalia Khuri

Former members:

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

Rockefeller University

Mike Rout
Javier Fernandez-Martinez
Loren Hough
John LaCava
Jody Franke
Jaclyn Novatt

Brian Chait
Matthew Sekedat
Rosemary Williams

John Aitchison (ISB)
David Stokes (NYSBC)
Chris Akey (BU)
Robert Stroud (UCSF)
Stephen Burley (Lilly)
Steven Almo (AECOM)
Hiro Tsuruta (Stanford)
John Tainer (BNL)

Wolfgang Baumeister (MPI)
Trisha Davis (Univ of Wash)
Tom Ferrin (UCSF)
Haim Wolfson (TAU)
David Agard (UCSF)
Wah Chiu (Baylor)
Joachim Frank (Columbia)
Nevan Krogan (UCSF)
Al Burlingame (UCSF)
Carol Robinson (Cambridge)

NIH, NSF
The Sandler Family Foundation
Human Frontiers Science Program
IBM, Intel, Hewlett-Packard, NetApps, Pfizer, Structural Genomix Pharmaceuticals, Mike Homer, Ron Conway