BMI 206

Integrative Structure Determination of Macromolecular Assemblies

11/14/16

Andrej Sali
(sali@salilab.org)

Topics

1. Introduction to integrative (hybrid) structure determination
2. Integrative structure determination of the Nuclear Pore Complex

F. Alber et al. "Determining the architectures of macromolecular assemblies". *Nature* **450**, 683-694, 2007.

F. Alber et al. “Integrating Diverse Data for Structure Determination of Macromolecular Assemblies”
Annual Review of Biochemistry **77**, 11.1-11.35, 2008.

D. Russel et al. “Putting the pieces together: integrative structure determination of macromolecular assemblies.”
PLoS Biol. **10**, e1001244, 2012.

A. Ward, A. Sali, I. Wilson. Integrative structural biology. *Science* **339**, 913-915, 2013

From: http://salilab.org/publications/

Topics

1. Introduction to integrative (hybrid) structure determination
2. Integrative structure determination of the Nuclear Pore Complex

F. Alber et al. "Determining the architectures of macromolecular assemblies". *Nature* **450**, 683-694, 2007.

F. Alber et al. “Integrating Diverse Data for Structure Determination of Macromolecular Assemblies”
Annual Review of Biochemistry **77**, 11.1-11.35, 2008.

D. Russel et al. “Putting the pieces together: integrative structure determination of macromolecular assemblies.”
PLoS Biol. **10**, e1001244, 2012.

A. Ward, A. Sali, I. Wilson. Integrative structural biology. *Science* **339**, 913-915, 2013

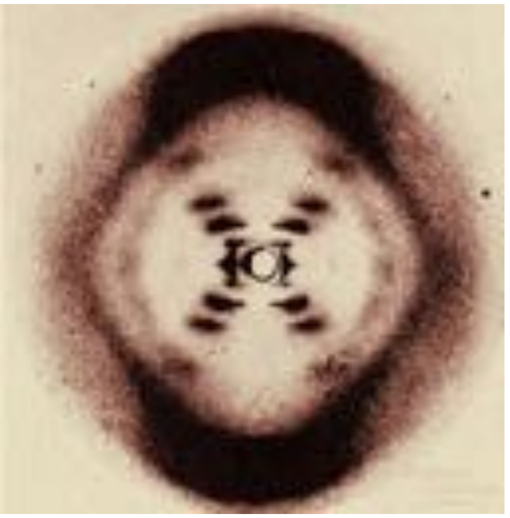
From: http://salilab.org/publications/

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

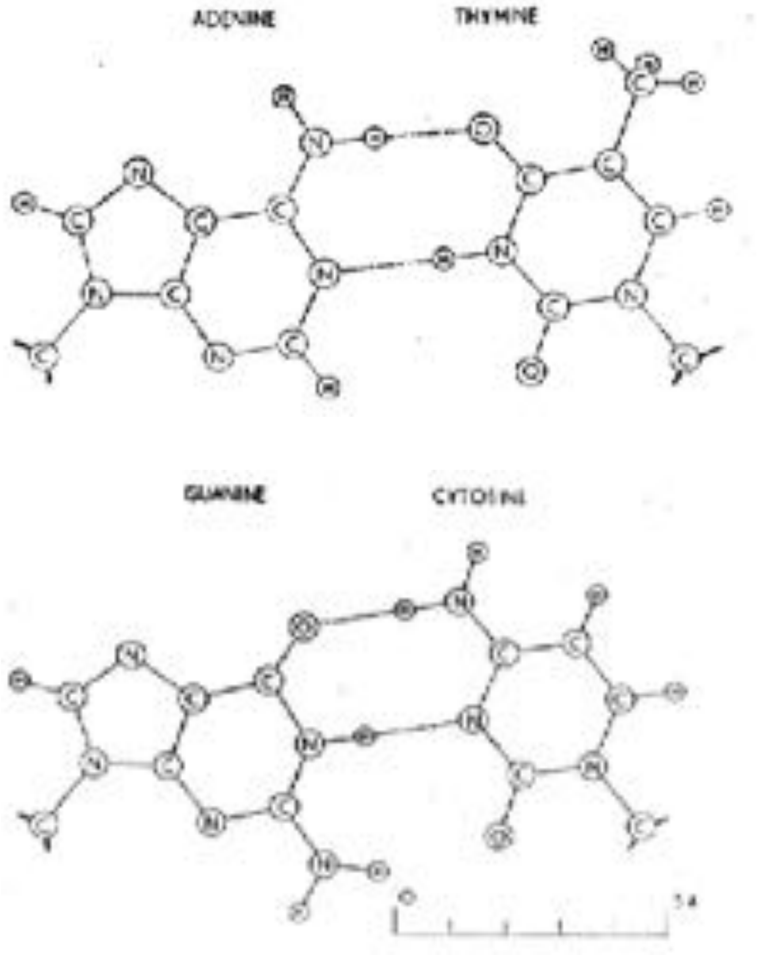
A Structure for Deoxyribose Nucleic Acid

J. D. WATSON
F. H. C. CRICK

No. 4356 April 25, 1953
NATURE



X-ray diffraction

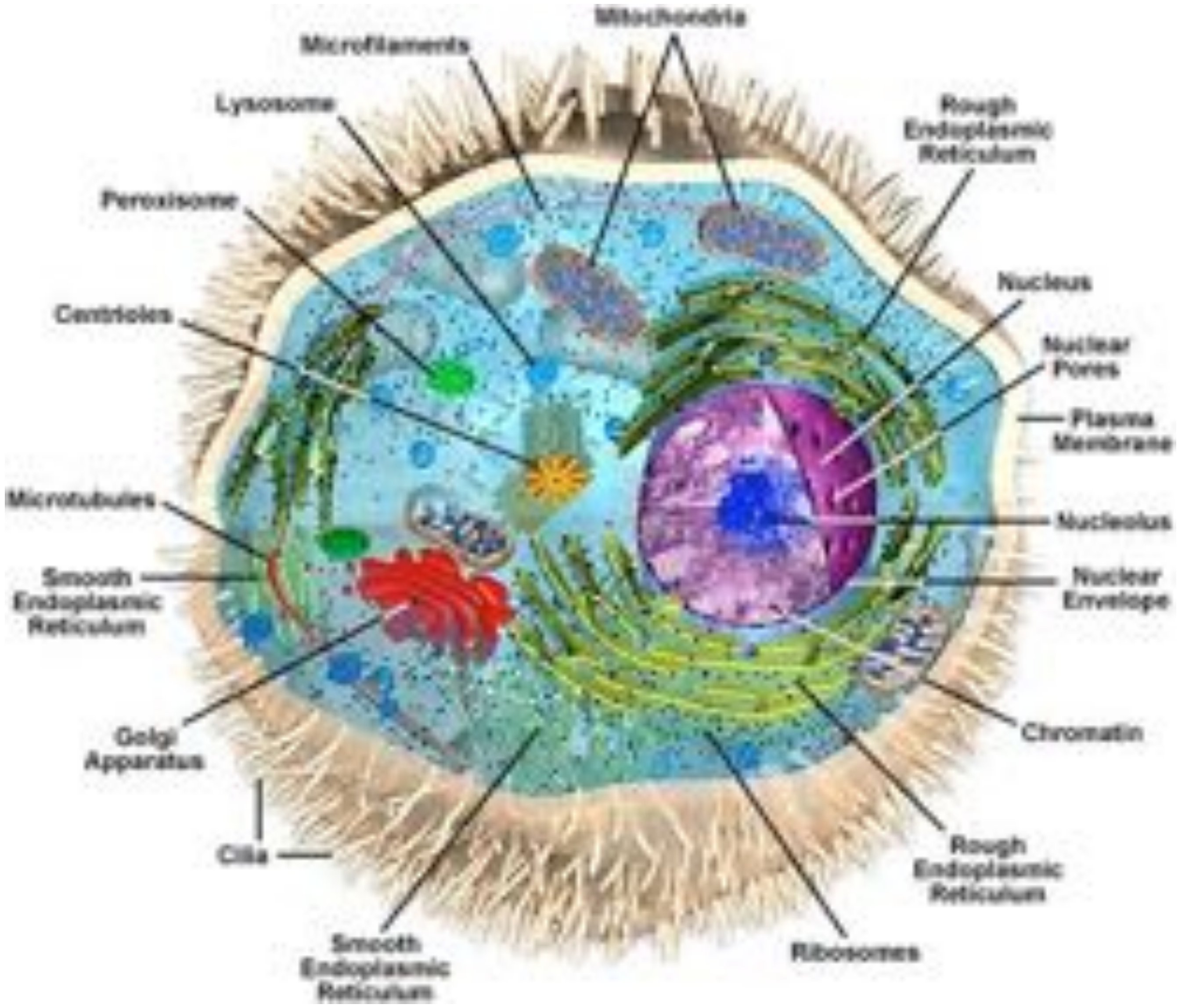


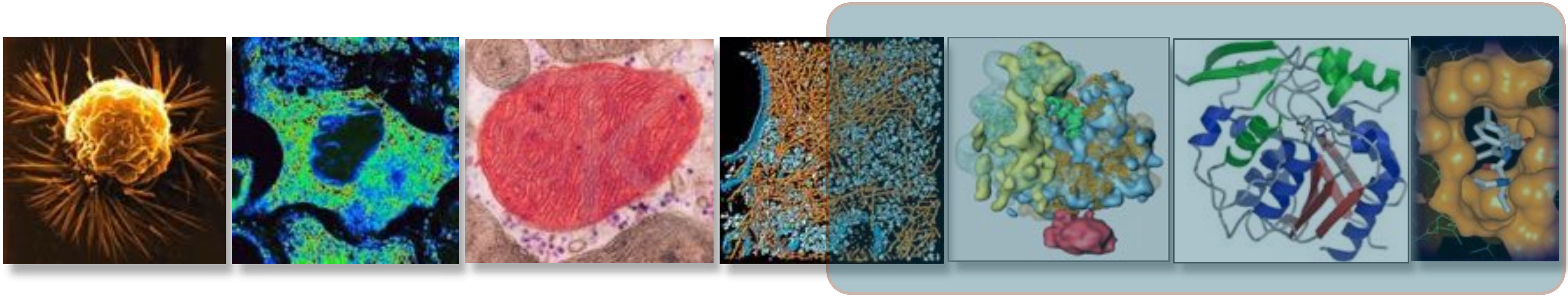
Composition
Stoichiometry
Chemical complementarity

To understand and modulate cellular processes, we need their models.

These models are best generated by considering all available information.

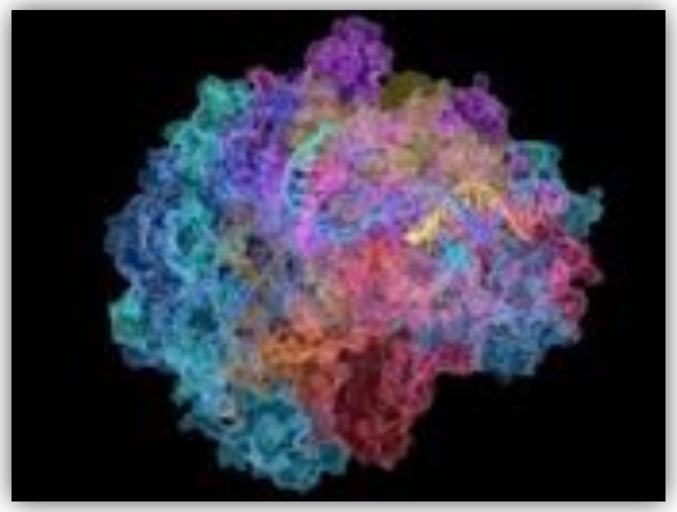
Towards a spatial, temporal, and logical model of the cell?

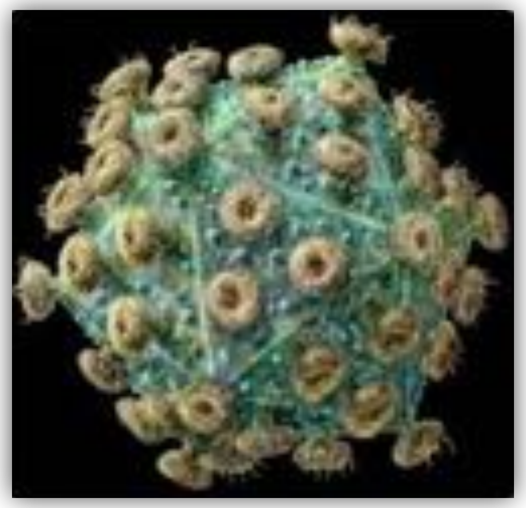




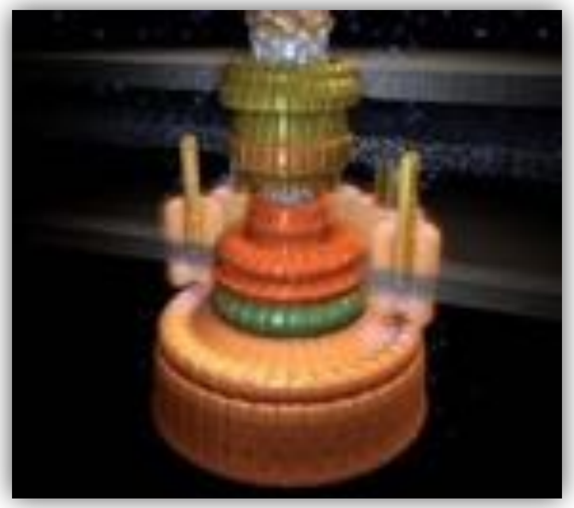
Structural biology:
Maximize accuracy, resolution, completeness, and efficiency of the structural coverage of macromolecular assemblies

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

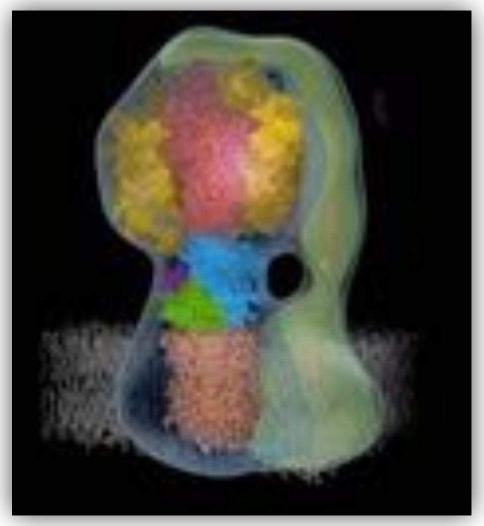

RNA polymerase II

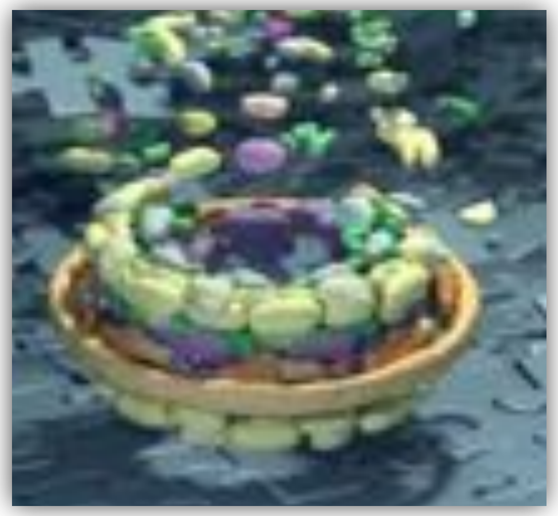

HIV virus

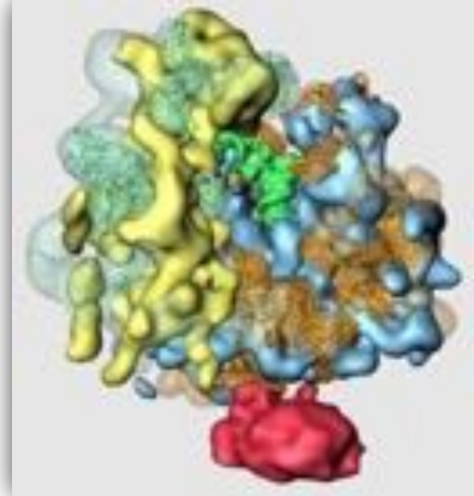

tRNA synthetase


flagellar motor


GroEL chaperonin


ATP synthase


nuclear pore complex

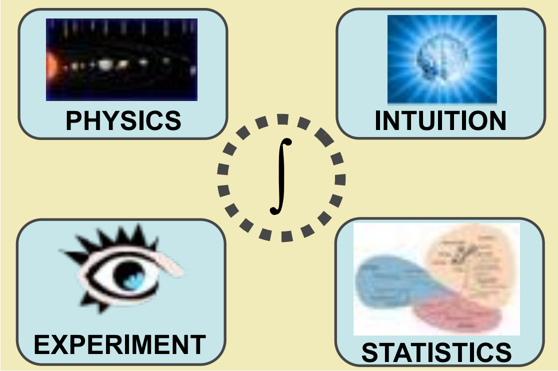

ribosome

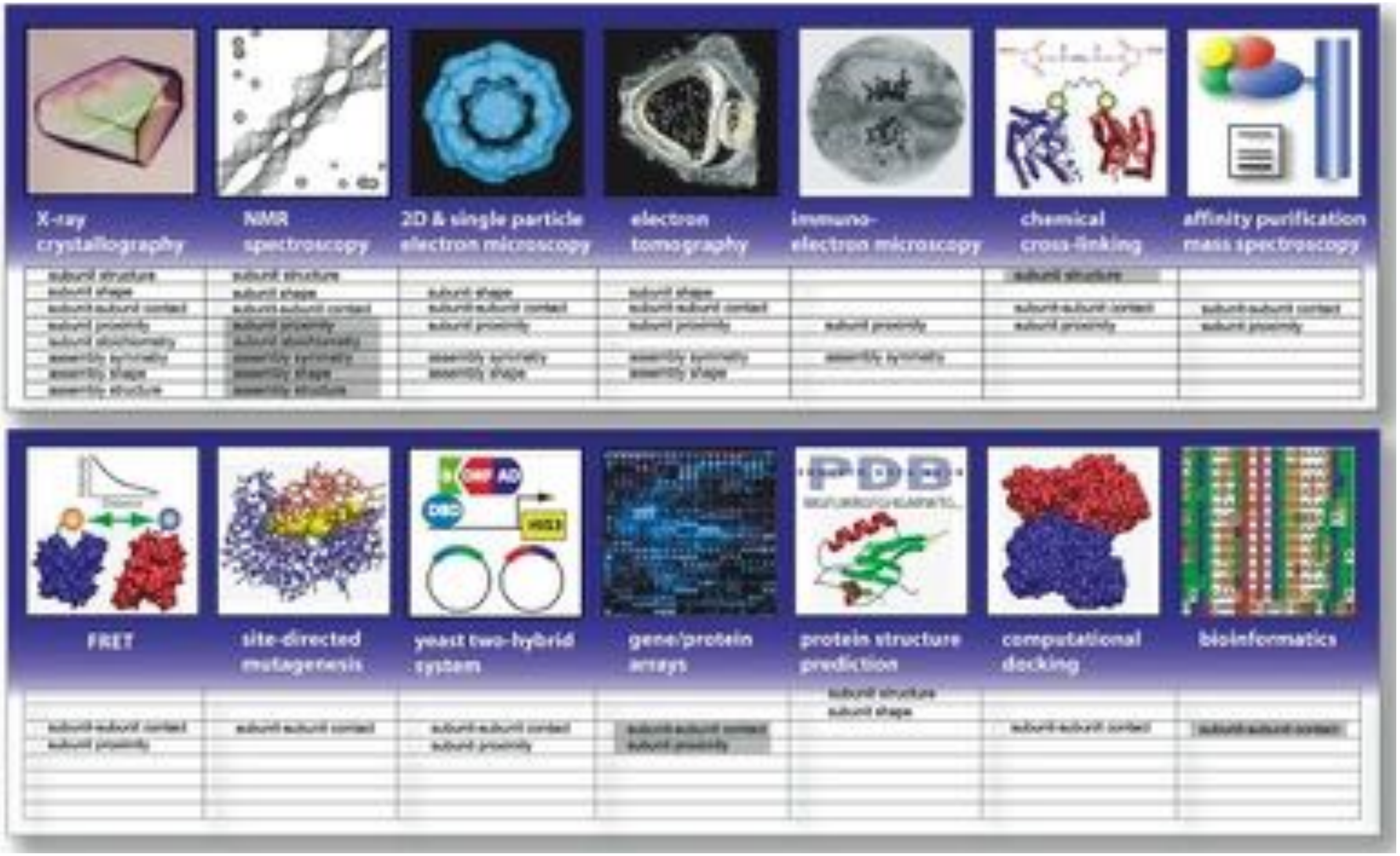
There may be thousands of biologically relevant macromolecular complexes whose structures are yet to be characterized, involved in a few hundred core biological processes.

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.

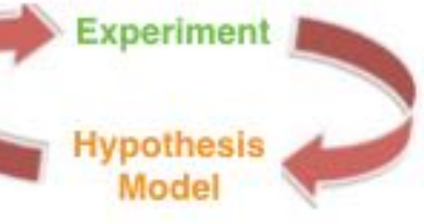


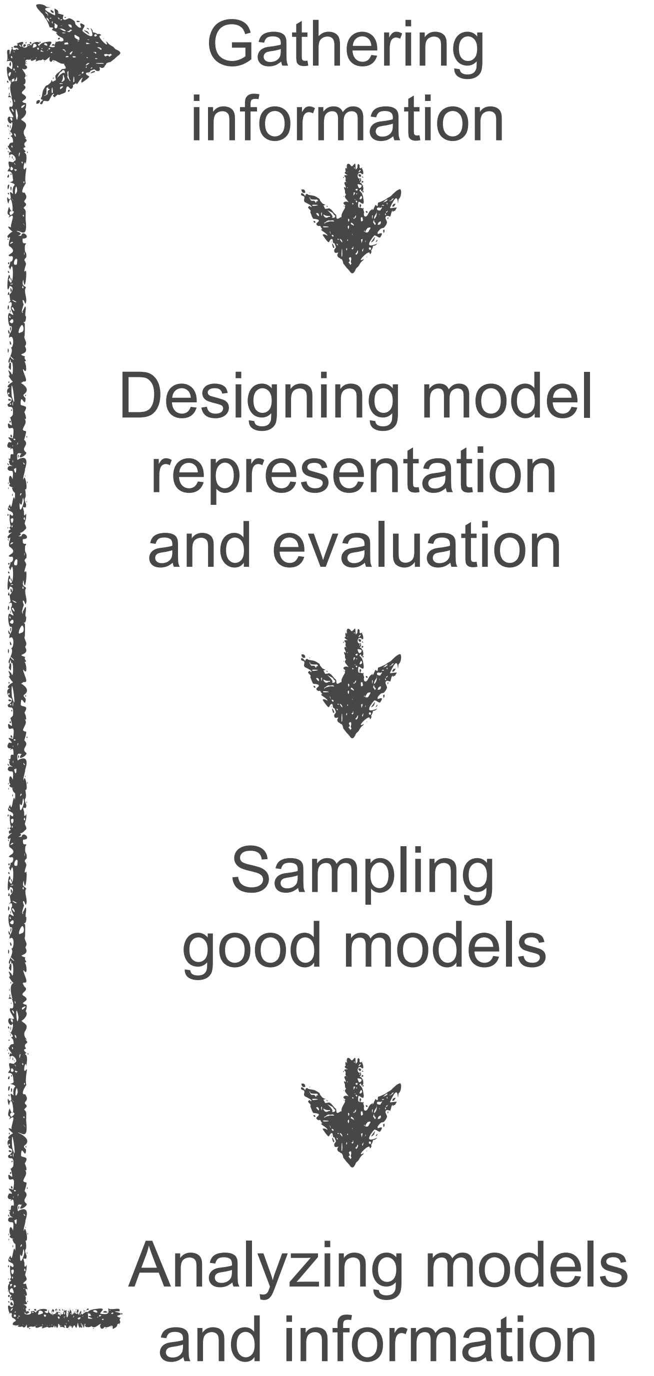


Sali A, Earnest T, Glaeser R, Baumeister W. From words to literature in structural proteomics. *Nature* 422, 216-225, 2003.
Ward A, Sali A, Wilson I. Integrative structural biology. *Science* 339, 913-915, 2013.

A description of integrative structure determination

Sali *et al*. *Nature* **422**, 216-225, 2003.
Alber *et al*. *Nature* **450**, 683-694, 2007
Robinson *et al*. *Nature* **450**, 974-982, 2007
Alber *et al*. *Ann.Rev.Biochem.* **77**, 11.1–11.35, 2008
Russel *et al*. *PLoS Biology* **10**, 2012
Ward *et al*. *Science* **339**, 913-915, 2013
Schneidman *et al*. *Curr.Opin.Str.Biol*., 2014.



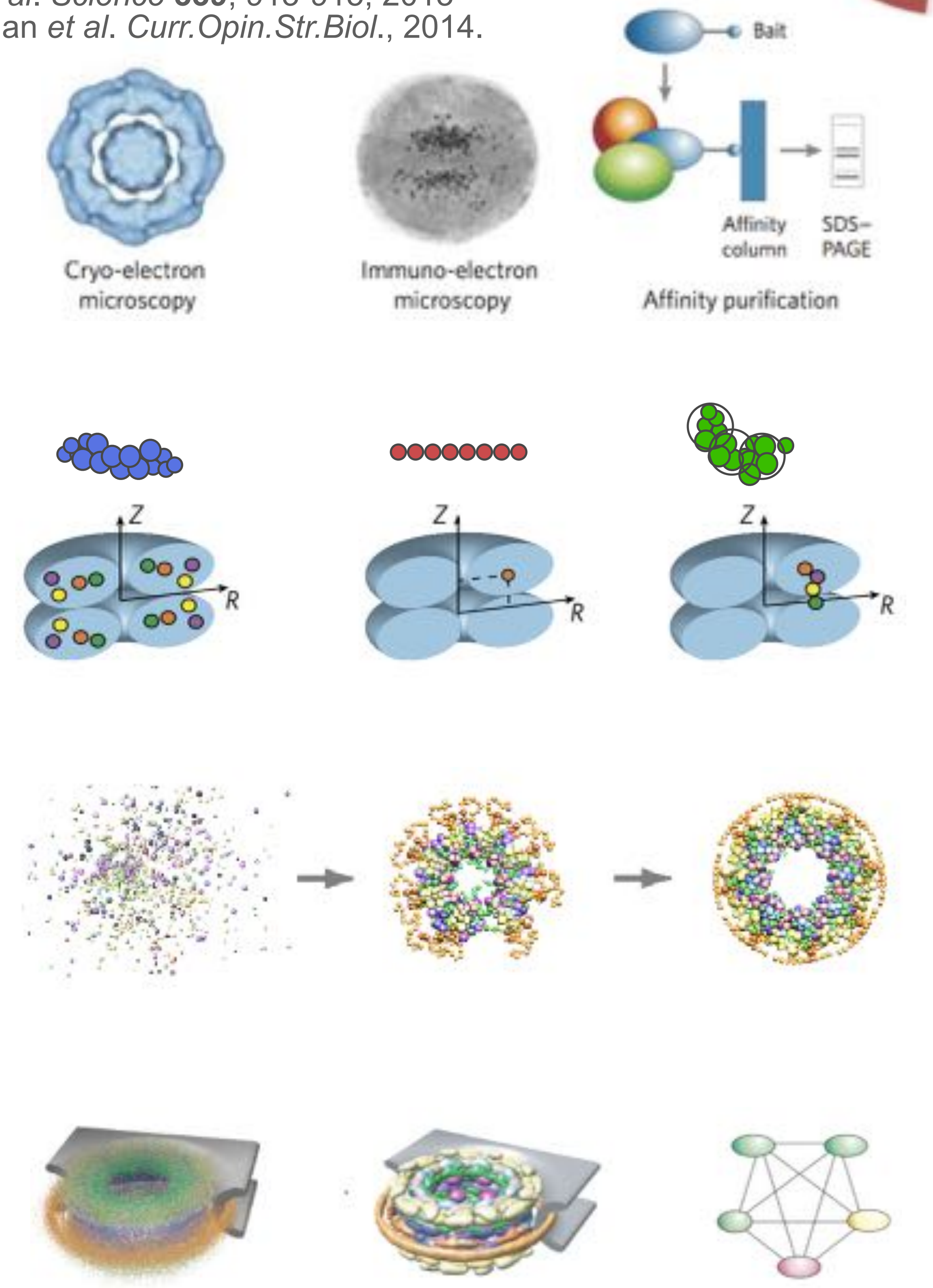


Gathering information

Designing model representation and evaluation

Sampling good models

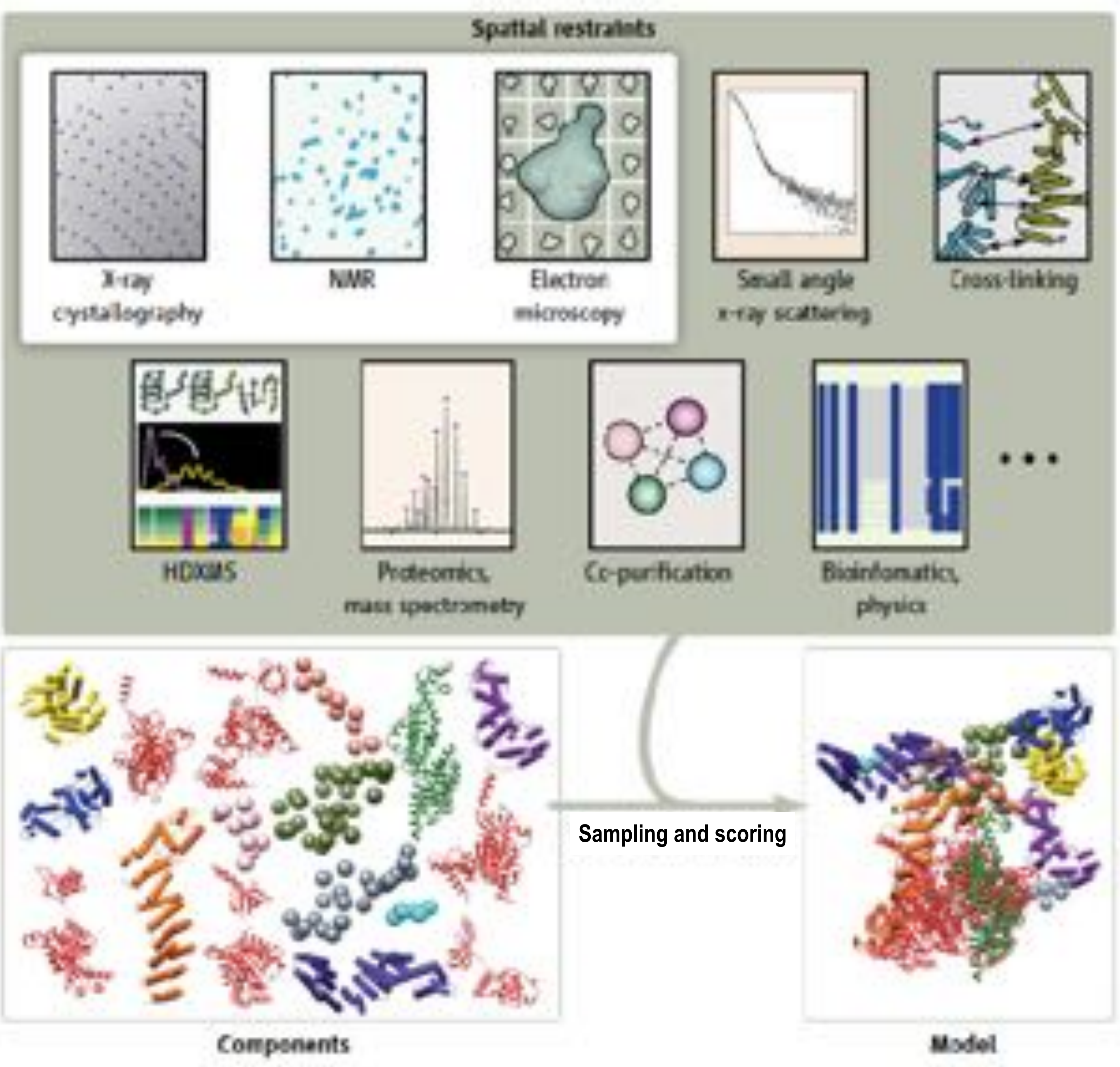
Analyzing models and information



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

Integrative structure determination

- Uses multiple types of information (experiments, physical theory, statistical inference).
- Maximizes accuracy, resolution, completeness, and efficiency of the structure determination.
- Finds all models whose computed data match the experimental data within an acceptable threshold.



Sali *et al*. *Nature* **422**, 216-225, 2003.
Alber *et al*. *Nature* **450**, 683-694, 2007
Robinson *et al*. *Nature* **450**, 974-982, 2007
Alber *et al*. *Ann.Rev.Biochem.* **77**, 11.1–11.35, 2008
Russel *et al*. *PLoS Biology* **10**, 2012
Ward *et al*. *Science* **339**, 913-915, 2013
Schneidman *et al*. *Curr.Opin.Str.Biol*., 96-104, 2014.
Sali *et al*. *Structure* **23**, 1156-1167, 2015.

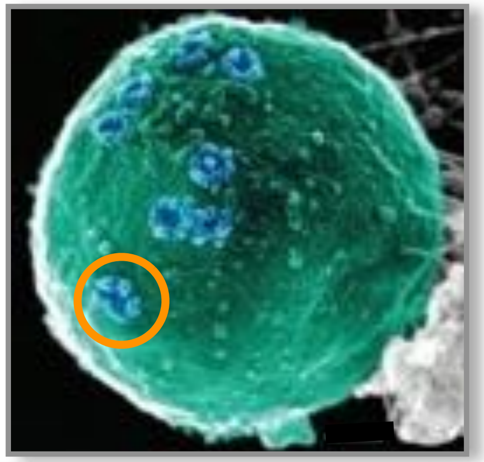
A model is built iteratively, contributes continuously.

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

Configuration of 456 proteins in the Nuclear Pore Complex

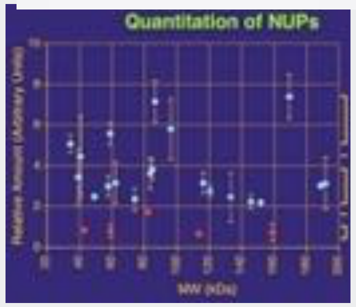
with M. Rout & B. Chait

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



Quantitative Immunoblotting

30 relative abundances



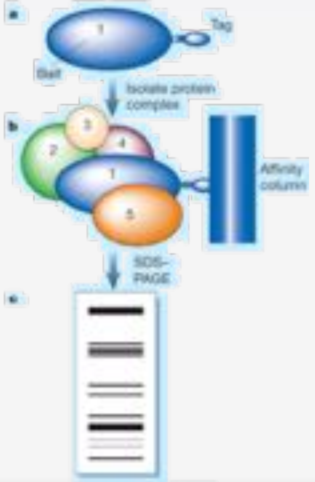
Protein Stoichiometry →

Affinity Purification

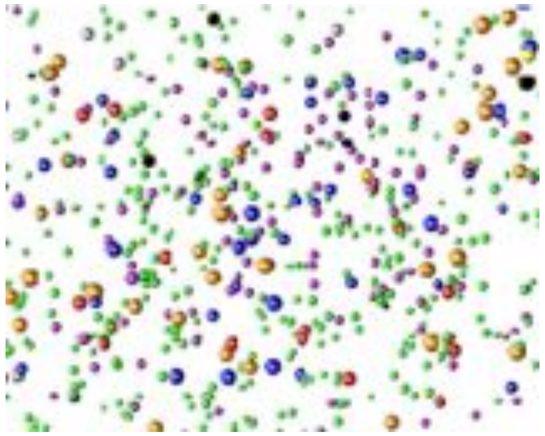
75 composites

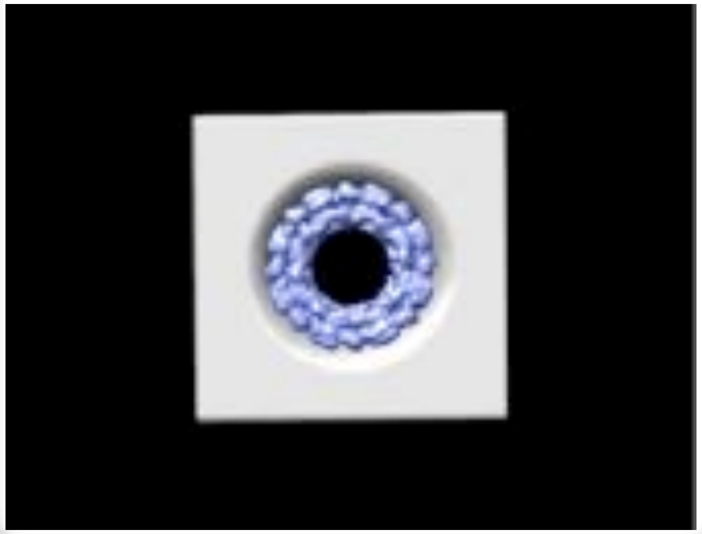
Overlay Assay

7 contacts



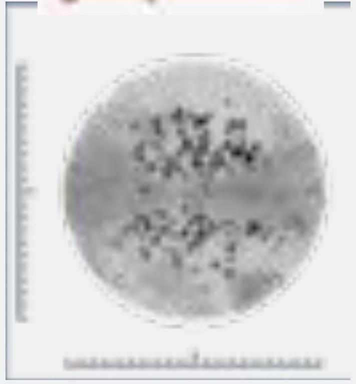
Protein-protein Proximities →





Immuno-Electron Microscopy

10,615 gold particles



← **Protein Localization**

Electron Microscopy

electron microscopy map

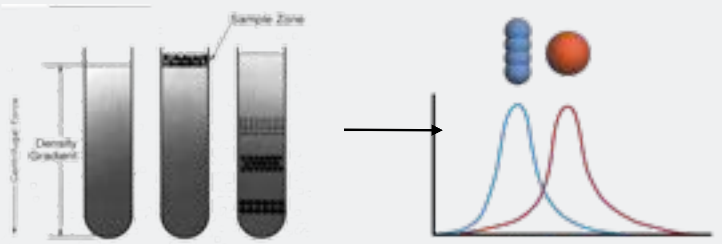
← **Symmetry**

Protein Shape ↑

Ultracentrifugation

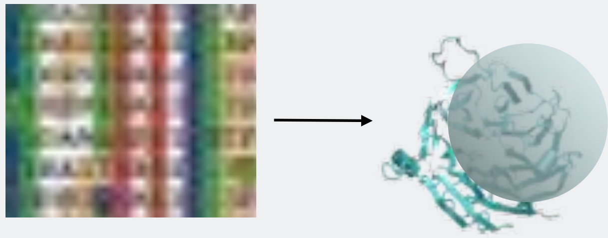
30 S-values

1 S-value

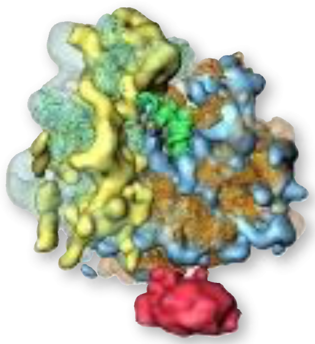


Bioinformatics and Membrane Fractionation

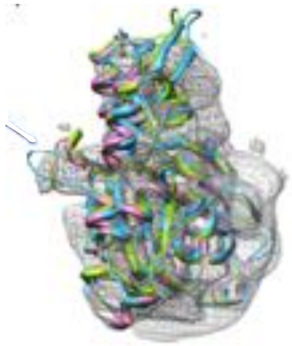
30 protein sequences



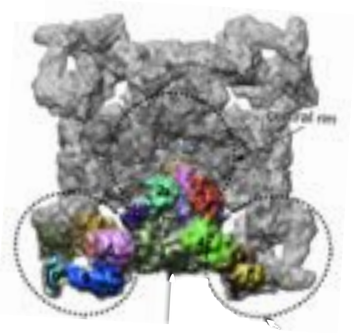
Integrative structure models from our lab

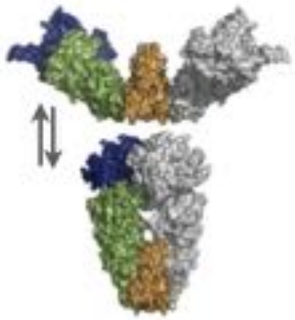

Ribosomes,
Frank, Akey

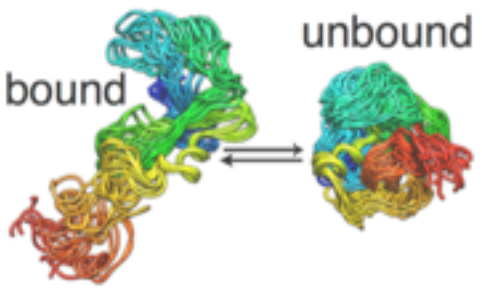

PCSK9-Fab,
Cheng, Agard,
Pons


Actin
Chiu

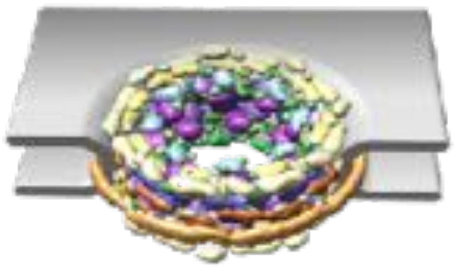

TRiC/CCC
Frydman, Chiu

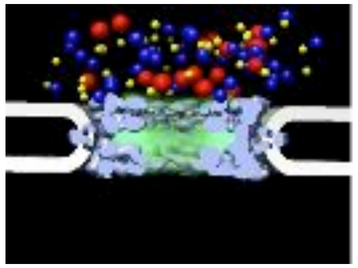

RyR channel
Serysheva, Chiu

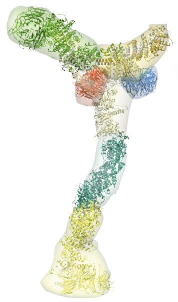

Hsp90 landscape
Agard



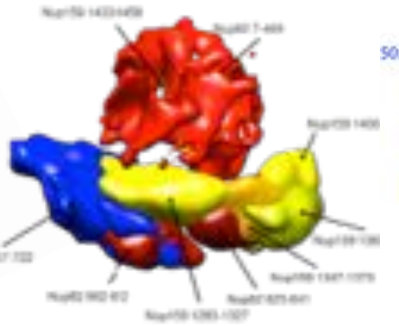
Substrate folding
by Hsp90
Agard

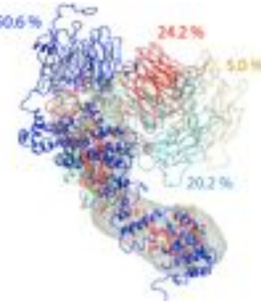

Nuclear Pore Complex,
Rout, Chait


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

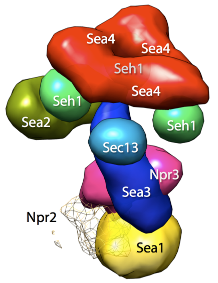

Nup84 complex,
Rout, Chait


Nup84 hub
Rout, Chait


Nup82 complex,
Rout, Chait

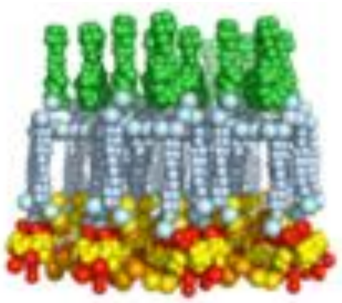


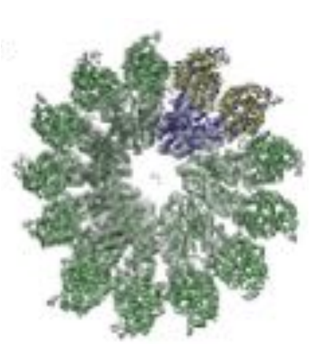
Nup133,
Rout, Chait

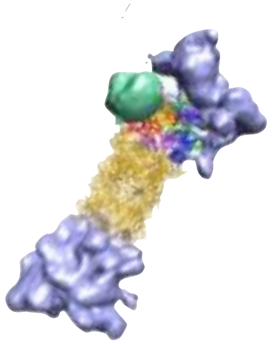


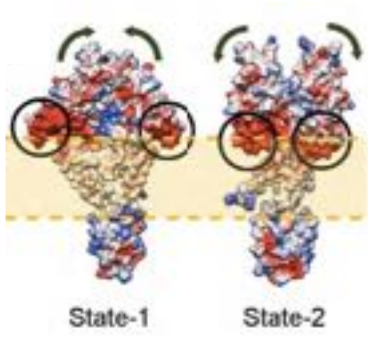
SEA complex
Rout, Chait,
Dokudovskaya


PDE6
Chu

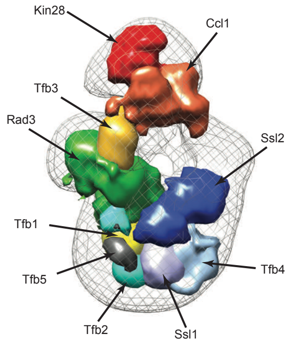

Spindle PoleBody
Davis, Muller


Microtubule
nucleation
Agard

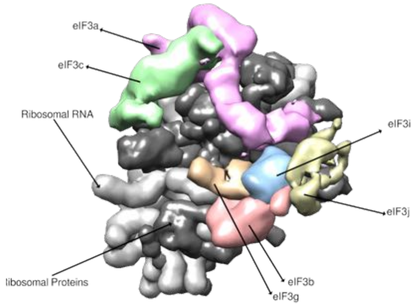

26 Proteasome
Baumeister



PhoQ His kinase
DeGrado



TFIIH
Ranish



40S-eIF1-eIF3
Aebersold,Ban

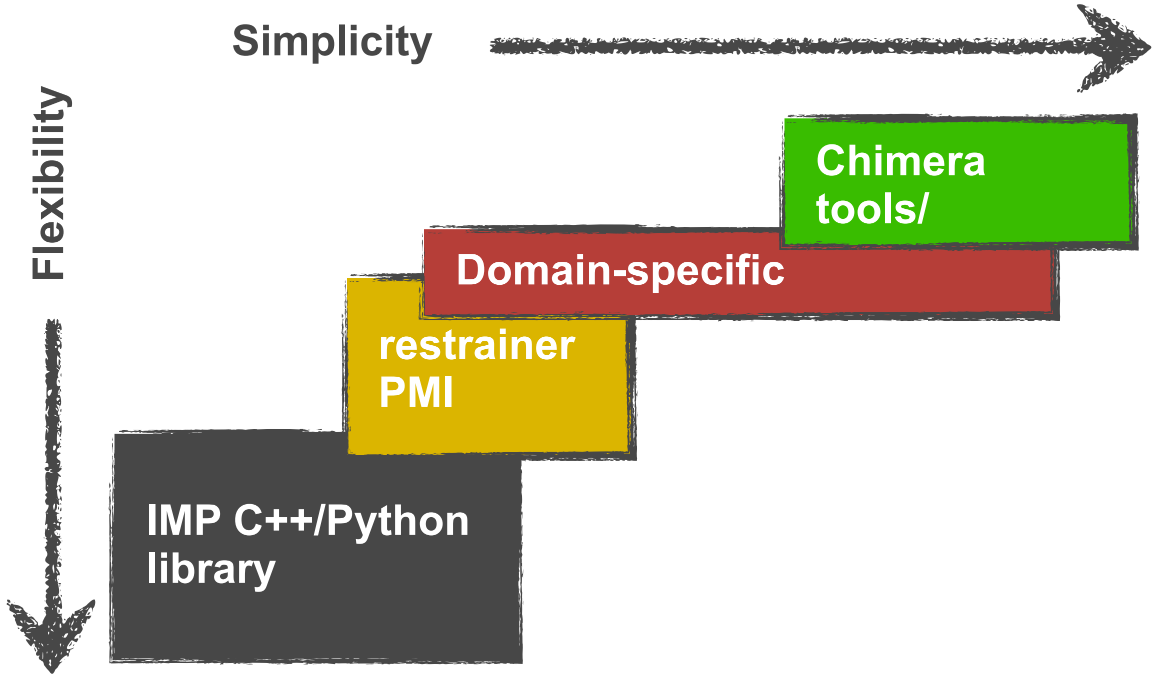

Prion aggregation
Prusiner

Integrative Modeling Platform (IMP)

http://integrativemodeling.org

D. Russel, K. Lasker, B. Webb, J. Velazquez-Muriel, E. Tjioe, D. Schneidman, F. Alber, B. Peterson, A. Sali, PLoS Biol, 2012.
R. Pellarin, M. Bonomi, B. Raveh, S. Calhoun, C. Greenberg, G.Dong, S.J. Kim, I. Chemmama, D. Saltzberg, S. Viswanath

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



Simplicity

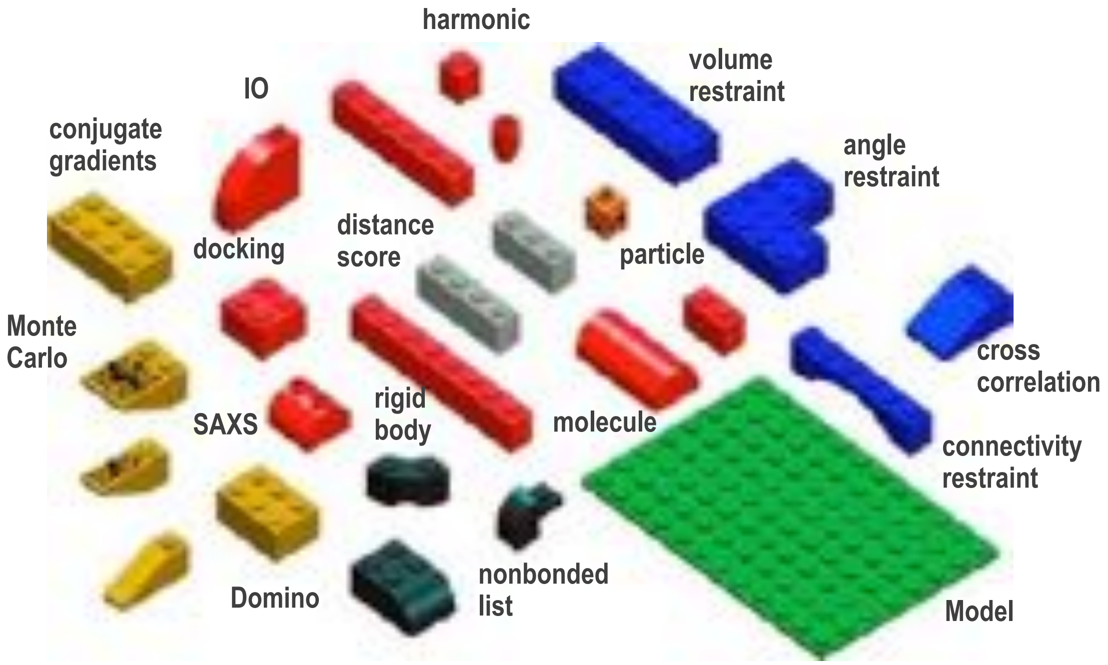
Flexibility

Chimera tools/

Domain-specific

restrainer PMI

IMP C++/Python library



Representation:

- Atomic
- Rigid bodies
- Coarse-grained
- Multi-scale
- Symmetry / periodicity
- Multi-state systems

Scoring:

- Density maps
- EM images
- Proteomics
- FRET
- Chemical and Cys cross-linking
- Homology-derived restraints
- SAXS
- H/D exchange
- Second harmonic generation
- Native mass spectrometry
- Genetic interactions
- Statistical potentials
- Molecular mechanics forcefields
- Bayesian scoring
- Library of functional forms (ambiguity, ...)

Sampling:

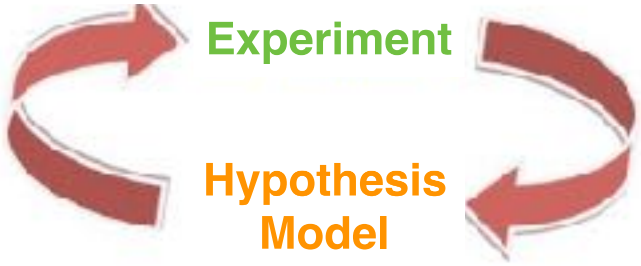
- Simplex
- Conjugate Gradients
- Monte Carlo
- Brownian Dynamics
- Molecular Dynamics
- Replica Exchange
- Divide-and-conquer enumeration

Analysis:

- Clustering
- Chimera
- Pymol
- PDB files
- Density maps

Integration across computational resources

Goal: Maximize accuracy, resolution, completeness, and efficiency of the structural coverage of macromolecules



RCSB PDB
PROTEIN DATA BANK

CCP4

autoPACK

VMD
Visual Molecular Dynamics

RosettaCommons

Phenix

DOCK

OpenMM

CHIMERA

PSI | The Protein Model Portal

CHARMM (Chemistry at HARvard Macromolecular Mechanics)

EMAN

Integrative Methods Task Force Workshop

Andrej Sali, Helen M. Berman, Torsten Schwede, Jill Trewhella, Gerard Kleywegt, Stephen K. Burley, John Markley, Haruki Nakamura, Paul Adams, Alexandre Bonvin, Wah Chiu, Tom Ferrin, Kay Grünewald, Aleksandras Gutmanas, Richard Henderson, Gerhard Hummer, Kenji Iwasaki, Graham Johnson, Cathy Lawson, Frank di Maio, Jens Meiler, Marc Marti-Renom, Guy Montelione, Michael Nilges, Ruth Nussinov, Ardan Patwardhan, Matteo dal Peraro, Juri Rappsilber, Randy Read, Helen Saibil, Gunnar Schröder, Charles Schwieters, Claus Seidel, Dmitri Svergun, Maya Topf, Eldon Ulrich, Sameer Velankar, and John D. Westbrook. *Structure* **23**, 1156-1167, 2015.

First Integrative Methods Task Force Workshop was held at the European Bioinformatics Institute in Hinxton, UK, on October 6 and 7, 2014:

What should be archived?

How should integrative models be represented?

How should the data and integrative models be validated?

How should the data and models be archived?

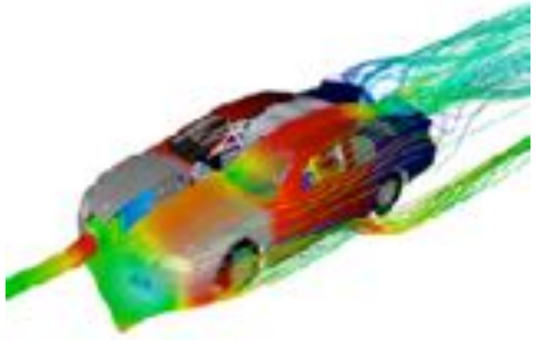
What information should accompany the publication of integrative models?

Pushing the envelope of structural biology by integration of all available information

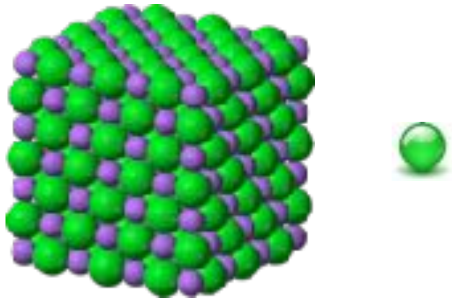
- Size

- Static systems in single and multiple states

- Dynamic systems

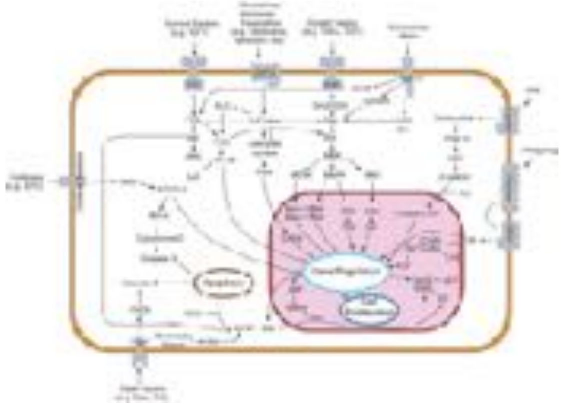


- Bulk and single molecule views



- Impure samples

- Overlapping with other domains such as systems biology

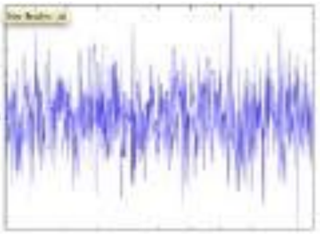


Challenges in interpreting the data in terms of a structural model

1. **Model representation**
2. **Sampling**
3. **Scoring function:**

- **Sparseness**, due to incompleteness of measurements

- **Error**, due to measurement and other imperfections



- **Ambiguity**, due to, *eg*, multiple copies of a protein in a system

- **Incoherence (mixture)**, due to multiple states of a system in a heterogenous sample

Scoring function

Rank models based on all available information:

1. Least-squares scoring function:

$$S(M) = \sum_i w_i [D_i - f_i(M)]^2$$

M	model
D	measured data point
f	computed data point (forward model)
w	weight of data point

2. Bayesian scoring function:

$$p(M|D,I) \propto p(D|M,I) \cdot p(M|I)$$

posterior **likelihood** **prior**

↑

$$D - f(M)$$

I prior information

Posterior is the probability density of model M, given data D and information I.

Model M can include coordinates of one or more structures as well as additional parameters (noise levels, weights, calibration parameters, ...).

Likelihood is the probability density of observing data D, given model M and prior information I (by relying on a model of noise and a forward model, which computes data D given model M).

Prior is the probability density of model M, given prior information I.

$$p(\mathrm{AB}) = p(\mathrm{BA}) = p(\mathrm{A}) \cdot p(\mathrm{B/A}) = p(\mathrm{B}) \cdot p(\mathrm{A/B})$$

Rieping, Habeck, Nilges. *Science*, 2005

Topics

1. Introduction to integrative (hybrid) structure determination
2. Integrative structure determination of the Nuclear Pore Complex

F. Alber et al. "Determining the architectures of macromolecular assemblies". *Nature* **450**, 683-694, 2007.

F. Alber et al. “Integrating Diverse Data for Structure Determination of Macromolecular Assemblies”
Annual Review of Biochemistry **77**, 11.1-11.35, 2008.

D. Russel et al. “Putting the pieces together: integrative structure determination of macromolecular assemblies.”
PLoS Biol. **10**, e1001244, 2012.

A. Ward, A. Sali, I. Wilson. Integrative structural biology. *Science* **339**, 913-915, 2013

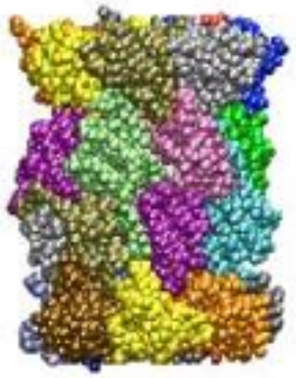
From: http://salilab.org/publications/

Very low-resolution modeling of large assemblies

Many times the structures of some subunits are not available.

In such cases, we can only model the **configuration** of the subunits in the complex.

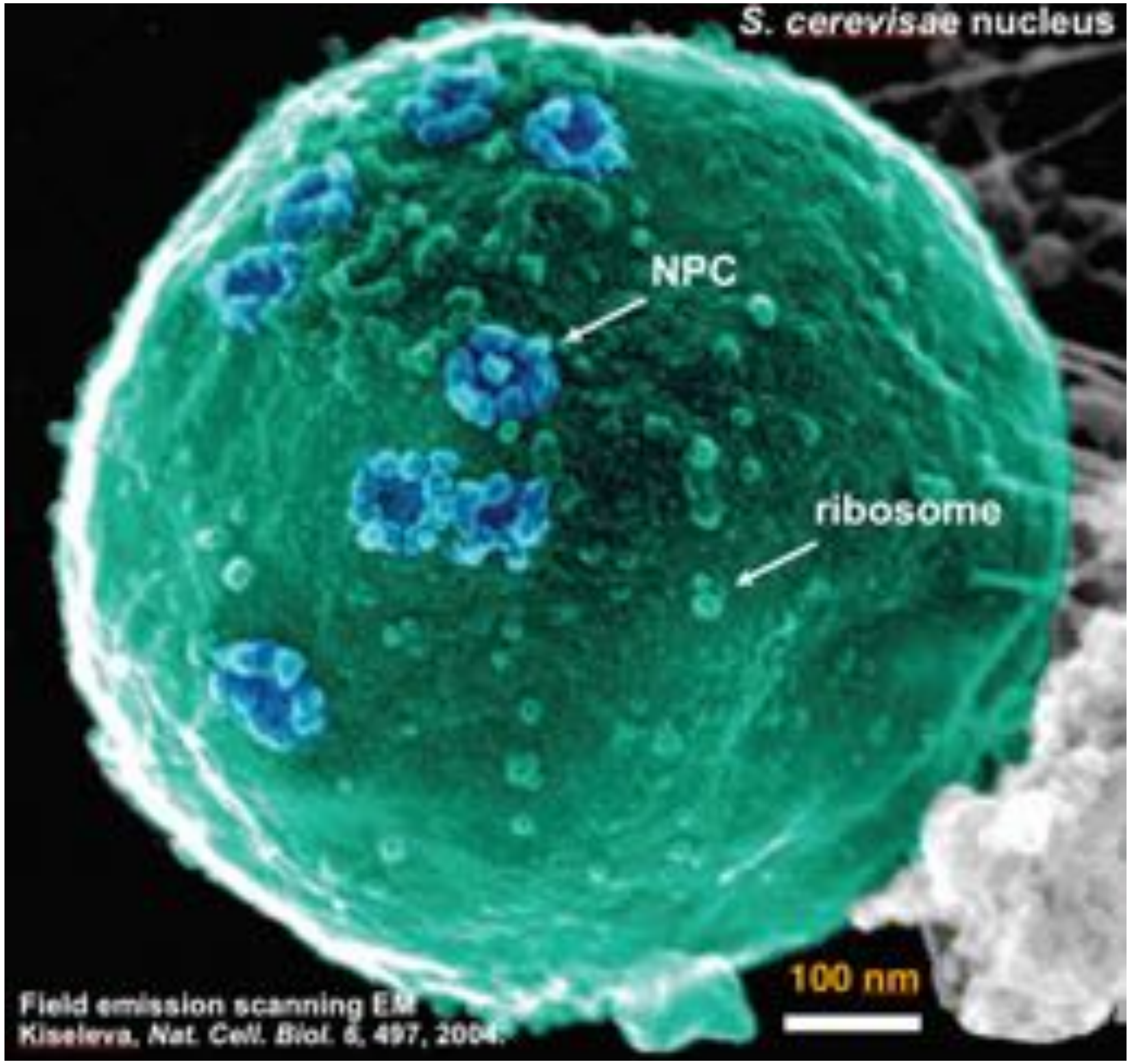
atoms



residues

proteins

Nuclear Pore Complex (NPC)



Field emission scanning EM
Kiseleva, *Nat. Cell. Biol.* 6, 497, 2004.

Consists of broadly conserved nucleoporins (nups).

50 MDa complex: ~480 proteins of 30 different types.

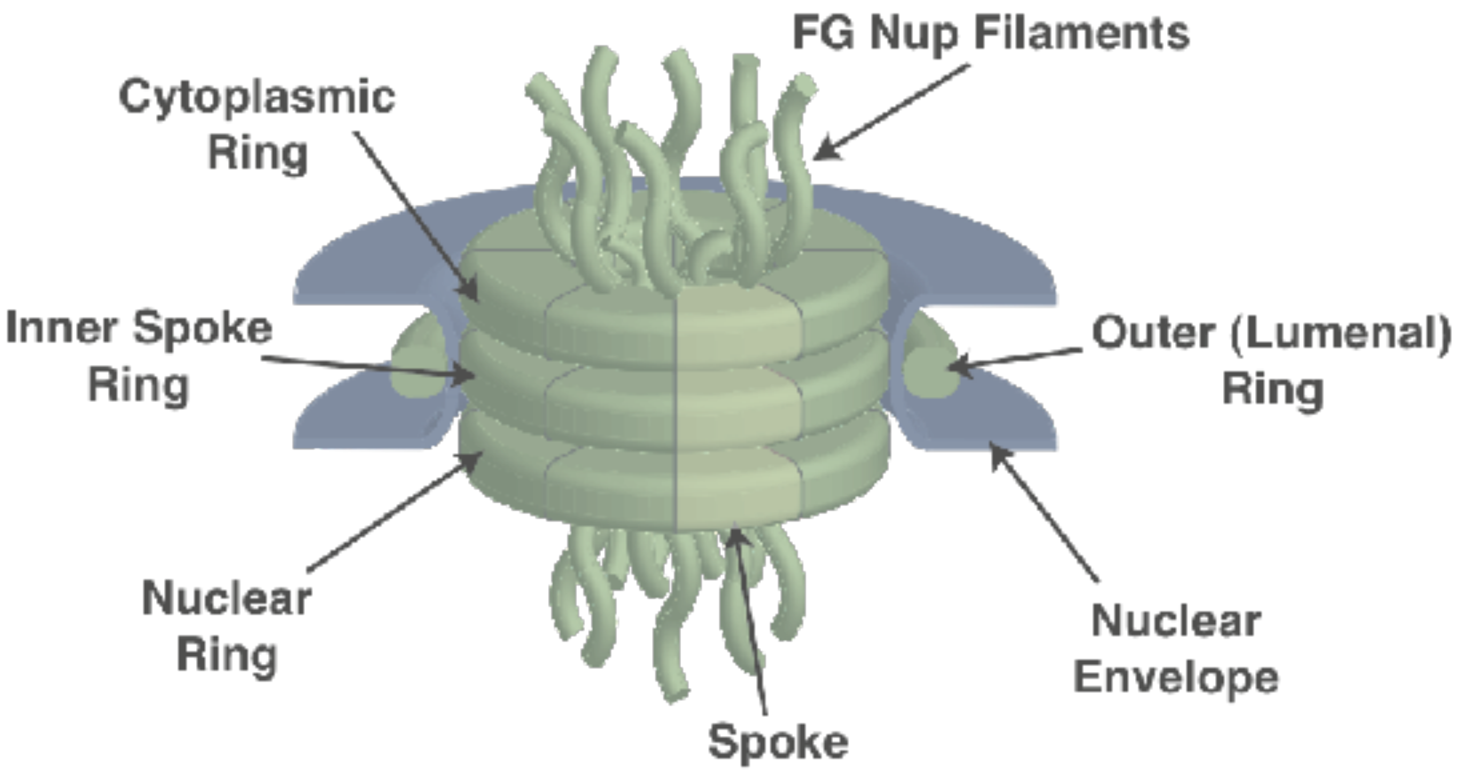
Mediates all known nuclear transport, *via* cognate transport factors (karyoferins or kaps)

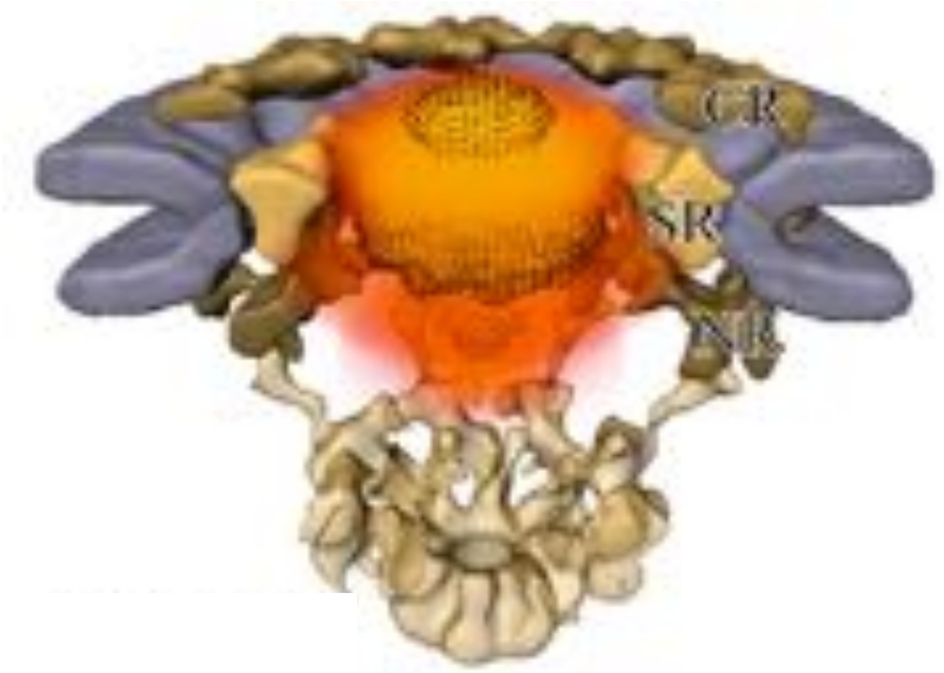
1. Structure
2. Evolution
3. Mechanism of transport
4. Mechanism of assembly
5. Interactions with other systems
6. Modulation and therapy

A large collaborative effort with Mike Rout and Brian Chait at Rockefeller University, also involving many other collaborators (Acknowledgments).

NCDIR *National Center for Dynamic Interactome Research* NIH TCNP

What was known about the NPC structure?





M. Beck, V. Lucic, F. Forster, W. Baumeister, O. Medalia
Nature 449, 611–615 (2007).

R. Milligan, W. Baumeister, O. Medalia, G. Blobel,
E. Hurt, U. Aebi, T. Schwartz, M. Stewart,
C. Akey, B. Chait, M. Rout, ...

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

↓

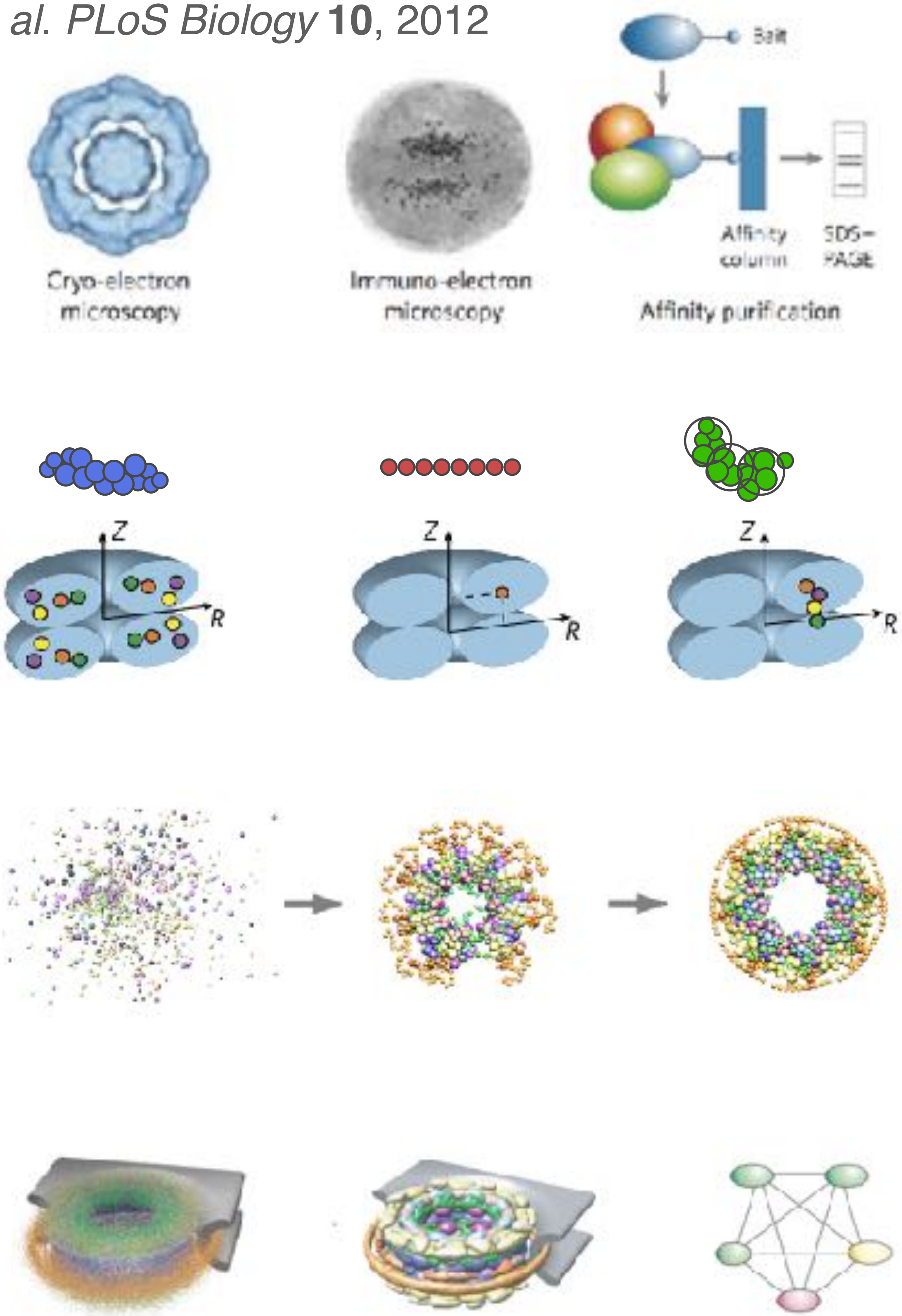
Designing model representation and evaluation

↓

Sampling good models

↓

Analyzing models and information

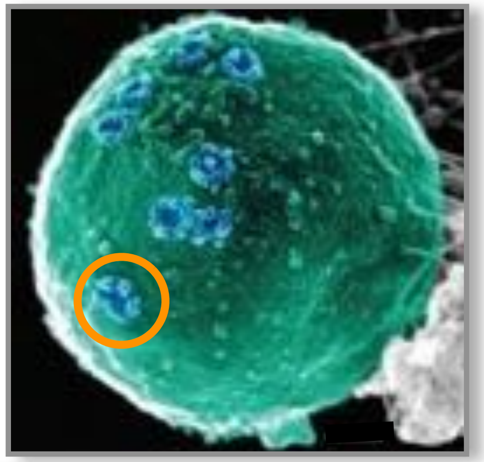


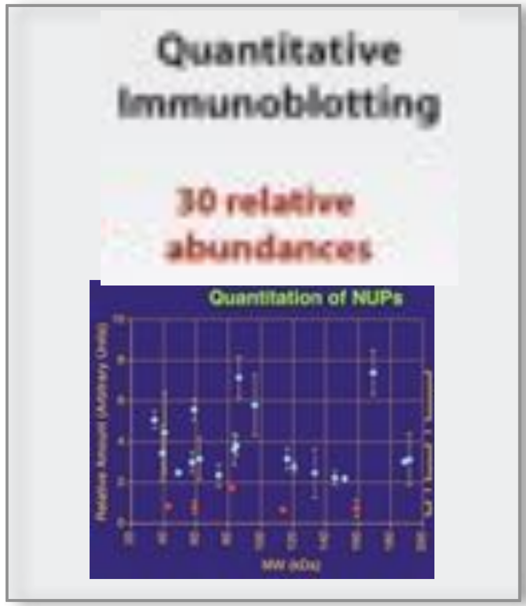
While it may be hard to live with generalization, it is inconceivable to live without it. *Peter Gay*, Schnitzler's Century (2002).

Configuration of 456 proteins in the Nuclear Pore Complex

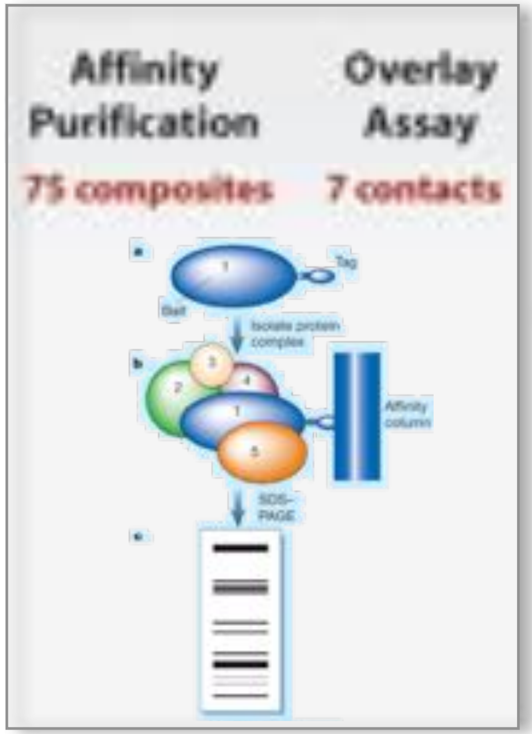
with M. Rout & B. Chait

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

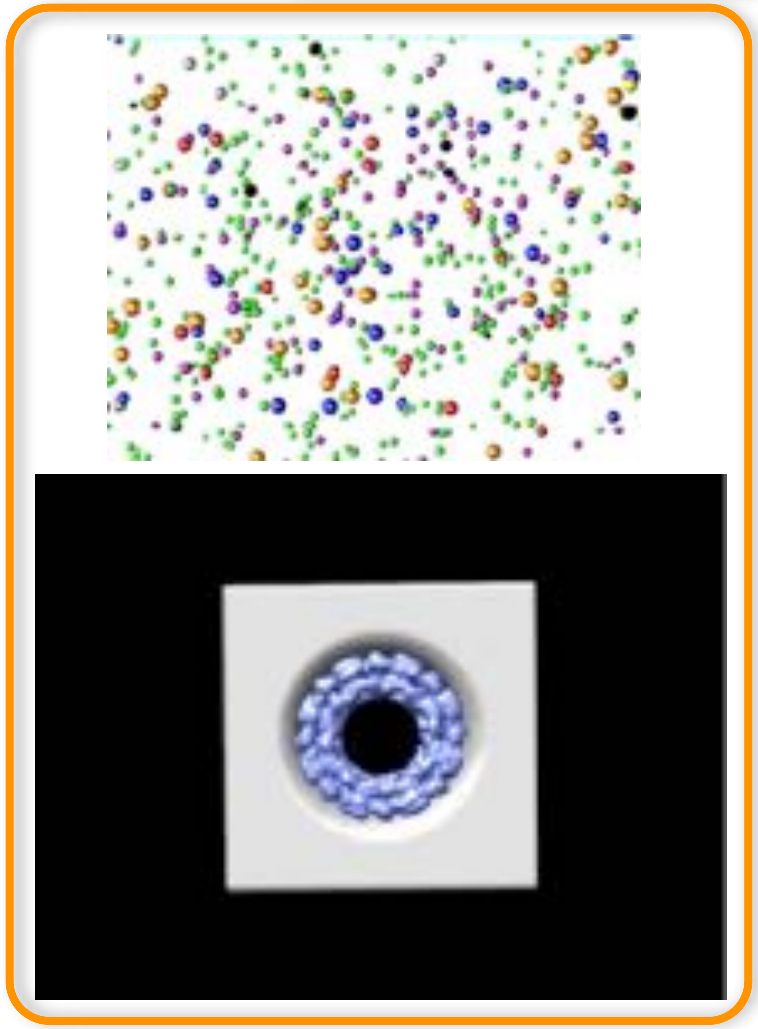




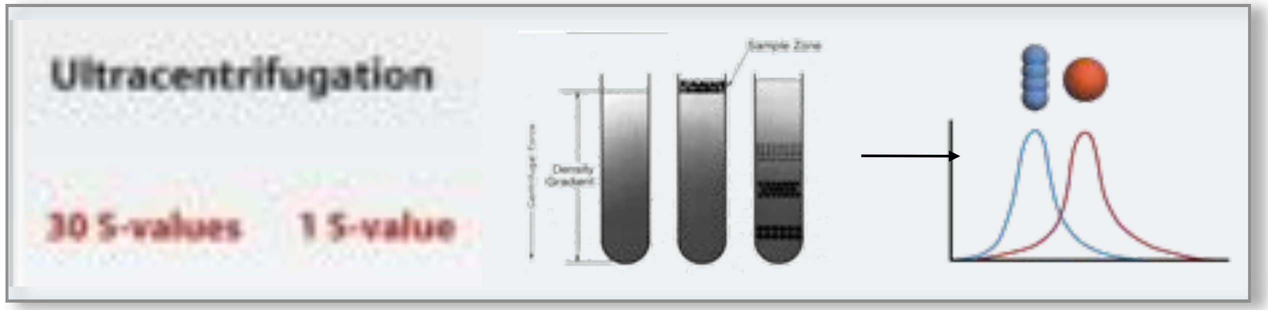
Protein Stoichiometry

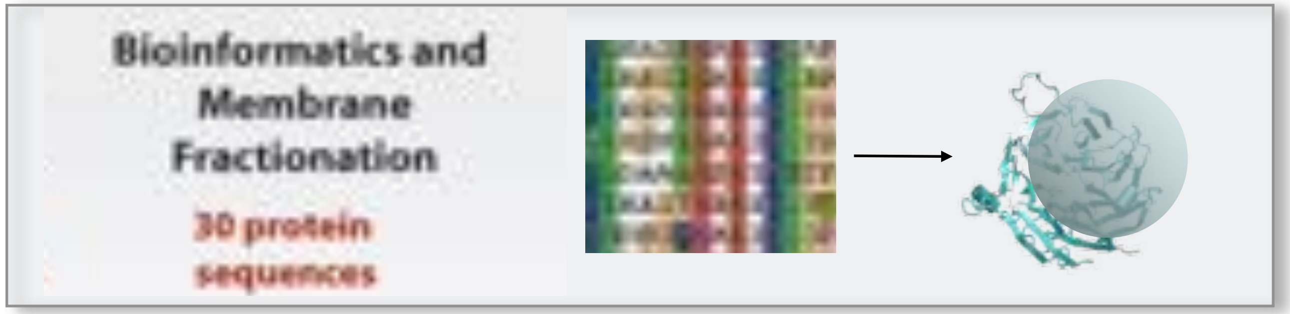


Protein-protein Proximities

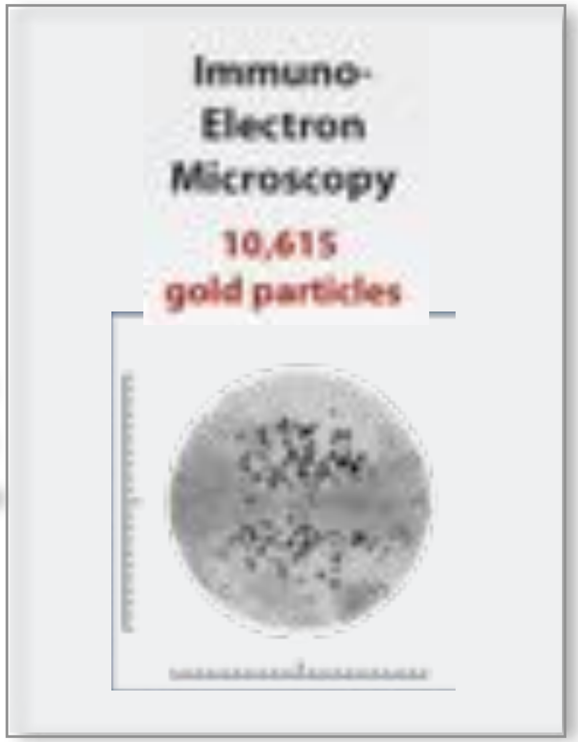


Protein Shape

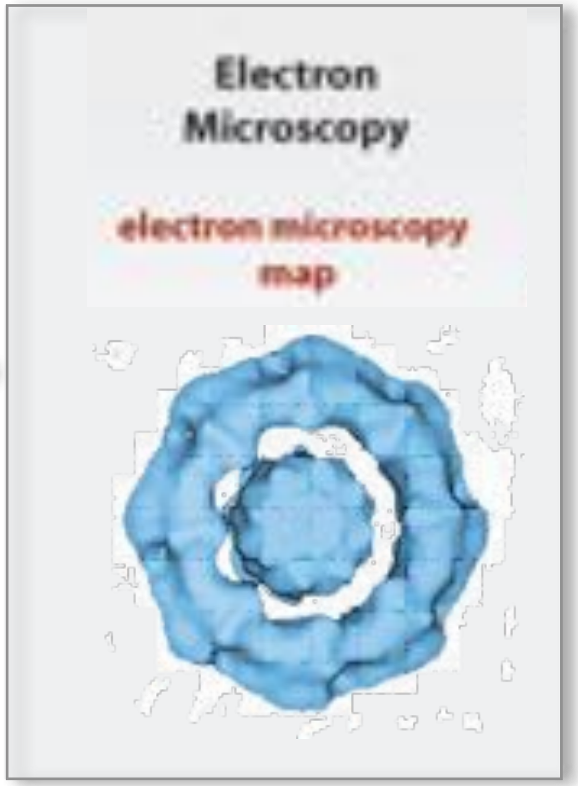




Protein Localization

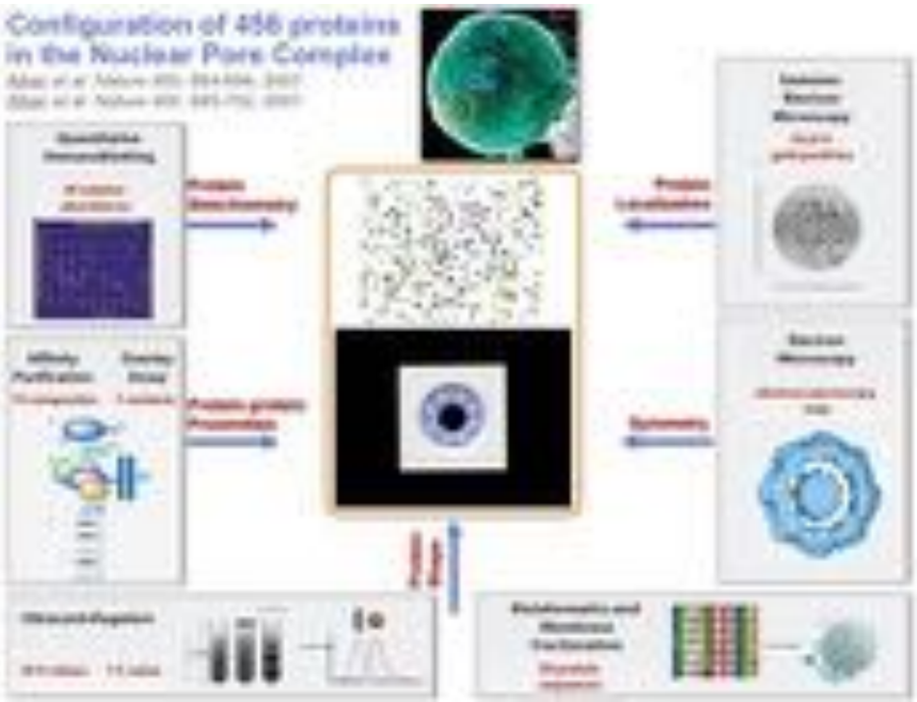


Symmetry

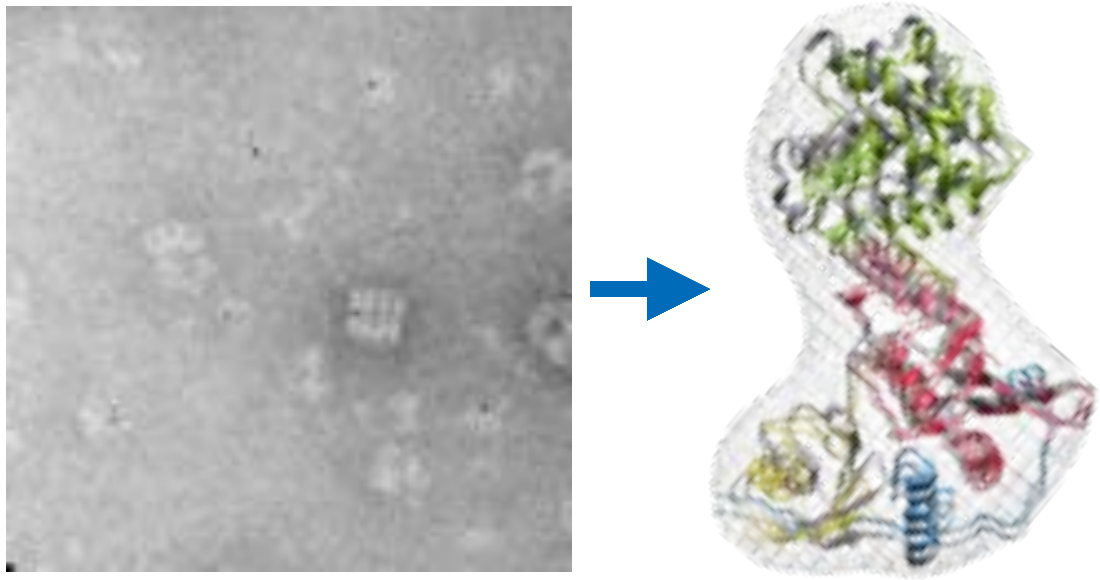


Determination by experiment *versus* prediction by modeling

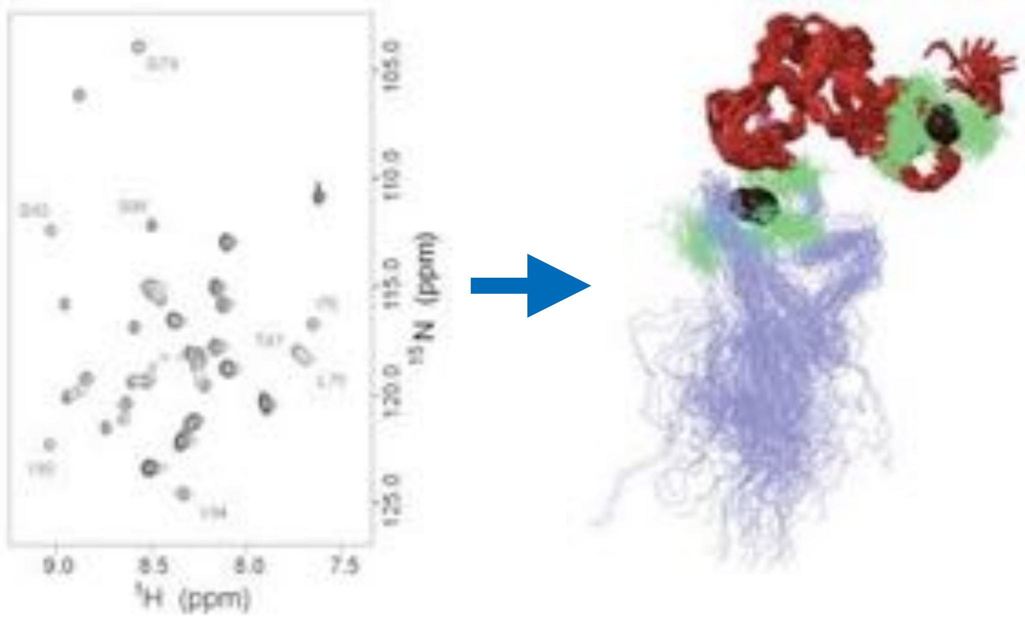
Integrative structure determination



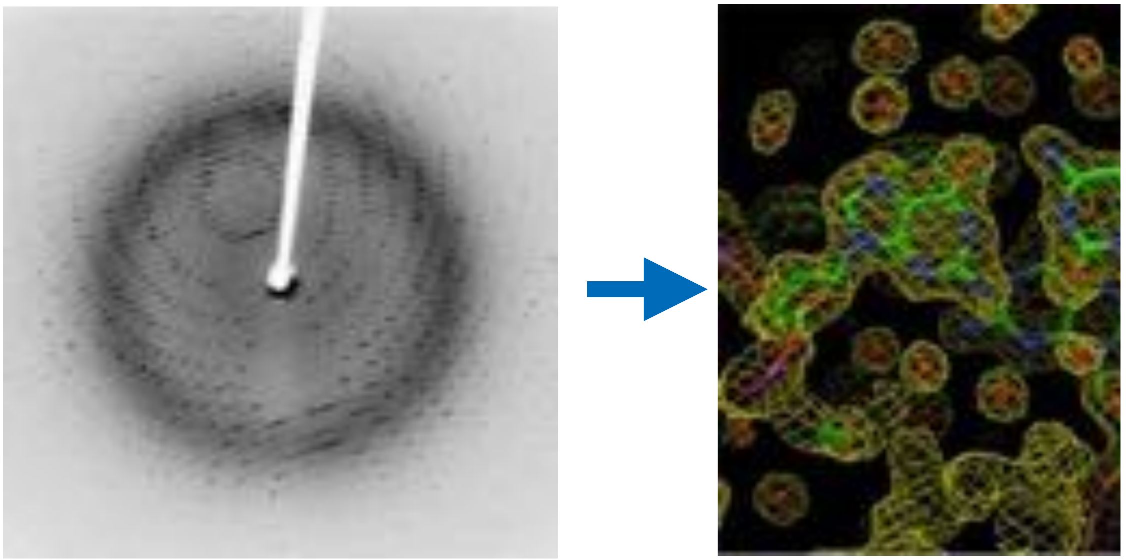
EM microscopy



NMR spectroscopy

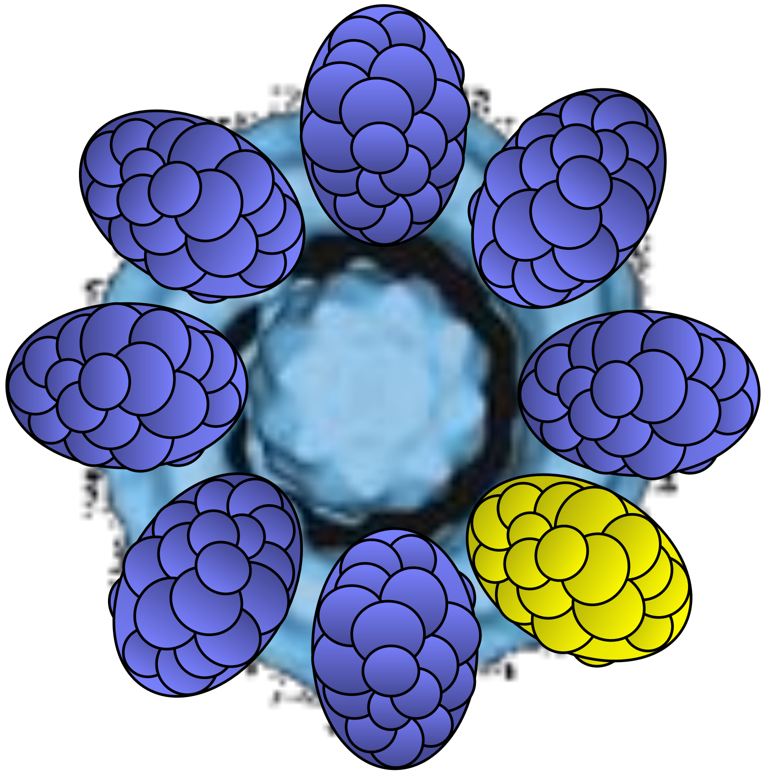


X-ray crystallography

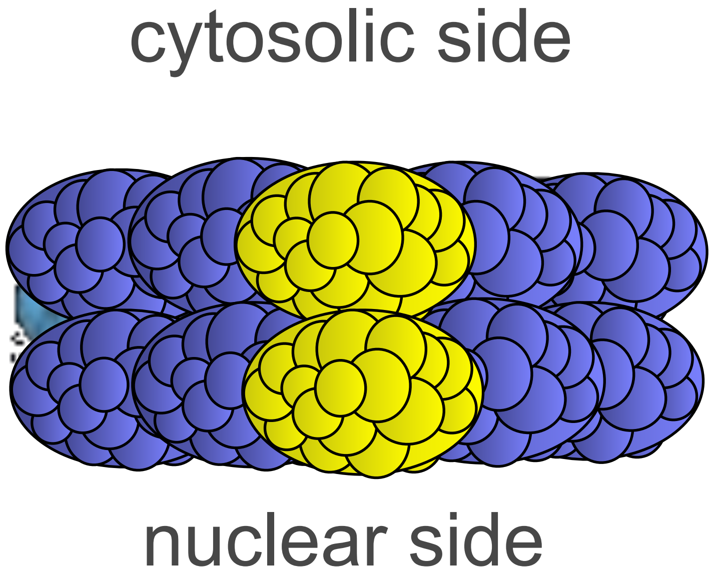


Symmetry Restraints

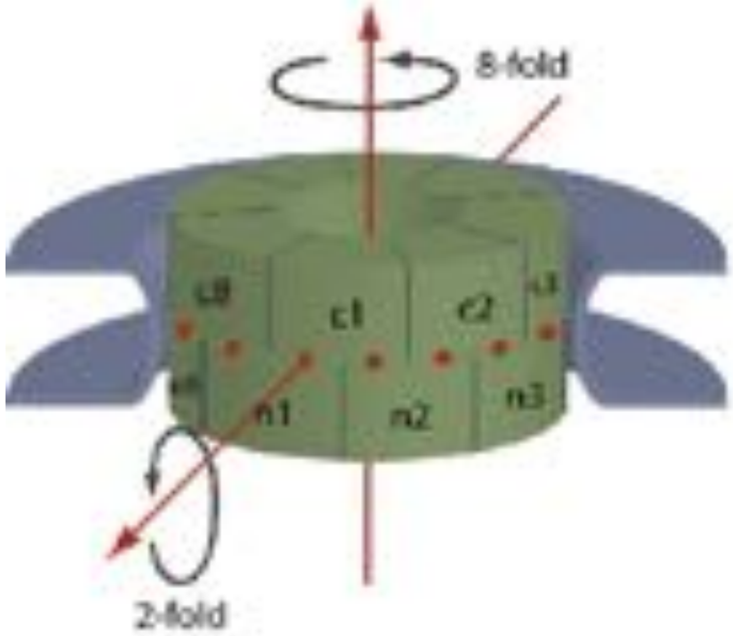
TOP VIEW



SIDE VIEW



SCHEMATIC



Yang, Rout, Akey, *Mol. Cell*. 1, 223, 1998.

half-spoke contains ~30 nucleoporin proteins (NUPs).

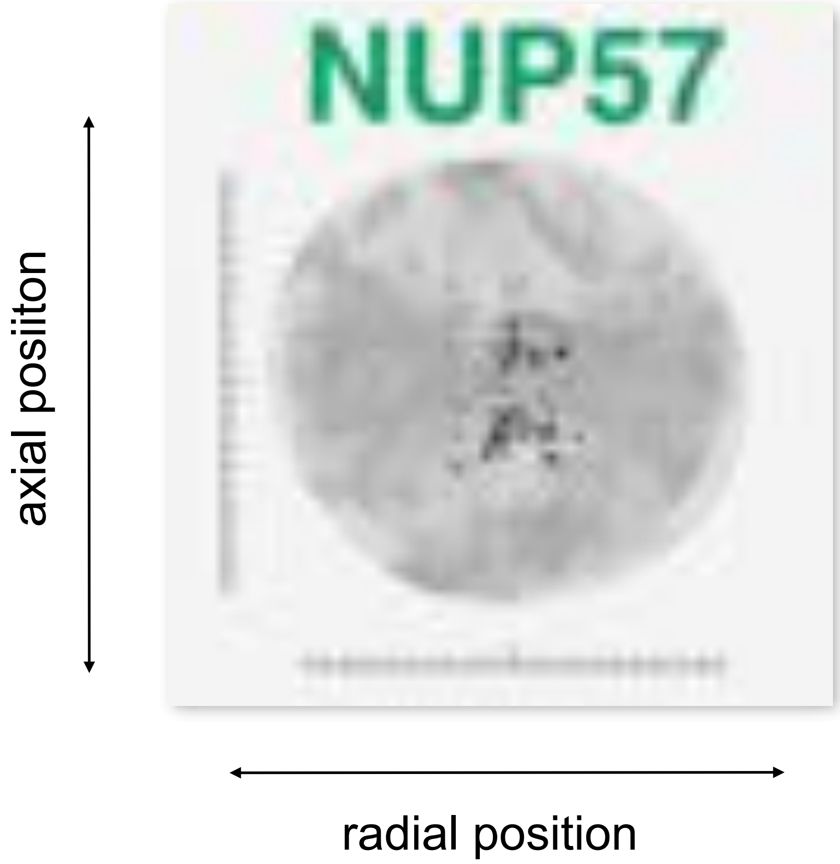
~480 NUPs in NPC.

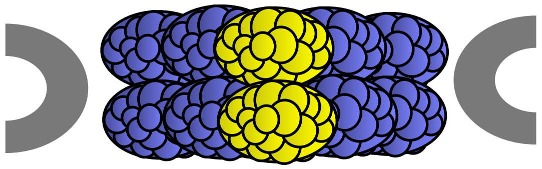
Configurations in spokes and rings are restrained to be similar to each other *via* a DRMS-type restraint.

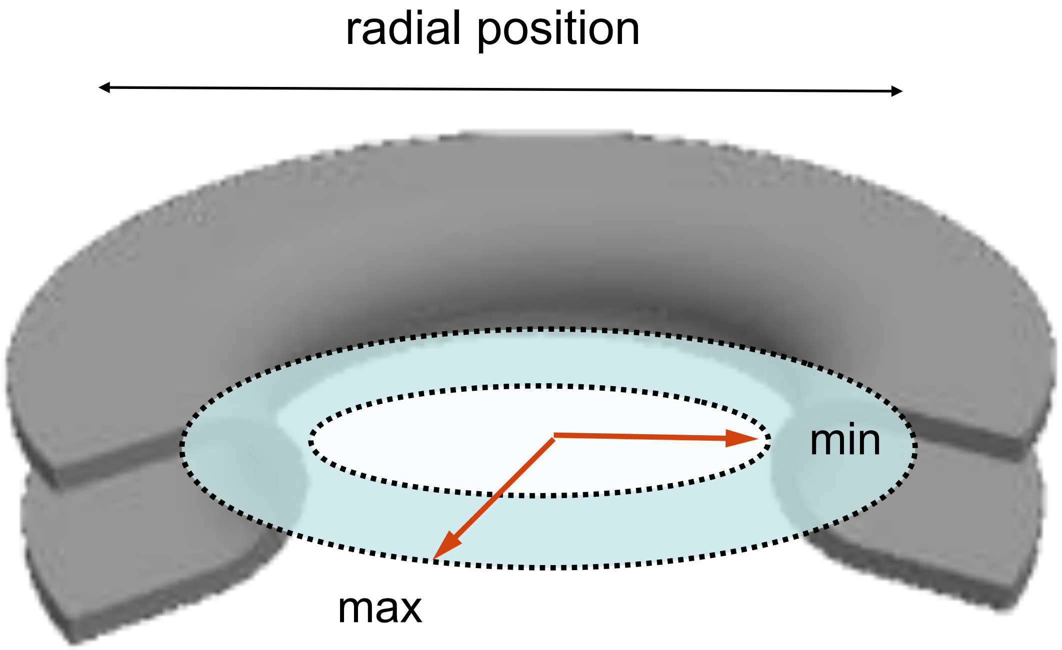
The same handedness of the half-spokes and rings is achieved *via* dihedral angle restraints on subsets of nucleoporins.

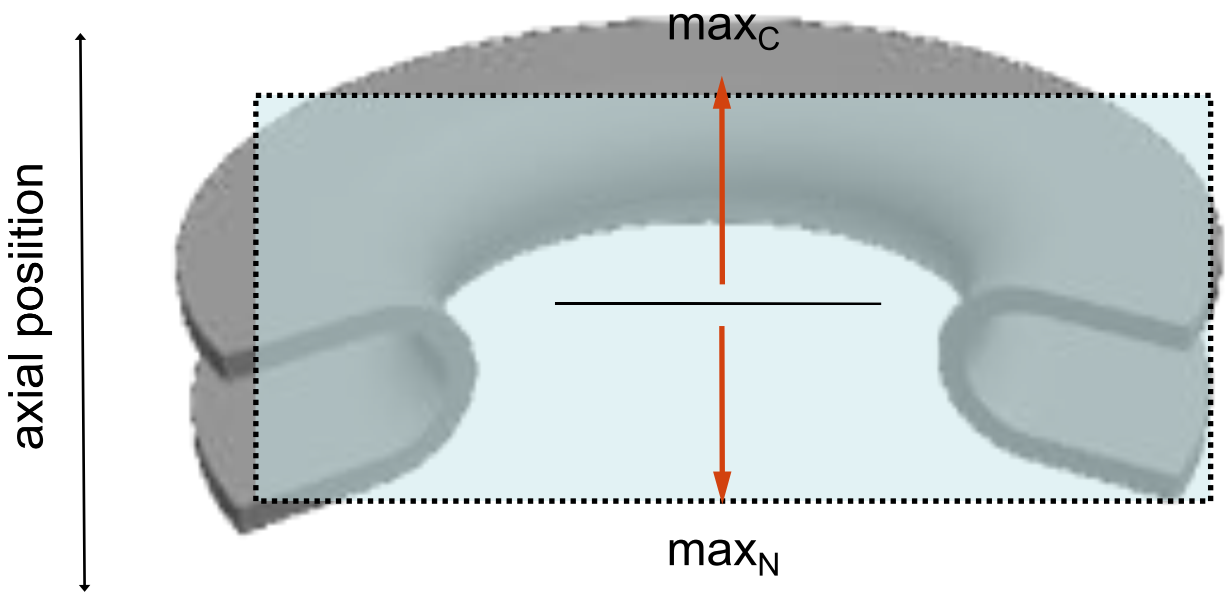
Axial and Radial Localization Restraints on C-terminal Protein Beads

IEM montage (side view)

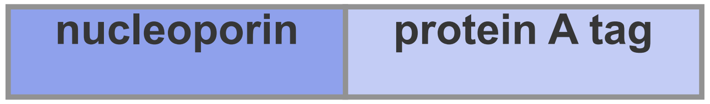


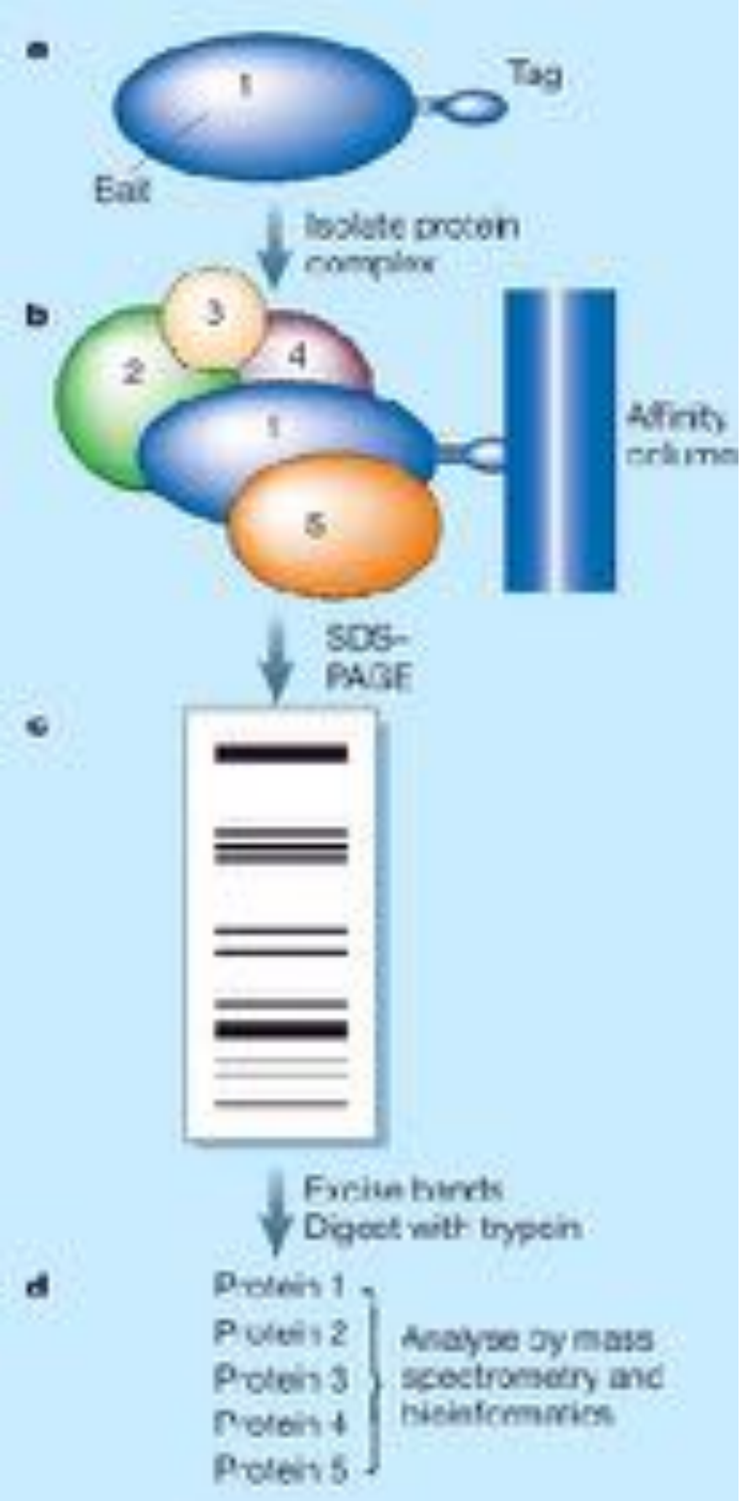




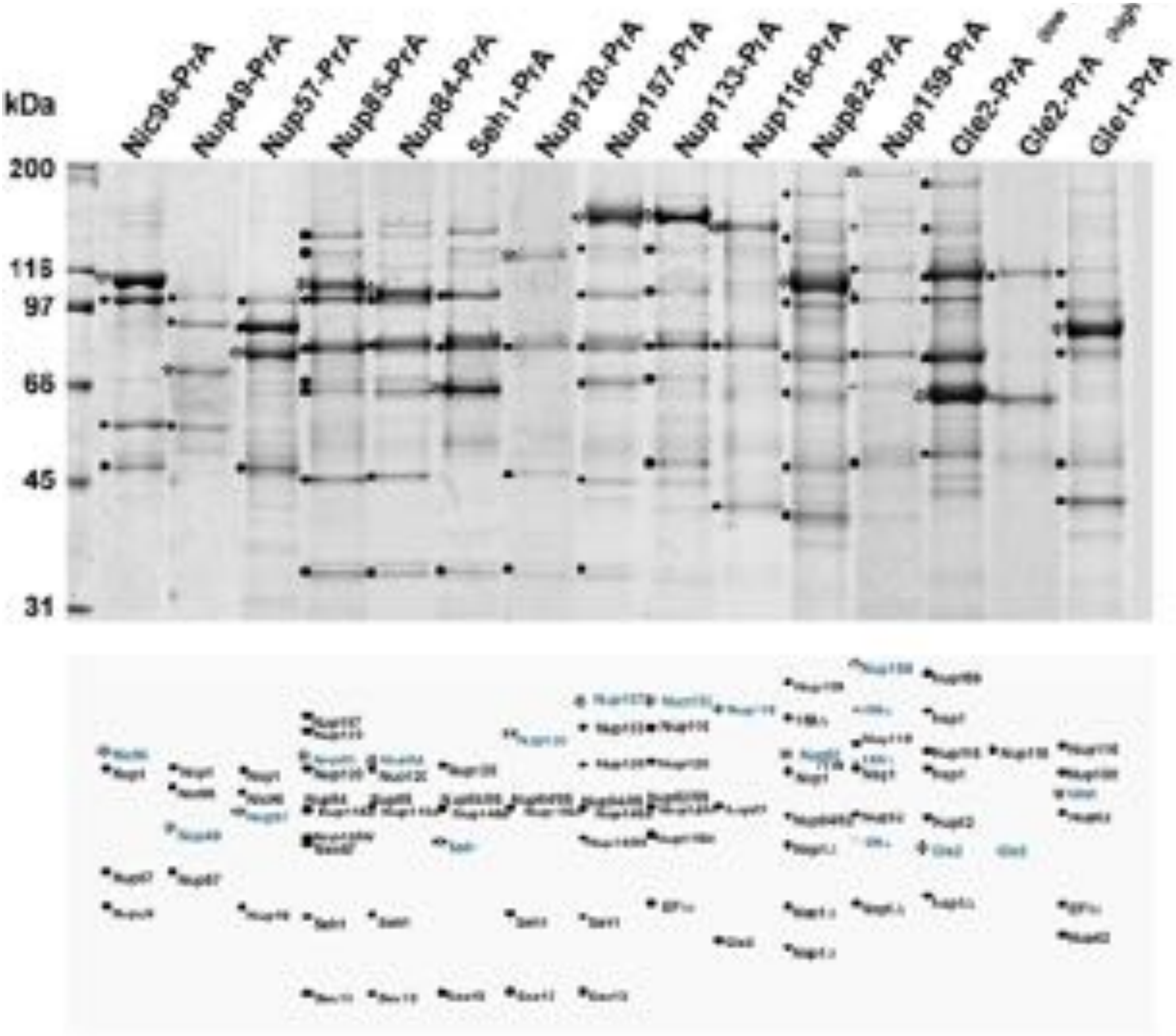


Tagging, Affinity Purification and Analysis of Nucleoporin "Composites"

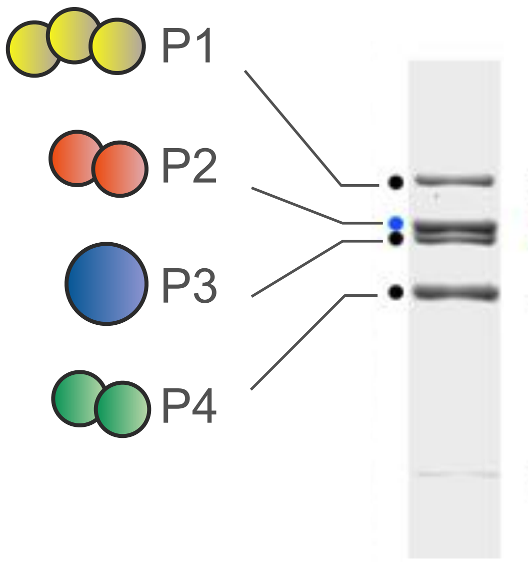




- several hundred "composites"
- ~1,300 protein bands identified by MS



Composites are informative structurally, but subject to assignment ambiguity



- A composite implies at least three direct protein interactions that connect all four protein types.
- But there is assignment ambiguity:
  - Which protein copies interact?
  - What domains interact?
- Many possible alternative restraint assignments are consistent with the composite data.

Alber *et al*. Nature 450, 683-694, 2007
Alber *et al*. Structure 13, 435-445, 2005

Optimization

- Start with a random configuration of protein centers.
- Minimize violations of input restraints by conjugate gradients and molecular dynamics with simulated annealing.
- Obtain an “ensemble” of many independently calculated models (~200,000).

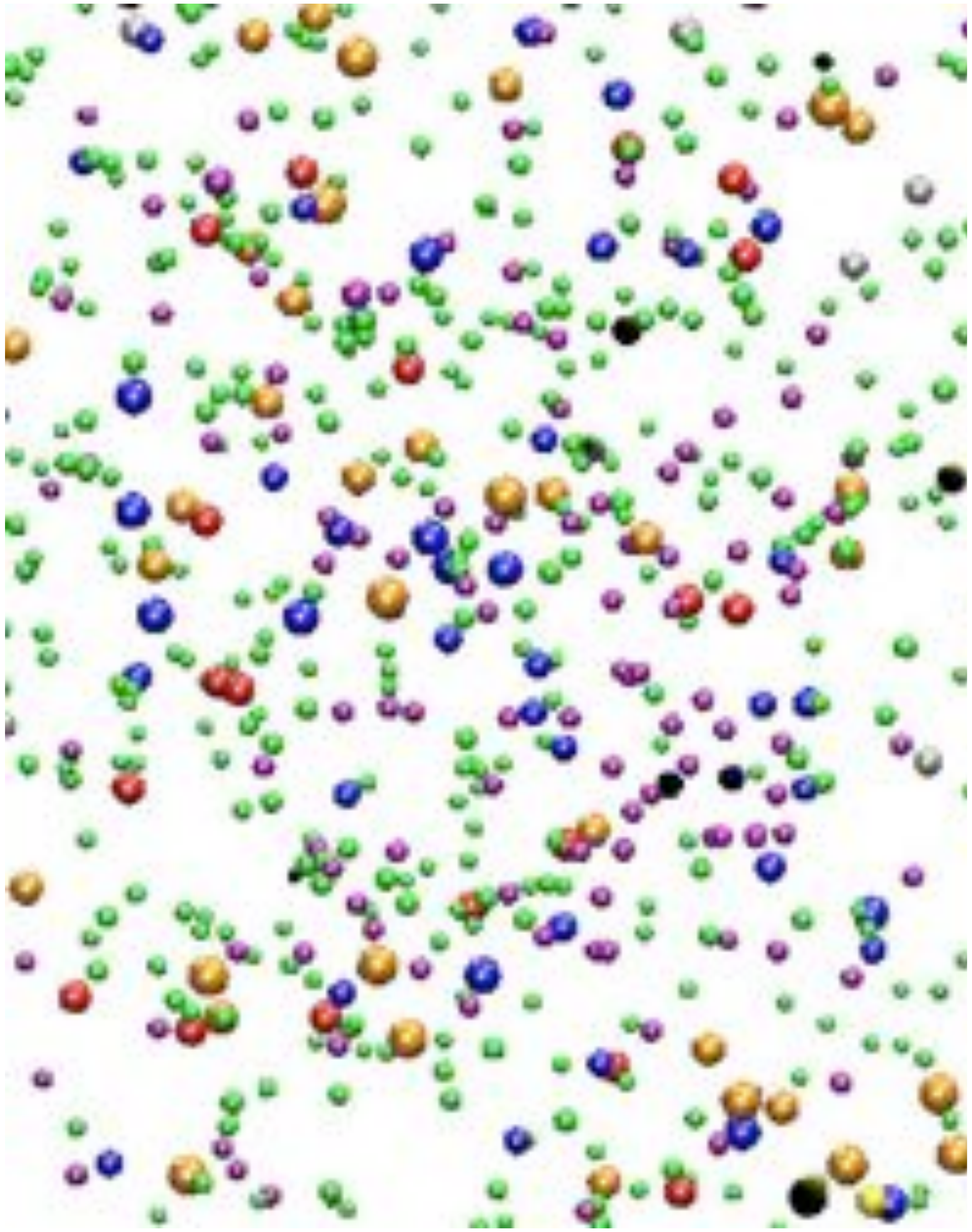
Membrane spanning proteins:
Pom152 Pom34
Ndc1

FG repeat proteins:
Nup159 Nup60
Nsp1 Nup59
Nup1 Nup57
Nup100 Nup53
Nup116 Nup49
Nup145N Nup42

Nup84 complex:
Nup84 Seh1
Nup85 Sec13
Nup120 Nup145C
Nup133

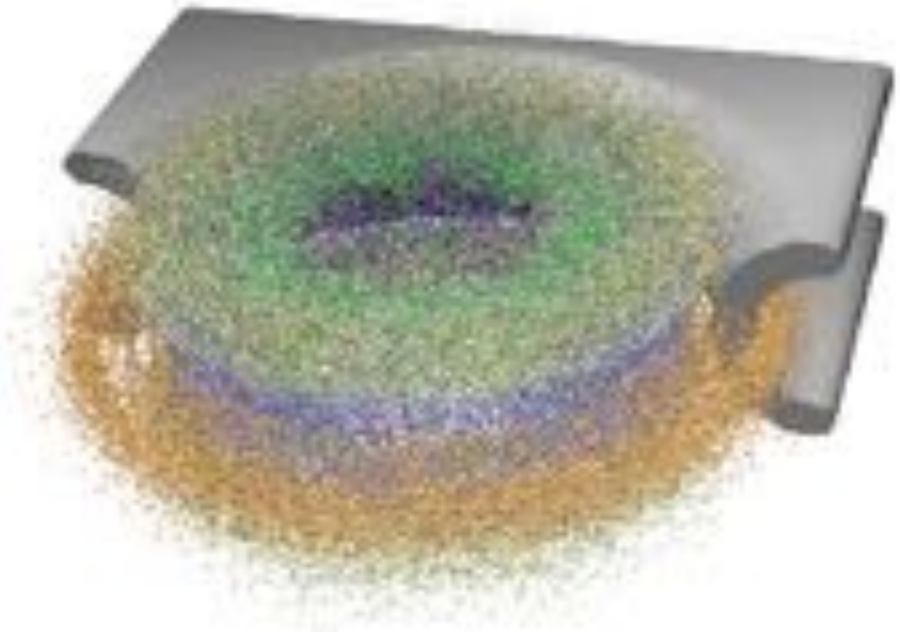
Large Core proteins:
Nup192 Nup170
Nup188 Nup157

Nup82
Nic96

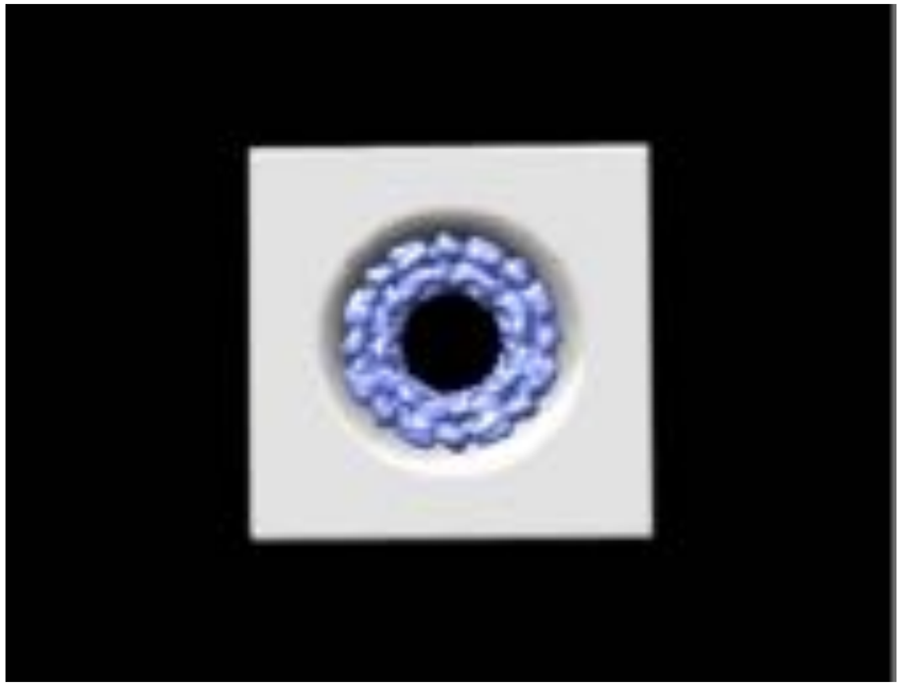


Protein Localization Probability and Volume

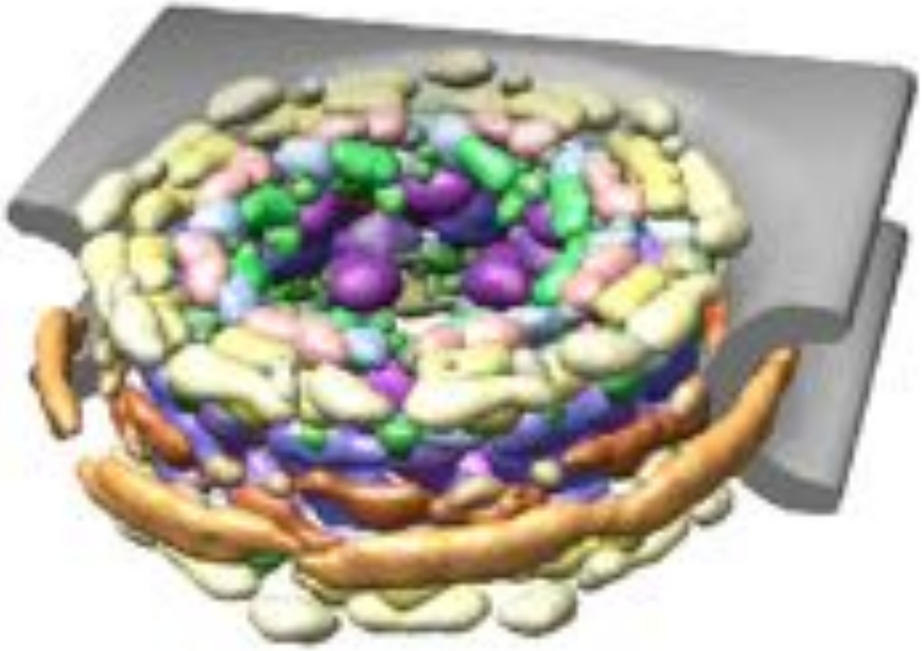
Calculated from the structural superposition of the ensemble of models that satisfy all input restraints



Ensemble of solutions



Animation



Protein localization

can see position of every NPC protein

How accurate is the structure of the NPC?

Assessing the well-scoring models

1. Self-consistency of independent experimental data.
2. Structural similarity among the configurations in the ensemble that satisfy the input restraints.
3. Simulations where a native structure is assumed, corresponding restraints simulated from it, and the resulting calculated structure compared with the assumed native structure.
4. Patterns emerging from a mapping of independent and unused data on the structure that are unlikely to occur by chance.
5. Experimental spatial data that were not used in the calculation of the structure.

Assessment 3/5:

Validation of the structure by a “simulated” model

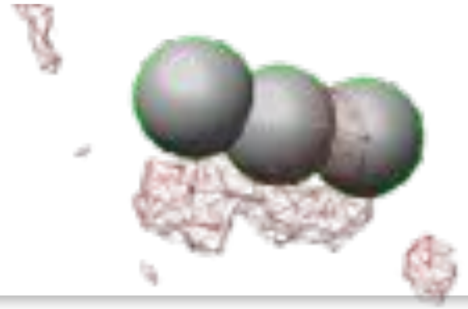
1. Define a structure of the NPC as the native structure.
2. Simulate the restraints, given the native structure.
3. Calculate the structure based on the restraints.
4. Compare the calculated structure with the native one.

Nup84

Nup192

Nup170

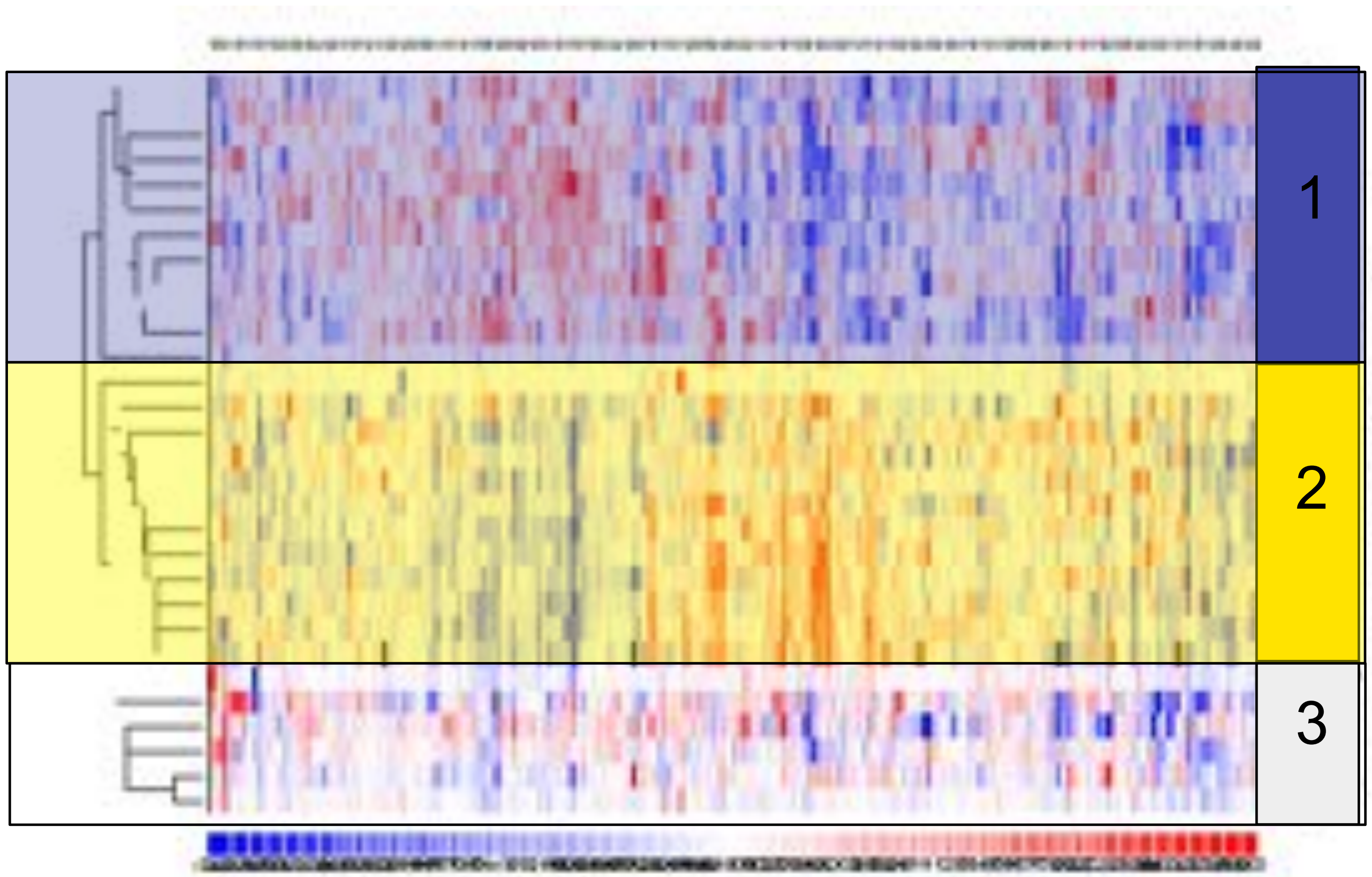
Nup85

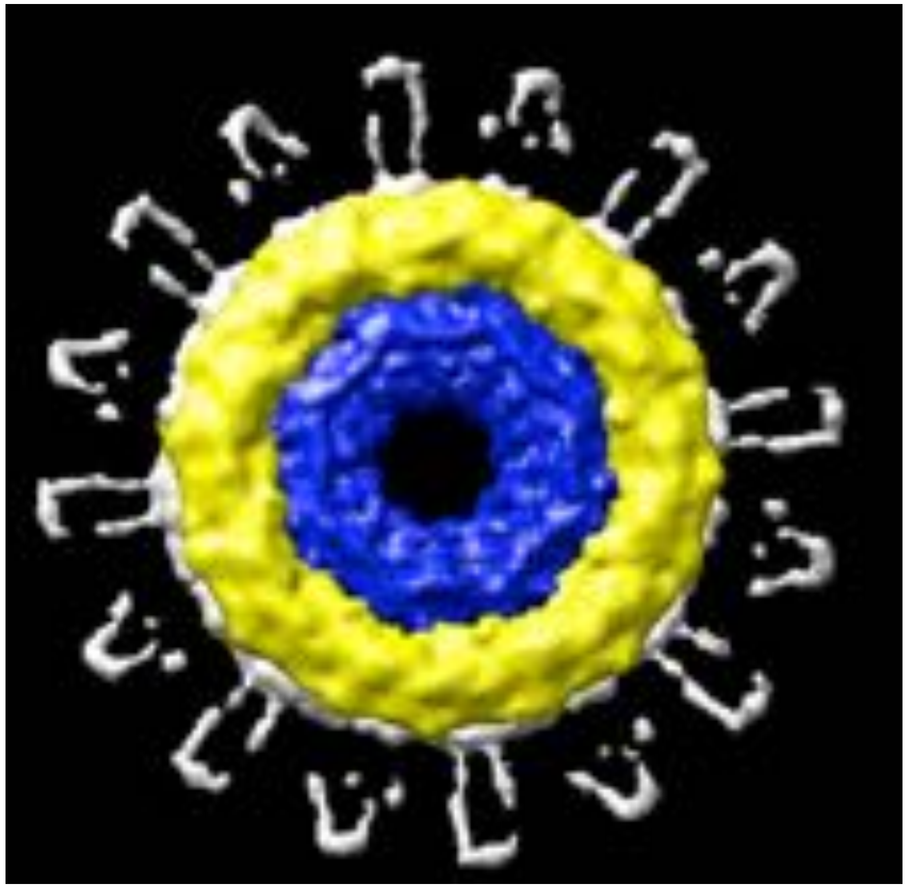


Nup157

Assessment 4/5:
Patterns that are unlikely to occur by chance

X. Zhou (USC):
clustering of nucleoporin expression profiles

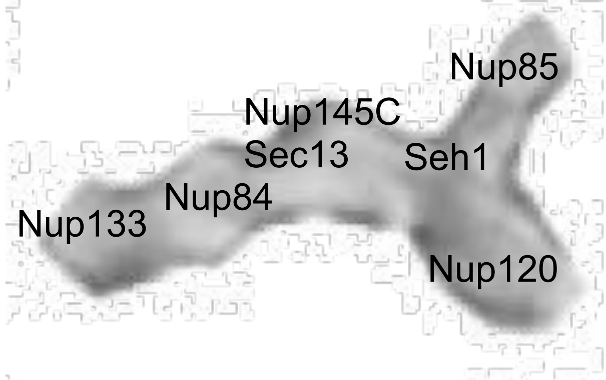




Assessment 5/5:

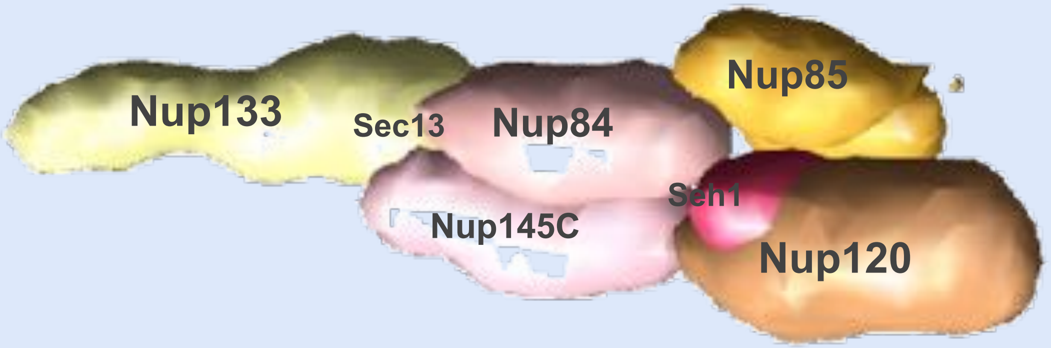
Experimental spatial data about the modeled structure that were not used in the calculation of the model

Nup84 Complex Topology



M. Lutzmann, R. Kunze, A. Buerer, U. Aebi & E. Hurt, *EMBO J*. 21, 387, 2002.

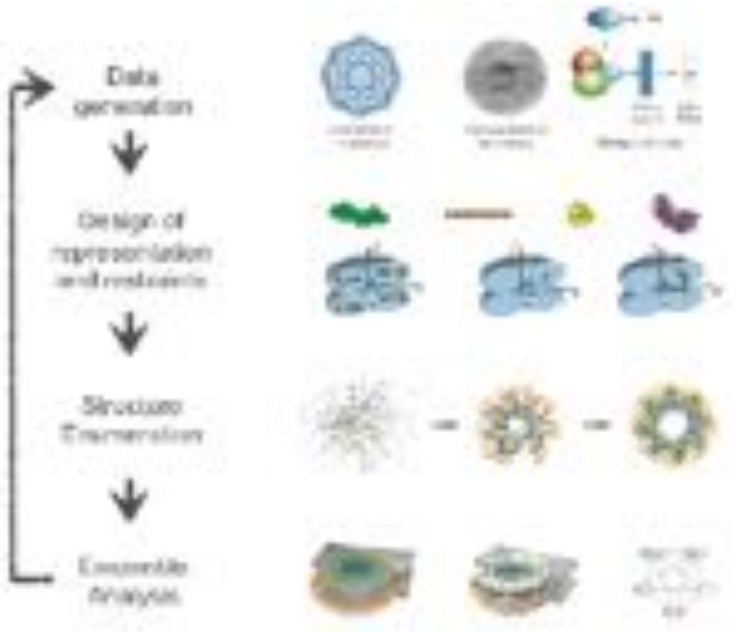
NPC Map is Consistent with Experimental Data *Not* Included in the Calculations

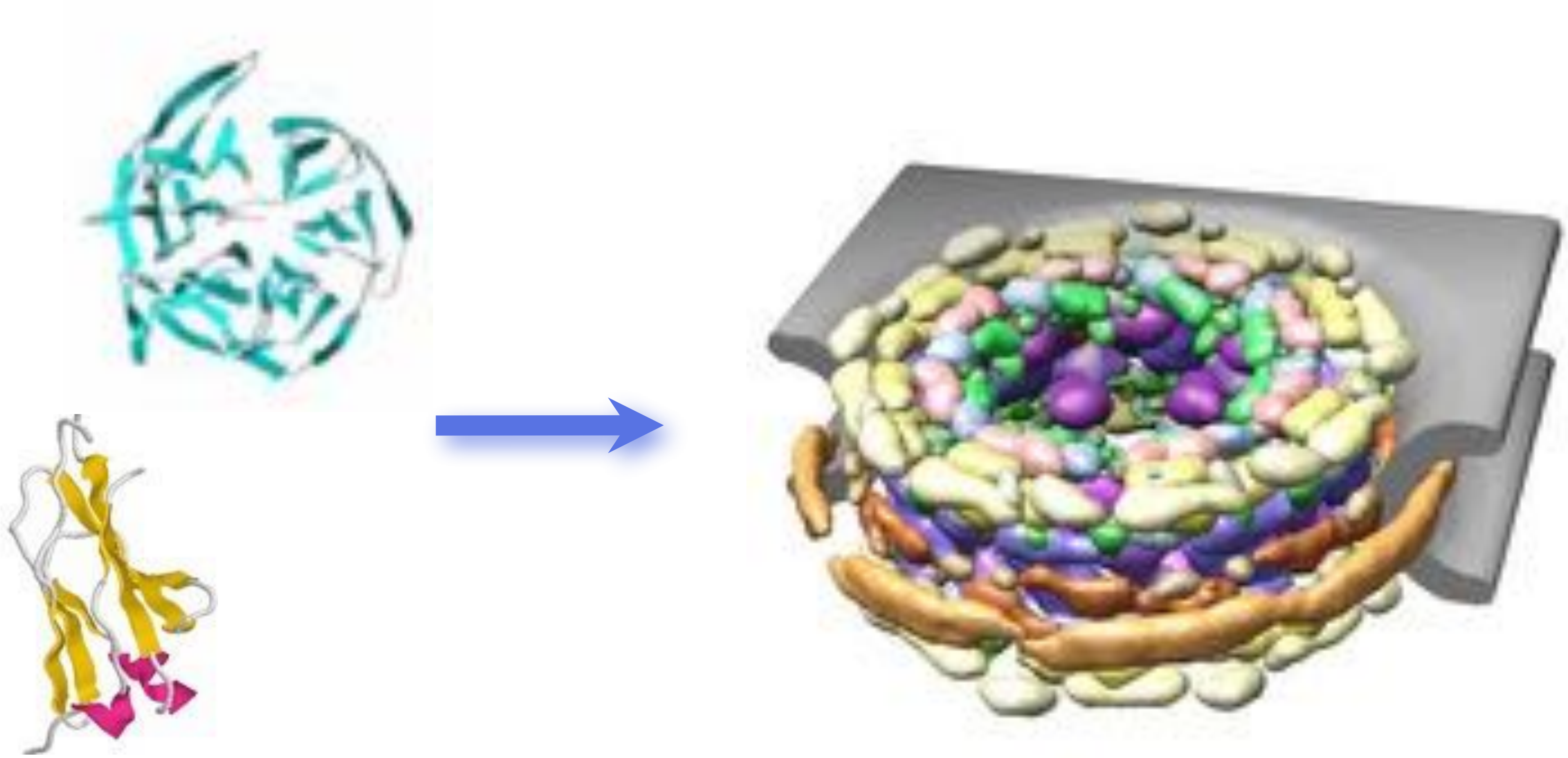


Our Structure

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.





Integrative structure determination of the Nup82 complex

Rout *et al*. Cell 2016, in press

Gathering of Data

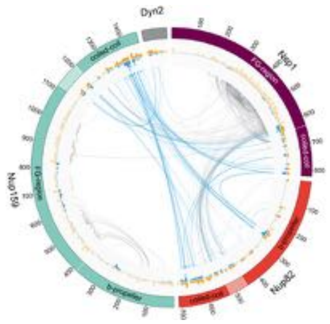
Representing and Translating Data Into Spatial Restraints

Configurational Sampling

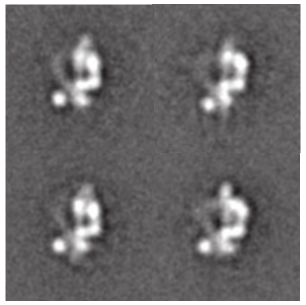
Analysis, Assessment, and Validation of the nsemble

Experimental data

Residue-specific chemical cross-linking
662 DSS (yeast), 126 EDC (yeast), and 343 DSS (hybrid) cross-links
Distance restraints (residue level)

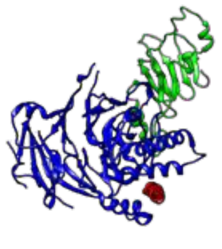


Negative stain electron microscopy
21 wild-type, 23 FG deletion, and 10 GFP-tagged class averages
EM 2D restraint (size and shapes)

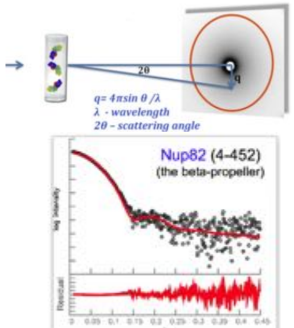


Stoichiometry
10 proteins
2 : 2 : 2 : 2 : 2
Number of subunits

X-ray crystallography
12 domains
Atomic structures



Small angle X-ray scattering
3 constructs of Nup82
Validation (size and shapes)

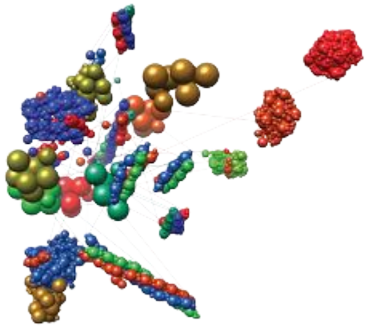


Statistical inference and physical principles

Comparative modeling
18 domains
Fold models of coiled-coils

Steric Effect
10 proteins
Excluded Volume

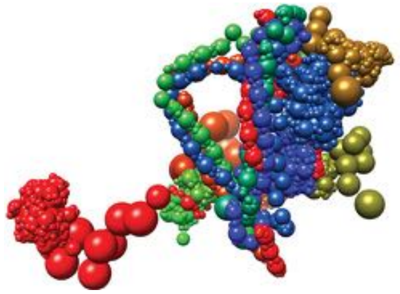
Initial random configurations



Initial
Replica Exchange Gibbs sampling

Using one of each EM 2D class, then applied the EM 2D Filter

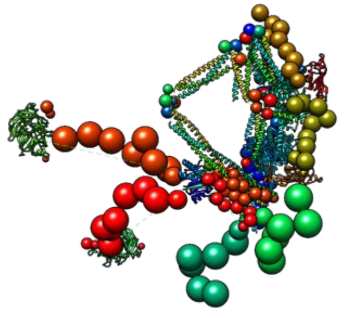
Filtered model structures



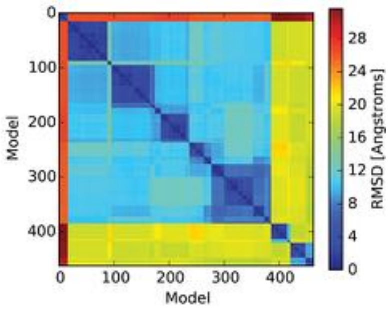
Refinement
Replica Exchange Gibbs sampling

Using all of the 21 EM 2D class averages

Final refined structures

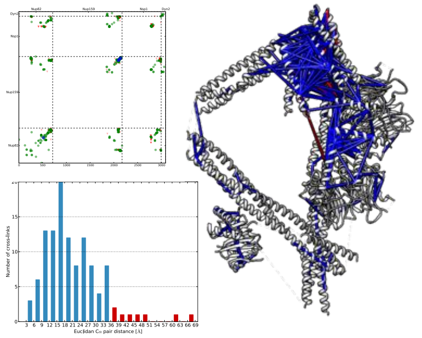


Clustering of 463 top-scoring solutions



Localization probability densities

Satisfaction of chemical cross-links

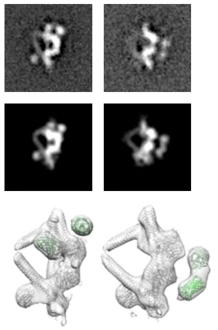


Validation by GFP mass-tagged EM

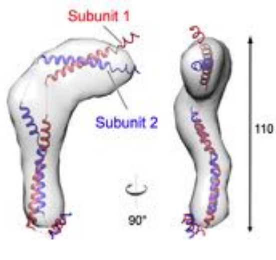
EM 2D class averages

Structure Projections

GFP mass-tagged structures



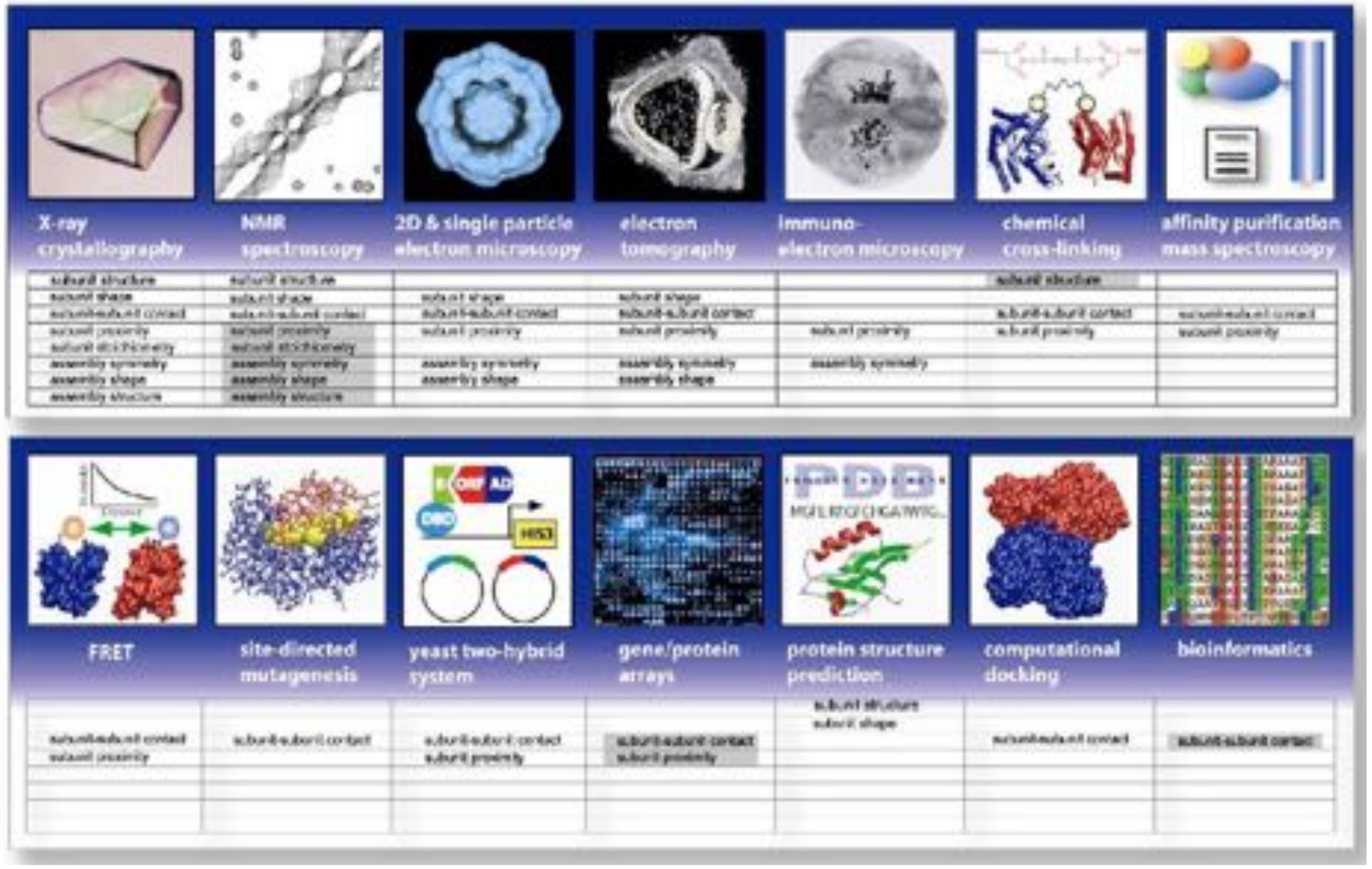
Validation by SAXS data



In Conclusion

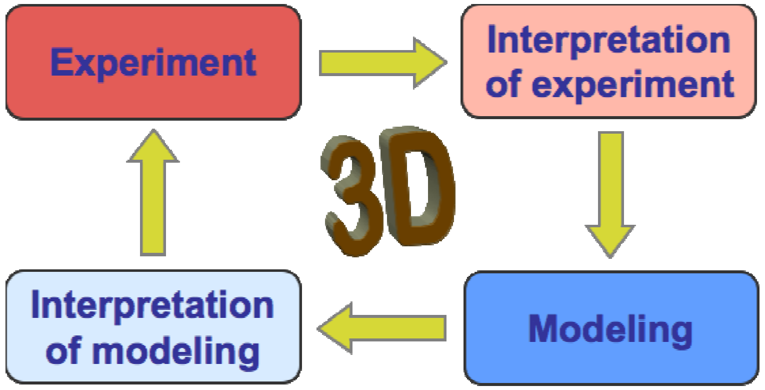
The goal is a comprehensive description of the multitude of interactions between molecular entities, which in turn is a prerequisite for the discovery of general structural principles that underlie all cellular processes.

This goal will be achieved by a ***formal*** integration of **experiment**, **physics**, and **statistical inference**, spanning all relevant size and time scales.



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 proximity	subunit-subunit contact	subunit-subunit contact subunit proximity	subunit-subunit contact subunit proximity		subunit-subunit contact	subunit-subunit contact



Experiment

Interpretation of experiment

3D

Modeling

Interpretation of modeling

Sali, Earnest, Glaeser, Baumeister. From words to literature in structural proteomics. *Nature* 422, 216-225, 2003.
Robinson, Sali, Baumeister. The molecular sociology of the cell. *Nature* 450, 974-982, 2007.
Alber, Foerster, Korkin, Topf, Sali. *Annual Reviews in Biochemistry* 77, 11.1–11.35, 2008.
D. Russel et al. “Putting the pieces together: integrative structure determination of macromolecular assemblies”. PLoS Biol., 2012.
A. Ward, A. Sali A, I. Wilson. Integrative structural biology. Science 339, 913-915, 2013.
Schneidman, Pellarin, Sali. Uncertainty in integrative structural modeling. Curr. Opin. Str. Biol., 96-104, 2014.

Topics

1. Introduction to integrative (hybrid) structure determination
2. Integrative structure determination of the Nuclear Pore Complex

F. Alber et al. "Determining the architectures of macromolecular assemblies". *Nature* **450**, 683-694, 2007.

F. Alber et al. “Integrating Diverse Data for Structure Determination of Macromolecular Assemblies”
Annual Review of Biochemistry **77**, 11.1-11.35, 2008.

D. Russel et al. “Putting the pieces together: integrative structure determination of macromolecular assemblies.”
PLoS Biol. **10**, e1001244, 2012.

A. Ward, A. Sali, I. Wilson. Integrative structural biology. *Science* **339**, 913-915, 2013

From: http://salilab.org/publications/